Research Square

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

Construction and Application of an AI-Agent-Based Workflow for the Rapid Survey of Building Damage and Functionality

Xu Zheng
Shandong Earthquake Agency
Jin Liu
lj-hit@hit.edu.cn
Harbin Institute of Technology

Article

Keywords: Post-earthquake damage survey, Building functionality, Seismic resilience, Bayesian network, Multi-agent workflow, Human-machine collaboration

Posted Date: August 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10658253/v1

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

Additional Declarations: No competing interests reported.

Abstract

For a building's occupants, the assessment of its post-earthquake function is the most precise feedback available, and it governs the recovery strategy that follows. It nevertheless remains manual and subjective, and both conventional appraisal and intelligent methods stop at component-level damage, leaving room-level availability unaddressed. This study constructs an AI-agent-based workflow carrying a survey from on-site evidence to a resilience-oriented functional conclusion. Perception, clause-grounded knowledge reasoning, Bayesian inference, assessment, and reporting are decoupled into agents orchestrated by LangGraph, with uncertainty-triggered human review. A Bayesian network propagates observed damage from components through subsystems to rooms using the functional-state definitions of an established hospital resilience framework, so functionality follows from field evidence rather than expert impression. A hospital operating-room case and a scenario matrix, both driven by expert-provided node-aligned inputs, show the inference decoupling physical damage from functional loss. Identical component-level severe damage leaves the room unusable when the nonstructural subsystem fails, but degraded and operable when one function subsystem fails. Rule-based aggregation collapses this distinction and overestimates the loss by one grade; the result is stable under Dirichlet perturbation of the parameters. By binding code-grounded field evidence to probabilistic inference, the workflow brings resilience-oriented assessment within reach of routine post-earthquake practice.Article Highlights Building functionality differs from component damage; resilience-based assessment requires the two to be combined. Component and subsystem damage are coupled; grading components alone cannot measure a building's true loss. Standardized site photographs, code clauses and expert review are brought together in a single workflow.

1. Introduction

With ongoing urbanization, a large stock of aged buildings whose seismic fortification levels lag behind current standards is concentrated in densely populated urban centers. Their post-disaster functionality is a central concern of resilience-based earthquake engineering. A building is itself a complex system coupling structural components, nonstructural components, equipment, and mechanical, electrical and plumbing (MEP) systems. For buildings with stringent functional-continuity requirements, such as hospitals, a purely structural safety appraisal cannot answer whether the building can still operate after an earthquake. The assessment must therefore advance from component damage to system functionality, and from a binary safe-or-dangerous judgment to a graded functional-loss classification. A substantial literature has modeled post-earthquake building and hospital functionality, functional recovery, and resilience (Liu et al. 2022; Chavez et al. 2025; Sunil and Kamatchi 2025; Pei et al. 2023; Song et al. 2024; Ding et al. 2025). These studies commonly follow a component-to-service chain. Fragility or observed damage establishes the state of structural, nonstructural, MEP, and medical-equipment components. Dependency models then aggregate those states to rooms, departments, or clinical pathways. Functionality-time curves, treatment capacity, or recovery time finally translate the result into resilience metrics (Bruneau et al. 2003; Cimellaro et al. 2010). Field surveys of healthcare facilities struck by recent earthquakes confirm that the loss of clinical function is frequently governed by damage to nonstructural components, MEP systems, and medical equipment rather than by structural collapse (Achour and Miyajima 2020; Santarsiero et al. 2019). Recent reviews consolidate the corresponding functional-quantification and resilience-enhancement strategies (Liu et al. 2026; Bayraktar et al. 2026). Bayesian networks suit this task because they retain probabilistic evidence while representing coupled failures and unequal functional criticality, and they have been applied to both conventional and base-isolated hospitals (Liu et al. 2022; Chavez et al. 2025). However, their inputs are usually derived from analytical fragility models or predefined earthquake scenarios rather than from the image-based observations available in a rapid survey. Consequently, these works target performance-based design or regional resilience, and they seldom bind component-level damage evidence to a field-usable functional-loss inference. A practical link between visual damage evidence, code-grounded component appraisal, and probabilistic functional inference therefore remains underdeveloped, and it is the gap this workflow addresses.

Building-damage identification and seismic damage assessment have long relied on manual visual inspection, local non-destructive testing, and global monitoring. All three demand substantial expertise, and each has inherent limitations. A deeper problem lies in standardization and consistency. For the same damage, different standards differ in sampling, judgment basis, and grading, and they remain largely qualitative. The Chinese Standard for Appraisal of Reliability of Civil Buildings, GB 50292 (MOHURD 2015), and the Standard for Dangerous Building Appraisal, JGJ 125 (MOHURD 2016), are two such cases. Similar buildings therefore receive mutually contradictory conclusions, and report quality is uneven. Hashim et al. (2025) attribute this subjectivity to the absence of objective quantitative indicators and unified judgment criteria. It is precisely this lack of standardization that has motivated the combination of artificial-intelligence methods with traditional appraisal, in order to establish a more objective, quantifiable, and reproducible assessment system.

Deep learning and computer vision have markedly advanced the automated identification of building damage. Surface-defect identification has shifted from image-level classification to pixel-level semantic and instance segmentation: encoder-decoder and hybrid Transformer-convolution networks have progressively improved crack segmentation accuracy and robustness against complex backgrounds (Kang and

Cha 2022; Wang and Su 2022; Zhang et al. 2023; Li et al. 2024; Qin et al. 2024; Wu et al. 2024; He et al. 2024), and these methods have been extended to multiple defect types, high-resolution cracks, and annotation-efficient (semi- and self-supervised) segmentation (Arafin et al. 2024; Chu and Chun 2024; Wang and Su 2021; Bel-Hadj et al. 2025), and combined with unmanned-aerial-vehicle (UAV) and robot-based multi-sensor acquisition for efficient three-dimensional localization (Waqas et al. 2024; Hu et al. 2024; Tse et al. 2024). At the seismic-assessment level, vision- and learning-based methods have been applied at the component and individual-building scale to residual-capacity and damage-degradation assessment (Ji et al. 2023; Miao et al. 2023; Dogan et al. 2023), at the regional and remote-sensing scale to large-area damage grading (Yu et al. 2024; Wang et al. 2023b), and as machine-learning surrogates for fragility curves, rapid visual screening, and street-view building-typology identification (Kazemi et al. 2023; Bhatta and Dang 2023; Ghimire et al. 2022; Bektaş and Kegyes-Brassai 2025; Gomez et al. 2025; Wang et al. 2023a). Image-based approaches have also estimated building-scale seismic vulnerability directly from photographs (for example the VULMA project, Ruggieri et al. 2021), and machine-learning and deep-reinforcement-learning surrogates are increasingly used for structural identification and model updating at scale (Bruno et al. 2026).

Despite this considerable progress, existing research still exhibits several structural shortcomings. First, model performance remains constrained by variations in illumination, occlusion, viewpoint, and scale, and seismic damage samples are moreover scarce, class-imbalanced, and subjectively labelled, which weakens cross-structure, cross-region, and cross-standard generalization (Bel-Hadj et al. 2025; Gu et al. 2024). Second, most methods output qualitative masks or damage categories, and they lack engineering indicators such as crack width and spalling area as well as any relation between those indicators and residual capacity (Chu and Chun 2024; Ji et al. 2023). Third, detection, quantification, assessment, and reporting remain mutually disconnected, and no integrated, field-deployable system outputs standard-compliant appraisal reports (Kunlamai et al. 2024). Most methods furthermore neither quantify predictive uncertainty nor provide a human-machine review mechanism, which leaves the error rate of safety-critical decisions difficult to control (Pyle et al. 2022; Hüllermeier and Waegeman 2021). Finally, and most critically, existing vision and machine-learning methods stop at component-level damage identification and grading, so they can hardly infer the system-level functional loss of a building and cannot answer the core question of post-earthquake building functionality (Ji et al. 2023; Yu et al. 2024).

The rise of large language models (LLMs) and foundation models offers a new opportunity to overcome these bottlenecks. LLMs such as GPT-4 (OpenAI 2023) and LLaMA (Touvron et al. 2023) exhibit strong reasoning, planning, and tool-use capabilities, while vision foundation models such as Segment Anything (Kirillov et al. 2023) possess excellent zero-shot transfer ability. These models have been rapidly introduced into civil engineering, from GPT across the construction life cycle and safety-text mining (Saka et al. 2024; Smetana et al. 2024) to parameter-efficient fine-tuning of vision foundation models for crack and structural-damage segmentation (Ge et al. 2024; Ye et al. 2024). Multimodal large models can already perform expert-level disaster-image analysis (Areerob et al. 2025), and visual question answering for bridge inspection can be pre-trained directly on the image-text pairs present in existing inspection reports (Kunlamai et al. 2024). Domain standards, inspection reports, and expert experience are themselves high-value knowledge resources for artificial intelligence.

Furthermore, the research paradigm is shifting from a single model toward multi-agent workflows. LLM agents adopt a general brain-perception-action architecture (Xi et al. 2023; Wang et al. 2024) and suppress cascading hallucinations through tool use (Toolformer, Schick et al. 2023), deliberate reasoning (Tree of Thoughts, Yao et al. 2023), and standard-operating-procedure (SOP)-driven multi-agent collaboration (MetaGPT, Hong et al. 2023; ChatDev, Qian et al. 2024); in chemistry and medicine, tool-augmented research agents have achieved closed-loop autonomy (Boiko et al. 2023; M. Bran et al. 2024; Qiu et al. 2024). In civil and structural engineering, multi-agent workflows have emerged rapidly over the past two years for code-compliant RC design, structural-engineering automation, building-information-modeling (BIM) coordination and compliance checking, resilience-oriented RC lifecycle management, construction-safety inspection, geotechnical design, and UAV-based infrastructure inspection (Chen and Bao 2025; Liang et al. 2025; Dong et al. 2025; Zaitsev 2026; Hussain et al. 2026; Iversen and Huang 2026; Xu et al. 2025; Liu et al. 2025). For seismic damage assessment, Jiang et al. (2026) proposed MT-SDAChat, a multitask unified large vision-language model. These works are oriented either toward design and routine inspection or toward seismic damage assessment with a single large model; a multi-agent workflow for the rapid post-earthquake survey of building damage and functionality, in particular one advancing the assessment from component damage to system functionality, has not yet emerged.

Structuring seismic-appraisal standards, damage mechanisms, and assessment rules into a computable earthquake-engineering knowledge base is the key link between the above perception models and reliable decision-making. Three reasons make it so. First, consolidating the testing requirements, damage grading, and judgment criteria dispersed across standards such as GB 50292 (MOHURD 2015) and JGJ 125 (MOHURD 2016) supplies unified and objective criteria, which addresses the standardization problem identified above. It also encodes multi-material, multi-damage-coupling mechanistic knowledge as reusable rule-based reasoning. Second, such a base is a prerequisite for constraining LLMs and suppressing their hallucinations. Anchoring model outputs to an external, updatable, and traceable standard database through retrieval-augmented generation (RAG, Gao et al. 2023) makes the assessment conclusions explainable,

auditable, and traceable to specific clauses, which is particularly necessary for safety-critical appraisal. Third, and most importantly for the goal of this study, the knowledge base can jointly model the relationships among damage types, assessment criteria, repair rules, and component or system functionality. It thereby supports the advancement from component damage to system-functionality assessment. Knowledge graphs, informed machine learning, and semantic damage modeling provide representation and fusion approaches (Ji et al. 2022; von Rueden et al. 2023; Hamdan et al. 2021). However, a domain knowledge base and ontology binding damage grading to component and system functionality is still largely lacking in structural engineering.

