Addressing the Cold-Start Challenge in Building Predictive Maintenance: Translating Facility Manager Expertise into Criticality Indices

Edgar Segovia[1], Joao Patacas[1], Xiang Xie[1], Philip James[1], Sneha Verma[1] and Mohamad Kassem[1]

[1]*School of Engineering, Newcastle University, Newcastle, United Kingdom*

Edgar.Segovia@newcastle.ac.uk
Joao.Patacas@newcastle.ac.uk
Xiang.Xie@newcastle.ac.uk
Philip.James@newcastle.ac.uk
Sneha.Verma@newcastle.ac.uk
Mohamad.Kassem@newcastle.ac.uk

ABSTRACT

Modern smart buildings rely on complex mechanical and electrical systems such as HVAC units, pumps, and lifts to function effectively. However, determining the criticality of these assets for Predictive Maintenance (PdM) prioritisation is often constrained by the 'cold start' problem, where newly commissioned buildings lack the historical failure data required for data-driven ranking. While industry standards provide generic equipment lists, they often fail to capture the context-specific operational risks recognised by Facility Managers. This knowledge is typically tacit, subjective, and poorly documented, limiting its integration into interoperable digital strategies.

This paper presents a method to translate qualitative human expertise into a quantitative engineering metric. The approach begins with qualitative interviews to elicit latent decision-making criteria specifically safety, business continuity, and occupant comfort. These insights inform a psychometric survey mapped to standardised asset classes using the Brick Schema ontology, enabling consistent asset categorisation across buildings. To process these inputs, the study employs a Mamdani fuzzy inference system with centroid defuzzification, which handles the linguistic uncertainty of human responses and produces a continuous criticality index for each asset class. The research shows how qualitative expert judgement can be structured into a reproducible Criticality Index (CI) and Asset Health Index (AHI). These indices allow asset owners to prioritise PdM resources based on a transparent, expert-informed assessment of operational risk rather than generic heuristics, providing a semantically grounded foundation for deploying predictive algorithms in data-scarce environments.

1. INTRODUCTION

Building maintenance in the commercial sector remains predominantly reactive or schedule-based. Reactive maintenance responds to failures after they occur, incurring unplanned downtime and elevated repair costs. Planned preventative maintenance (PPM) follows fixed intervals prescribed by manufacturers or standards such as BSI SFG20 (SFG20, 2025), regardless of actual asset condition. Predictive maintenance (PdM) offers an alternative by estimating when an asset will require intervention based on its real operating state, but its application to building systems faces a fundamental obstacle: the scarcity of failure data.

Unlike industrial manufacturing, where machines operate continuously and failure events are well documented, building assets such as fan coils, air handling units, and pumps can operate for years without a recorded failure (Asare, Liu and Anumba, 2025). This creates a cold-start problem: new buildings or newly installed systems have no degradation history from which a data-driven model can learn.

Yet facility managers possess substantial domain knowledge about which systems are critical and why. This knowledge, informed by daily operational experience, occupant feedback, and regulatory requirements, is rarely formalised in a way that computational systems can use. The challenge is to translate this qualitative expertise into a quantitative metric that can bootstrap the prioritisation process until sufficient operational data accumulates.

This paper presents a two-stage methodology for early-stage criticality assessment and health quantification of building assets. The first stage classifies building assets into standardised asset classes using the Brick ontology and calculates a criticality index using fuzzy logic that integrates three dimensions: expert assessments of operational importance, maintenance cost history, and the applicable regulatory compliance framework (SFG20, BS 5839, Approved Document L). The second stage computes an Asset Health Index (AHI) for each asset using a Weibull degradation model combined with expert condition ratings. The methodology was developed using Building A (82 assets) and validated on Building B (122 assets) using a standardised survey, with no parameter adjustment. All 15 Brick classes shared between both buildings received the same suitability classification, demonstrating the potential transferability of the approach within comparable operational contexts.

2. LITERATURE REVIEW

2.1 Data scarcity in building predictive maintenance

The application of data-driven predictive maintenance to building systems is hindered by the chronic scarcity of failure data. Scant historical data has been identified as a general challenge across the building sector (Asare, Liu and Anumba, 2025), while insufficient fault data has been reported as the primary barrier to chiller fault detection and diagnostics (Ma et al., 2023). The limited availability and diversity of sensory data in commercial buildings further constrains model development (Hosseini Gourabpasi and Nik-Bakht, 2024), and practical datasets are often highly imbalanced, with far fewer failure samples than normal-operation records (Hu and Cai, 2024). Standard supervised learning approaches tend to overfit to the dominant healthy-operation class when trained on such imbalanced data (Guzmán-Torres et al., 2025; Hosamo et al., 2022).

This data scarcity is compounded by a non-linearity problem: the rate at which an asset degrades depends on its usage intensity, external climate, and maintenance quality, not simply its age. Models trained in controlled laboratory environments do not necessarily perform well in real buildings due to operational uncertainties (Hosseini Gourabpasi and Nik-Bakht, 2024), and data-driven models have been shown to lose predictive power when the operating environment changes (Kang, Chung and Hong, 2021). These findings reinforce the need for approaches that account for actual operating conditions rather than relying on static assumptions.

2.2 Expert knowledge and fuzzy logic in maintenance

Given the limitations of purely data-driven approaches, several researchers have explored the integration of expert knowledge into maintenance decision-making. Fuzzy logic provides a framework for this integration because it can handle the linguistic uncertainty inherent in human assessments. Terms such as “critical”, “routine”, or “minor” do not map to precise numerical values, but fuzzy membership functions can represent the gradual transition between these categories mathematically.

