SUPPLEMENTARY DATASET

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

---

Authors: Hassan Gbran[1,2*],

Corresponding Author: Hassan Gbran
Email: gbranhassan882@gmail.com

Version: 3.0 (Final Data Release – Fully Verified)
Date: 2026-07-16
DOI: https://doi.org/10.5281/zenodo.17333842
OSF: https://doi.org/10.17605/OSF.IO/GM47X

TABLE OF CONTENTS

VOLUME I: STUDY DESIGN AND METHODOLOGY

1. STUDY OVERVIEW AND SCOPE
2. SITE CHARACTERISATION – COMPREHENSIVE DETAILS
3. SENSOR DEPLOYMENT AND HARDWARE SPECIFICATIONS
4. PHOTOGRAMMETRIC DATA ACQUISITION
5. RADIOMETRIC CORRECTION AND PREPROCESSING
6. FEATURE EXTRACTION AND ENGINEERING
7. LABELLING AND EXPERT ANNOTATION PROTOCOL
8. HIERARCHICAL CLUSTERING METHODOLOGY AND OUTPUTS
9. RANDOM FOREST CLASSIFICATION METHODOLOGY

VOLUME II: RESULTS AND ANALYSIS

10. ENVIRONMENTAL MONITORING RESULTS
11. HIERARCHICAL CLUSTERING PERFORMANCE
12. RANDOM FOREST CLASSIFICATION PERFORMANCE
13. FEATURE IMPORTANCE AND ABLATION STUDY
14. SPATIAL EVALUATION (IoU ANALYSIS)
15. LEAD TIME ANALYSIS
16. PREDICTIVE FORECASTING VALIDATION
17. CROSS-SITE GENERALISABILITY
18. HERITAGE VALUES INDEX
19. STATISTICAL MODEL OUTPUTS

VOLUME III: QUALITY ASSURANCE AND SUPPLEMENTARY MATERIALS

20. DATA QUALITY AND COMPLETENESS
21. SENSOR CALIBRATION AND DRIFT CORRECTION
22. SUPPLEMENTARY TABLES AND FIGURES
23. APPENDICES A–H
24. DECLARATION OF DATA AVAILABILITY

VOLUME I: STUDY DESIGN AND METHODOLOGY

---

1. STUDY OVERVIEW AND SCOPE

1.1 Research Context

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

Research Objective: To develop and empirically validate an integrated AI-IoT framework that combines continuous environmental sensing with multi-stage machine learning (unsupervised hierarchical clustering + supervised Random Forest) to provide early warning of chromatic deterioration on heritage building façades in tropical coastal environments.

1.2 Geographic and Climatic Context

Location: Semarang, Central Java, Indonesia
Coordinates: 6°58′S, 110°25′E
Climate Classification: Köppen Af (Tropical Rainforest Climate)

Climatic Characteristics:

- Annual rainfall: 2,500–3,200 mm
- Wettest months: November–March (monsoon season)
- Driest months: June–September
- Mean annual temperature: 27.8°C (range: 25–34°C)
- Mean annual relative humidity: 78.2% (range: 65–92%)
- Daily RH rarely drops below 70% even during dry season
- Prevailing wind direction: North-west (monsoon) and South-east (dry season)
- Tidal influence: Semarang experiences semi-diurnal tides with range 0.5–1.2 m

Justification for Site Selection:
Semarang was selected due to:

1. High density of heritage buildings from the Dutch colonial period
2. Severe environmental pressures (coastal location, high humidity, flooding)
3. Active conservation challenges documented in prior studies
4. Accessibility and institutional support for long-term monitoring

1.3 Temporal Scope

Monitoring Period: 14 months (1 March 2025 – 30 April 2026)

Seasonal Breakdown:

- Monsoon season (wet): November 2025 – March 2026
- Dry season: June 2025 – September 2025
- Transition periods: April–May 2025, October 2025

1.4 Key Quantities

Parameter	Value
Heritage sites	3 (Lawang Sewu, Pasar Johar, Tawang Station)

IoT sensor nodes (primary analysis)	16 (DHT22)
IoT sensor nodes (additional monitoring)	2 (LS-09, LS-10)
Total IoT sensor nodes deployed	18
Logging interval	15 minutes
Total raw observations (16 primary nodes)	517,632 readings
Total raw observations (18 nodes total)	584,064 readings
Valid readings after QC (16 primary nodes)	509,632 (98.4% completeness)
Valid readings after QC (18 nodes total)	574,176 (98.3% completeness)
Photogrammetric images	1,200 (Nikon D5600, 24 MP, RAW)
Expert-labelled pixels	48,830
Annotated tiles	3,750 (sizes: 32×32, 64×64, 128×128)
Decay classes	5 (Moisture Discoloration, Biological Growth, Surface Accumulations, Chromatic Alteration, Unaltered)
Expert annotators	3 (8–20 years experience)
Heritage Values Index participants	25 (modified Delphi)

Note on Node Counts:

- The primary analysis published in the main manuscript is based on 16 sensor nodes (8 at Lawang Sewu, 4 at Pasar Johar, 4 at Tawang Station), yielding 517,632 raw observations.
- Two additional nodes (LS-09 at Lawang Sewu Tower Base, LS-10 at Lawang Sewu Roof Void) were deployed for supplementary monitoring after the primary analysis was completed. These nodes are included in this supplementary dataset for completeness but did not influence the primary results reported in the manuscript.
- All results reported in Sections 10–19 of this dataset are based on the 16 primary nodes unless explicitly stated otherwise.

---

2. SITE CHARACTERISATION – COMPREHENSIVE DETAILS

2.1 Site 1: Lawang Sewu

2.1.1 Historical and Architectural Context

Year of Construction: 1904
Original Name: Nederlandsch-Indische Spoorweg Maatschappij (NIS) Hoofdkantoor
Original Function: Administrative headquarters of the Dutch East Indies Railway Company
Architect: Prof. Jacob F. Klinkhamer (with C. van der Wal as executing architect)
Architectural Style: Neo-Renaissance with Art Deco influences, adapted to tropical climate

Architectural Significance:
Lawang Sewu (Javanese for "Thousand Doors") is one of Indonesia's most significant colonial heritage buildings. It was the tallest building in Semarang at the time of construction and represented the technological sophistication of Dutch colonial engineering. The building's name derives from its numerous doorways and large windows (actual count: 62 doors and 522 windows, not 1,000).

Climatic Adaptation Features:

1. **Deep verandas:** 3–4 m wide on all sides providing shade and rain protection
2. **Twin ventilation towers:** Harness passive stack effect for natural ventilation
3. **High ceilings:** 5–6 m to promote air circulation
4. **Operable louvre windows:** Allow adjustable cross-ventilation
5. **Thick masonry walls:** 50–70 cm with cavity construction for thermal mass
6. **Elevated ground floor:** 1.2 m above street level to mitigate flooding

2.1.2 Structural and Material Characteristics

Structural System:

- Foundation: Deep pile foundation (timber piles, 8–12 m depth)
- Substructure: Brick masonry with lime mortar
- Superstructure: Load-bearing brick masonry walls (50–70 cm)
- Floor construction: Timber joists with tile flooring (ground floor), timber floorboards (upper floors)
- Roof construction: Timber trusses with clay tile roofing
- Internal partitions: Timber stud walls with plaster finish

Material Composition:

Material	Location	Condition at Study Start
Brick	External walls	Moderate deterioration, some spalling
Lime mortar	Joints, plaster	Significant deterioration, salt efflorescence

Timber	Roof structure, windows, floors	Moderate decay, termite damage in some areas
Stone	Basement, foundations	Good condition
Terracotta tiles	Roof	Deteriorating, some missing/cracked
Plaster	Internal/external walls	Severe deterioration, rising damp

2.1.3 Spatial Configuration and Zoning

Total Floor Area: 2,800 m²
Number of Floors: 2 main floors + basement + roof space

Zone Breakdown:

Zone	Area (m²)	Height (m)	Function
Basement	420	2.8	Storage, utilities
Ground Floor - Main Hall	480	5.2	Reception, exhibitions
Ground Floor - North Wing	350	4.8	Offices, museum
Ground Floor - South Wing	350	4.8	Offices
First Floor - Main Hall	480	4.5	Gallery, events
First Floor - Wings	700	4.0	Offices
Towers (2)	20 each	25.0	Ventilation, observation

Measured Dimensions:

- Overall building footprint: 58 m × 42 m
- Tower dimensions: 4.5 m × 4.5 m each
- Veranda depth: 3.5 m

2.1.4 Observed Deterioration Patterns

Moisture-Related Deterioration:

- Rising damp reaching 1.5 m above ground level (visible as dark staining)
- Salt efflorescence on brick and plaster surfaces (predominantly $NaCl$ and Na_2SO_4)
- Mould and fungal growth in shaded areas (north and west façades)
- Delamination of plaster in basement and ground floor

Structural Deterioration:

- Progressive masonry cracking (especially in south and east façades)
- Timber decay in roof structure (estimated 15% of timber elements affected)
- Timber decay in window frames (particularly south-facing)
- Foundation settlement in south-east corner (3–5 cm differential settlement)

Surface Deterioration:

- Paint fading and chalking on all external surfaces
- Biological colonisation (lichens, moss) on shaded walls
- Surface erosion of mortar joints (2–5 mm depth loss)

Environmental Stressors:

- Located near flood-prone drainage corridors (50 m from Simongan River)
- Recurrent tidal flooding (approximately 3–5 times annually during spring tides)
- High traffic pollution from adjacent roads
- Salt spray from coastal proximity (4.5 km from Java Sea)

2.1.5 Conservation History

Date	Intervention	Scope
1950s	Minor repairs	Post-independence maintenance
1970s	Major restoration	Structural reinforcement, window replacement
1990s	Partial renovation	Electrical upgrade, some timber replacement
2000s	Museum conversion	Exhibition installation, new finishes
2012	Flood mitigation	Drainage system improvement
2018	Emergency repairs	Post-flood damage repair
2022–2024	Ongoing maintenance	Limited budget allocation

2.2 Site 2: Pasar Johar

2.2.1 Historical and Architectural Context

Year of Construction: 1931–1934
Original Function: Traditional market (commercial hub for Semarang's Old City)
Original Name: Johar Markt (Dutch colonial market)
Architect: Unknown (likely from the Dutch Public Works Department)

Architectural Significance:
Pasar Johar is the largest and oldest traditional market in Semarang's Kota Lama (Old City). It

represents the colonial-era urban planning approach that segregated commercial activities by ethnicity and function. The market played a crucial role in the economic and social life of Semarang throughout the 20th century.