One research gap therefore stands out. No workflow carries a post-earthquake survey through the full chain, from on-site damage determination, through functional assessment, to resilience evaluation. Existing vision and machine-learning methods are mostly single-task, component-level damage identification, and detection, quantification, assessment, and reporting remain disconnected stages (Ji et al. 2023; Ye et al. 2024; Liu et al. 2025). Functional and resilience assessment has meanwhile developed on analytical fragility models and predefined earthquake scenarios, at a remove from what a surveyor actually observes in the field. Closing that distance would quantify building functionality from field evidence rather than from expert impression, which is where the subjectivity of conventional survey mainly enters, and would bring resilience-oriented functional assessment within reach of routine engineering practice.

To address this gap, this study constructs an AI-agent-based workflow that carries a post-earthquake survey from on-site evidence, through functional assessment, to a resilience-oriented conclusion and a standard-compliant appraisal report. Its contributions are threefold:

1. A workflow spanning the full chain from on-site damage determination to functional and resilience evaluation is proposed. Perception, clause-grounded knowledge reasoning, functional inference, assessment, and report generation are decoupled into collaborating agents orchestrated with LangGraph over a shared state object, so that field evidence reaches a functional conclusion without leaving one traceable pipeline.
2. Building functionality is quantified directly from field evidence. A clause-indexed code knowledge base supplies traceable, individually attributed criteria, and a Bayesian network propagates the observed damage from components through subsystems to rooms using the functional-state definitions of an established hospital seismic-resilience framework (Liu et al. 2022). The room-level output is therefore a resilience-oriented functionality state rather than an aggregated component score, and it rests on code clauses rather than on the surveyor's impression.
3. The assessment is made auditable and field-deployable. Every conclusion is traceable to the clause and the evidence that produced it, and an uncertainty-triggered review routes only the least-confident cases to an engineer, which bounds the error rate of the automated output at a quantified cost in manual work (Section 4.4).

The remainder of the paper is organized as follows. Section 2 presents the framework, the key methods, and the module interface contracts, and Section 3 demonstrates the workflow on a hospital operating-room case. Section 4 then reports the perception measurement, the scenario-matrix experiments, and the discussion, and Section 5 concludes.

2. Proposed Approach

2.1 Overall Framework and System Layering

The proposed workflow adopts a layered multi-agent design. Collaborating agents undertake each stage, a unified state is passed among them, and LangGraph orchestrates the state transitions so that the modules communicate only through that state. This integrates the originally disconnected detection, quantification, assessment, and reporting stages into a traceable closed loop with human-in-the-loop intervention. A knowledge layer couples the code knowledge base with a Bayesian network, so that the assessment is grounded in code clauses while the functional loss of the system is inferred from component damage. Uncertainty measurement in turn drives the human-machine collaboration: high-uncertainty samples are transferred to manual review, and the expert corrections are returned to an annotated sample pool. The active-learning loop that would consume this pool is designed but not implemented or validated here (Section 4.5).

From input to report output, the system is divided into the following layers, each undertaken by dedicated agents with the unified state passed among them (Fig. 1):

- **Input layer**: receives on-site photos and videos (from phones, UAVs, or handheld devices) with their metadata (GPS, time, view angle, and reference scale), and can import existing documents such as historical appraisal reports, design drawings, and codes.
- **Task-scheduling agent**: the unified entry point. It normalizes the text and image inputs, classifies the query as a damage-identification or functional-assessment task using a language-model classifier with a keyword fallback, dispatches the extracted parameters to the specialized agents, and manages the task lifecycle.

- **Perception** layer: damage detection, grading, scale estimation, and confidence reporting are carried out in a single structured request to the vision backend, which returns the component identifier, damage category, severity, scale, and confidence for every component it reports. The backend is a general multimodal model called through an OpenAI-compatible interface and is replaceable behind that contract; the model used and measured in this study is given in Section 4.2.
- **Knowledge-base layer (code knowledge base plus Bayesian network).** On one hand, seismic-appraisal codes such as GB 50292, JGJ 125, FEMA, ASCE 41, Japan's emergency guidelines and Italy's AeDES, together with damage-grading standards and repair recommendations, are held as individual clauses. Each clause is labelled with its standard, jurisdiction, clause number, and the component types and damage types it governs. The clauses applicable to a detection are selected through that index, and further standards are incorporated by appending clauses without altering any other module. The language model renders the selected clause text into the assessment narrative and the report, but it does not decide which clause applies. Clause attribution therefore stays deterministic and auditable, and a mis-ranked retrieval cannot silently change an appraisal conclusion. On the other hand, a Bayesian network models damage propagation and functional dependency from components through subsystems (structural, nonstructural, equipment, MEP) to rooms and buildings, as detailed in Section 2.2. The candidate components offered to the perception agent are drawn from this taxonomy, narrowed by the building ledger, and canonicalized so that members the codes do not treat as independent are folded into the member that governs them (Section 4.2). Because these standards differ in scope, legal status, and damage-classification systems, they are kept as separately attributed clauses rather than a single interchangeable scale, and no cross-regional equivalence between them is claimed. The corpus used here is an illustrative seed of ten clauses spanning nine standards and five jurisdictions, sized to exercise the mechanism rather than to cover the standards in force.
- **Assessment agent**: fuses the perception results, code rules, and Bayesian posterior into the functional-loss grading from component damage to the room or building level, and into the repair priority. It marks the sample as pending manual review when the uncertainty exceeds a preset threshold.
- **Report-generation agent**: fills an existing appraisal-report template with evidence images, damage location, severity, confidence, recommended measures, and the corresponding code clauses, and exports a structured, auditable report as JSON and HTML, with PDF rendering available as an option.
- **Learning loop (designed, not yet implemented)**: the results of manual review are written back to an annotated sample pool, which is intended to feed incremental training under an uncertainty-prioritized active-learning strategy. Only the corrections-to-pool interface is realized here.

2.2 Key Methods

Multi-agent collaboration and orchestration. LangGraph defines each of the above stages as a node in a state graph and passes the unified state object among the nodes, ensuring the traceability and recoverability of the process. The roles are heterogeneous rather than uniform: M1 and M2 are language- and vision-model calls, M3 to M5 are deterministic services (the clause index, the Bayesian inference service, and template rendering), and M6 is a human step, so the workflow is agentic in its role decomposition and its contracts rather than in free-form dialogue between models. Consistency between the two assessment paths is checked deterministically: when a component is graded Severe while every subsystem remains below Moderate, the divergence is recorded and raises the review flag. Borrowing MetaGPT's idea of encoding standard operating procedures into multi-agent collaboration (Hong et al. 2023), the framework constrains the standardized appraisal steps into the module sequence to suppress cascading hallucinations. The choice of an agentic, LLM-driven architecture over a single monolithic model or a fixed script is motivated by the need for tool use, SOP-constrained role separation, natural-language grounding in heterogeneous codes and reports, and independently replaceable modules with graceful degradation. Two of these components are measured against a simpler alternative. Section 4.2 quantifies what the code-derived scaffolding of the perception stage contributes over a naive single prompt, and Section 4.4 quantifies the error rate that the uncertainty-triggered review buys per unit of manual work. An end-to-end comparison of the orchestration itself against a purely scripted pipeline is not attempted here.

Coupled knowledge-base and Bayesian-network system-functionality assessment. This method is the core of the advancement from component damage to system functionality. The Bayesian-network modeling develops from the authors' probabilistic framework for the seismic resilience of hospital buildings (Liu et al. 2022), which revealed the interdependencies among departments, rooms, and internal components for the calculation of their availabilities. The present study extends that framework into an online functional-inference layer, driven by perception evidence and embedded in the multi-agent workflow. The network takes components as bottom-level nodes, subsystems (structural, nonstructural, equipment, MEP) as intermediate nodes, and room or whole-building functionality as top-level nodes. Component damage states follow Appendix E of the Chinese Standard for Seismic Resilience Assessment of Buildings, with medical equipment following Appendix A of the Chinese Standard for Seismic Resilience Design of Healthcare Buildings, and structural components following GB 50292 and JGJ 125. The room functional-state definition is adopted from Liu et al. (2022), and the subsystem

conditional probabilities are generated by a transparent weakest-link rule. Appendix A documents the full structure, priors, and conditional probability tables (CPTs) in Tables A1 and A2. Given the component-damage evidence from the perception layer, the network infers the posterior probability that the room or building functionality is in each state, namely intact, slight, moderate or severe damage. It thereby aggregates discrete component-damage information into a system-level functional-loss conclusion. The structure can be configured per building type. For medical buildings, key-department equipment and MEP subsystems can be incorporated to assess functional continuity. Bayesian networks with expert-elicited or rule-based CPTs have been widely applied to the post-earthquake functionality and resilience assessment of hospitals and healthcare networks (Liu et al. 2022; Chavez et al. 2025; Sunil and Kamatchi 2025; Pei et al. 2023; Song et al. 2024; Ding et al. 2025), which supports the modeling choices adopted here.

Scale quantification and uncertainty-driven human-machine collaboration. Existing methods tend to emphasize qualitative description while neglecting quantitative indices. The perception layer therefore estimates crack width, spalling area, and the like against reference objects visible in the image, which shifts the severity judgment from classification to quantification. A confidence is returned with each detection, and when the aggregated uncertainty exceeds a threshold the sample is automatically transferred to manual review. The single-pass confidence used here is nevertheless not calibrated against correctness (Section 4.2), and ensemble-based uncertainty quantification (Pyle et al. 2022) is the intended remedy rather than part of the present implementation. The review results correct the current assessment and are fed back to an annotated sample pool (Bel-Hadj et al. 2025; Andriotis and Papakonstantinou 2021).

2.3 Unified State Object and Task-Scheduling Mechanism

This workflow decouples agent collaboration and task scheduling through a unified state object combined with state-graph orchestration. Borrowing the blackboard architecture, a unified state object (State) runs through the entire process as the shared data contract. Each agent takes only the State as input and returns incremental updates to it, without being coupled to any other agent. Module replaceability and process evolvability follow directly from that contract. The main State fields are the task identifier and type (task_id, task_type), images and metadata (image_meta), perception detections (detections), code-retrieval evidence (kb_evidence), Bayesian results (bn_result), assessment conclusions (assessment), the review flag (need_review), the report (report), and the processing trace (history). Specifically, history uses an accumulative reducer that merges the execution records appended by each node, which makes the reasoning process traceable and auditable.

At the orchestration layer, each agent is registered as a node in a LangGraph state graph, with directed edges defining the scheduling relationships. The resulting topology is the module chain and conditional review branch of Fig. 1. The regular path runs from task scheduling (M1) through perception (M2) and knowledge-base and Bayesian inference (M3) to assessment (M4). A conditional edge after M4 then routes on the need_review flag. When the uncertainty is below the threshold and no mandatory review is required, the flow proceeds directly to report generation (M5); otherwise it passes through manual review (M6) before returning to M5. This review-on-demand routing controls the assessment error rate while preserving throughput.