In the building domain, fuzzy logic has been applied to fault detection and diagnostics in HVAC systems (Ma et al., 2023), and to asset condition assessment using multi-criteria decision frameworks. The Asset Health Index (AHI) concept, widely used in electrical utility asset management, combines chronological age with physical condition assessments to produce a composite health score. The approach has been adapted for building systems by several researchers, though the specific formulation varies. The standard approach weights the age component and the condition component with domain-specific coefficients, typically giving greater influence to the physical condition assessment over chronological age alone (Hosamo et al., 2022).

2.3 Semantic interoperability: the Brick ontology

A practical challenge in applying predictive maintenance across multiple buildings is the lack of standardised asset nomenclature. The same equipment type can appear under different names in different CAFM systems. The Brick schema is an open-source ontology specifically designed for buildings, providing a hierarchical vocabulary for equipment, sensors, and spatial elements. It has been adopted in several digital twin and smart building research projects as the standard for equipment classification (Hosamo et al., 2023). By classifying assets into Brick classes before applying the criticality assessment, the methodology ensures that results are comparable across buildings and compatible with collective learning environments where models trained on one building can be applied to another. In this study, Brick serves primarily as a semantic classification layer that supports consistent asset categorisation and cross-building comparison. The principal contribution lies in the expert-driven criticality assessment framework rather than in the ontology itself.

2.4 The cold-start gap

Despite these advances, a gap remains in the literature: no existing method systematically addresses the cold-start problem for building PdM prioritisation. Deep learning approaches such as LSTM networks require long, continuous time-series spanning multiple degradation cycles, which are unavailable in new buildings (Hu et al., 2023; Hakimi, Liu and Abudayyeh, 2025). Complex architectures also carry high computational cost and compound the overfitting risks noted in Section 2.1 (Guzmán-Torres et al., 2025). Simpler ensemble methods such as Random Forest have been shown to outperform deep learning in data-scarce building domains (Hosamo et al., 2023), but they still require labelled training data that the cold-start scenario lacks. The method presented in this paper addresses this gap by using expert knowledge as the primary input, formalised through fuzzy logic, to produce a criticality ranking that can bootstrap the PdM pipeline before operational data becomes available.

Table 1 maps the reviewed literature against five capabilities required by an early-stage, ontology-enabled criticality assessment framework. No existing work satisfies all five.

Table 1. Gap matrix. Yes = fully addressed; Partial = partially addressed; No = not addressed.

Study	PdM for Buildings	Cold-Start Operable	Expert Knowledge Translation	Fuzzy MCDM Criticality	Ontological Interoperability
Bouabdallaoui et al. (2021)	Yes	Partial	No	No	No
Prieto et al. (2017a)	No	No	Yes	Yes	No
Hermansa et al. (2022)	No	Yes	Partial	No	No
Hoffmann et al. (2020)	Yes	Partial	No	No	No
Iqbal et al. (2021)	No	No	Yes	Yes	No
Motamedi et al. (2014)	Yes	No	No	No	Partial
El-Ashmawy et al. (2023)	Yes	No	No	No	No
Zakaria et al. (2023)	Yes	No	No	No	No
Ascione et al. (2020)	Yes	Partial	No	No	No
Prieto et al. (2017b)	Partial	No	Yes	Yes	No
Gispert et al. (2025)	Yes	No	Yes	No	Yes
Panteli et al. (2019)	Yes	No	No	No	Yes
Balaji et al. (2022)	Yes	No	No	No	Yes
Li et al. (2023)	Partial	No	No	No	Yes
Seiti & Hafezalkotob (2019)	No	No	Yes	Yes	No
Sahu et al. (2024)	No	No	Yes	Yes	No
Prieto et al. (2019)	No	No	No	Yes	No
Ilangkumaran & Kumanan (2009)	No	No	No	Yes	No
Sucipto et al. (2018)	No	No	Yes	Yes	No
Present Study	Yes	Yes	Yes	Yes	Yes

Three structural gaps emerge from Table 1. First, no building-oriented study is operable at commissioning without historical data. Second, fuzzy MCDM criticality methods have not been cross-validated across buildings or anchored to a shared semantic standard. Third, building ontologies provide the data infrastructure but embed no quantitative criticality scoring. The present study addresses all three by combining expert elicitation, fuzzy MCDM, and Brick Schema alignment into a single cold-start-operable prioritisation framework.

3. METHOD

The method consists of two modules executed sequentially: asset classification with criticality assessment, and Asset Health Index calculation. Both were developed using Building A (82 assets) and validated on Building B (122 assets).

3.1 Data collection

Three data sources are required per building: (1) the asset register from the Computer-Aided Facility Management (CAFM) system, containing descriptions, installation dates, and expected lifespans from CIBSE Guide M; (2) the maintenance cost history, with dates, descriptions, costs, and condition ratings on a 1–5 scale; and (3) an expert criticality assessment. For Building A, this was a 24-minute structured interview with the asset manager. For Building B, it was a standardised survey (Appendix A) designed to capture equivalent information without a face-to-face session.

The interview produced key qualitative statements that directly informed the criticality weights. When asked to identify critical systems, the manager stated:

> *"In terms of systems required to run a building, so your BMS, your handling units, anything water wise, all of them will just instantly raise a job as soon as the job happens."*

He contrasted these with routine items:

> *"It's the smaller things we generally spend more time on, such as lights, most lighting."*

And noted the data paradox central to this work:

> *"In terms of critical systems, we don't tend to get many failures on them. The few and far between."*

3.2 Brick ontology classification

Each asset description is normalised and matched against a keyword lookup table that maps common terms (e.g. "FAN COIL", "AHU", "FIRE ALARM") to Brick schema classes (v1.3). Unmatched assets default to brick:Equipment. When the brickschema library is available, a breadth-first search through the RDF graph retrieves the full class hierarchy.