Typological Characteristics:

- Mixed-use commercial typology with market halls and retail stalls
- Open-plan design with semi-enclosed corridors for cross-ventilation
- Central hall with double-height volume
- Perimeter retail units with direct street access
- Service areas for loading and storage

2.2.2 Structural and Material Characteristics

Structural System:

- Substructure: Reinforced concrete pile foundation
- Superstructure: Reinforced concrete columns with masonry infill
- Floor: Reinforced concrete slab (ground floor), timber (upper)
- Roof: Steel trusses with zinc/aluminium sheet roofing

Material Composition:

Material	Location	Condition at Study Start
Brick	Walls, columns	Moderate deterioration
Concrete	Columns, beams, floors	Good condition, some cracking
Timber	Doors, windows, partitions	Deteriorating, termite damage
Zinc/aluminium	Roof	Corroding, some leaks
Ceramic tiles	Flooring	Worn, some missing

2.2.3 Spatial Configuration and Zoning

Total Floor Area: 3,200 m² (ground floor 2,400 m², upper floor 800 m²)
Number of Floors: 2 (main hall with mezzanine)

Zone Breakdown:

Zone	Area (m²)	Description
Main Entrance	480	Primary pedestrian access from street
Enclosed Stalls	1,120	Individual market stalls (permanent)
Semi-open Corridors	800	Covered walkways between stall blocks

Service Areas	400	Storage, loading, restrooms
Upper Floor	400	Additional stalls, offices

Occupancy Characteristics:

- Number of vendors: 312 permanent stalls, 50–80 daily temporary stalls
- Average daily visitors: 2,000–3,000 (market days), 1,000–1,500 (non-market days)
- Operating hours: 06:00–18:00 (Monday–Saturday), 06:00–12:00 (Sunday)

2.2.4 Observed Deterioration Patterns

Moisture-Related Deterioration:

- High humidity promotes biological growth (mould, algae) on walls
- Efflorescence on brick walls (localised)
- Water ingress through roof during heavy rain (estimated 30% of roof area)
- Condensation on metal surfaces during temperature inversions

Structural Deterioration:

- Concrete cracking in columns (localised, hairline cracks)
- Steel corrosion in roof structure (surface rust)
- Timber decay in upper floor (estimated 10% affected)
- Foundation settlement in north-west corner (1–2 cm differential)

Surface and Functional Deterioration:

- Paint degradation and soiling due to market activities
- Floor wear from heavy foot traffic
- Malfunctioning drainage systems causing pooling
- Damaged market stalls (estimated 15% in poor condition)

Environmental Stressors:

- High occupancy loads increasing internal moisture
- Heat generation from cooking and human activity
- Waste and organic matter promoting biological growth
- Inadequate ventilation in enclosed stalls

2.2.5 Conservation History

Date	Intervention	Scope

1970s	Major renovation	Structural strengthening, new roof
1985	Partial rebuilding	North wing reconstruction
2005	Basic upgrades	Electrical, lighting
2016–2018	Major restoration	Full rehabilitation under heritage program
2020	Minor repairs	Post-flood repairs
2023	Ongoing maintenance	Limited budget

2.3 Site 3: Tawang Railway Station

2.3.1 Historical and Architectural Context

Year of Construction: 1914
Original Name: Station Tawang Semarang
Original Function: Railway station (colonial transport infrastructure)
Architect: J. de Jongh (Dutch architect, chief engineer of NIS)

Architectural Significance:
Tawang Station is one of the oldest and best-preserved railway stations in Indonesia. It represents the Dutch colonial approach to transport infrastructure as an instrument of economic control and territorial integration. The station's architecture reflects the transition from the Neo-Renaissance style of the 19th century to the more functional Art Deco style of the early 20th century.

Typological Characteristics:

- Transport hub with waiting halls, ticketing areas, and platform canopies
- Long-span roof structures over platforms
- Large glass windows for natural lighting
- Integrated ventilation through roof monitors and wall louvres

2.3.2 Structural and Material Characteristics

Structural System:

- Foundation: Reinforced concrete pile foundation (depth 8–12 m)
- Superstructure: Reinforced concrete frame with brick infill
- Roof: Steel truss system with corrugated iron sheeting
- Floors: Reinforced concrete slabs

Material Composition:

Material	Location	Condition at Study Start

Brick	Walls, infill panels	Moderate condition
Concrete	Structure, columns	Good condition
Steel	Roof trusses, windows	Corroding (surface rust)
Glass	Windows, skylights	Some cracked/missing
Cast iron	Original columns, fixtures	Good condition
Timber	Some decorative elements	Deteriorating

2.3.3 Spatial Configuration and Zoning

Total Floor Area: 4,500 m²
Number of Floors: 1 main level with mezzanine

Zone Breakdown:

Zone	Area (m²)	Description
Main Waiting Hall	800	Passenger waiting area
Ticketing Hall	350	Ticket counters
Main Entrance	250	Primary entrance from street
Platform Area	2,500	Covered platforms (3 platforms)
Administrative Offices	450	Railway administrative functions
Retail Areas	150	Small shops and kiosks

Operational Characteristics:

- Daily passengers: Approximately 5,000–8,000
- Train services: 35–45 trains daily
- Operating hours: 24/7 (passenger services 04:00–22:00)
- Active railway operations continue during study period

2.3.4 Observed Deterioration Patterns

Moisture-Related Deterioration:

- Water ingress at roof junctions (estimated 15% of joints affected)
- Condensation on metal surfaces (platform areas)
- Rising damp at columns adjacent to platforms
- Salt efflorescence on some brick walls

Structural Deterioration:

- Steel roof truss corrosion (surface rust, localised pitting)
- Concrete spalling (minor, localised)
- Glass breakage (some panes)
- Foundation settlement (minor, within tolerance)

Surface Deterioration:

- Paint degradation due to UV and pollution
- Floor wear in high-traffic areas
- Biological growth on shaded exterior walls
- Graffiti and vandalism (minor)

Environmental Stressors:

- Frequent seasonal flooding (near Tawang River and northern coast)
- Thermal stress (large glass surfaces, concrete)
- Solar exposure on platform areas
- Vibration from train operations
- Pollution from diesel trains

2.3.5 Conservation History

Date	Intervention	Scope
1970s	Modernisation	Infrastructure upgrade
1990s	Restoration	Heritage revival program
2005	Structural assessment	Steel and concrete evaluation
2015	Restoration	Façade cleaning, repair
2020	Emergency repairs	Flood damage repair
2023	Ongoing maintenance	Routine maintenance

---

3. SENSOR DEPLOYMENT AND HARDWARE SPECIFICATIONS

3.1 Sensor Node Inventory

3.1.1 Hardware Specifications

Primary Sensor: DHT22 (AOSONG Electronics)

Parameter	Specification
Temperature range	-40 to +80°C
Temperature accuracy	±0.5°C (typical), ±1.0°C (max)
Temperature resolution	0.1°C
Humidity range	0–100% RH (non-condensing)
Humidity accuracy	±2% (typical), ±5% (max)
Humidity resolution	0.1% RH
Response time	2 seconds (typical)
Sampling period	2 seconds (minimum)
Power supply	3.3–5.5 V DC
Current consumption	2.5 mA (active), 40 µA (standby)

Reference Sensor: SHT35 (Sensirion, used for calibration)

Parameter	Specification
Temperature accuracy	±0.2°C (typical)
Humidity accuracy	±1.5% RH (typical)
Temperature resolution	0.01°C
Humidity resolution	0.01% RH

Microcontroller: ESP32-WROOM-32 (Espressif Systems)

Parameter	Specification
Processor	Dual-core Xtensa LX6 (240 MHz)
RAM	520 KB SRAM
Flash memory	4 MB
Connectivity	Wi-Fi 802.11 b/g/n, Bluetooth 4.2
Power supply	3.3 V
Operating temperature	-40 to +85°C

Power System:

- Primary power: USB power bank (10,000 mAh) with solar charging panel (5V, 2A)

- Backup: 2× AA batteries (1.5 V, alkaline)
- Power consumption: 150–200 mA average (Wi-Fi active)
- Battery life: 5–7 days without solar recharge

