applied sciences

MDPI

Article

A Five-Step MCDM Framework for AR Use Case Selection in Railway Maintenance

Tayyarat Oumaima [1,*], Abdeslam Ahmadi [1], Sedki Mohamed [2] and Hicham El Kimi [3]

[1] Laboratory of Modelling and Multiphysics Engineering, École Nationale Supérieure d'Arts et Métiers, Moulay Ismail University, Meknes 50000, Morocco; ab.ahmadi@umi.ac.ma
[2] Department of Innovation Platform and Digital Material, Office National des Chemins de Fer, Rabat 10090, Morocco; mohamed.sedki@oncf.ma
[3] Direction of Engineering and Projects, Société Marocaine de Maintenance des Rames à Grande Vitesse, Tanger 90000, Morocco
* Correspondence: o.tayyarat@edu.umi.ac.ma

Abstract

Despite the growing adoption of Augmented Reality (AR) in industrial maintenance, no structured methodology exists to systematically identify which operations are best suited for effective AR deployment. This study addresses this gap by proposing a five-step, Multi-Criteria Decision-Making (MCDM)-based selection framework for determining AR-compatible maintenance operations in high-speed railway systems. The framework—applied under the AFNOR FD X 60-000 standard—integrates maintenance-level compatibility analysis, multi-criteria filtering across five dimensions (operational frequency, execution complexity, safety impact, traceability, and scalability), and expert validation involving 100 railway maintenance professionals. Applied to 12 candidate operations at a high-speed railway maintenance facility in Morocco, the framework identified OP10 (insulating oil level verification of the Main Transformer) as the optimal pilot use case, confirming expert consensus (Kruskal–Wallis: $H = 18.479$, $p < 0.001$). The selected operation was subsequently integrated into a hybrid AR–Deep Reinforcement Learning architecture employing a Deep Q-Learning (DQL) agent for adaptive decision support, deployed on a Magic Leap 2 head-mounted device via a Unity-based rendering pipeline with hybrid marker-based and markerless computer vision tracking through Vuforia Engine. Experimental validation conducted over three months under simulated and semi-operational conditions yielded a 34–47% reduction in intervention time, a 55–70% decrease in human error rates, and a 28–42% decline in failure-related costs. While results are currently limited to a single-site context, the proposed methodology is directly transferable to any asset-intensive, regulated maintenance environment beyond the railway sector.

Keywords: railway maintenance; augmented reality; use case selection; MCDM; AHP; maintenance optimization; deep reinforcement learning

Check for updates

Academic Editor: Alexandre Carvalho

Received: 10 May 2026
Revised: 28 May 2026
Accepted: 5 June 2026
Published: 4 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

In high-speed railway maintenance, the gap between technological potential and operational reality is striking. Digital tools exist, AR solutions have been tested, and their economic viability has been demonstrated—yet on the ground, technicians continue to rely on paper-based inspection sheets, manual measurement tools, and experience-dependent cross-referencing procedures, navigating dark engine compartments while mentally cross-

referencing tolerance tables. The problem is not the absence of solutions. It is the absence of a structured method for deciding where to start.

This disconnect is rooted in three fundamental problems that persist across the sector. First, maintenance costs in railway systems are notoriously high—accounting for approximately 70% of the total lifecycle cost of train-sets [1]—and traditional visual inspection procedures remain time-consuming, experience-dependent, and prone to omission errors, particularly in complex multi-step operations carried out in constrained environments [2]. Second, knowledge transfer represents a critical and growing vulnerability: as experienced technicians retire, they carry with them tacit operational knowledge that is rarely formalized, creating skill gaps that degrade both consistency and safety [3]. This limitation has been systematically documented across asset-intensive industries: spatial computing technologies—including AR-based overlay systems and mixed reality environments—have emerged as inherently coherent responses to knowledge transfer limitations by externalizing tacit procedural knowledge into reusable, context-anchored digital guidance [4,5]. In the present system, this externalization mechanism is implemented through the Magic Leap 2 AR environment, which overlays procedure-anchored digital guidance directly onto the physical equipment in real time—converting tacit expert knowledge into a structured, replayable AR guidance layer accessible to any technician regardless of prior experience. Third, while digital tools have been progressively introduced in maintenance contexts, their adoption remains fragmented—deployed in isolation, disconnected from existing workflows, and rarely scaled beyond isolated pilot experiments [6,7].