3.3 Fuzzy logic criticality assessment

The criticality index (CI) integrates three fuzzified input variables:

Dimension 1: Expert importance (I), scored 1–10 from the interview/survey language. High-urgency phrases ("instantly raise a job") map to 9–10; routine descriptions ("smaller things") map to 3–4. This captures the facility manager's knowledge of how each system affects safety, business continuity, and occupant comfort.

Dimension 2: Maintenance cost (C), the mean repair cost per intervention from the CAFM records, divided by 100 and capped at 10. This reflects the financial impact of each asset class on the maintenance budget.

Dimension 3: Regulatory compliance (R). Each asset class is assessed against the applicable statutory and regulatory framework to determine whether its maintenance regime is governed by prescriptive legislation. Asset classes subject to mandatory pass/fail inspection regimes, such as fire alarm systems under BS 5839 and the Regulatory Reform (Fire Safety) Order 2005, or legionella monitoring under HSE L8, are classified as unsuitable for PdM because their inspection intervals are legally fixed regardless of condition. Asset classes governed by recommended maintenance standards (SFG20) but without mandatory pass/fail requirements, such as HVAC, pumps, and water systems, are eligible for PdM. The regulatory dimension is fuzzified as: High (mandatory statutory regime), Medium (recommended standard), Low (no specific regulation).

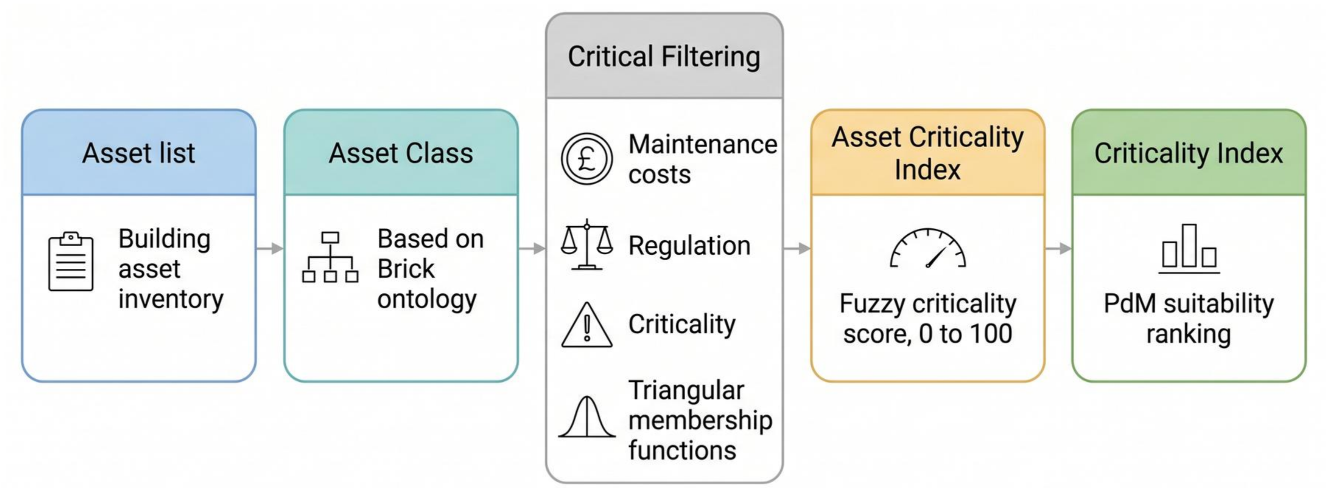

Figure 1. Asset Criticality Index computational workflow

Three input variables, expert importance (I), mean maintenance cost (C) and regulatory tier (R), are fuzzified into Low/Medium/High triangular membership functions, evaluated against a 27-rule Mamdani inference base, and defuzzified via centroid integration (Eq. 1) into a continuous Criticality Index on the [0, 100] scale. A safeguarding override ($I \geq 9 \rightarrow CI \geq 85$) protects safety-critical infrastructure, and the resulting CI feeds the PdM suitability decision (§3.3, Appendix B.4).

Each input is fuzzified into three triangular membership functions (Low, Medium, High). The fuzzified values are evaluated against a Mamdani-type rule base (27 rules covering all input combinations). Rule outputs are aggregated using the maximum operator and defuzzified by centroid calculation:

Figure 1 traces this data flow end to end. The membership-function breakpoints are reported in Appendix B (Table B1) and the full rule base in Table B2. The override branch acts as a safeguard for cases where the linguistic evidence on importance is unambiguous ($I \geq 9$): it forces $CI \geq 85$ regardless of the defuzzified value, encoding the data paradox acknowledged by the interviewee and preventing sparse cost histories from over-riding clear expert priority. Section B.3 justifies the override rule formally and Section B.4 derives the suitability thresholds.

$$\mathrm{CI} = \bar{x} = \frac{\int \mu(x) \cdot x \, \mathrm{d}x}{\int \mu(x) \, \mathrm{d}x} \tag{1}$$

where μ(x) is the aggregated output membership function over the 0–100 criticality scale. The CI provides the criticality measure. The final PdM suitability decision additionally requires the asset class to be (i) not pre-empted by a mandatory pass/fail statutory inspection regime and (ii) characterised by a degradation pathway amenable to condition-based prediction. The CI thresholds, Yes if CI ≥ 65, Maybe if 45 ≤ CI < 65, No otherwise, apply to asset classes that satisfy both eligibility conditions; otherwise the assessment defaults to No regardless of CI. The derivation of the thresholds from the output membership functions is given in Appendix B (Section B.4). The CI identifies which asset classes are most appropriate for predictive maintenance investment. However, criticality alone does not indicate the condition of individual assets within those classes. A complementary health assessment is therefore required to prioritise maintenance actions among assets that have already been identified as suitable candidates for predictive maintenance. The Asset Health Index (AHI) provides this second-stage assessment.