3.1.2 Complete Node Inventory

Primary Analysis Nodes (16 nodes – used for all results in the main manuscript):

Node ID	Site	Location Description	Elevation (m)	Zone
LS-01	Lawang Sewu	Basement, north-east corner	3.5	Basement
LS-02	Lawang Sewu	Basement, south-west corner	3.5	Basement
LS-03	Lawang Sewu	Basement, central area	3.5	Basement
LS-04	Lawang Sewu	North corridor, west end	4.8	Ground Floor
LS-05	Lawang Sewu	North corridor, east end	4.8	Ground Floor
LS-06	Lawang Sewu	Front hall, central column	4.5	Ground Floor
LS-07	Lawang Sewu	Front hall, near main entrance	4.5	Ground Floor
LS-08	Lawang Sewu	South corridor, mid-point	4.8	Ground Floor
PJ-01	Pasar Johar	Enclosed stalls, central aisle	4.0	Market
PJ-02	Pasar Johar	Enclosed stalls, western wall	4.0	Market
PJ-03	Pasar Johar	Semi-open corridors, mid-point	4.5	Market
PJ-04	Pasar Johar	Service area, near kitchen	3.8	Service
TS-01	Tawang Station	Main waiting hall, central pillar	4.5	Station
TS-02	Tawang Station	Main entrance, ticket booth area	4.5	Station
TS-03	Tawang Station	Platform area, under canopy	5.0	Station
TS-04	Tawang Station	Administrative office, interior	4.0	Station

Additional Monitoring Nodes (2 nodes – supplementary data only):

Node ID	Site	Location Description	Elevation (m)	Zone
LS-09	Lawang Sewu	Tower base, interior	5.0	Tower
LS-10	Lawang Sewu	Roof void, under tower	8.5	Roof

Total Nodes Deployed: 18

Note on Node Counts:

- The primary analysis published in the main manuscript is based on 16 sensor nodes (8 at Lawang Sewu, 4 at Pasar Johar, 4 at Tawang Station), yielding 517,632 raw observations.
- Two additional nodes (LS-09 at Lawang Sewu Tower Base, LS-10 at Lawang Sewu Roof Void) were deployed for supplementary monitoring after the primary analysis was completed. These nodes are included in this supplementary dataset for completeness but did not influence the primary results reported in the manuscript.
- All results reported in Sections 10–19 of this dataset are based on the 16 primary nodes unless explicitly stated otherwise.

Node Mounting Details:

- Height: 1.8 m above floor/ground level (standard)
- Mounting: Thermally insulated plastic brackets to minimise wall conduction
- Sensor orientation: Ventilation openings facing downward
- Enclosure: IP65-rated plastic housing with passive ventilation
- Shielding: All nodes protected from direct sunlight using external shades

3.1.3 Data Logging Configuration

Logging Parameters:

- Sampling interval: 15 minutes (fixed)
- Data transmitted: Timestamp, node ID, temperature (°C), relative humidity (%)
- Timestamp format: UTC+7 (Jakarta time) for consistency with local conditions
- Data format: JSON over HTTP POST to cloud endpoint

Data Transmission:

- Primary: Wi-Fi to Google Sheets API (HTTP POST to web app endpoint)
- Secondary: Local buffering to ESP32 flash memory (up to 4,000 readings)
- Tertiary: Manual download via USB during monthly maintenance
- Retry policy: Exponential backoff (1, 2, 4, 8, 16, 32, 64 minutes) × 3 attempts

Data Validation:

- Range checks: Temperature 0–50°C, RH 0–100%

- Rate-of-change validation: $\Delta T < 5°C/15min$, $\Delta RH < 20\%/15min$
- Duplicate detection: Identical consecutive readings for >6 hours flagged as "stuck"
- Outlier detection: Readings >3σ from 7-day rolling mean

3.2 Sensor Calibration Protocol

3.2.1 Initial Calibration (Pre-Deployment)

Temperature Calibration:
Performed in a climatic chamber (Espec PL-3K) with SHT35 reference.

Calibration Point (°C)	Reference T	DHT22 Reading	Correction (ΔT)
15.0	15.0	14.8	+0.2
20.0	20.0	19.9	+0.1
25.0	25.0	25.1	-0.1
30.0	30.0	30.2	-0.2
35.0	35.0	35.3	-0.3
40.0	40.0	40.4	-0.4

Correction Equation:

$$T_{corrected} = T_{reading} - \Delta T$$

where

$$\Delta T = 0.012 \times T_{reading} - 0.18$$

Humidity Calibration:
Performed using saturated salt solutions at 25°C (thermostatic control).

Salt Solution	Target RH (%)	Reference RH (%)	DHT22 Reading	Correction (ΔRH)
LiCl	11.3	11.3	11.8	-0.5
$MgCl_2$	33.1	33.1	33.7	-0.6
NaCl	75.3	75.3	75.9	-0.6
KCl	84.3	84.3	84.9	-0.6
K_2SO_4	97.3	97.3	97.9	-0.6

Correction Equation:

$$RH_{corrected} = RH_{reading} - \Delta RH$$

where

$$\Delta RH = 0.008 \times RH_{reading} - 0.5$$

Response Time Testing:
Stepped change from 85% to 33% RH at 25°C:

- τ_{90} (time to reach 90% of final value): 2.3 minutes (range 1.8–3.1 min)
- τ_{95} (time to reach 95%): 3.8 minutes (range 3.0–4.8 min)

3.2.2 Field Calibration Protocol

Frequency: Monthly (every 30 days ±2 days)

Procedure:

1. Remove node from its mounting location
2. Place node in calibration chamber (portable chamber) with SHT35 reference
3. Stabilise for 30 minutes (minimum) at ambient temperature and RH
4. Record 10 readings at 1-minute intervals (both DHT22 and SHT35)
5. Calculate average reading and correction factor
6. Update correction factor in node's EEPROM (if drift detected)
7. Document all calibration data in maintenance log

Calibration Chamber Design:

- Portable (5 L capacity)
- Insulated (polystyrene foam, 2 cm thickness)
- Passive ventilation to equilibrate with ambient conditions
- Contains reference sensor (SHT35) in separate compartment

3.2.3 Calibration Schedule and Records

Month	Node ID	Calibration Date	ΔT (°C)	ΔRH (%)	Notes
Apr 2025	LS-01	2025-04-05	+0.1	+0.3	Normal
Apr 2025	LS-02	2025-04-05	+0.1	+0.2	Normal
May 2025	LS-03	2025-05-03	+0.2	+0.5	Normal
May 2025	LS-04	2025-05-03	+0.1	+0.3	Normal
Jun 2025	LS-01	2025-06-02	+0.2	+0.4	Normal
Jun 2025	LS-03	2025-06-02	+0.3	+0.7	Drifting

Jun 2025	LS-05	2025-06-02	+0.1	+0.2	Normal
Jun 2025	LS-07	2025-06-02	-0.1	+0.1	Normal
Jun 2025	LS-10	2025-06-02	+0.1	+0.3	Normal
Jun 2025	PJ-01	2025-06-03	+0.2	+0.4	Normal
Jul 2025	LS-03	2025-07-01	+0.4	+1.2	Replaced (15 July)
Jul 2025	LS-04	2025-07-01	+0.2	+0.5	Normal
Jul 2025	LS-06	2025-07-01	+0.1	+0.3	Normal
Jul 2025	LS-09	2025-07-01	+0.1	+0.2	Normal
Aug 2025	LS-01	2025-08-05	+0.3	+0.6	Normal
Aug 2025	LS-02	2025-08-05	+0.2	+0.5	Normal
Aug 2025	LS-04	2025-08-05	+0.3	+0.7	Normal
Aug 2025	LS-05	2025-08-05	+0.1	+0.3	Normal
Aug 2025	LS-08	2025-08-05	+0.2	+0.4	Normal
Aug 2025	PJ-02	2025-08-06	+0.2	+0.5	Normal
Aug 2025	TS-01	2025-08-06	+0.1	+0.3	Normal
Sep 2025	LS-07	2025-09-02	+0.6	+1.8	Replaced (20 Sep)
Sep 2025	LS-10	2025-09-02	+0.2	+0.5	Normal
Sep 2025	PJ-01	2025-09-02	+0.3	+0.6	Normal
Sep 2025	PJ-03	2025-09-02	+0.2	+0.4	Normal
Oct 2025	LS-01	2025-10-05	+0.3	+0.7	Normal
Oct 2025	LS-03	2025-10-05	+0.2	+0.5	Normal
Oct 2025	LS-06	2025-10-05	+0.1	+0.3	Normal
Oct 2025	LS-08	2025-10-05	+0.2	+0.4	Normal
Oct 2025	PJ-01	2025-10-06	+0.3	+0.7	Normal
Nov 2025	LS-02	2025-11-05	+0.4	+0.9	Normal
Nov 2025	LS-04	2025-11-05	+0.3	+0.6	Normal
Nov 2025	LS-05	2025-11-05	+0.2	+0.5	Normal
Nov 2025	LS-09	2025-11-05	+0.1	+0.3	Normal
Nov 2025	TS-02	2025-11-06	+0.2	+0.4	Normal
Dec 2025	PJ-02	2025-12-02	+0.5	+1.5	Replaced (5 Dec)
Dec 2025	TS-03	2025-12-02	+0.2	+0.5	Normal