For asynchronous, resumable on-site processing, the system uses task_id as the LangGraph thread_id and persists each task's state through an SQLite checkpointer, so that tasks run asynchronously in the background while the client polls the result by task_id. The manual-review stage adopts an interrupt-resume mechanism in which the M6 node pauses the execution of the whole graph and exposes the content to be reviewed. After the expert submits corrections, those corrections are injected with Command(resume) and execution continues from the breakpoint, which completes the closed loop without losing context.

2.4 Module Functions and Interface Contracts

To ensure loose coupling and replaceability, the modules collaborate through explicit input/output contracts, so that the internal implementation of a module (the perception backend, the clause index, or the inference engine) can evolve independently of the overall topology. Table 1 lists the core functions and interface contracts.

Each key module also carries a graceful-degradation strategy. When the commercial multimodal interface is unavailable the perception module falls back to a deterministic default output, and when the Bayesian-inference service is unreachable the assessment module replaces posterior inference with rule-based aggregation. The failure of any single external dependency therefore does not interrupt the graph, and the workflow still produces reports under unstable on-site connectivity.

Table 1
Core functions and interface contracts of each module.

Module	Core function	Input	Output (written to State)
M1 Task scheduling	Parse input; task classification (identify/assess); parameter validation and dispatch; lifecycle management	User query; image and metadata	task_type; normalized image_meta
M2 Perception	Damage detection and grading; scale estimation and confidence reporting	image_meta	detections (component, type, severity, scale, confidence, uncertainty)
M3 Knowledge base + Bayesian	Clause selection by component and damage type; component → subsystem → room functional posterior inference	detections; building type	kb_evidence; bn_result
M4 Assessment	Fuse posterior and rules; functional loss grading; repair priority; review decision	detections; kb_evidence; bn_result	assessment; need_review
M5 Report generation	Render a structured report with evidence images, grades, code basis, and recommendations	full State	report (JSON, HTML; PDF optional)
M6 Human review	Expose low-confidence samples; receive expert corrections; write back to the annotated sample pool	need_review samples	corrected assessment (with need_review cleared)

3. Case Study

3.1 Case Setup and Positioning

This section demonstrates the multi-agent workflow of Section 2 on a realistic medical-building seismic damage scenario. The case is followed module by module along the full chain. The focus is the core capability of advancing from component damage to room functional loss, and the feasibility and interpretability of that step.

To focus on the core innovation and ensure reproducibility, the perception stage (M2) in this case adopts damage-detection results given by experts from on-site interpretation. Component identifiers and severities are strictly aligned to the node definitions and state spaces of the Bayesian network, which makes them equivalent to the output of an ideal perception module. The measured performance of the perception module itself is reported in Section 4.2. Code retrieval, Bayesian functional inference, damage and functional assessment, and report generation (M3-M5) are all executed by real system modules, and the Bayesian inference is computed by an independent service (pgmpy variable elimination). One constraint of the available case data should be noted. Existing hospital appraisal reports are mostly computation-based seismic load-bearing appraisals, and their conclusions are not of the same semantics as the room functional loss this workflow outputs. This section is therefore positioned as a feasibility and interpretability demonstration rather than a quantitative validation against expert ground truth. That validation is left to future work using case data with functional conclusions (Section 4.5).

3.2 Pipeline Demonstration

The case object is a hospital Class-1000 laminar-flow operating room (the room node Class-1000 OR-1 in the Bayesian-network template), in a post-earthquake rapid survey in which the room exhibits moderate damage. Each stage is presented in the module order of Fig. 1.

Input and task scheduling (M1). On-site photos, videos, and metadata, together with existing documents, enter the system through the input layer. The task-scheduling agent classifies this query as a functional-assessment task, extracts the room identifier and building type, dispatches them to the specialized agents, and registers the task lifecycle.

Perception (M2). The component-level damage detections are shown in Table 2. Five items in total cover key components of the three subsystems, namely nonstructural, MEP, and medical-equipment. The component identifier of each detection is aligned with the Bayesian-network nodes, and the severity is taken from the state space of the corresponding node. As noted in Section 3.1, these detections are expert-given, that is, an ideal M2 output, in order to isolate the assessment stage and ensure reproducibility. In real deployment M2 is produced by the vision backend measured in Section 4.2.

Table 2
Expert-given component-level damage detections of the case. The numeric suffix of each component is its Bayesian-network node identifier (Appendix A).

Component (node)	Damage type	On-site severity	Scale	Confidence
Masonry infill wall_325	Crack	Moderate	Crack width 4.5 mm	0.91
Suspended ceiling_324	Spalling	Moderate	Spalling area 180 cm²	0.86
Door/window_327	Displacement	Intact	n/a	0.88
Medical-gas pipeline_321	Displacement	Intact	n/a	0.80
Anesthesia machine_315	Displacement	Intact	n/a	0.83

Knowledge retrieval and Bayesian functional inference (M3). The knowledge-base agent selects, by component and damage type, the governing clauses from GB 50292 (MOHURD 2015) (§ 5.4.5, § 5.4.6 and Table 6.4.3) for the crack grading of masonry walls. It also selects the damage-state definitions of Appendix E of the resilience standard for the nonstructural components. The Bayesian network then takes the component-level detections as hard evidence at the root nodes and infers along the component-subsystem-room propagation structure (Fig. 2). The functional-loss posterior of Class-1000 OR-1 concentrates at Moderate (P = 0.976), with an expected loss of 0.675. At subsystem level the nonstructural system is Moderate. In contrast, the MEP and medical-equipment systems, which carry the core functions of the operating room, remain Intact, because the observed displacements reach no damage state defined for those components.

Damage and functional assessment (M4). The assessment agent fuses the perception results, code rules, and Bayesian posterior. At the component level, the rule engine checks the measured scale against the thresholds of GB 50292. A non-load-bearing crack wider than 1.5 mm already fails the serviceability grade bs (Table 6.4.3), whereas a load-bearing wall is downgraded only once the crack exceeds 5 mm (§ 5.4.6). A masonry crack of 1.5-5 mm therefore corresponds to Moderate, and the infill-wall crack of 4.5 mm is confirmed as Moderate. At the room level, the infill wall and the ceiling are the most damaged, but both are nonstructural. The medical-equipment and MEP subsystems that carry the operating room's core functions are undamaged. The room functional loss is therefore graded Moderate rather than Severe: component physical damage does not equal loss of room function. The repair priority, in descending order of severity and uncertainty, is masonry infill wall > suspended ceiling > medical-gas pipeline > anesthesia machine > door/window.

Report generation (M5). The report-generation agent fills the appraisal-report template with the evidence images, damage location, severity and confidence, the code-clause basis above, the subsystem and room functional-loss grading, and the repair recommendations, and exports a structured, auditable report as JSON and HTML. Both the component-level conservative grade and the room-level functional conclusion are recorded with their code provenance, which facilitates manual verification and compliance traceability.

Human review (M6). Here the maximum detection uncertainty is below the threshold and the Bayesian and rule conclusions do not diverge, so the workflow completes fully automatically; the review mechanism and its triggering conditions are analyzed in Section 4.4.

This case runs the proposed workflow on a realistic medical-building seismic damage scenario, mapping each stage onto the Section 2 architecture and producing the component-level grading, the subsystem and room functional-loss grading, the code basis, and the repair priority. The results preliminarily verify the feasibility and interpretability of the coupled knowledge-base and Bayesian-network method, and in particular the ability of Bayesian functional inference to distinguish component physical damage from room functional availability. Section 4 examines the workflow systematically under different inputs, the incremental contribution of the Bayesian inference over rule degradation, and the triggering of human review.

4. Experiments and Results

4.1 Experimental Design

Research questions and positioning. This section reports two experiments. The perception module is executed on field photographs and measured against expert annotation (Section 4.2), and the functional-inference layer is scanned over a controlled scenario matrix (Section 4.3). They address two questions. RQ1 asks whether the workflow can produce, for a realistic post-earthquake hospital room scenario, the component-level grading, the subsystem and room functional-loss posterior, the code basis, and the disposal recommendations; Section 3 already demonstrates this. RQ2 asks whether the component-subsystem-room Bayesian inference yields finer-grained conclusions, more faithful to functional availability, than grading components by rules and then coarsely aggregating them. As stated in Section 3.1, the

functional-inference stage is positioned as a proof-of-concept and systematic scenario analysis rather than a quantitative validation against expert ground truth (Section 4.5).

Experimental object and Bayesian network. The experimental object is a hospital Class-1000 operating room, corresponding to the room node Class-1000 OR-1 in the Bayesian-network template. Following the hierarchical modeling of the authors' previous work (Liu et al. 2022), fourteen component leaf nodes, grouped into the nonstructural, MEP and medical-equipment subsystems, feed the three subsystem nodes and the four-state room node. Each component carries only the damage states the standards define for it, so the state spaces are not uniform, and Appendix A gives the full node inventory, priors, and CPTs in Tables A1 and A2. Consistent with Section 3, the perception stage (M2) is given expert node-aligned detections while M3-M5 are run by real modules, with the Bayesian inference computed by an independent service (pgmpy variable elimination) and the demonstration scripts provided as supporting code for reproducibility.

Subsystem functional-aggregation convention. This paper adopts a unified and transparent weakest-link rule: the functional state of a subsystem takes the state of its most severely damaged component. In a post-earthquake rapid-assessment context, whether a subsystem maintains its function often depends on its most severely damaged critical component, and a single severe failure of the medical-gas pipeline can for example disable the gas-supply subsystem, which makes this a conservative and engineering-common assumption. The deterministic CPT nevertheless does not make the network deterministic, because undetected components are marginalized over their priors and detection confidence propagates as evidence, so the subsystem and room outputs remain probability distributions. The room-level CPT, in contrast, treats the three subsystems asymmetrically, and thereby carries the professional judgment of which subsystem failure is more fatal to the operating-room function.

Scenario-matrix design. A single case is only one point in the input space. The workflow's behavior is therefore examined by a controlled scan along two dimensions: the subsystem in which the damage is concentrated (nonstructural / MEP / medical-equipment) and the severity tier. The available tiers differ by subsystem because the standards define different state sets. Moderate exists only for the nonstructural subsystem, and Slight does not exist for medical equipment, so the medical-equipment row has a Severe scenario only. Two composite scenarios are added: D1, the real case of Section 3, and D2, several subsystems severe simultaneously. Each scenario is subjected to real Bayesian inference (Table 4 and Fig. 3).

4.2 Measured Performance of the Perception Module

Sections 3 and 4.1 supply the perception stage with expert-provided, node-aligned detections in order to isolate the functional-inference layer. This section reports what the perception agent produces when it is executed on field photographs. The accuracy of the first stage of the chain, and the manner in which its errors propagate downstream, can then be stated quantitatively rather than assumed.