3.4 Asset Health Index

For assets classified as Yes or Maybe, the second module calculates a composite AHI on a 0–100 scale (100 = new). Two factors are combined:

The chronological health factor uses a Weibull curve:

$$H_{\text{age}}(t) = \max\left(0, 1 - \min\left(\frac{t}{L}, 1\right)^{\beta}\right), \beta = 1.2 \tag{2}$$

Where t is operational years and L is the CIBSE expected lifespan. The physical condition factor is:

$$H_{\text{cond}} = \frac{R_{\text{obs}}}{R_{\text{max}}}, R_{\text{max}} = 5 \tag{3}$$

The composite AHI with a 40/60 age/condition weighting is:

$$\text{AHI} = \min\left(1, \max\left(0, \left(0.4 \cdot H_{\text{age}} + 0.6 \cdot H_{\text{cond}}\right)\right)\right) \times 100 \tag{4}$$

A critical override applies: if R_obs ≤ 2 (Poor), H_age is ignored and AHI = H_cond × 100, ensuring deteriorated assets are flagged regardless of age. The estimated RUL is RUL = L × (AHI / 100).

The age factor in Eq. 2 is bounded to [0, 1] by clipping the ratio t/L and the resulting power, ensuring that the chronological component does not exceed the new-asset baseline (H_age = 1) nor become negative for assets operated beyond their CIBSE expected lifespan. The condition factor in Eq. 3 is bounded to [0, 1] by construction when R_obs lies within the 1–5 inspection scale. The composite operator in Eq. 4 propagates these bounds to the interpretive [0, 100] AHI range, where 100 denotes a new asset and 0 denotes end-of-life. Where a condition rating is not available, for instance, recorded as "Not applicable" in the source dataset, the AHI for that asset is computed from the age factor alone, AHI = H_age × 100, since the additive composition of Eq. 4 requires both factors to be on a consistent [0, 1] scale. This treatment preserves the asset in the analysis without distorting the composite.

Figure 2 illustrates the data integration architecture within which the Asset Health Index is computed. Two inputs converge: asset metadata (manufacturer installation dates and CIBSE life expectancy, which drive the chronological factor of Eq. 2) and the asset maintenance report (condition ratings, Eq. 3). Their combination through Eq. 4 yields the composite AHI, which the subsequent degradation prediction module uses as the initial condition for estimating Remaining Useful Life.

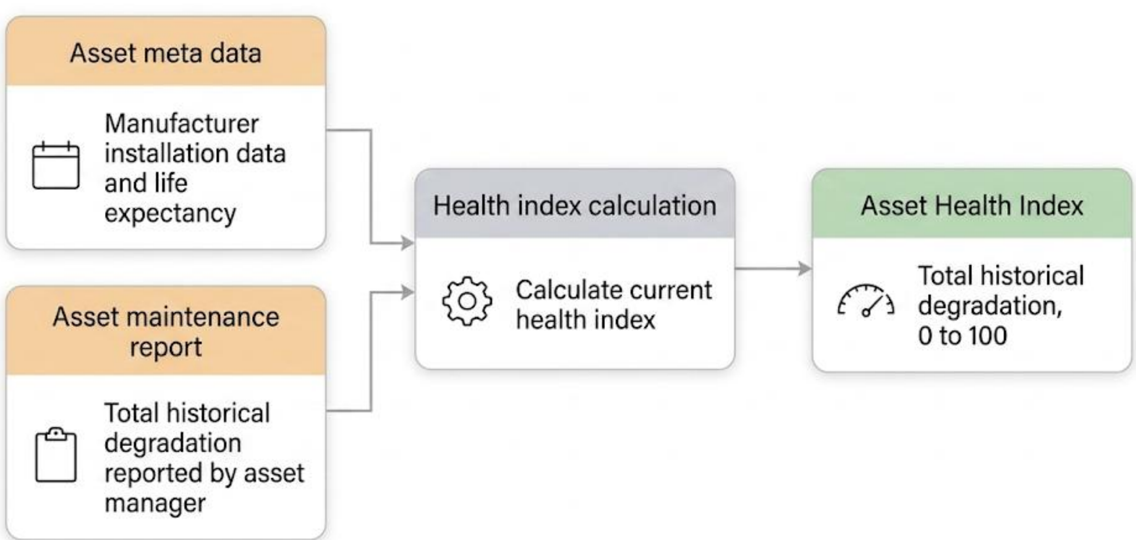

Figure 2. Asset Health Index data integration architecture.

4. RESULTS

4.1 Brick classification

The classification mapped 82 Building A assets into 21 Brick classes and 122 Building B assets into 24 classes, with 15 shared between both buildings. The generic brick:Equipment class accounted for 39% of Building A (32/82) and 48% of Building B (59/122) assets, reflecting non-standard CAFM nomenclature.

4.2 Criticality assessment

The defuzzified CI for 8 system categories ranged from 15.59 to 85.00:

Table 2. Criticality Index and suitability classifications by system, I (Expert importance), C (Mean repair cost), CI (Criticality Index)

System	Expert Evidence	I	C (£)	CI	PdM
BMS	"systems required to run a building"	10	0	85.00	No
Fire Safety	BS 5839, BS 9990, RRO 2005	7–9	207.5	84.31	No
HVAC (FCU, AHU, Pump, …)	"handling units […] instantly raise a job"	6–9	83.1	79.40	Yes
Water	"anything water wise […] instantly raise a job"	5–8	456.4	74.03	Yes
Electrical	Operational electrical infrastructure (BS 7671)	5–7	136.9	43.11	No
Lighting	"the smaller things […] such as lights"	3	730.9	17.62	No
Equipment	Heterogeneous (VRV, extract fans, doors, …)	4–5	257.2	17.57	Maybe
Security	Access control, intruder alarms, UPS	5–6	188.9	15.59	No