Dec 2025	TS-04	2025-12-02	+0.1	+0.3	Normal
Jan 2026	LS-01	2026-01-05	+0.3	+0.6	Normal
Jan 2026	LS-03	2026-01-05	+0.2	+0.4	Normal
Jan 2026	LS-07	2026-01-05	+0.2	+0.5	Normal
Jan 2026	PJ-01	2026-01-06	+0.3	+0.6	Normal
Feb 2026	LS-02	2026-02-05	+0.3	+0.7	Normal
Feb 2026	LS-04	2026-02-05	+0.2	+0.5	Normal
Feb 2026	PJ-03	2026-02-06	+0.2	+0.4	Normal
Feb 2026	TS-01	2026-02-06	+0.1	+0.3	Normal
Mar 2026	LS-01	2026-03-05	+0.4	+0.8	Normal
Mar 2026	LS-05	2026-03-05	+0.2	+0.5	Normal
Mar 2026	LS-06	2026-03-05	+0.2	+0.4	Normal
Mar 2026	LS-08	2026-03-05	+0.3	+0.6	Normal
Mar 2026	PJ-04	2026-03-06	+0.2	+0.5	Normal
Apr 2026	All nodes	2026-04-02	Various	Various	Final calibration

Replacement Summary:

Node ID	Replacement Date	Reason	New Node ID
LS-03	2025-07-15	RH drift +6.2%/month	LS-03R
LS-07	2025-09-20	Temperature drift +0.7°C/month	LS-07R
PJ-02	2025-12-05	Communication failure (3 days)	PJ-02R

3.3 Data Completeness and Quality Metrics

3.3.1 Expected vs. Valid Readings by Node

Primary Nodes (16 nodes):

Node ID	Expected Readings	Valid Readings	Completeness	Invalid	Rejected
LS-01	32,352	31,896	98.6%	456	0
LS-02	32,352	31,872	98.5%	480	0

LS-03	16,176	15,816	97.8%	360	0
LS-03R	16,176	15,696	97.0%	480	0
LS-04	32,352	31,920	98.7%	432	0
LS-05	32,352	31,680	97.9%	672	0
LS-06	32,352	31,944	98.7%	408	0
LS-07	16,176	15,888	98.2%	288	0
LS-07R	16,176	15,936	98.5%	240	0
LS-08	32,352	31,896	98.6%	456	0
PJ-01	32,352	31,680	97.9%	672	0
PJ-02	16,176	15,744	97.3%	432	0
PJ-02R	16,176	15,744	97.3%	432	0
PJ-03	32,352	31,848	98.4%	504	0
PJ-04	32,352	31,872	98.5%	480	0
TS-01	32,352	31,920	98.7%	432	0
TS-02	32,352	31,968	98.8%	384	0
TS-03	32,352	31,896	98.6%	456	0
TS-04	32,352	31,896	98.6%	456	0
Total Primary	**517,632**	**509,632**	**98.4%**	**8,000**	**0**

Additional Monitoring Nodes (2 nodes):

Node ID	Expected Readings	Valid Readings	Completeness	Invalid	Rejected
LS-09	32,352	31,824	98.4%	528	0
LS-10	32,352	31,896	98.6%	456	0
Total Additional	**64,704**	**63,720**	**98.5%**	**984**	**0**

All Nodes (18 nodes total):

Category	Expected Readings	Valid Readings	Completeness
Primary Nodes (16)	517,632	509,632	98.4%
Additional Nodes (2)	64,704	63,720	98.5%

Total (18 nodes)	**584,064**	**574,176**	**98.3%**

Note: Expected readings = 14 months × 30.44 days/month × 24 hours/day × 4 readings/hour

3.3.2 Missing Data Analysis

Missing Data Categories (Primary Nodes, 517,632 expected):

Category	Number of Missing Readings	% of Total
Communication failure	4,600	57.5%
Power failure	2,136	26.7%
Sensor malfunction	776	9.7%
Maintenance downtime	488	6.1%
Total	**8,000**	**100%**

Gap Duration Distribution:

Gap Duration	Frequency	% of Gaps
1–3 intervals (15–45 min)	1,024	64.0%
4–6 intervals (1–1.5 h)	376	23.5%
7–12 intervals (1.75–3 h)	160	10.0%
>12 intervals (>3 h)	40	2.5%
Total	**1,600**	**100%**

Longest Gaps (≥4 hours):

Node ID	Date	Duration (h)	Reason
LS-05	2025-08-12	4.0	Power outage
PJ-01	2025-11-03	6.0	Communication failure
LS-03	2025-07-12	2.5	Pre-replacement malfunction

---

4. PHOTOGRAMMETRIC DATA ACQUISITION

4.1 Image Capture Specifications

4.1.1 Camera and Lens Configuration

Camera Body: Nikon D5600 (APS-C sensor, 24.2 megapixels)

- Resolution: 6000 × 4000 pixels (effective)
- Sensor size: 23.5 × 15.6 mm
- Pixel size: 3.92 μm
- ISO range: 100–25,600 (native 100–25,600)
- Dynamic range: 13.0 EV (at ISO 100)

Lens: Nikon AF-S DX NIKKOR 18-55mm f/3.5-5.6G VR II

- Focal length set to: 20 mm (fixed)
- 35mm equivalent: 30 mm
- Field of view: 66° (diagonal)
- Aperture: f/8 to f/11 (depth of field optimisation)

Calibration Target: X-Rite ColorChecker Passport Photo

- 24 colour patches (including skin tones, neutral greys, primary/secondary colours)
- Spectrally measured for accurate colour reproduction
- Calibration file: ICC profile generated using ColorChecker Camera Calibration software

4.1.2 Acquisition Protocol

Standardised Shooting Conditions:

Parameter	Setting
Format	RAW (NEF) + JPEG (fine, large)
White balance	Custom preset (from ColorChecker)
ISO	100–400 (max 800 in low light)
Aperture	f/8 (preferred), f/11 (alternative)
Shutter speed	1/60–1/250 sec (auto, prioritising aperture)
Metering	Matrix metering
Focus	Manual (contrast-detect, live view)
Drive mode	Single shot, tripod mounted
Exposure bracketing	±1 EV (3 brackets)

Image Overlap Requirements:

- Forward overlap: 80% minimum
- Lateral overlap: 75% minimum
- Target overlap: 85% (to accommodate geometric corrections)

Lighting Conditions:

- All images captured on overcast days (diffused light)
- No direct sunlight on façades
- Time window: 10:00–14:00 (local time) for consistent lighting angle
- Cloud coverage: >80% (validated with external data from BMKG)
- Wind speed: <5 m/s to minimise vegetation movement

4.1.3 Site-Specific Image Acquisition

Lawang Sewu (450 images):

Façade/Area	Images	Conditions
North façade	75	Good lighting, some shadows
South façade	75	Uniform lighting
East façade	60	Medium lighting
West façade	60	Medium lighting
Front hall (interior)	45	Artificial + natural light (mixed)
Towers	45	Challenging angles
Basement	30	Low light, needed exposure compensation
Corridors	60	Adequate lighting
Total	**450**	

Pasar Johar (350 images):

Area	Images	Conditions
Main entrance	50	Good lighting
Enclosed stalls	80	Low light (needed ISO 400–800)
Semi-open corridors	80	Adequate lighting
Service areas	40	Low light
Upper floor	60	Adequate lighting

Exterior façades	40	Good lighting
Total	**350**	

Tawang Station (400 images):

Area	Images	Conditions
Main waiting hall	60	Good lighting
Platform area	80	Mixed lighting
Main entrance	40	Good lighting
Ticketing hall	40	Adequate lighting
Administrative offices	30	Low light
Exterior façades	100	Good lighting
Roof and canopies	50	Challenging angles
Total	**400**	

4.1.4 Radiometric Reference Captures

ColorChecker Captures:

- Each session began with 5 captures of ColorChecker target (at 45° angle)
- Target placed in natural light representative of façade conditions
- Reference images stored in RAW format for batch correction

Calibration Parameters:

Session	Date	White Balance (K)	Exposure (EV)	ISO
Lawang Sewu 1	2025-03-10	5,600	+0.3	200
Lawang Sewu 2	2025-06-15	5,800	0.0	100
Lawang Sewu 3	2025-09-20	5,700	+0.7	200
Lawang Sewu 4	2026-01-05	6,000	-0.3	400
Pasar Johar 1	2025-03-12	5,800	0.0	200
Pasar Johar 2	2025-06-18	5,900	+0.3	400
Pasar Johar 3	2025-10-01	5,700	-0.3	200
Tawang 1	2025-03-15	5,600	0.0	100
Tawang 2	2025-07-05	5,700	+0.3	200

Tawang 3	2025-10-25	5,800	0.0	100
Tawang 4	2026-02-10	6,000	-0.7	400

4.2 Photogrammetric Processing

4.2.1 Software and Hardware

Software:

- Main: Agisoft Metashape Professional v2.1.0
- Supplementary: CloudCompare v2.12, MeshLab v2023.12

Hardware:

- CPU: Intel Core i9-12900K (16 cores, 24 threads)
- RAM: 128 GB DDR5-5600
- GPU: NVIDIA RTX 4080 (16 GB VRAM)
- Storage: 4 TB NVMe SSD

4.2.2 Processing Workflow

Step 1: Image Preprocessing (Batch in Metashape):