Data and reference. Forty photographs from a post-earthquake field survey of rural masonry dwellings were used (Laiwu and neighboring villages, Jinan, September 2022). Three were set aside as non-survey material, namely a demolished building, a working photograph, and a wide-angle overview with no legible component-level damage, which leaves 37. Every photograph was annotated before any model was run. The annotation records component type, damage type, damage state, and a measured or estimated quantity wherever a reference object was visible. It follows the same code criteria the workflow applies: § 5.4 and Table 6.4.3 of GB 50292 for masonry walls, § 5.5 for timber roof members, § 5.2 for concrete ring beams and constructional columns, and Appendix E of the resilience standard for applied finishes.

Because the initial annotation recorded the salient damage of each photograph rather than an exhaustive inventory, precision could not be read from it directly. Every model claim that did not coincide with the annotation was therefore adjudicated by the same expert. The three outcomes are damage that is present but was not recorded initially, which counts as a true positive; a component present but undamaged; and a component absent from the photograph. The latter two are false positives of distinct kinds. Recall is reported against the initial annotation alone, which was produced without sight of any model output and is therefore unbiased. The adjudication only reclassifies false positives and cannot recover a missed component.

Configuration. The perception agent is a general multimodal model (Qwen-VL-Max, commercial API) used zero-shot, without task-specific fine-tuning. It was given the component taxonomy of Section 2.1, in which each component carries only the damage states the codes define for it. Two properties of that taxonomy matter here. First, gable walls and door/window openings are not independent component classes. JGJ 125 § 5.7.1 lists the door/window lintel among the load-bearing envelope members, § 5.7.3 requires it to be assessed under the masonry provisions, and a gable is itself a masonry bearing wall. Both are therefore folded into the masonry bearing wall, whose criteria retain their characteristic phenomena. Second, the candidate list is restricted to the component types recorded in the building ledger for this construction class. Those are single-storey masonry walls, a timber pitched roof, ring beams and constructional columns, floor slabs, infill walls, and applied finishes. Parapets, frames, curtain walls, lifts, MEP systems, and medical equipment are physically

absent here and are not offered as candidates. Images were reduced to a long side of 1536 px before transmission. Thirty-eight requests were issued in total: the 37 evaluation photographs plus one whose reference status was left unresolved and which is excluded from every score. All of them completed without failure. Inference runs against the commercial API and uses no local GPU. Averaged over the requests, one photograph consumed 4.8 k prompt and 0.2 k completion tokens and 13.2 s of wall-clock time, or 182.0 k tokens and 502 s in total.

Component-level detection. Taking one component type on one photograph as the unit of evaluation, the perception agent attains a precision of 0.689 (95% CI 0.543–0.805), a recall of 0.646 (0.504–0.766) and an F1 of 0.667 (Table 3). The composition of the errors is more informative than the aggregate. Once the ledger restricts the candidate list, no detection refers to a component that is absent from the photograph, and every remaining false positive is a real component on which damage has been over-reported. Twenty-two claims did not appear in the initial annotation. The adjudication confirmed eight of them as genuine damage that the annotator had not recorded, and classified the remaining fourteen as components that were present but undamaged. Part of the apparent imprecision therefore reflects the density of the reference rather than the behavior of the model.

Damage-state grading. Among the component types on which the model and the annotation agree, the damage state is reproduced exactly in 30.4% of cases and falls within one tier in 95.7% (0.790–0.992). A single grading, 4.3%, is displaced by more than one tier. The errors are directionally consistent: the model grades downward, placing six of the nine components the expert judged Severe at Moderate. Precision rises monotonically with the graded severity: 0.562 for Slight (n = 64), 0.682 for Moderate (n = 22) and 1.000 for Severe (n = 8, CI 0.676-1.000). These three figures are counted per individual detection, 94 in all, and are therefore not commensurable with the headline precision. No component was falsely reported as Severe in this sample. Both the over-reporting and the under-grading therefore concentrate in the tiers whose consequences for triage are smallest.

Scale estimation. Where the annotation records a measured crack width, the model underestimates it in every one of the seven comparable cases. The mean absolute error is 5.97 mm and the mean relative error 64.5%. This is the proximate cause of the downward grading rather than a failure of the grading logic itself. The thresholds are applied correctly to the width the model believes it has measured, and a width that is too small yields a state that is too favourable. Part of the deficit is attributable to the pre-transmission reduction of the images, which shortens a 3 mm crack in a 4032 px photograph from roughly four pixels to fewer than two.

Confidence calibration. The confidence returned by a single inference pass is not calibrated against correctness. In the highest bin the mean reported confidence is 0.916 while the observed accuracy of the damage state is 0.316, and the expected calibration error over the sample is 0.591. The confidence nevertheless carries ordinal information: precision increases monotonically across the confidence bins, and discarding detections below 0.75 raises precision to 0.789 while retaining 30 of the 31 true positives. That threshold was selected on this evaluation set and is therefore reported as a sensitivity curve rather than adopted as part of the configuration; the full sweep is given in Table 3.

Ablation of the domain scaffolding. The perception stage is supplied with a component taxonomy, a ledger-restricted candidate list, and a per-component state space, all of which cost prompt length and tokens. To measure what that scaffolding contributes, the same 37 photographs were processed a second time with a naive single prompt. That prompt keeps the output schema and the four-level severity vocabulary but supplies no component list, no state spaces, and no code criteria, which shortens the system prompt from 5508 to 294 characters. Both arms are scored against the initial annotation alone, so the recall quoted here for the constrained arm, 0.575, is lower than the adjudication-based 0.646 reported above. Two differences follow. First, none of the 41 claims returned by the naive prompt names a component in the vocabulary the codes define. It reports colloquial classes such as wall, roof, and window, where the standards distinguish masonry bearing walls, roof systems, and openings assessed under the masonry provisions. No such claim can be bound to a clause or entered as evidence at a network node, so the functional inference cannot run on that output at all. Second, the deficit is not merely lexical. Under a deliberately charitable manual mapping of those colloquial names onto the code classes, the naive arm still recalls 0.375 of the annotated components against 0.575, and it examines 1.11 component types per photograph against 2.76. The scaffolding therefore raises recall by about one half and widens the range of components considered, at the cost of 3.5 times the input tokens (182.0 k against 51.4 k) and roughly a third more wall-clock time. The precision of the naive arm is not reported, because establishing it would require the same clause-by-clause expert adjudication applied to the constrained arm.

Implications for the workflow. Three consequences follow. First, the perception stage, not the functional inference, is the accuracy bottleneck of the present chain, and the dominant term within it is metric estimation rather than recognition or grading. Second, the concentration of error in the low-consequence tiers, together with the absence of false Severe calls, is consistent with the intended role of the workflow as preliminary screening and triage. Third, the poor calibration of single-pass confidence argues directly for the uncertainty-

triggered review of Section 4.4 and for the ensemble-based uncertainty estimate identified in Section 2.2 as future work. A confidence that does not track correctness cannot by itself gate a safety-critical conclusion.

Limitations of this evaluation. The sample is small and drawn from a single construction class in a single region, so the confidence intervals above are wide and the grading statistics rest on 23 matched components. The annotation and the adjudication were performed by the same expert, so inter-rater agreement is not established, and localization is not measured because the reference contains no bounding boxes. A larger and more varied photograph set, independent double annotation, and a comparison across vision backends are listed in Section 4.5.

Table 3
Measured performance of the perception module on 37 field photographs of rural masonry dwellings (Qwen-VL-Max; 95% confidence intervals by the Wilson score method).

Quantity	Value	95% CI	n
Precision	0.689	0.543–0.805	45
Recall (independent annotation)	0.646	0.504–0.766	48
F1	0.667	n/a	n/a
False positives: component present, damage over-reported	14	n/a	n/a
False positives: component absent from photograph	0	n/a	n/a
Damage state reproduced exactly	0.304	0.156–0.509	23
Damage state within one tier	0.957	0.790–0.992	23
Precision, detections graded Slight	0.562	0.441–0.677	64
Precision, detections graded Moderate	0.682	0.473–0.836	22
Precision, detections graded Severe	1.000	0.676–1.000	8
Crack width, mean absolute error	5.97 mm	n/a	7
Crack width, mean signed error	-5.97 mm	n/a	7
Expected calibration error	0.591	n/a	23
Confidence threshold	Precision	Recall	F1
none	0.689	0.646	0.667
≥ 0.70	0.705	0.646	0.674
≥ 0.75	0.789	0.638	0.706
≥ 0.85	0.793	0.535	0.639
≥ 0.95	0.778	0.171	0.280

4.3 Scenario-Matrix Results

Table 4
Scenario-matrix results (real Bayesian inference).

Scenario	Damaged subsystem·severity	Room posterior [Intact,Slight,Moderate,Severe]	Room level	Expected loss	Damaged subsystem level	Review
A1	Nonstructural·Slight	[0, 0.729, 0.234, 0.037]	Slight	0.436	Nonstructural=Slight	No
A2	Nonstructural·Moderate	[0, 0, 0.963, 0.037]	Moderate	0.679	Nonstructural=Moderate	No
A3	Nonstructural·Severe	[0, 0, 0, 1.000]	Severe	1.000	Nonstructural=Severe	No
B1	MEP·Slight	[0, 0.665, 0.281, 0.054]	Slight	0.463	MEP=Slight	No
B2	MEP·Severe	[0, 0, 0.775, 0.225]	Moderate	0.742	MEP=Severe	No
C2	Medical-equip.·Severe	[0, 0, 0.914, 0.086]	Moderate	0.695	Medical-equip.=Severe	No
D1	Composite·Real case	[0, 0, 0.976, 0.024]	Moderate	0.675	Nonstr.=Moderate; MEP/equip.=Intact	No
D2	Composite·Multi-severe	[0, 0, 0, 1.000]	Severe	1.000	All three subsystems=Severe	Yes

Within the same subsystem, the room functional loss rises monotonically with the component severity. The nonstructural subsystem goes from Slight to Moderate to Severe (A1 → A2 → A3), with expected losses of 0.436, 0.679 and 1.000. The MEP subsystem goes from Slight to Severe (B1 → B2), with 0.463 and 0.742. The medical-equipment subsystem has no Slight state defined in the standard and therefore appears only at Severe (C2, 0.695). The inference thus responds correctly and interpretably to the input severity. To test how far these conclusions depend on the expert-assumed parameters, the fourteen component priors and the room-level CPT were jointly perturbed by Dirichlet resampling (N = 1000). Figure 3 summarizes the resulting distribution. Softening the one-hot room CPT moves probability mass toward the lower states, so the perturbed mean lies slightly below the nominal value, which is therefore drawn separately. No scenario changes its qualitative room-functionality grade under this perturbation, the standard deviation of the expected loss staying at or below 0.040. Moreover, the separation between severe nonstructural damage (A3, 0.88–0.97) and a single severe function subsystem (B2/C2, 0.64–0.74) persists throughout. The subsystem-discrimination behavior is therefore not an artifact of the particular CPT values.