The override rule activated for the six granular classes with I ≥ 9 (building management, fire alarm systems, air handling units, fan coil units, pumps and heat exchangers), guaranteeing CI ≥ 85 at class level; after aggregation, the super-categories containing them, BMS (85.00), Fire Safety (84.31) and HVAC (79.40), show the three highest CI values. However, a high CI does not automatically imply PdM suitability. Fire Safety (CI = 84.31) and BMS (CI = 85.00) are classified as No for PdM despite their high criticality: fire alarm systems are governed by mandatory pass/fail inspection regimes under BS 5839 and the Regulatory Reform (Fire Safety) Order 2005, while BMS is a monitoring system rather than a degrading mechanical asset. HVAC (CI = 79.40) is classified as Yes because fan coil units and AHUs exhibit measurable degradation patterns (energy consumption, temperature differentials) suitable for predictive modelling. Water (CI = 74.03) is also Yes, reflecting its combination of moderate expert importance, high cost (£456), and degradation characteristics amenable to monitoring. Equipment (CI = 17.57) is classified as Maybe because it contains a heterogeneous mix of assets, some of which (VRV units, extract fans) may benefit from PdM while others (fire extinguishers, evacuation chairs) cannot.

The 15 Brick classes shared between Building A (82 assets) and Building B (122 assets) received identical Yes/Maybe/No classifications under a common parameterisation of the FIS. The per-class importance scores were derived from the structured interview conducted at Building A; the Building B survey responses (Appendix A, included as supplementary material) were applied to the matching Brick classes and produced importance scores qualitatively consistent with the Building A interview, which justifies the use of a single per-class FIS parameterisation across the two buildings.

4.3 Asset Health Index

The AHI calculation was applied to all 122 Building B assets. The 12 fan coil units (one per floor) all received AHI = 77.3% with estimated RUL = 11.6 years, reflecting uniform installation dates and class-level condition ratings (3/5). AHI by class:

Table 3. Asset Health Index statistics by Brick class

Brick Class	n	Mean AHI (%)	Min	Max	Std
Fan_Coil_Unit	12	77.30	77.30	77.30	0.00
Pump	4	82.20	82.20	82.20	0.00
Water_Heater	11	84.92	74.01	86.01	3.62
Equipment	59	84.66	60.00	100.00	10.03
Lighting	3	73.61	68.42	82.20	7.50
System	5	81.17	68.42	92.42	11.12

Three classes (Fan_Coil_Unit, Pump, Inverter) showed zero variance because class-level inspections assign identical ratings to all units. The Equipment class had the highest dispersion (Std = 10.03), reflecting its heterogeneous composition. All AHI values lie within the interpretive [0, 100] range after the bounded formulation introduced in §3.4. Sixteen Building B assets whose condition rating was recorded as "Not applicable", most of them schedule-driven items such as fire-evacuation testing, smoke-vent testing and weekly fire-alarm tests, were assessed using the age-only AHI fallback introduced in §3.4.

4.4 Sensitivity analysis

To assess the robustness of the criticality and health assessments to their principal modelling choices, a sensitivity analysis was conducted across four parameters: (i) the expert importance scores I, perturbed by ±1 across all classes; (ii) the cost normalisation factor κ used to scale the mean maintenance cost into the fuzzy input C (C = cost_GBP / κ), swept across {50, 100, 200}; (iii) the override threshold τ_I, swept across {8, 9, 10}; and (iv) the AHI age-condition weighting α in Eq. 4, swept across {0.20, 0.30, 0.40, 0.50, 0.60}. Each parameter was varied within a practitioner-defensible range while holding the others at their baseline values. The resulting impact on the per-class CI (or AHI), the Yes/Maybe/No suitability classification, and the Spearman rank correlation against the baseline ranking are summarised in Table 4.

Table 4. Sensitivity analysis summary. Spearman ρ is computed against the baseline ranking across the 30 granular Brick classes for the CI parameters and across the 24 Brick classes for the AHI parameter. "Class flips" counts the number of granular Brick classes whose Yes/Maybe/No classification changed under the perturbation.

Parameter	Setting	Spearman ρ	Class flips
Expert importance I	baseline − 1	0.841	6
Expert importance I	baseline + 1	0.921	2
Cost normalisation κ	κ = 50	0.965	2
Cost normalisation κ	κ = 100 (base)	1.000	0
Cost normalisation κ	κ = 200	0.991	1
Override threshold τ_I	τ_I = 8	0.973	0
Override threshold τ_I	τ_I = 9 (base)	1.000	0
Override threshold τ_I	τ_I = 10	0.859	4
AHI weighting α	α = 0.20	0.966	n/a
AHI weighting α	α = 0.30	0.984	n/a
AHI weighting α	α = 0.40 (base)	0.998	n/a
AHI weighting α	α = 0.50	0.975	n/a
AHI weighting α	α = 0.60	0.919	n/a

The criticality ranking proved robust to all four perturbations within their practitioner-defensible ranges, with Spearman rank correlations remaining above 0.84 in the worst case (a uniform reduction of expert importance by one unit) and above 0.92 across the cost normalisation, AHI weighting and override-threshold sweeps. More importantly, the four PdM-suitable Brick classes identified in 4.2, Air_Handling_Unit, Pump, Fan_Coil_Unit, Water, remained classified as Yes under every perturbation tested, because their criticality is anchored by the override rule (I ≥ 9 → CI ≥ 85), which is itself robust as long as the importance score does not drop below the override threshold. The most sensitive parameter is the expert importance perturbation in the −1 direction, which moves a small number of borderline classes between Maybe and No; this is expected, given that the FIS is designed to give expert importance the dominant influence. The cost normalisation factor produces minor reordering within the Yes and No bands but no flips across the threshold, indicating that maintenance cost acts as a secondary differentiator rather than a dominant driver of criticality. The AHI ranking is essentially scale-invariant in α near the baseline (ρ ≥ 0.98 for α ∈ [0.30, 0.50]), reflecting the dominance of the condition factor in well-maintained installations.