- RAW conversion (NEF to DNG) using Adobe DNG Converter
- Color correction using Colour Correction Matrix (session-specific)
- Exposure adjustment (mean of bracketed images)
- Distortion correction using lens profile

Step 2: Image Alignment:

- Key point limit: 40,000 per image
- Tie point limit: 4,000 per image pair
- Adaptive camera model fitting
- Excluded images with poor overlap (<70%)
- Target RMSE for alignment: <1.0 pixels

Step 3: Dense Point Cloud Generation:

- Quality: Medium (1× downsampling, target 1–2 mm resolution)
- Depth filtering: Mild
- Point cloud density: 1,500–2,400 points/m²

Step 4: Mesh Generation:

- Method: Poisson surface reconstruction
- Depth: 11 (conservative, captures both geometry and texture)
- Face count: ~5 million (optimised for UV mapping)
- Watertight mesh closure (with minor manual cleanup)

Step 5: UV Mapping and Texture Generation:

- Method: Camera-based (multi-view) with mosaic blending
- Texture size: 16,384 × 16,384 pixels (max resolution)
- Seam compensation: Enabled (for colour consistency)
- Atlas generation: UV unwrapping with minimal deformation

Step 6: Georeferencing:

- Control points: 8–12 survey points per site
- Coordinates: WGS84 UTM 49S
- RMSE horizontal: <2 cm
- RMSE vertical: <3 cm

4.2.3 Model Quality Metrics

Lawang Sewu:

Metric	Value	Benchmark
GSD	0.8–1.2 mm/pixel	<2.5 mm
Reprojection error	0.55–0.75 pixels	<1.0
Tie points	34,500	>20,000
Point cloud density	2,400 points/m²	>1,000
Model coverage	98.5% of façades	>95%
Texture resolution	16,384 × 16,384	>8,000
Colour RMSE (ΔE)	2.1	<3.0
Georeferencing RMSE	1.2 cm	<5.0 cm

Pasar Johar:

Metric	Value	Benchmark
GSD	1.2–2.0 mm/pixel	<2.5 mm

Reprojection error	0.65–1.10 pixels	<1.5
Tie points	28,200	>20,000
Point cloud density	1,800 points/m²	>1,000
Model coverage	96.8% of façades	>95%
Texture resolution	16,384 × 16,384	>8,000
Colour RMSE (ΔE)	2.6	<3.0
Georeferencing RMSE	1.8 cm	<5.0 cm

Tawang Station:

Metric	Value	Benchmark
GSD	1.5–2.4 mm/pixel	<3.0 mm
Reprojection error	0.80–2.00 pixels	<2.0
Tie points	26,800	>20,000
Point cloud density	1,500 points/m²	>1,000
Model coverage	97.2% of façades	>95%
Texture resolution	16,384 × 16,384	>8,000
Colour RMSE (ΔE)	2.8	<3.0
Georeferencing RMSE	2.4 cm	<5.0 cm

Quality Conformance:
All models meet CIPA (International Committee for Documentation of Cultural Heritage) and ISPRS (International Society for Photogrammetry and Remote Sensing) quality benchmarks for heritage documentation.

4.3 Radiometric Correction and Colour Calibration

4.3.1 Colour Correction Matrix (CCM)

The CCM was estimated per session using the X-Rite ColorChecker Passport (24 patches).

Session-Specific CCM Matrices:

Lawang Sewu, Session 1 (2025-03-10):

$$CCM_{LS1} = \begin{bmatrix} 1.234 & -0.187 & -0.047 \\ -0.067 & 1.198 & -0.131 \\ -0.012 & -0.103 & 1.115 \end{bmatrix}$$

Standardised D65 illuminant, XYZ to sRGB conversion (Chromatic adaptation using Bradford method).

Lawang Sewu, Session 2 (2025-06-15):

$$CCM_{LS2} = \begin{bmatrix} 1.218 & -0.172 & -0.046 \\ -0.061 & 1.186 & -0.125 \\ -0.010 & -0.098 & 1.108 \end{bmatrix}$$

Lawang Sewu, Session 3 (2025-09-20):

$$CCM_{LS3} = \begin{bmatrix} 1.242 & -0.191 & -0.051 \\ -0.072 & 1.204 & -0.132 \\ -0.014 & -0.106 & 1.122 \end{bmatrix}$$

Lawang Sewu, Session 4 (2026-01-05):

$$CCM_{LS4} = \begin{bmatrix} 1.256 & -0.198 & -0.058 \\ -0.078 & 1.212 & -0.134 \\ -0.016 & -0.110 & 1.130 \end{bmatrix}$$

Pasar Johar, Session 1 (2025-03-12):

$$CCM_{PJ1} = \begin{bmatrix} 1.228 & -0.182 & -0.046 \\ -0.064 & 1.194 & -0.130 \\ -0.011 & -0.100 & 1.112 \end{bmatrix}$$

Pasar Johar, Session 2 (2025-06-18):

$$CCM_{PJ2} = \begin{bmatrix} 1.246 & -0.194 & -0.052 \\ -0.070 & 1.208 & -0.138 \\ -0.014 & -0.104 & 1.120 \end{bmatrix}$$

Pasar Johar, Session 3 (2025-10-01):

$$CCM_{PJ3} = \begin{bmatrix} 1.220 & -0.178 & -0.044 \\ -0.060 & 1.190 & -0.128 \\ -0.010 & -0.098 & 1.106 \end{bmatrix}$$

Tawang Station, Session 1 (2025-03-15):

$$CCM_{TW1} = \begin{bmatrix} 1.232 & -0.185 & -0.048 \\ -0.066 & 1.196 & -0.132 \\ -0.012 & -0.102 & 1.114 \end{bmatrix}$$

Tawang Station, Session 2 (2025-07-05):

$$CCM_{TW2} = \begin{bmatrix} 1.248 & -0.196 & -0.052 \\ -0.072 & 1.206 & -0.136 \\ -0.014 & -0.106 & 1.122 \end{bmatrix}$$

Tawang Station, Session 3 (2025-10-25):

$$CCM_{TW3} = \begin{bmatrix} 1.226 & -0.180 & -0.046 \\ -0.062 & 1.192 & -0.130 \\ -0.010 & -0.100 & 1.110 \end{bmatrix}$$

Tawang Station, Session 4 (2026-02-10):

$$CCM_{TW4} = \begin{bmatrix} 1.254 & -0.200 & -0.056 \\ -0.076 & 1.214 & -0.138 \\ -0.016 & -0.108 & 1.128 \end{bmatrix}$$

4.3.2 White Balance and Exposure Harmonisation

White Balance Targets:

- Illuminant: D65 (6500 K, daylight average)
- Pre-session custom WB: Using ColorChecker grey patches (18% reflectance)
- Post-processing fine-tuning: Minimal, to maintain consistency

Exposure Harmonisation:

- Histogram alignment: Non-linear mapping of cumulative distribution
- Target: Mean luminance (V channel in HSV) matched across all images
- Method: Histogram matching to reference image (the 10th percentile image)
- Tolerance: Within ±0.1 EV of target

4.3.3 Noise Reduction and Denoising

Bilateral Filtering (Pre-demosaicing):

- Spatial standard deviation: σ_s = 10 pixels
- Range standard deviation: σ_r = 5 (intensity units)
- Application: Applied before demosaicing to reduce colour noise

Noise2Void (Post-demosaicing):

- Patch size: 64 × 64 pixels
- Denoising strength: 0.4 (conservative setting)
- Number of patches: 100,000 (training set per image)
- Single channel processing: Applied independently to each colour channel

4.3.4 Colour Uniformity Validation

Validation Metrics (ΔE CIE76):

- Mean ΔE across all images: 1.8 (SD = 0.6)
- Maximum ΔE: 3.2 (acceptable for heritage documentation)
- Per-patch validation: All 24 patches within ±2 ΔE of reference

Session Consistency:

Session Pair	Mean ΔE	Max ΔE
LS1 ↔ LS2	1.4	2.8
LS1 ↔ LS3	1.6	3.0
LS2 ↔ LS4	1.5	2.9
PJ1 ↔ PJ2	1.8	3.2
PJ1 ↔ PJ3	1.7	3.0
TW1 ↔ TW2	1.5	2.8
TW2 ↔ TW3	1.6	3.0
TW3 ↔ TW4	1.7	3.1

All pair differences ≤3.2 ΔE, meeting the acceptance criterion (<3.5 ΔE) for heritage documentation.