The most illuminating result of the matrix is the horizontal comparison at the Severe tier. For the same component-level Severe damage, different damaged subsystems yield markedly different room conclusions. Severe nonstructural damage (A3) makes the room Severe (expected loss 1.000), that is, physically unsafe and unoccupiable. A single severe function subsystem, B2 for MEP or C2 for equipment, makes it only Moderate (0.742 and 0.695), degraded but still partly operable. Only when several function subsystems are severe simultaneously (D2) does the room become Severe again. This difference stems not from the subsystem-aggregation rule, which is a uniform weakest link, but from the asymmetric room-level CPT. That table encodes three things: severe nonstructural damage threatens occupiability, a single function-subsystem failure still allows degraded operation, and only multiple function failures cause total loss of room function. The Bayesian network thereby decouples component physical damage from room functional loss.

To quantify the incremental contribution of the Bayesian functional inference, each scenario is run on two paths. The first is real Bayesian inference, and the second is the rule-aggregation fallback, which propagates the single most severe component to the whole-room level. Take the severe MEP scenario (B2) as an example (Fig. 4). The rule path propagates the severe medical-gas pipe straight to the whole room and judges it Severe, whereas the Bayesian path recognizes that a single failed function subsystem still leaves the room in degraded but partial operation and gives Moderate. Across all scenarios (Fig. 5), the rule degradation gives Severe in B2 and C2 where the Bayesian inference gives the finer Moderate, while elsewhere, in A1, A2, A3, B1, D1 and D2, the two conclusions agree. Rule aggregation is thus conservative but distorted, since it systematically overestimates the room functional loss by one grade at B2 and C2 and reflects neither the subsystem to which the component belongs nor functional availability. The two paths are nevertheless complementary rather than mutually exclusive: the component-level rule check against the code thresholds is retained as a quantitative safety floor, while the room and subsystem functional conclusions follow the Bayesian posterior.

4.4 Uncertainty-Triggered Human Review

The human-machine collaboration of the workflow (M6) triggers review under three conditions: an explicit forced review, the maximum detection uncertainty exceeding the threshold (default 0.30), or a divergence between the Bayesian and rule conclusions. In the scenario matrix the rule fallback is not executed in parallel, so only the uncertainty condition is active. Only D2 is flagged, because that scenario contains a high-uncertainty detection, the anesthesia machine graded severe at uncertainty 0.35. The divergence condition is exercised

separately in Section 4.3, where B2 and C2 would also be flagged in a deployment that runs both paths. The workflow pauses with interrupt. After the manual correction it resumes via Command(resume) and continues to generate the report, while the other scenarios are produced fully automatically.

Error rate against review workload. The single-pass confidence is not calibrated, but it carries ordinal information (Section 4.2). The review threshold therefore sets an explicit trade-off between the error rate of the automated output and the volume of manual work. Reading the threshold sweep of Table 3 as a routing rule makes that trade-off quantitative: detections below the threshold are sent to M6 instead of being discarded. Routing the least-confident 11.7% of detections, at a threshold of 0.75, raises the precision of the automatically accepted set from 0.689 to 0.789 at almost no cost in recall, which falls only from 0.646 to 0.638. All eight detections graded Severe remain in the automated set. A threshold of 0.85 raises the workload to 57.4% of detections for a further precision gain of only 0.004. At 0.95 the workload reaches 89.4% while three of the eight Severe detections are diverted, so the automated output loses precisely the cases that matter most for triage. The deployed default, uncertainty 0.30 or equivalently confidence 0.70, routes 1.1% of detections and is therefore deliberately conservative. These thresholds were read off the evaluation set and are reported as a sensitivity curve rather than a calibrated operating point.

Effect of the expert correction. To make the consequence of a review explicit, scenario D2 was re-run after the reviewing engineer acted on the flagged detection (Table 5). The anesthesia machine was re-judged as Intact, since the observed displacement reaches none of the damage states the code defines for it. That correction returns the medical-equipment subsystem to Intact, yet the room conclusion stays Severe. Under the asymmetric room-level CPT (Appendix A.4), a severe nonstructural subsystem alone already renders the room unoccupiable, so the correction changes the diagnosis and the repair priority without changing the room grade. A second correction re-measures the infill-wall crack at 4.5 mm, inside the 1.5-5 mm moderate band of GB 50292 rather than above the severe threshold. This one does change the room conclusion, from Severe to Moderate, and the expected loss falls from 1.000 to 0.728. Human review is therefore not cosmetic. It acts on the same evidence channel as the perception layer and is propagated by the same inference. Each correction is appended to State.history with its author and timestamp and written to the annotated sample pool, so every published conclusion remains traceable to the evidence actually used. The incremental retraining that would consume that pool is designed but not implemented (Section 2.1).

Table 5
Effect of the expert correction on scenario D2.

Case state	Nonstructural	MEP	Medical equipment	Room posterior [Intact, Slight, Moderate, Severe]	Room level	Expected loss
M4 output (review triggered)	Severe	Severe	Severe	[0, 0, 0, 1.000]	Severe	1.000
After correction 1: anesthesia machine re-judged Intact (no damage state met)	Severe	Severe	Intact [0.849, 0.151]	[0, 0, 0, 1.000]	Severe	1.000
After correction 2: infill-wall crack re-measured 4.5 mm, below the 5 mm severe threshold	Moderate [0, 0, 0.961, 0.039]	Severe	Intact [0.849, 0.151]	[0, 0, 0.816, 0.184]	Moderate	0.728

4.5 Discussion

Scope of the comparison. The reference used in Section 4.3 is the rule-aggregation fallback of the workflow itself, because no established procedure maps the same input to the same output. Post-earthquake usability protocols such as ATC-20 and AeDES assign a safety-based usability tag to an entire building rather than a functional state to a room. FEMA P-58 in turn derives losses and downtime from analytical fragility functions driven by ground-motion intensity rather than from observed component damage, and HAZUS operates at regional scale. None of these yields a room-level functional state conditioned on observed component damage, so an input- and output-aligned baseline does not exist. Constructing one would first require a shared definition of room-level post-earthquake functionality, and the present work adopts that of Liu et al. (2022).

Limitations and future work. Toward more sufficient validation and generalization, the next phase will pursue five directions. First, it will extend the perception evaluation of Section 4.2 beyond a single construction class and a single vision backend, with independent double annotation and localization ground truth, and reduce the metric-estimation error that presently dominates it. Second, it will conduct quantitative validation with case data containing functional conclusions, covering accuracy, missed-detection rate, and time cost against manual survey. Third, it will calibrate the conditional-probability parameters with historical appraisal cases and expert-calibrate the scale-rule thresholds. Fourth, it will expand multi-building-type Bayesian-network templates and multi-country code corpora for cross-type and

cross-region adaptability. Fifth, it will implement and evaluate the incremental retraining that consumes the annotated sample pool, together with its comparison against random sampling. In addition, the weakest-link subsystem rule is a conservative simplification. It does not represent functional redundancy, such as parallel or backup components, and can therefore overestimate loss for redundant subsystems. Redundancy-aware aggregation, by noisy-OR or k-of-n, and its comparison with the present rule are left to future work.

Safety-critical use and model reliability. Post-earthquake building and hospital assessment is safety-critical. The intended role of the workflow is therefore preliminary rapid screening and triage rather than a substitute for formal engineering appraisal. Its outputs are advisory and should be verified by a qualified engineer before entering any official assessment report, and professional and legal responsibility remains with the licensed appraiser. The clause grounding and the uncertainty-triggered review mitigate but do not eliminate large-language-model hallucination and error. Residual risks include mis-grounded clause selection, miscalibrated confidence, and the propagation of undetected-component uncertainty to the room-level conclusion. Reliable deployment therefore also requires calibrated uncertainty, audit trails linking every conclusion to its code provenance, and disciplined versioning of the knowledge base as standards are revised.

Real-world deployment challenges. Field use raises practical issues this proof-of-concept does not yet address. They include variability in image quality, illumination, occlusion, and viewpoint; incomplete or inaccessible observations; and the computational requirements of the perception and inference backends, to which intermittent on-site connectivity adds a possible demand for offline or edge operation. The privacy and custody of on-site imagery, regulatory acceptance and liability, and adaptation to different national codes and damage-classification systems raise further questions, and all of these are prerequisites for moving from demonstration to operational deployment.

5. Conclusion

Existing intelligent seismic damage assessment stops at component-level damage and can hardly answer system-level functional availability. This paper proposes and implements an AI multi-agent workflow for the rapid survey of building damage and functionality, with the following conclusions.

(1) The workflow-based multi-agent architecture is feasible and maintainable. With LangGraph orchestrating the unified state flow, task scheduling, perception, knowledge reasoning, assessment, and report generation are decoupled into collaborating agents. Together with the unified state object, review-on-demand routing, and graceful degradation, the workflow produces the component grading, the subsystem and room functional loss, the code basis, and the repair priority, and it generates a structured, auditable appraisal report automatically.

(2) The coupled knowledge-base and Bayesian-network method achieves the advancement from component damage to system functionality. The code knowledge base provides traceable, individually attributed criteria through a clause-level index. The Bayesian network propagates functional loss from components through subsystems to rooms, decoupling component physical damage from room functional availability. The scenario matrix shows a monotonic and interpretable response to input severity. It also shows the subsystem-discrimination capability that physical damage does not equal functional loss: for the same component-level severe damage, different damaged subsystems yield markedly different room-functionality conclusions, which pure rule aggregation cannot express.

(3) The uncertainty-driven human-machine collaboration balances efficiency and reliability. Manual review is triggered only when the detection uncertainty exceeds the threshold or the Bayesian and rule conclusions diverge, and the remaining cases are produced fully automatically.

The limitations of the current stage and the corresponding future work are detailed in Section 4.5. In summary, this paper provides a scalable, interpretable, and auditable agentic-workflow paradigm for the component-damage-to-system-functionality assessment of post-earthquake building functionality. The scope of the results should be read carefully. The perception module has been measured on field photographs of a single construction class (Section 4.2), whereas the functional-inference layer was exercised with expert-provided, node-aligned inputs in order to isolate it. The room-level conditional probabilities remain rule-based, while the component damage states follow the national standards. The statements on practical applicability are therefore evidence of feasibility and interpretability rather than of validated accuracy. The contributions of the perception scaffolding and of the uncertainty-triggered review are measured here (Sections 4.2 and 4.4). Comprehensive quantitative validation nevertheless remains essential future work: real-image end-to-end perception, multi-case functional ground truth, conditional-probability calibration, and an end-to-end comparison of the orchestration and the clause index against simpler alternatives.

Declarations

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by Science and Technology Innovation Team Training Special of Shandong Earthquake Agency of China (Disaster Risk Assessment Research Team) (RW2603).

Author Contribution

Xu Zheng: conceptualization, methodology, software, validation, investigation, resources, data curation, writing – original draft. Jin Liu: formal analysis, writing – review and editing, visualization, supervision.