AR thus represents a coherent response to all three challenges simultaneously. Prior work in our research program has established its economic viability in industrial maintenance contexts and mapped the adoption strategies most likely to succeed across different operational environments [8].

Yet a critical prerequisite question has remained largely unaddressed in the AR maintenance literature: before deploying AR, how should one determine which maintenance operation to target first, and according to what criteria? A review of peer-reviewed studies on AR in railway and industrial maintenance—including the most cited contributions in this field [9–12]—reveals a consistent pattern: AR is applied to a pre-selected operation without explicit justification of that choice. Operation selection is treated as self-evident rather than as a research problem in its own right.

This gap has direct industrial consequences: selecting the wrong target operation leads to poor return on investment, low technician adoption, and integration architectures that resist generalization beyond a single pilot [13]. As demonstrated by Bock, Bohné and Tadeja (2024) [14], the value of AR as an assistance system is highly dependent on task-specific characteristics—error-proneness, procedural complexity, and data availability—criteria that are rarely made explicit in current deployment decisions. Without a systematic selection methodology, organizations risk investing in high-visibility but low-leverage applications, missing the operations where AR could generate the greatest systemic impact.

This article addresses that gap directly. It makes three distinct contributions not found in existing work: (1) a structured, replicable five-step decision methodology for selecting the most suitable AR-compatible maintenance operation, grounded in Multi-Criteria Decision-Making principles and adapted to regulated railway environments, with filtering criteria—operational frequency, execution complexity, safety impact, traceability, and scalability—derived from the classical cost–time–quality performance framework [15] and adapted to the specific constraints of AR deployment in regulated maintenance environments—unlike existing AHP applications in AR industrial contexts [14,16], which optimize technology selection within a pre-defined operation scope, this framework determines which operation

to target before any technological deployment begins; (2) an expert-validated industrial application of this methodology to a real high-speed train maintenance facility in Morocco operating under the AFNOR FD X 60-000 standard [17], resulting in the identification and full characterization of OP10 as the highest-priority candidate for AR integration; and (3) a preliminary validation of the selected operation within a hybrid AR–Deep Reinforcement Learning architecture—combining Vuforia Engine-based tracking, Unity rendering, and a Deep Q-Learning agent for adaptive decision support on Magic Leap 2—demonstrating measurable performance gains under simulated and semi-operational conditions, with full architectural specification reported in the companion study [18].

Unlike previous studies that apply AR to a convenient or visible operation, this work treats operation selection as the primary research contribution—one that conditions the scalability, transferability, and return on investment of any subsequent technological development. It explicitly positions itself as the third component of a multi-article research program: where the first contribution established the economic viability of AR adoption and the second mapped optimal adoption strategies across operational environments [8], the present study provides the critical bridge between those strategic findings and their concrete industrial application.

While recent vision transformer architectures—including Vision Transformer (ViT), Distillation with NO labels (DINO), and Segment Anything Model (SAM)—have demonstrated strong performance in industrial anomaly detection and component recognition [19,20], their deployment in operational maintenance environments requires extensive annotated training datasets and yields probabilistic outputs that cannot guarantee detection reliability in safety-critical contexts. The present system addresses this constraint through a deterministic model-based tracking approach: the pre-existing 3D OBJ model of the Main Transformer enables Vuforia Engine to perform geometry-anchored spatial registration without any training phase [21–23], supplemented by marker-based tracking.

The remainder of this article is organized as follows. Section 2 reviews the role of use case selection in AR-based maintenance and the current state of AR in the railway sector. Section 3 presents the proposed five-step selection methodology. Section 4 applies it to identify and validate OP10. Section 5 defines the technological stack and implementation framework through a validated use case. Section 6 presents the discussion. Section 7 concludes with key findings and future research directions.