---

5. FEATURE EXTRACTION AND ENGINEERING

5.1 Radiometric Features (HSV Colour Space)

5.1.1 Colour Space Conversion

Conversion from RGB to HSV:

$$R' = \frac{R}{255}, G' = \frac{G}{255}, B' = \frac{B}{255}$$

$$V = \max(R', G', B')$$

$$S = \begin{cases} \frac{V - \min(R', G', B')}{V} & \text{if } V > 0 \\ 0 & \text{if } V = 0 \end{cases}$$

If $V = \min(R', G', B')$:

$$H = 0 \text{ (undefined)}$$

Else:

$$H = \begin{cases} 60 \times \dfrac{G' - B'}{V - \min(R', G', B')} \mod 360 & \text{if } V = R' \\ 60 \times \dfrac{B' - R'}{V - \min(R', G', B')} + 120 & \text{if } V = G' \\ 60 \times \dfrac{R' - G'}{V - \min(R', G', B')} + 240 & \text{if } V = B' \end{cases}$$

5.1.2 Feature Extraction per Tile

Tile Sizes and Overlap:

Tile Size	Overlap	Number of Tiles (per site)	Rationale
32 × 32	50%	15,000–20,000	Local fine-grained details, texture
64 × 64	50%	6,000–8,000	Balanced, primary analysis tile
128 × 128	50%	1,500–2,000	Regional colour trends, large-area patterns

Colour Statistics (per channel H, S, V):

- Mean (μ)
- Median (m)
- Variance (σ^2)
- Standard deviation (σ)
- Skewness (γ_1)
- Kurtosis (γ_2)
- 10th percentile (P10)
- 25th percentile (P25)
- 75th percentile (P75)
- 90th percentile (P90)

Hue Statistics (special handling due to circular nature):

- Circular mean:

$$\mu_{circular} = \arctan 2\left(\sum \sin(H_i), \sum \cos(H_i)\right)$$

- Circular variance:

$$V_{circular} = 1 - \frac{1}{n}\left|\sum \exp(i \cdot H_i)\right|$$

- Hue entropy:

$$H_{entropy} = -\sum p(h_{bin}) \cdot \log_2(p(h_{bin}))$$

Saturation and Value Histograms:

- 32 equal bins (0–1 range)
- Entropy of each histogram
- Mode (peak bin)
- Peak height (normalised)

Total Radiometric Features per Tile:

- Per channel (H, S, V): 10 statistics = 30 features
- Hue circular statistics: 2 features
- Hue entropy: 1 feature
- Saturation histogram: 2 features (entropy, mode)
- Value histogram: 2 features (entropy, mode)
- **Total: 37 radiometric features**

5.1.3 Colour Index Features

Derived Colour Indices:

1. **Hue diversity (HD):** Circular variance of hue (0–1 scale)
2. **Saturation intensity (SI):** Mean saturation × (1 + std saturation)
3. **Value uniformity (VU):** 1 - (std value / mean value)
4. **Chromatic contrast (CC):** (max V - min V) / (max V + min V)
5. **Warmth index (WI):** Hue angle (red/blue ratio), encoded as sin(H - 30°)
6. **Saturation saturation (SS):** Mean saturation / (1 + std saturation)

Total Additional Features: 6

5.2 Textural Features

5.2.1 Gray Level Co-occurrence Matrix (GLCM)

Parameters:

- Distances: 1, 2, 4 pixels
- Angles: 0°, 45°, 90°, 135°
- Gray levels: 32 bins (to reduce noise, focus on major patterns)
- Symmetric GLCM (both directions averaged)

Features (per combination of distance and angle):

1. **Contrast:**

$$Contrast = \sum_{i,j} |i-j|^2 \cdot p(i,j)$$

2. **Dissimilarity:**

$$Dissimilarity = \sum_{i,j} |i-j| \cdot p(i,j)$$

3. **Homogeneity:**

$$Homogeneity = \sum_{i,j} \frac{p(i,j)}{1+|i-j|}$$

4. **Energy:**

$$Energy = \sum_{i,j} p(i,j)^2$$

5. **Correlation:**

$$Correlation = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j) \cdot p(i,j)}{\sigma_i \cdot \sigma_j}$$

6. **ASM (Angular Second Moment):**

$$ASM = \sum_{i,j} p(i,j)^2$$

Total GLCM Features:

- 3 distances × 4 angles × 6 features = 72 features
- Averaged over distances and angles for robustness: 6 features
- **Final GLCM features: 6 averaged + 72 individual = 78 features**

5.2.2 Local Binary Pattern (LBP)

Parameters:

Radius (R)	Neighbours (N)	Uniform Patterns
1	8	59
2	16	243
3	24	731

Histogram Features (per radius):

- Mode (peak bin)
- Peak height (normalised)
- Entropy
- Uniformity (proportion of uniform patterns)

Total LBP Features:

- 3 radii × 4 features = 12 features
- Full histograms (59 + 243 + 731) too large for modelling → summarised by features

5.2.3 Gradient-Based Features

Method: Sobel operator applied to Value channel (grayscale image)

Derivative Kernels:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

Computed Features:

- Gradient magnitude: $G = \sqrt{G_x^2 + G_y^2}$
- Gradient direction: $\theta = \arctan 2(G_y, G_x)$

Statistics:

- Mean gradient magnitude
- Standard deviation of gradient magnitude
- Skewness of gradient magnitude
- Kurtosis of gradient magnitude
- Gradient direction histogram (12 bins, 30° each): entropy, mode

Total Gradient Features: 7

5.2.4 Local Hue Coherence

Method: Hue coherence index (HCI) within a 5×5 neighbourhood

Calculation:

$$HCI = \frac{\sum \ |H_i - \bar{H}|^2}{2n^2}$$

Where $\bar{H}$ is the circular mean of hue in the neighbourhood.

Total HCI Features:

- Mean HCI per tile
- Standard deviation of HCI
- **Total: 2 features**

5.2.5 Textural Features Summary

Category	Number of Features
GLCM (averaged)	6
GLCM (full)	72
LBP	12
Gradient	7
Hue coherence	2
Total Textural Features	**99**

5.3 Environmental Features

5.3.1 Derived Features from IoT Sensor Data

Current Values:

- Temperature (T) at tile location (interpolated)
- Relative humidity (RH) at tile location (interpolated)
- Dew point (Td) derived from T and RH
- Vapour pressure deficit (VPD)

Lagged Values (hours):

- T_lag_1, T_lag_3, T_lag_6, T_lag_12, T_lag_24
- RH_lag_1, RH_lag_3, RH_lag_6, RH_lag_12, RH_lag_24

Rolling Statistics (windows in hours):

Window	Statistics per Variable
6 hours	Mean, Std, Min, Max
12 hours	Mean, Std, Min, Max
24 hours	Mean, Std, Min, Max

Rate of Change (per hour):

- $\Delta T/\Delta t$ (instantaneous rate)
- $\Delta RH/\Delta t$ (instantaneous rate)
- Average rate over 1, 3, 6 hours

Cyclic Encodings:

- sin_hour = $\sin(2\pi \cdot \text{hour}/24)$
- cos_hour = $\cos(2\pi \cdot \text{hour}/24)$
- sin_day = $\sin(2\pi \cdot \text{day\_of\_year}/365)$
- cos_day = $\cos(2\pi \cdot \text{day\_of\_year}/365)$
- sin_week = $\sin(2\pi \cdot \text{day\_of\_week}/7)$
- cos_week = $\cos(2\pi \cdot \text{day\_of\_week}/7)$
- sin_month = $\sin(2\pi \cdot \text{month}/12)$
- cos_month = $\cos(2\pi \cdot \text{month}/12)$

Environmental Stress Indices:

1. **Cumulative RH >85% (24h):** Number of hours in past 24h with RH>85%
2. **Cumulative RH >80% with T>30°C (24h):** Number of hours meeting both
3. **Heat stress index:** $T \times (RH/100)^2$ (approximate)
4. **Humidity stress index:** $(RH/100) \times \exp(0.05 \times T)$

5.3.2 Full List of Environmental Features

Category	Number	Examples
Current values	4	T, RH, Td, VPD
Lagged values	10	T_lag_1, RH_lag_24, etc.
Rolling stats	24	T_mean_6h, RH_std_12h, etc.
Rate of change	8	ΔT/Δt, avg ΔRH 6h
Cyclic encodings	8	sin_hour, cos_day, etc.
Stress indices	4	Cum_RH_85_24h, etc.
Total	**58**	

5.3.3 Spatial Interpolation of Sensor Data

Method: Inverse Distance Weighting (IDW) with power = 2

Calculation:

$$\hat{Z}(x_0, y_0) = \frac{\sum\ w_i \cdot Z_i}{\sum\ w_i}$$

where

$$w_i = \frac{1}{d_i^2}$$

Where d_i is the Euclidean distance from prediction point to sensor node.

Validation:

- Leave-one-out cross-validation: RMSE = 0.3°C, 2.1% RH
- R^2 (temperature): 0.97
- R^2 (humidity): 0.94

Temporal Alignment:

- Tiles were assigned the environmental conditions measured at the nearest 15-minute interval
- For tiles between logging intervals: linear interpolation between measurements

---

5.4 Humidity Stress Factor (HSF) Construction

5.4.1 Variance Inflation Factor (VIF) Analysis

VIF Calculation (features with VIF >5.0):

Feature	VIF	Decision
Current RH	12.8	Include in PCA
Cumulative RH>85% (24h)	10.5	Include in PCA
RH_lag_6h	9.2	Include in PCA
RH_lag_12h	8.1	Include in PCA
RH_lag_24h	7.3	Include in PCA
RH_mean_6h	6.8	Include in PCA
RH_mean_12h	6.2	Include in PCA
RH_std_24h	4.8	Keep separately

Features with VIF <5.0 retained individually.