Data Availability

The field photographs used in Section 4.2 and the scripts that reproduce the reported results are available from the corresponding author on reasonable request

References

1. Achour, N. & Miyajima, M. Post-earthquake hospital functionality evaluation: the case of Kumamoto Earthquake 2016. *Earthq. Spectra*. **36**, 1670–1694. https://doi.org/10.1177/8755293020926180 (2020).
2. Andriotis, C. & Papakonstantinou, K. Deep reinforcement learning driven inspection and maintenance planning under incomplete information and constraints. *Reliab. Eng. Syst. Saf.* **212**, 107551. https://doi.org/10.1016/j.ress.2021.107551 (2021).
3. Arafin, P., Billah, A. M. & Issa, A. Deep learning-based concrete defects classification and detection using semantic segmentation. *Struct. Health Monit.* **23**, 383–409. https://doi.org/10.1177/14759217231168212 (2024).
4. Areerob, K. et al. Multimodal artificial intelligence approaches using large language models for expert-level landslide image analysis. *Comput. Aided Civ. Infrastruct. Eng.* **40**, 2900–2921. https://doi.org/10.1111/mice.13482 (2025).
5. Bayraktar, A., Hökelekli, E., Akköse, M. & Taş, Y. Evidence-based seismic performance assessment of non-structural components in fixed-base and base-isolated hospital buildings using monitoring data. *Bull. Earthq. Eng.* **24**, 4459–4484. https://doi.org/10.1007/s10518-026-02400-4 (2026).
6. Bektaş, N. & Kegyes-Brassai, O. Developing a machine learning-based rapid visual screening method for seismic assessment of existing buildings on a case study data from the 2015 Gorkha, Nepal earthquake. *Bull. Earthq. Eng.* **23**, 4981–5019. https://doi.org/10.1007/s10518-024-01924-x (2025).
7. Bel-Hadj, Y., Weijtjens, W. & Devriendt, C. Structural health monitoring in a population of similar structures with self-supervised learning: a two-stage approach for enhanced damage detection and model tuning. *Struct. Health Monit.* https://doi.org/10.1177/14759217251324194 (2025).
8. Bhatta, S. & Dang, J. Seismic damage prediction of RC buildings using *machine learning*. *Earthq Eng. Struct. Dyn.* **52**, 3504–3527. https://doi.org/10.1002/eqe.3907 (2023).
9. Boiko, D. A., MacKnight, R., Kline, B. & Gomes, G. Autonomous chemical research with large language *models*. *Nature* **624**, 570–578. https://doi.org/10.1038/s41586-023-06792-0 (2023).
10. Bruneau, M. et al. A framework to quantitatively assess and enhance the seismic resilience of communities. *Earthq. Spectra*. **19**, 733–752. https://doi.org/10.1193/1.1623497 (2003).
11. Bruno, G., Parisi, F., Ruggieri, S., Chatzi, E. & Uva, G. Model updating of structures by combining reduced order modelling and deep reinforcement learning. *Mech. Syst. Signal. Process.* **248**, 114002. https://doi.org/10.1016/j.ymssp.2026.114002 (2026).
12. Chavez, J. P., Murcia-Delso, J. & López-Almansa, F. Probabilistic framework for post-earthquake functionality assessment of seismically isolated hospital buildings using Bayesian networks. *Eng. Struct.* **328**, 119633. https://doi.org/10.1016/j.engstruct.2025.119633 (2025).
13. Chen, J. & Bao, Y. Multi-agent large language model framework for code-compliant automated design of reinforced concrete structures. *Autom. Constr.* **177**, 106331. https://doi.org/10.1016/j.autcon.2025.106331 (2025).

14. Chu, H. & Chun, P. Fine-grained crack *segmentation for high-resolution images* via a multiscale cascaded network. *Comput. Aided Civ. Infrastruct. Eng.* **39**, 575–594. https://doi.org/10.1111/mice.13111 (2024).
15. Cimellaro, G. P., Reinhorn, A. M. & Bruneau, M. Seismic resilience of a hospital system. *Struct. Infrastruct. Eng.* **6**, 127–144. https://doi.org/10.1080/15732470802663847 (2010).
16. Ding, J., Song, Z., Zhai, C. & Wen, W. An approach for improving post-earthquake functionality *of hospital buildings* with fluid viscous damper. *J. Build. Eng.* **105**, 112498. https://doi.org/10.1016/j.jobe.2025.112498 (2025).
17. Dogan, G., Hakan Arslan, M. & Ilki, A. Detection of damages caused by *earthquake and* reinforcement corrosion in RC buildings with deep transfer learning. *Eng. Struct.* **279**, 115629. https://doi.org/10.1016/j.engstruct.2023.115629 (2023).
18. Dong, Y., Zhan, Z., Hu, Y., Doe, D. M. & Han, Z. AI BIM coordinator for non-expert interaction in *building design* using LLM-driven multi-agent systems. *Autom. Constr.* **180**, 106563. https://doi.org/10.1016/j.autcon.2025.106563 (2025).
19. Gao, Y. et al. Retrieval-augmented generation for large language models: A *survey.* arXiv:2312.10997 (2023).
20. Ge, K. et al. Fine-tuning vision foundation model for crack segmentation in civil infrastructures. *Constr. Build. Mater.* **431**, 136573. https://doi.org/10.1016/j.conbuildmat.2024.136573 (2024).
21. Ghimire, S., Guéguen, P., Giffard-Roisin, S. & Schorlemmer, D. Testing machine learning models for seismic damage prediction at a regional scale *using building-damage* dataset compiled after the 2015 Gorkha Nepal earthquake. *Earthq. Spectra*. **38**, 2970–2993. https://doi.org/10.1177/87552930221106495 (2022).
22. Gomez, D., Charris, D., Percybrooks, W. & Arteta, C. A. Automating building typology identification for seismic risk assessment using deep learning. *Earthq. Spectra*. **41**, 1833–1862. https://doi.org/10.1177/87552930251327435 (2025).
23. Gu, J., Xie, Z., Zhang, J. & He, X. Advances in rapid damage identification methods for post-disaster *regional buildings based* on remote sensing images: A survey. *Buildings* **14**, 898. https://doi.org/10.3390/buildings14040898 (2024).
24. Hamdan, A. H. et al. A semantic modeling approach for the automated detection and interpretation of structural *damage*. *Autom. Constr.* **128**, 103739. https://doi.org/10.1016/j.autcon.2021.103739 (2021).
25. Hashim, M. H. M. et al. A review towards a standardized approach in building structure inspection for maintenance management. *Int. J Res Sci. Innov.* **12**, 3443–3450 (2025).
26. He, Z., Chen, W., Zhang, J. & Wang, Y. H. Crack segmentation on steel structures using boundary guidance model. *Autom. Constr.* **162**, 105354. https://doi.org/10.1016/j.autcon.2024.105354 (2024).
27. Hong, S. et al. MetaGPT: Meta programming for a multi-agent collaborative *framework.* arXiv:2308.00352 (2023).
28. Hu, K., Chen, Z., Kang, H. & Tang, Y. 3D vision technologies for a self-developed structural external crack damage recognition robot. *Autom. Constr.* **159**, 105262. https://doi.org/10.1016/j.autcon.2023.105262 (2024).
29. Hussain, R. et al. Vision-language model-based intelligent assistant for onsite construction safety inspection. *Autom. Constr.* **182**, 106728. https://doi.org/10.1016/j.autcon.2025.106728 (2026).
30. Hüllermeier, E. & Waegeman, W. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Mach. Learn.* **110**, 457–506. https://doi.org/10.1007/s10994-021-05946-3 (2021).
31. Iversen, O. & Huang, L. Leveraging large language models for BIM-based automated compliance checking. *Autom. Constr.* **182**, 106707. https://doi.org/10.1016/j.autcon.2025.106707 (2026).
32. Ji, S., Pan, S., Cambria, E., Marttinen, P. & Yu, P. S. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Trans. Neural Netw. Learn. Syst.* **33**, 494–514. https://doi.org/10.1109/tnnls.2021.3070843 (2022).
33. Ji, X., Zhuang, Y., Miao, Z. & Cheng, Y. Vision-based seismic damage detection and residual capacity assessment for an RC shaking table test structure. *Earthq. Eng. Struct. Dyn.* **52**, 806–827. https://doi.org/10.1002/eqe.3788 (2023).
34. Jiang, Y., Wang, J., Shen, X., Dai, K. & Ge, Q. Multitask unified large vision-language model for post-earthquake structural damage assessment of buildings. *Autom. Constr*. **182**, 106720. https://doi.org/10.1016/j.autcon.2025.106720 (2026).
35. Kang, D. H. & Cha, Y. J. Efficient attention-based deep encoder and decoder for automatic crack segmentation. *Struct. Health Monit*. **21**, 2190–2205. https://doi.org/10.1177/14759217211053776 (2022).
36. Kazemi, F., Asgarkhani, N. & Jankowski, R. Machine learning-based seismic fragility and seismic vulnerability assessment of reinforced concrete structures. *Soil. Dyn. Earthq Eng*. **166**, 107761. https://doi.org/10.1016/j.soildyn.2023.107761 (2023).
37. Kirillov, A. et al. Segment anything. In: Proceedings of the IEEE Conference on *Computer Vision* (ICCV), Paris, France, 2–6 October 2023, pp 3992–4003. (2023). https://doi.org/10.1109/ICCV51070.2023.00371
38. Kunlamai, T., Yamane, T., Suganuma, M., Chun, P. & Okatani, T. Improving visual question answering for bridge inspection by pre-training with external data of image–text pairs. *Comput. Aided Civ. Infrastruct. Eng.* **39**, 345–361. https://doi.org/10.1111/mice.13086 (2024).