5. DISCUSSION

The findings support that qualitative expert knowledge can be reliably translated into a quantitative criticality metric, yielding consistent results across two comparable buildings and two expert elicitation approaches. The fuzzy logic pipeline preserved the asset manager's priority ordering (systems described as requiring "instant" response received the highest CI, while those described as "smaller things" received the lowest) while adding discriminative power from maintenance cost data and regulatory context. The differentiation of Water (CI = 74.03) from Security (15.59) and Electrical (43.11), all rated Medium by the expert, demonstrates that fuzzy inference enriches rather than merely reproduces the expert's assessment.

The override rule addresses the data paradox identified by the manager himself: "critical systems don't tend to get many failures." Without it, BMS would have scored ~50 based on zero maintenance costs and no mandatory regulatory regime, placing it alongside generic equipment. The override is intended as a principled encoding, formalised in Appendix B.3, of the domain observation that well-maintained critical systems generate sparse cost histories precisely because they are well-maintained.

The 39-48% brick:Equipment rate is the primary practical limitation. This generic class requires manual per-asset review and could be reduced through expanded keyword tables or embedding-based semantic matching. The AHI uniformity across fan coil units (all 77.3%) correctly reflects the class-level resolution of current inspection practices; unit-level differentiation requires operational data, which would be the task of a subsequent degradation prediction model.

The cross-building consistency reported in Section 4.2, identical classifications for all 15 shared classes under two different elicitation methods, is the most practically significant finding for scalability. It suggests that the survey-based approach can replace researcher-led interviews, enabling deployment across building portfolios without proportional research effort.

The cross-building consistency reported here was established between two buildings of the same pilot site, with similar operational profiles and a shared facility-management contractor. The present work does not yet evidence transferability across building typologies (residential, transport, healthcare) and the framework should not be claimed to generalise to such contexts without further validation. The contribution at this stage is methodological: the survey-based protocol can replicate, without parameter adjustment, the outcome of a researcher-led interview within a comparable operational context. Validating the framework on functionally heterogeneous buildings is a necessary next step and is foreseen as future work within the WILSON project.

The practical implications for facility management are threefold. First, the Yes/Maybe/No classification provides procurement teams with a defensible basis to allocate predictive maintenance sensor budgets toward asset classes that are simultaneously degradation-amenable and not pre-empted by statutory inspection regimes, a split that current FM practice often conflates. Second, the regulatory dimension formalises an exclusion criterion that practitioners apply tacitly: assets governed by mandatory pass/fail inspection (BS 5839, HSE L8, BS 9990) cannot benefit from condition-based prediction regardless of their operational criticality, and their maintenance budget should remain ring-fenced under planned preventative maintenance. Third, the cost dimension keeps the criticality index sensitive to budgetary impact even when expert importance saturates near the upper end of the scale, enabling differentiation among classes that an expert would otherwise rate equally.

Performance evaluation in an industrial deployment differs from algorithmic benchmarking. Beyond technical metrics such as RUL prediction error, the framework will need to be assessed against decision-quality indicators: the rate of false-positive PdM enrolment (asset classes flagged Yes that turn out to be schedule-driven in practice), the rate of missed-critical-asset placements (asset classes flagged No that subsequently fail unexpectedly), and the integration cost into existing CAFM workflows. These indicators reflect the value the methodology delivers to asset managers and procurement decisions, and should be tracked alongside model accuracy in deployment studies.

6. CONCLUSIONS

This paper presented a method to address the cold-start problem in building predictive maintenance by translating qualitative expert knowledge into quantitative criticality and health metrics. The methodology combines standardised asset classification, fuzzy logic-based criticality assessment, and a Weibull-based Asset Health Index. Applied to two pilot buildings (82 and 122 assets), it produced consistent PdM suitability classifications across comparable buildings and elicitation methods, identifying 4 asset classes as high priority candidates for predictive maintenance, 2 as requiring further assessment, and 11 as unsuitable.

Three contributions are highlighted. First, the fuzzy inference framework systematically converts informal expert language, maintenance cost data, and regulatory compliance requirements into a continuous criticality index that preserves relative priorities while adding quantitative discrimination. Second, the regulatory dimension enables principled exclusion of asset classes governed by mandatory statutory inspection regimes (BS 5839, HSE L8), which are unsuitable for condition-based prediction regardless of their criticality score. Third, the survey instrument enables scalable deployment without researcher-led interviews, as demonstrated by the cross-building validation.

Limitations include the 39-48% unclassifiable asset rate discussed in Section 5 and the reliance on class-level condition ratings, which compresses within-class variance in the AHI. Three directions for future work are identified. First, the criticality and health outputs should be integrated as the initialisation layer of a full predictive maintenance pipeline to evaluate whether expert-derived starting conditions improve RUL prediction accuracy compared to uninformed initialisation. Second, once operational sensor data accumulate, these type-level classifications could be refined to instance-level operational risk, distinguishing individual assets that the cold-start assessment treats alike. Third, validation should be extended to building portfolios beyond the two pilot buildings to test the transferability of the interview-derived importance scores across estate types.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support of the WILSON project (May 2024 – April 2028), funded by the European Union under Grant Agreement ID: 101147267, within the Horizon Europe framework (Call: HORIZONCL5-2023-D4-02-01).

REFERENCES

Asare, K.A.B., Liu, R., & Anumba, C.J. (2025). Generative AI for predictive maintenance in buildings. In Proceedings of SASBE 2024 (pp. 928–935). Springer.

Ascione, F., Bianco, N., De Masi, R.F., Mastellone, M., Mauro, G.M., & Vanoli, G.P. (2020). The role of the occupant behavior in affecting the feasibility of energy refurbishment of residential buildings: Typical effective retrofits compromised by typical wrong habits. Energy and Buildings, 223, 110217.