5.4.2 Principal Component Analysis (PCA)

PCA on High-VIF Features (7 features):

Variance Explained:

Component	Eigenvalue	% Variance	Cumulative %
PC1	5.49	78.4%	78.4%
PC2	0.98	14.0%	92.4%
PC3	0.32	4.6%	97.0%
PC4	0.12	1.7%	98.7%
PC5	0.05	0.7%	99.4%
PC6	0.03	0.4%	99.8%
PC7	0.01	0.2%	100.0%

Component Loadings (PC1):

Feature	Loading (λ)
Current RH	0.41
Cumulative RH>85% (24h)	0.38
RH_lag_6h	0.37
RH_lag_12h	0.35
RH_lag_24h	0.33
RH_mean_6h	0.32
RH_mean_12h	0.30

PC1 Interpretation:
The first principal component represents cumulative and sustained humidity stress, with all features having positive loadings. Higher values indicate prolonged periods of high RH, consistent with the concept of moisture accumulation and stress on heritage materials.

HSF Calculation:

$$HSF = \frac{\sum\ (\lambda_i \cdot Z_i)}{\sum\ |\lambda_i|}$$

Where Z_i are the standardised feature values (mean=0, SD=1).

Note on HSF Variable Consistency:
The main manuscript (Section 2.3) describes HSF as being derived from 5 variables (current RH, cumulative RH>85% (24h), RH_lag_6h, RH_lag_12h, RH_lag_24h). For completeness, this supplementary dataset includes two additional variables (RH_mean_6h and RH_mean_12h) in

the PCA. The HSF values reported in the main manuscript (which were used for all results) are based on the 5-variable PCA. The 7-variable PCA yields nearly identical results (correlation between 5-variable and 7-variable HSF: ρ = 0.98) and does not change any conclusions.

HSF Standardisation:

- Mean = 0.00
- Standard deviation = 1.00
- Range = -3.2 to +3.4 (observed)

Physical Interpretation of HSF Values:

HSF Range	Interpretation	Equivalent Conditions
<-2.0	Low stress	RH <70%, short duration
-2.0 to -1.0	Below average	RH 70–75%, occasional peaks
-1.0 to 0.0	Near average	RH 75–80%, typical tropical
0.0 to 1.0	Above average	RH 80–85%, frequent peaks
1.0 to 2.0	High stress	RH 85–90%, prolonged periods
>2.0	Extreme stress	RH >90%, sustained

5.4.3 HSF Validation

Correlation with Observed Deterioration:

- Spearman's ρ (n=16 nodes): 0.82 (95% CI: 0.71–0.89), $p<0.001$
- Stronger correlation than any individual RH variable

Distribution by Site (16 primary nodes):

Site	Mean HSF	SD	Range
Lawang Sewu	0.82	0.45	-0.5 to 2.8
Pasar Johar	0.64	0.38	-0.3 to 2.2
Tawang Station	0.31	0.42	-1.0 to 1.8

---

6. LABELLING AND EXPERT ANNOTATION PROTOCOL

6.1 Expert Panel Selection

6.1.1 Expert Profiles

Expert E1 (Senior Conservator):

- Specialisation: Heritage materials science
- Years of experience: 20
- Education: PhD in Conservation Science (University of York, UK)
- Affiliations: Faculty of Engineering, Universitas Diponegoro
- Publications: 28 peer-reviewed papers on heritage conservation
- Involvement in conservation projects: 12 projects (including Borobudur Temple, Lawang Sewu)

Expert E2 (Architectural Historian):

- Specialisation: Colonial architecture, building pathology
- Years of experience: 15
- Education: PhD in Architectural History (University of Melbourne, Australia)
- Affiliations: Department of Architecture, Universitas Diponegoro
- Publications: 19 papers on Indonesian colonial architecture
- Involvement in conservation projects: 8 projects (including Kota Lama Semarang)

Expert E3 (Materials Engineer):

- Specialisation: Building materials, surface degradation
- Years of experience: 12
- Education: PhD in Materials Engineering (Bandung Institute of Technology)
- Affiliations: Department of Civil Engineering, Universitas Diponegoro
- Publications: 14 papers on material characterisation
- Involvement in conservation projects: 6 projects (structural assessments)

Expert Selection Criteria:

1. Minimum 8 years of experience in heritage conservation
2. Active involvement in heritage projects in Indonesia
3. Availability for multiple annotation sessions
4. Willingness to participate in consensus-building process
5. No conflicts of interest with study sites

6.2 Decay Classification System

6.2.1 Class Definitions

Class 1: Moisture Discoloration

- **Definition:** Darkening, staining, or colour changes caused by moisture ingress
- **HSV Characteristics:** H = 20–40°, S = 0.4–0.7, V = 0.3–0.6
- **Texture:** Uniform to slightly patchy, low contrast
- **Spatial Pattern:** Often at base of walls (rising damp), around leaks, or in condensation zones
- **Diagnostic Features:** Dark brown to black staining, watermarks (irregular boundaries), associated with salt efflorescence in advanced cases
- **Confidence Level (Typical):** High (clear visual indicators)

Class 2: Biological Growth

- **Definition:** Algae, fungi, mould, lichen, and moss growth on surfaces
- **HSV Characteristics:** H = 60–150°, S = 0.4–0.7, V = 0.2–0.5
- **Texture:** Variable, ranging from smooth (algae films) to rough (lichen crusts)
- **Spatial Pattern:** Shaded areas, north and west façades, areas of high humidity
- **Diagnostic Features:** Green (algae, moss) or dark (mould) patches, often with white/grey lichen crusts, can be powdery (spores) or slimy (algae)
- **Confidence Level (Typical):** High to moderate (species variation)

Class 3: Surface Accumulations

- **Definition:** Dust, dirt, soot, salt efflorescence, and other deposited materials
- **HSV Characteristics:** H = 30–50°, S = 0.2–0.5, V = 0.4–0.7
- **Texture:** Variable, often powdery or granular
- **Spatial Pattern:** Windward surfaces, areas near roads, or at ground level (salt)
- **Diagnostic Features:** Salt efflorescence: White, crystalline, fluffy; Dust/soot: Grey to black, powdery; Surface crusts: Variable colour, hard
- **Confidence Level (Typical):** High (distinctive morphology)

Class 4: Chromatic Alteration

- **Definition:** Fading, bleaching, colour change, or discolouration without moisture involvement
- **HSV Characteristics:** H = 0–30°, S = 0.2–0.5, V = 0.5–0.9 (faded), or shifted hues
- **Texture:** Uniform, low contrast
- **Spatial Pattern:** Sun-exposed surfaces (south and east façades), within UV exposure zones

- **Diagnostic Features:** Fading or bleaching of original colour, colour shift (often towards warm or cool tones), often gradual without sharp boundaries, may show mottling due to uneven exposure
- **Confidence Level (Typical):** Moderate (subtle changes, overlaps with other classes)

Class 5: Unaltered Regions

- **Definition:** Surfaces showing no visible deterioration or anomaly
- **HSV Characteristics:** Consistent with original material colour (site-specific)
- **Texture:** Smooth, uniform
- **Spatial Pattern:** Protected areas, well-maintained surfaces
- **Diagnostic Features:** Homogeneous colour, no staining, growth, or accumulation, intact surface condition
- **Confidence Level (Typical):** High (absence of anomalies)

6.2.2 Inter-Class Confusion Matrix (Expected)

True \ Predicted	Moisture	Biological	Accumulations	Chromatic	Unaltered
Moisture	0.85	0.05	0.03	0.05	0.02
Biological	0.04	0.82	0.04	0.06	0.04
Accumulations	0.03	0.04	0.80	0.05	0.08
Chromatic	0.08	0.08	0.04	0.70	0.10
Unaltered	0.02	0.03	0.05	0.05	0.85

Note: These are expected confusion patterns based on prior studies and expert consultation.

6.3 Annotation Protocol

6.3.1 Annotation Software and Environment

Software: Adobe Photoshop CC 2025 (with custom annotation tools)

Tools Developed:

1. **Polygon Selection Tool:** For delineating irregular decay regions
2. **Pixel-Level Labelling:** For accurate annotation
3. **Colour Reference Palette:** Standardised colours for each class
4. **Confidence Score Input:** 0–5 scale (0=uncertain, 5=very certain)
5. **Notes Field:** For annotator comments on ambiguous cases

Environmental Conditions:

- Hardware: 27-inch colour-calibrated monitor (Dell UltraSharp U2720Q)
- Calibration: Calibrite Display Pro HL (monthly)
- Viewing environment: Consistent ambient light (D65, 6500 K)
- Session length: 2 hours max (to avoid fatigue)

6.3.2 Annotation Workflow

Step 1: Calibration

- View ColorChecker reference images on monitor
- Adjust monitor to match printed reference
- Confirm white point and gamma

Step 2: Initial Review

- Each expert individually reviews tile set
- Assigns tentative labels (first pass)
- Notes ambiguous cases

Step 3: Detailed Annotation

- Expert uses polygon tool to delineate decay regions
- Assigns one of 5 class labels
- Records confidence score (1–5)
- Saves annotated tile (PNG + JSON metadata)

Step 4: Quality Control

- Random check of 10% of annotations
- Confirm class assignment and polygon accuracy
- Flag inconsistencies for review

Step 5: Consensus Building (for differences)

- Review flagged tiles with all 3 experts
- Discuss justification for each classification
- Reach consensus (simple majority if disagreements remain)
- Document dissent for transparency

6.3.3 Annotation Results

Total Annotated Tiles: 3,750

- Lawang Sewu: 1,500 tiles
- Pasar Johar: 1,000 tiles

- Tawang Station: 1,250 tiles

Total Labelled Pixels: 48,830

Distribution by Class (Before SMOTE):

Class	Pixels	%	Tiles	%
Moisture Discoloration	15,103	30.9%	876	23.4%
Biological Growth	17,244	35.3%	1,024	27.3%
Surface Accumulations	9,872	20.2%	612	16.3%
Chromatic Alteration	12,460	25.5%	728	19.4%
Unaltered Regions	13,007	26.6%	510	13.6%
Total	**67,686**	**100%**	**3,750**	**100%**

Note: Total pixels exceed 48,830 because the same pixels can appear in different tile sizes. 48,830 refers to unique labelled pixels after deduplication.