39. Li, H. et al. Automatic crack detection on concrete and asphalt surfaces using semantic segmentation *network with* hierarchical Transformer. *Eng. Struct.* **307**, 117903. https://doi.org/10.1016/j.engstruct.2024.117903 (2024).
40. Liang, H., Zhou, Y., Talebi-Kalaleh, M. & Mei, Q. Automating structural engineering workflows with large language *model agents.* arXiv:2510.11004 (2025).
41. Liu, J., Li, H., Chai, C., Chen, K. & Wang, D. A LLM-informed multi-agent AI system for drone-based visual inspection for infrastructure. *Adv. Eng. Inf.* **68**, 103643. https://doi.org/10.1016/j.aei.2025.103643 (2025).
42. Liu, J., Song, Z., Zhai, C. & Ding, J. A review of hospital seismic resilience enhancement strategies and technologies: State of the art, challenges, and future. *Reliab. Eng. Syst. Saf.* **273**, 112285. https://doi.org/10.1016/j.ress.2026.112285 (2026).
43. Liu, J., Zhai, C. & Yu, P. A probabilistic *framework* to evaluate seismic resilience of hospital buildings using Bayesian networks. *Reliab. Eng. Syst. Saf.* **226**, 108644. https://doi.org/10.1016/j.ress.2022.108644 (2022).
44. Bran, M. et al. Augmenting large language models with chemistry tools. *Nat. Mach. Intell.* **6**, 525–535. https://doi.org/10.1038/s42256-024-00832-8 (2024).
45. Miao, Z., Ji, X., Wu, M. & Gao, X. Deep learning-based evaluation for mechanical property degradation of seismically damaged *RC columns*. *Earthq. Eng. Struct. Dyn.* **52**, 2498–2519. https://doi.org/10.1002/eqe.3749 (2023).
46. Ministry of Housing and Urban-Rural Development of the People's Republic of China (MOHURD). *Standard for appraisal of reliability of civil buildings (GB 50292 – 2015)* (China Architecture & Building, 2015). (in Chinese).
47. Ministry of Housing and Urban-Rural Development of the People's Republic of China (MOHURD). *Standard for dangerous building appraisal (JGJ 125–2016)* (China Architecture & Building, 2016). (in Chinese).
48. OpenAI. GPT-4 technical report. arXiv:2303.08774 (2023).
49. Pei, S., Zhai, C., Hu, J., Liu, J. & Song, Z. Seismic *functionality of healthcare network considering traffic congestion and hospital malfunctioning*: A medical accessibility approach. *Int. J. Disaster Risk Reduct.* **97**, 104019. https://doi.org/10.1016/j.ijdrr.2023.104019 (2023).
50. Pyle, R. J., Hughes, R. R., Ali, A. A. S. & Wilcox, P. D. Uncertainty quantification for deep learning in ultrasonic crack characterization. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*. **69**, 2339–2351. https://doi.org/10.1109/tuffc.2022.3176926 (2022).
51. Qian, C. et al. *() ChatDev: communicative agents for software development. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Bangkok, Thailand, 11–16 August* 2024, pp 15174–15186. 2024, pp 15174–15186. (2024). https://doi.org/10.18653/v1/2024.acl-long.810
52. Qin, S., Qi, T., Deng, T. & Huang, X. Image segmentation using vision Transformer for tunnel defect assessment. *Comput. Aided Civ Infrastruct. Eng*. **39**, 3243–3268. https://doi.org/10.1111/mice.13181 (2024).
53. Qiu, J. et al. LLM-based agentic systems *in medicine and* healthcare. *Nat. Mach. Intell.* **6**, 1418–1420. https://doi.org/10.1038/s42256-024-00944-1 (2024).
54. Ruggieri, S., Cardellicchio, A., Leggieri, V. & Uva, G. Machine-learning based vulnerability analysis of existing buildings. *Autom. Constr.* **132**, 103936. https://doi.org/10.1016/j.autcon.2021.103936 (2021).
55. Saka, A. et al. GPT models in construction industry: Opportunities, limitations, and a use case validation. *Dev. Built Environ.* **17**, 100300. https://doi.org/10.1016/j.dibe.2023.100300 (2024).
56. Santarsiero, G. et al. Performance of the healthcare facilities during the 2016–2017 Central Italy seismic sequence. *Bull. Earthq. Eng.* **17**, 5701–5727. https://doi.org/10.1007/s10518-018-0330-z (2019).
57. Schick, T. et al. Toolformer: Language models can teach themselves to use tools. arXiv:2302.04761 (2023).
58. Smetana, M., Salles de Salles, L., Sukharev, I. & Khazanovich, L. Highway construction safety analysis using large language models. *Appl. Sci.* **14**, 1352. https://doi.org/10.3390/app14041352 (2024).
59. Song, Z., Zhai, C., Ma, Y., Wang, Z. & Pei, S. Multi-stage and multi-objective design optimization for improving resilience of base-isolated *hospital* buildings. *Eng. Struct.* **304**, 117644. https://doi.org/10.1016/j.engstruct.2024.117644 (2024).
60. Sunil, J. C. & Kamatchi, P. *Quantification of seismic resilience of a hospital building in high seismic prone region of* India. *J. Inst. Eng. India Ser. A*. **106**, 737–747. https://doi.org/10.1007/s40030-025-00896-8 (2025).
61. Touvron, H. et al. LLaMA: Open and efficient foundation language models. (2023). arXiv:2302.13971.
62. Tse, K. W., Pi, R., Yang, W., Yu, X. & Wen, C. Y. Advancing UAV-based inspection system: The USSA-Net segmentation approach to crack *quantification*. *IEEE Trans Instrum. Meas.* **73**, 1–14. https://doi.org/10.1109/tim.2024.3418073 (2024).
63. von Rueden, L. et al. Informed machine learning – a taxonomy and survey *of integrating prior* knowledge into learning systems. *IEEE Trans. Knowl. Data Eng.* **35**, 614–633. https://doi.org/10.1109/TKDE.2021.3079836 (2023).

64. Wang, C., Hornauer, S., Yu, S. X., McKenna, F. & Law, K. H. Instance segmentation of soft-story buildings from *street-view images* with semiautomatic annotation. *Earthq. Eng. Struct. Dyn.* **52**, 2520–2532. https://doi.org/10.1002/eqe.3805 (2023a).
65. Wang, L. et al. A survey on large language model based autonomous agents. *Front. Comput. Sci.* **18**, 186345. https://doi.org/10.1007/s11704-024-40231-1 (2024). Lin Y.
66. Wang, W. & Su, C. Semi-supervised semantic segmentation network for surface crack detection. *Autom. Constr.* **128**, 103786. https://doi.org/10.1016/j.autcon.2021.103786 (2021).
67. Wang, W. & Su, C. Automatic concrete crack segmentation model based on transformer. *Autom. Constr.* **139**, 104275. https://doi.org/10.1016/j.autcon.2022.104275 (2022).
68. Wang, Y. et al. Geometry-guided semantic segmentation for post-earthquake buildings using optical *remote sensing* images. *Earthq. Eng. Struct. Dyn.* **52**, 3392–3413. https://doi.org/10.1002/eqe.3966 (2023b).
69. Waqas, A., Kang, D. & Cha, Y. J. Deep learning-based obstacle-avoiding autonomous UAVs with fiducial marker-based localization for structural health monitoring. *Struct. Health Monit*. **23**, 971–990. https://doi.org/10.1177/14759217231177314 (2024).
70. Wu, Y. et al. Dual attention transformer network for pixel-level concrete crack *segmentation considering* camera placement. *Autom. Constr.* **157**, 105166. https://doi.org/10.1016/j.autcon.2023.105166 (2024).
71. Xi, Z. et al. The rise and potential of large language model based agents: A survey. (2023). arXiv:2309.07864.
72. Xu, H. R., Zhang, N., Yin, Z. Y. & Njock, P. G. A. Multimodal framework integrating multiple large language model agents for intelligent geotechnical design. *Autom. Constr.* **176**, 106257. https://doi.org/10.1016/j.autcon.2025.106257 (2025).
73. Yao, S. et al. Tree of thoughts: Deliberate problem solving with large language models. (2023). arXiv:2305.10601.
74. Ye, Z., Lovell, L., Faramarzi, A. & Ninić, J. Sam-based instance segmentation models for the automation of structural damage detection. *Adv. Eng. Inf.* **62**, 102826. https://doi.org/10.1016/j.aei.2024.102826 (2024).
75. Yu, X. et al. Intelligent assessment of building damage of 2023 Turkey-Syria Earthquake *by multiple remote sensing approaches*. *npj Nat. Hazards*. **1**, 3. https://doi.org/10.1038/s44304-024-00003-0 (2024).
76. Zaitsev, A. Resilience-oriented lifecycle automation of structural RC *elements via* a multi-agent system framework. *J. Intell. Constr.* **4**, 1–19. https://doi.org/10.26599/JIC.2026.9180111 (2026).
77. Zhang, E., Shao, L. & Wang, Y. Unifying transformer and convolution for dam crack detection. *Autom. Constr.* **147**, 104712. https://doi.org/10.1016/j.autcon.2022.104712 (2023).

Figures

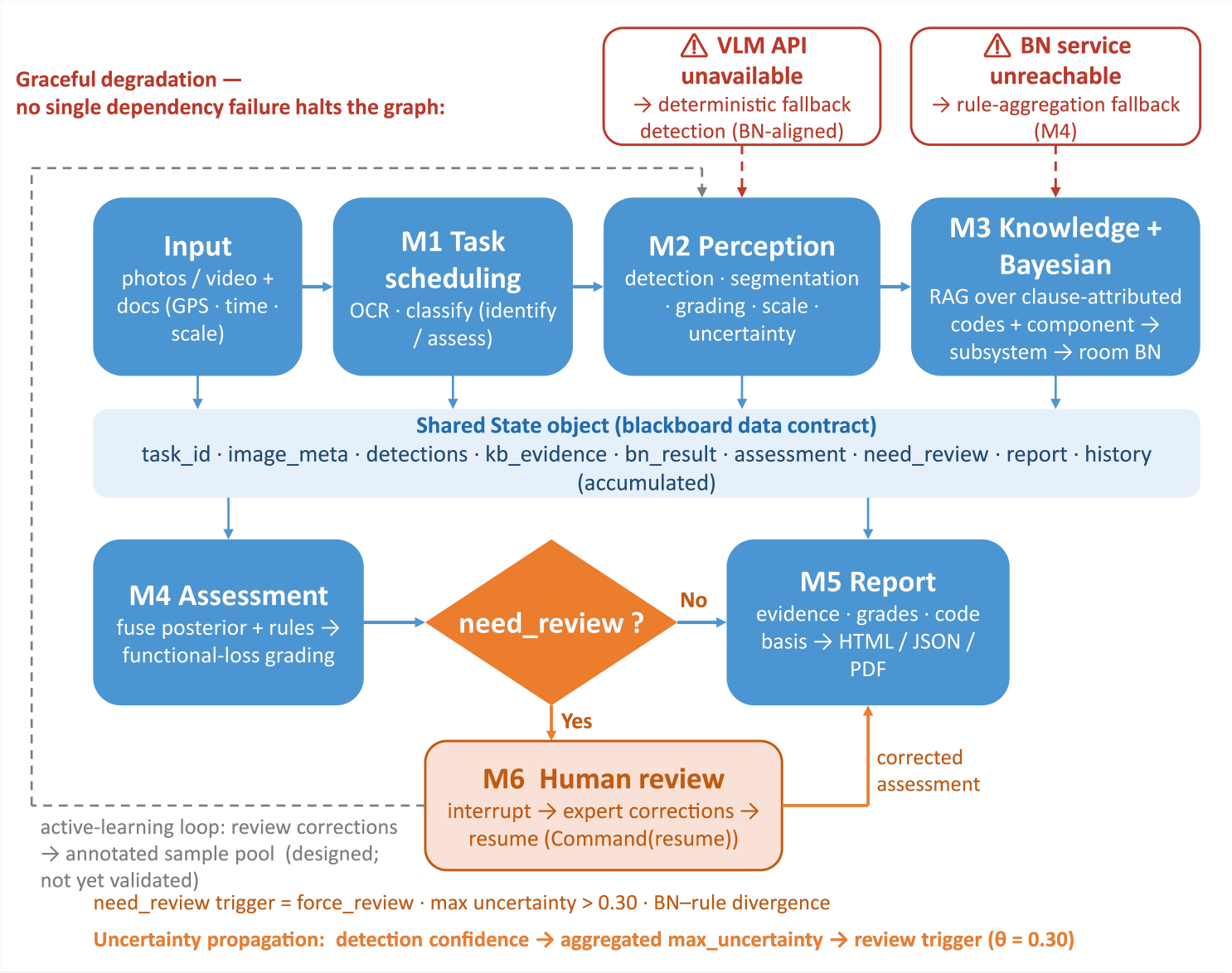


Figure 1

Layered multi-agent workflow for the rapid survey of building damage and functionality