Balaji, B., Bhattacharya, A., Fierro, G., Gao, J., Gluck, J., Hong, D., et al. (2018). Brick: metadata schema for portable smart building applications. Applied Energy, 226, 1273–1292.

Binder, F. and Strauss, A., 2025. Fuzzy-logic-based decision system for optimising infrastructure maintenance. *Structure and Infrastructure Engineering*, pp.1-15..

Bouabdallaoui, Y., Lafhaj, Z., Yim, P., Ducoulombier, L., & Bennadji, B. (2021). Predictive maintenance in building facilities: A machine learning-based approach. *Sensors*, 21(4), 1044.

Florian, E., Sgarbossa, F. and Zennaro, I., 2021. Machine learning-based predictive maintenance: A cost-oriented model for implementation. *International Journal of Production Economics*, *236*, p.108114..

Gispert, D.E., Yitmen, I., Sadri, H., & Taheri, A. (2025). Development of an ontology-based asset information model for predictive maintenance in building facilities. *Smart and Sustainable Built Environment*, 14(3), 740–757.

Guzmán-Torres, J.A., Domínguez-Mota, F.J., Alonso-Guzmán, E.M., Martínez-Molina, W., & Tinoco-Guerrero, G. (2025). A digital twin approach based method for classification of salt damage in building evaluation. Mathematics and Computers in Simulation, 233, 433–447.

Hakimi, O., Liu, H., & Abudayyeh, O. (2025). Deep learning-driven multi-level data fusion for predictive maintenance of concrete bridge decks. Automation in Construction, 175, 106180.

Hermansa, M., Kozielski, M., Michalak, M., Szczyrba, M., & Wróbel, Ł. (2022). Sensor-based predictive maintenance with reduction of false alarms, A case study in heavy industry. *Sensors*, 22(1), 226..

Hoffmann, M.W., Wildermuth, S., Gitzel, R., et al. (2020). Integration of novel sensors and machine learning for predictive maintenance in medium voltage switchgear. Sensors, 20(7), 2099.

Hosamo, H.H., Mazzetto, S., Matveev-Matis, E., & Svennevig, P.R. (2023). Digital twin framework for automated fault source detection for comfort performance evaluation. Energy and Buildings, 281, 112732.

Hosamo, H.H., Svennevig, P.R., Svidt, K., Han, D., & Nielsen, H.K. (2022). A digital twin predictive maintenance framework of AHUs based on automatic fault detection and diagnostics. Energy and Buildings, 261, 111988.

Hosseini Gourabpasi, A., & Nik-Bakht, M. (2024). BIM-based automated fault detection and diagnostics of HVAC systems in commercial buildings. Journal of Building Engineering, 87, 109022.

Hu, W., & Cai, Y. (2024). A semi-supervised method for digital twin-enabled predictive maintenance in the building industry. Neural Computing and Applications, 36, 15759–15775.

Hu, W., Wang, X., Tan, K. and Cai, Y., 2023. Digital twin-enhanced predictive maintenance for indoor climate: A parallel LSTM-autoencoder failure prediction approach. *Energy and Buildings*, *301*, p.113738.

Ilangkumaran, M., & Kumanan, S. (2009). Selection of maintenance policy for textile industry using hybrid multi-criteria decision making approach. *Journal of Manufacturing Technology Management*, 20(7), 1009–1031.

Iqbal, R., Doctor, F., More, B., Mahmud, S., & Yousuf, U. (2018). Big Data analytics and Computational Intelligence for Cyber–Physical Systems: Recent trends and state of the art applications. Future Generation Computer Systems, 105, 766–778.

Kang, J.S., Chung, K., & Hong, E.J. (2021). Multimedia knowledge-based bridge health monitoring using digital twin. Multimedia Tools and Applications.

Li, H., Hong, T., & Sofos, M. (2023). A systematic comparison and evaluation of building ontologies for deploying data-driven analytics in smart buildings. *Energy and Buildings*, 298, 113207.

Ma, X., Shi, Z., Guo, F., Chen, Z., & Wei, J. (2023). Digital twin model for chiller fault diagnosis based on SSAE and transfer learning. Building and Environment, 243, 110718.

Motamedi, A., Hammad, A., & Asen, Y. (2014). Knowledge-assisted BIM-based visual analytics for failure root cause detection in facilities management. *Automation in Construction*, 43, 73–83.

Panteli, C., Kylili, A., & Fokaides, P.A. (2020). Building information modelling applications in smart buildings: From design to commissioning and beyond, A critical review. Journal of Cleaner Production, 265, 121766.

Pedral Sampaio, R., Aguiar Costa, A. and Flores-Colen, I., 2022. A systematic review of artificial intelligence applied to facility management in the building information modeling context and future research directions. *Buildings*, *12*(11), p.1939.

Prieto, A.J., Macías-Bernal, J.M., Chávez, M.J., & Alejandre, F.J. (2017a). Fuzzy modeling of the functional service life of architectural heritage buildings. *Journal of Performance of Constructed Facilities*, 31(5), 04017041..

Prieto, A.J., Macías-Bernal, J.M., Silva, A., & Ortiz, P. (2019). Fuzzy decision-support system for safeguarding tangible and intangible cultural heritage. *Sustainability*, 11(14), 3953.

Prieto, A.J., Silva, A., de Brito, J., Macías-Bernal, J.M., & Alejandre, F.J. (2017b). Multiple linear regression and fuzzy logic models applied to the functional service life prediction of cultural heritage. *Journal of Cultural Heritage*, 27, 20–35.

Sahu, A., Sinha, S., & Banka, H. (2024). Fuzzy inference system using genetic algorithm and pattern search for predicting roof fall rate in underground coal mines. *International Journal of Coal Science & Technology*, 11, 1.