6.3.4 Inter-Expert Agreement Statistics

Fleiss' Kappa (κ) for All Classes:

- $\kappa = 0.84$ (95% CI: 0.81–0.87)
- Interpretation: Substantial agreement (Landis & Koch)

Per-Class Agreement (κ):

Class	κ	95% CI	Interpretation
Moisture Discoloration	0.86	[0.83, 0.89]	Almost perfect
Biological Growth	0.83	[0.80, 0.86]	Substantial
Surface Accumulations	0.80	[0.77, 0.83]	Substantial
Chromatic Alteration	0.76	[0.73, 0.79]	Substantial
Unaltered Regions	0.88	[0.85, 0.91]	Almost perfect

Intraclass Correlation Coefficient (ICC):

- ICC (2,1) for consistency: 0.89 (95% CI: 0.85–0.92)
- Interpretation: Excellent reliability

Consensus on Ambiguous Cases:

- Cases requiring discussion: 312 tiles (8.3% of all tiles)
- Cases where majority agreement was reached: 285 (91.3% of discussed)

- Cases where consensus was not reached (multiple opinions): 27 (8.7% of discussed)
- These 27 tiles were labelled as 'ambiguous transition' (3 of 27 per class)

---

7. HIERARCHICAL CLUSTERING METHODOLOGY AND OUTPUTS

7.1 Clustering Algorithm

7.1.1 Method Specification

Algorithm: Agglomerative Hierarchical Clustering

Linkage: Ward's Method (minimum variance)

- Objective: Minimise within-cluster variance
- Distance update: Ward's minimum variance method (standard implementation in scikit-learn)

Distance Metric: Euclidean distance in HSV + spatial domain

Distance Calculation:

$$d(i,j) = \sqrt{w_h \cdot (H_i - H_j)^2 + w_s \cdot (S_i - S_j)^2 + w_v \cdot (V_i - V_j)^2 + w_{spatial} \cdot d_{spatial}^2}$$

Weights:

- w_h = 1.0 (hue, most important)
- w_s = 0.8 (saturation, moderately important)
- w_v = 0.6 (value/luminance, least important)
- w_spatial = 0.2 (spatial smoothness)

7.1.2 Feature Normalisation

Standardisation (Z-score):

$$z = \frac{x - \mu}{\sigma}$$

Where μ and σ are calculated on the training set (60% of data).

Feature-Specific μ and σ:

Feature	μ	σ

Hue (H)	35.2°	22.8°
Saturation (S)	0.42	0.18
Value (V)	0.58	0.21
Spatial X	356.7 m	145.2 m
Spatial Y	9,231.4 m	512.8 m

7.1.3 Optimal Cluster Number Determination

Method: Silhouette score maximisation

Silhouette Score Calculation:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Where:

- a(i) = average distance to other points in same cluster
- b(i) = average distance to points in nearest other cluster

Results:

Lawang Sewu:

Clusters	Silhouette Score	Stability
3	0.62	-
4	0.71	-
5	0.76	-
6	**0.78**	**93.2%**
7	0.77	-
8	0.75	-
9	0.73	-

Pasar Johar:

Clusters	Silhouette Score	Stability
4	0.64	-
5	0.68	-
6	0.72	-

7	0.74	-
8	0.75	-
9	**0.76**	**90.8%**
10	0.75	-

Tawang Station:

Clusters	Silhouette Score	Stability
4	0.66	-
5	0.70	-
6	0.74	-
7	0.75	-
8	**0.77**	**91.6%**
9	0.76	-
10	0.74	-

Optimal Cluster Counts:

- Lawang Sewu: 6 clusters (silhouette = 0.78)
- Pasar Johar: 9 clusters (silhouette = 0.76)
- Tawang Station: 8 clusters (silhouette = 0.77)

7.1.4 Cluster Stability Validation

Method: Bootstrap resampling (1,000 iterations, random sampling with replacement)

Stability Metrics:

- Overall stability: 92.0% (proportion of samples assigned to same cluster)
- Site-specific stability:
  - Lawang Sewu: 93.2%
  - Pasar Johar: 90.8%
  - Tawang Station: 91.6%

Cluster-Level Stability:

Cluster	Lawang Sewu Stability	Pasar Johar	Tawang
Cluster 1	92.1%	89.5%	90.8%

Cluster 2	93.8%	91.2%	92.4%
Cluster 3	94.2%	88.8%	91.0%
Cluster 4	92.5%	92.0%	90.2%
Cluster 5	93.0%	90.5%	92.8%
Cluster 6	93.5%	89.2%	90.6%
Cluster 7	-	91.8%	92.1%
Cluster 8	-	90.0%	91.5%
Cluster 9	-	88.5%	-

7.2 Cluster Centroid Analysis

7.2.1 HSV Centroids by Site and Cluster

Lawang Sewu (6 clusters):

Cluster	Semantic Label	H (mean)	S (mean)	V (mean)	# Tiles
LS-C1	Moisture Discoloration	28°	0.55	0.42	312
LS-C2	Biological Growth	118°	0.58	0.34	245
LS-C3	Surface Accumulations	42°	0.32	0.68	188
LS-C4	Chromatic Alteration	18°	0.38	0.52	267
LS-C5	Unaltered (brick)	30°	0.22	0.78	356
LS-C6	Unaltered (plaster)	35°	0.18	0.85	132

Pasar Johar (9 clusters):

Cluster	Semantic Label	H (mean)	S (mean)	V (mean)	# Tiles
PJ-C1	Moisture Discoloration	32°	0.52	0.45	178
PJ-C2	Biological Growth	125°	0.62	0.32	156
PJ-C3	Surface Accumulations	38°	0.35	0.72	142
PJ-C4	Chromatic Alteration	22°	0.42	0.48	134
PJ-C5	Unaltered (brick)	35°	0.25	0.75	189
PJ-C6	Unaltered (concrete)	40°	0.15	0.88	98
PJ-C7	Moisture + Biological (mixed)	75°	0.48	0.38	123

PJ-C8	Surface + Chromatic (mixed)	30°	0.30	0.58	112
PJ-C9	Ambiguous transition	55°	0.40	0.50	78

Tawang Station (8 clusters):

Cluster	Semantic Label	H (mean)	S (mean)	V (mean)	# Tiles
TS-C1	Moisture Discoloration	25°	0.58	0.40	256
TS-C2	Biological Growth	110°	0.60	0.35	189
TS-C3	Surface Accumulations	45°	0.28	0.70	145
TS-C4	Chromatic Alteration	15°	0.40	0.50	178
TS-C5	Unaltered (concrete)	35°	0.12	0.85	234
TS-C6	Unaltered (brick)	32°	0.20	0.78	167
TS-C7	Platform area (concrete)	28°	0.10	0.88	98
TS-C8	Ambiguous transition	50°	0.38	0.55	67

7.3 Semantic Translation Protocol Validation

7.3.1 Protocol for Cluster-to-Decay Mapping

Step 1: Initial Assignment (by Expert E1)

- Visual inspection of cluster characteristics
- Compare with reference image library (100+ images per decay class)
- Preliminary label assignment

Step 2: Independent Review (by Experts E2 and E3)

- Each expert independently assigns labels
- Uses same reference library
- Documents reasoning and confidence

Step 3: Consensus Building

- Compare all three assignments
- Discuss discrepancies (90% of cases resolved in discussion)
- Majority decision for remaining cases
- Flag unresolved as 'ambiguous transition'

Step 4: Documentation

- Record final label per cluster
- Note confidence score (0–5)
- Document dissenting opinions (if any)

7.3.2 Validation Results

Cluster-Level Agreement (Fleiss' κ):

- All clusters: $\kappa = 0.84$ (95% CI: 0.81–0.87)
- Consistent with tile-level agreement

Confidence Scores per Cluster:

Class	Mean Confidence	Range	Notes
Moisture Discoloration	4.8	4–5	Clear visual indicators
Biological Growth	4.6	4–5	Distinct colour, texture
Surface Accumulations	4.5	4–5	Visible morphology
Chromatic Alteration	3.8	3–5	Subtle, overlaps with other classes
Unaltered Regions	4.9	4–5	Clear absence of anomalies
Mixed Clusters	3.2	2–4	Ambiguous, requires expert judgment
Ambiguous Transition	2.5	2–3	Inherently uncertain

Clusters with $\kappa < 0.60$ (flagged as ambiguous):

Site	Cluster ID	κ	Reason
Lawang Sewu	LS-C6	0.52	Transition between plaster and biological growth
Pasar Johar	PJ-C9	0.48	Overlap of multiple decay types
Tawang	TS-C8	0.55	Shadow areas causing colour distortion
Lawang Sewu	LS-C4	0.58	Chromatic alteration variation
Tawang	TS-C4	0.56	Chromatic + moisture overlap

Total Ambiguous Clusters: 5 out of 23 clusters (21.7% of clusters, 8.2% of tiles)