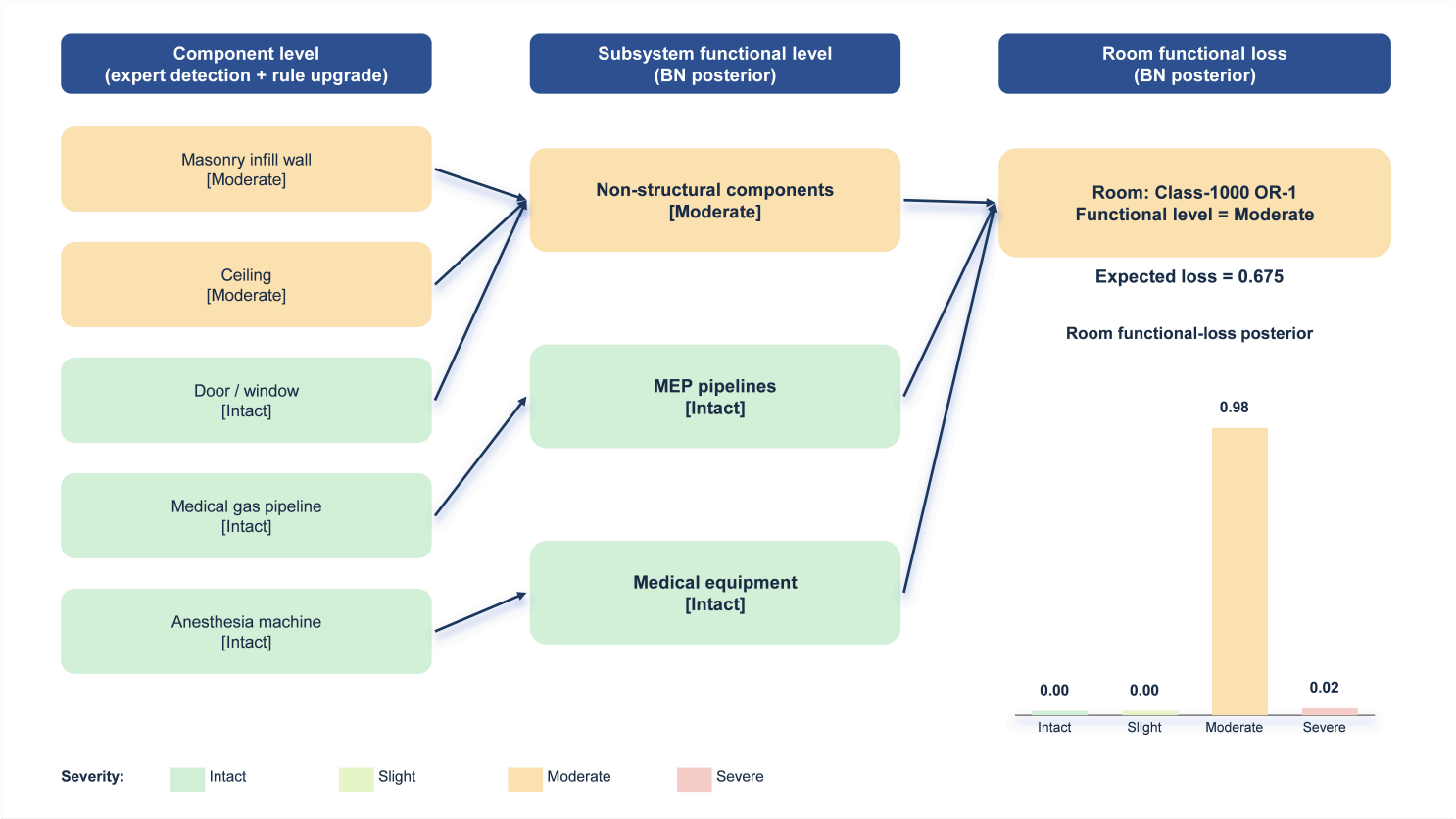


Figure 2

Case: Bayesian functional loss posterior propagation (component-subsystem-room)

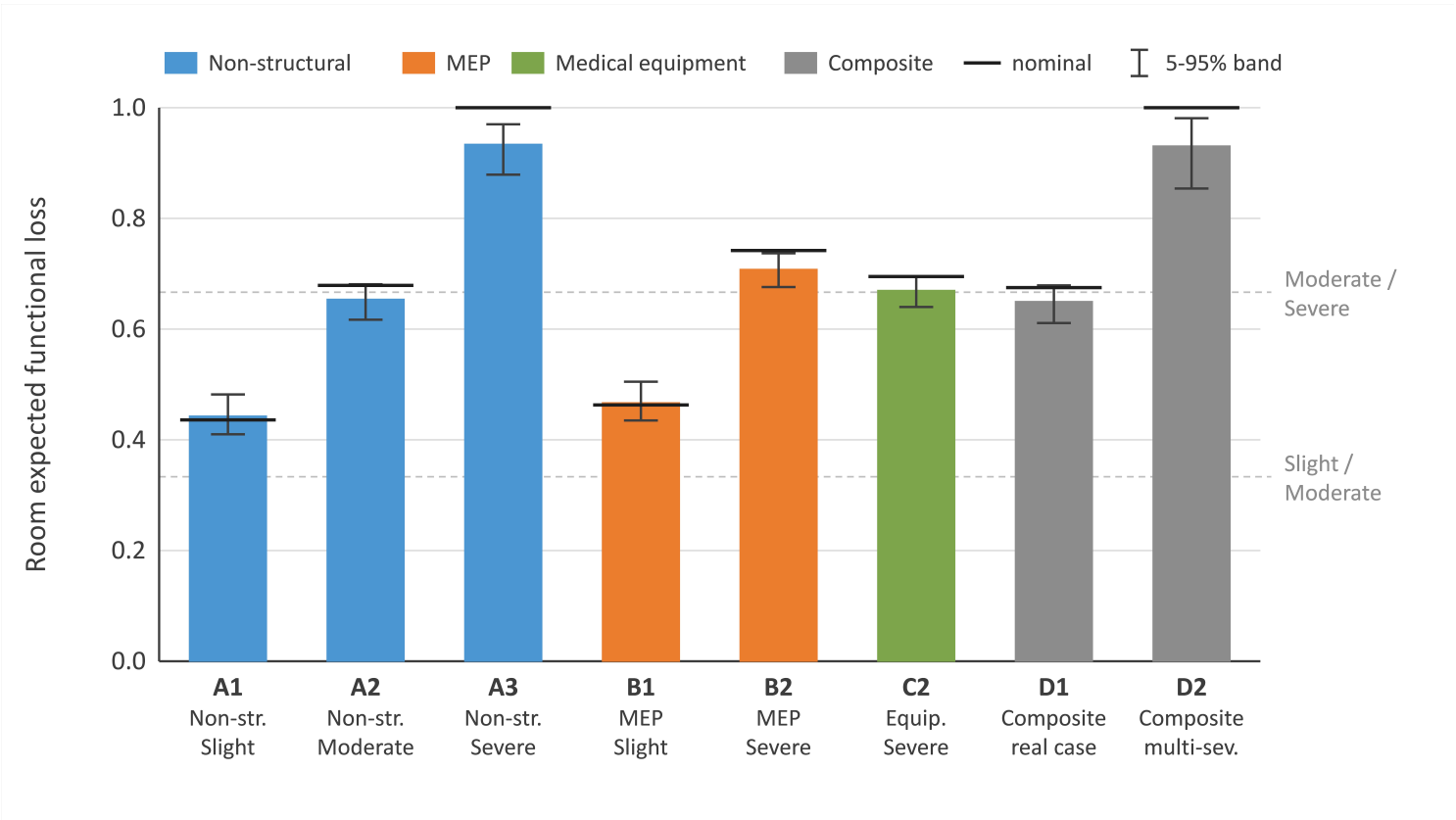


Figure 3

Room expected functional loss by damaged subsystem and severity (bars: mean over N = 1000 joint perturbations of the component priors and the room CPT; whiskers: 5–95% band; markers: nominal values of Table 4)

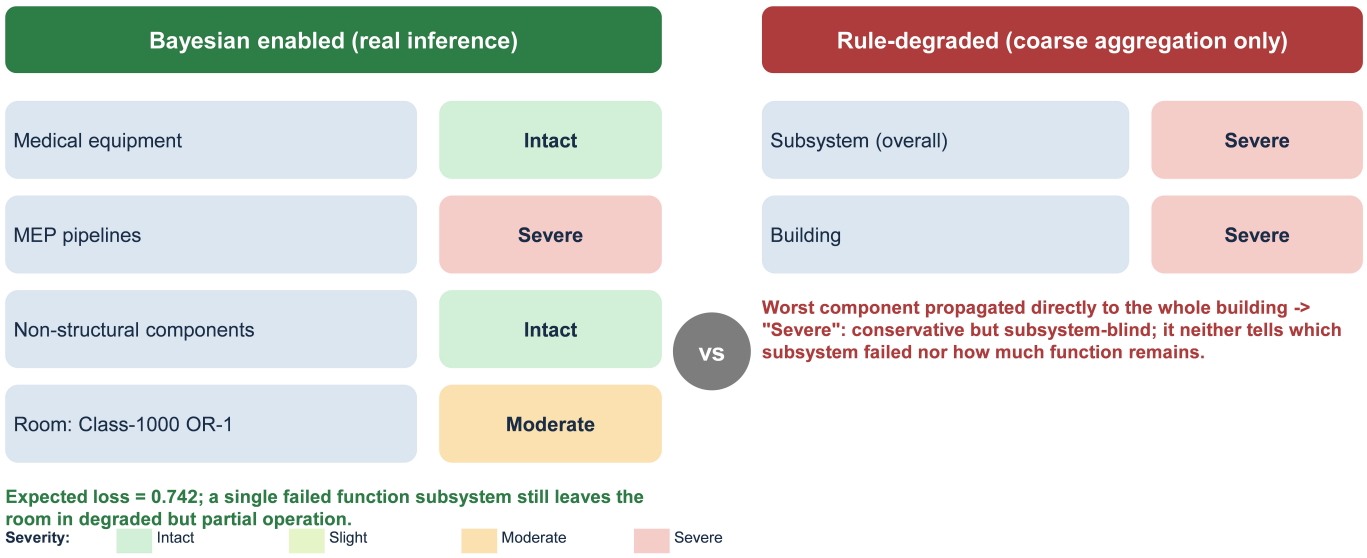


Figure 4

Bayesian inference vs. rule-degraded functional-assessment conclusions

Scenario	Input (subsystem · severity)	BN room level	Rule-degraded room level	Divergence
A1	Non-structural · Slight	Slight	Slight	
A2	Non-structural · Moderate	Moderate	Moderate	
A3	Non-structural · Severe	Severe	Severe	
B1	MEP · Slight	Slight	Slight	
B2	MEP · Severe	Moderate	Severe	★
C2	Medical-equipment · Severe	Moderate	Severe	★
D1	Composite (real case)	Moderate	Moderate	
D2	Composite (multi-severe)	Severe	Severe	

Severity: Intact Slight Moderate Severe

Rule degradation gives "Severe" via worst-component aggregation wherever a single component reaches its top tier; the Bayesian path gives the finer "Moderate" at B2 / C2 (★ = divergence). C1 is absent because the codes define no slight damage state for the medical-equipment items.

Figure 5

Bayesian inference vs. rule degradation: room functional level per scenario

Supplementary Files

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

- AppendixA.docx