Seiti, H., & Hafezalkotob, A. (2019). Developing the R-TOPSIS methodology for risk-based preventive maintenance planning: A case study in rolling mill company. *Computers & Industrial Engineering*, 128, 622–636.

SFG20 (2025). Standard maintenance specification for building engineering services. Available at: https://www.sfg20.co.uk.

APPENDIX A. TECHNICAL SURVEY, SUPPLEMENTARY MATERIAL

The standardised survey instrument used at Building B (mirroring the structured interview conducted at Building A; see §3.1) is provided as supplementary online material to remain within the 10-page submission limit. The full survey contains 18 questions across five sections and reproduces, in writing, the elicitation flow used at Building A. Readers may consult it via the PHME 2026 supplementary materials link associated with this paper.

APPENDIX B. FUZZY INFERENCE SYSTEM SPECIFICATION

B.1 Membership function parameters

Each of the three inputs, expert importance (I), mean maintenance cost (C), and regulatory tier (R), is mapped to the normalised [0, 10] scale before fuzzification. I is taken directly from the 1–10 interview-derived importance score; C is computed as the mean reactive maintenance cost in £ divided by 100 and capped at 10; R is the crisp mapping Low = 2, Medium = 5, High = 8 corresponding to the regulatory categories defined in 3.3. The output CI is defined on [0, 100]. All membership functions are triangular, parameterised by their (a, b, c) corner points.

Table B1. Triangular membership function parameters.

Variable	Universe	Linguistic level	(a, b, c)
I (Expert importance)	[0, 10]	Low	(0, 0, 4)
I (Expert importance)	[0, 10]	Medium	(3, 5, 7)
I (Expert importance)	[0, 10]	High	(6, 10, 10)
C (Maintenance cost)	[0, 10]	Low	(0, 0, 4)
C (Maintenance cost)	[0, 10]	Medium	(3, 5, 7)
C (Maintenance cost)	[0, 10]	High	(6, 10, 10)
R (Regulatory tier)	[0, 10]	Low	(0, 0, 4)
R (Regulatory tier)	[0, 10]	Medium	(3, 5, 7)
R (Regulatory tier)	[0, 10]	High	(6, 10, 10)
CI (Criticality Index)	[0, 100]	Low	(0, 0, 40)
CI (Criticality Index)	[0, 100]	Medium	(30, 50, 70)
CI (Criticality Index)	[0, 100]	High	(60, 100, 100)

B.2 27-rule Mamdani rule base

The rule base follows the design principle that expert importance is the dominant criterion, with cost and regulatory tier acting as supporting evidence. Encoding L = 0, M = 1, H = 2, the consequent level for each (I, C, R) antecedent is determined by the score $s = 2 \cdot I\_lvl + C\_lvl + R\_lvl$, mapped to: Low if $s \leq 2$, Medium if $3 \leq s \leq 5$, High if $s \geq 6$. This yields the 7 Low / 13 Medium / 7 High rule distribution shown in Table B2.

Table B2. The 27-rule Mamdani rule base.

I	C	R = Low	R = Medium	R = High
Low	Low	Low	Low	Low
Low	Medium	Low	Low	Medium
Low	High	Low	Medium	Medium
Medium	Low	Low	Medium	Medium
Medium	Medium	Medium	Medium	Medium
Medium	High	Medium	Medium	High
High	Low	Medium	Medium	High
High	Medium	Medium	High	High
High	High	High	High	High

B.3 Override rule justification

The override rule "if $I \geq 9$ then $CI \geq 85$" formalises an asymmetry observed during expert elicitation: critical systems do not accumulate failure cost histories precisely because they are kept under tight maintenance discipline. Without intervention, an asset described as "instantly raise a job" by the facility manager, and therefore mapped to I in the High range, but with low recorded maintenance cost (C in the Low range) and a non-mandatory regulatory tier (R in Low or Medium) would defuzzify to a CI in the 40–70 band. This would under-classify infrastructure that domain expertise unambiguously identifies as critical. The override sets a CI floor at the lower edge of the High output membership function, ensuring that unambiguous expert evidence on importance is not overridden by sparse cost histories. The threshold $I \geq 9$ corresponds to the activation boundary of the I "High" membership function ($a = 6$, $b = c = 10$); within this range the High membership grade exceeds 0.75. The floor CI = 85 corresponds to the centroid of the High output rule under a maximally activated singleton consequent.

B.4 Suitability threshold derivation

The Yes/Maybe/No partition of the [0, 100] CI scale is anchored on the edges of the output membership functions. The threshold CI ≥ 65 corresponds approximately to the lower edge of the "High" output membership function, interpretable as "the defuzzified evidence is consistent with a High-criticality classification". The threshold CI < 45 corresponds approximately to the upper edge of the "Low" output membership function, interpretable as "evidence is consistent with Low criticality". The intermediate band 45 ≤ CI < 65 captures cases where rule activations span both adjacent output regions without committing to either, warranting a "Maybe" status that triggers manual review rather than automated enrolment in or exclusion from the predictive maintenance pipeline. The "Maybe" band therefore serves a procedural rather than algorithmic purpose: it identifies asset classes for which the quantitative evidence is inconclusive and a per-asset judgement is required. In practice the CI band sets the default classification, which the final suitability decision then adjusts in two ways consistent with the suitability rule of Section 3: a class governed by a mandatory pass/fail statutory inspection regime is recorded as No irrespective of its CI (for example Fire Safety and BMS), and a class that is either internally heterogeneous (for example Equipment, whose members range from predictively monitorable plant to inspection-only items) or sits at the boundary of a statutory regime (for example water heaters under HSE L8) is recorded as Maybe, flagging it for per-asset review even when its aggregate CI falls outside the 45 to 65 band. The classification therefore reflects the combination of the CI band, the regulatory filter, and class homogeneity, rather than the CI band alone.