2. Background

2.1. Why Use Case Selection Matters in AR-Based Maintenance

The deployment of AR in industrial maintenance is not a monolithic decision—it is a sequence of interdependent choices, of which operation selection is arguably the most consequential. Poor use case selection consistently results in suboptimal return on investment, resistance to adoption among technicians, and deployment architectures that fail to scale beyond isolated pilots. Yet this selection step has received remarkably little attention as a standalone research problem.

The broader maintenance engineering literature has long recognized that not all operations benefit equally from technological augmentation. Decision frameworks such as multi-criteria analysis, failure mode prioritization, and risk-based maintenance planning all share a common premise: the value of a technology depends critically on where it is applied [24]. In the context of AR, this premise is reinforced by the technology's own constraints—AR adds value primarily where operations are visually guided, procedurally structured, and cognitively demanding, but offers little benefit in operations that are either too simple or too complex for current AR capabilities [8].

The question of technology adoption in industrial contexts has been extensively studied through behavioral frameworks, most notably the Technology Acceptance Model (TAM) originally proposed by Davis (1989) [25], which establishes perceived usefulness and perceived ease of use as the two primary determinants of technology adoption intention. In maintenance engineering contexts, TAM has been extended to account for safety culture, regulatory compliance, and workflow integration—dimensions that are particularly salient in railway environments [24]. Importantly, TAM-based analyses consistently show that adoption fails not when technology is unavailable, but when it is deployed in contexts where its perceived usefulness is low—a conclusion which directly reinforces the strategic importance of selecting the right operation before any deployment begins.

In complex sociotechnical systems such as railway infrastructure, this challenge is further compounded by safety and regulatory constraints. Any AR integration must be compatible with existing maintenance standards, traceable through documented procedures, and robust enough to withstand operational variability [26]. These constraints make the selection of a pilot use case not merely a practical convenience but a methodological necessity—one that conditions the scalability and transferability of the entire AR deployment framework.

Despite this, the dominant pattern in the AR maintenance literature is to select an operation based on convenience or visibility rather than on systematic evaluation criteria. This gap between the strategic importance of use case selection and the absence of structured methodologies for performing it represents the central motivation of the present study.

The pairwise comparison matrix was constructed through consensus by the four authors, all holding active positions in railway maintenance research and operations. A single consensus matrix was produced rather than aggregating individual matrices, following the direct consensus protocol recommended for small expert groups in AHP applications [27,28]. No individual CR threshold exclusion was required given the single-matrix approach; the resulting CR of 0.004 is well below the 0.10 acceptance threshold.

2.2. AR in Railway Maintenance: State of the Art

The global railway industry has increasingly embraced AR technologies as a means of improving efficiency across Maintenance, Repair, and Overhaul (MRO) operations, within a broader Industry 4.0 transformation that has repositioned digital technologies as core enablers of industrial competitiveness [29]. This transformation has steadily introduced cyber-physical systems, Internet of Things sensors, and data-driven decision tools into maintenance workflows—with AR emerging as one of the most operationally promising interfaces between the digital and physical layers of maintenance execution.

A structured review of publications from 2013 to 2025 was conducted following the PRISMA guidelines [30], using ScienceDirect, Scopus, Google Scholar, and IEEE Xplore as primary databases. The search employed the keywords “AR,” “maintenance,” and “railway”. The initial search identified 312 records across the four databases. After the removal of 89 duplicates, 223 records were screened based on titles and abstracts. Studies unrelated to AR in maintenance or railway contexts, hardware-only AR evaluations, and non-peer-reviewed publications were excluded at this stage (n = 152). The remaining 71 full-text articles were assessed for eligibility. Following the evaluation of the introduction and conclusion sections, 48 studies were excluded for the following reasons: virtual reality or hardware-only evaluation (n = 18), non-peer-reviewed status (n = 12), non-maintenance industrial context (n = 11), and insufficient methodological quality (n = 7). This process resulted in a final corpus of 23 peer-reviewed studies. Inclusion criteria required: (1) peer-reviewed publication status, (2) direct relevance to AR deployment in industrial or railway maintenance contexts, and (3) full-text availability in English. Figure 1 presents the PRISMA

flow diagram summarizing the selection process. This review reveals three dominant research themes: human factors and challenges (48%), training and knowledge transfer (24%), and real-time and remote assistance (28%), as summarized in Table 1.

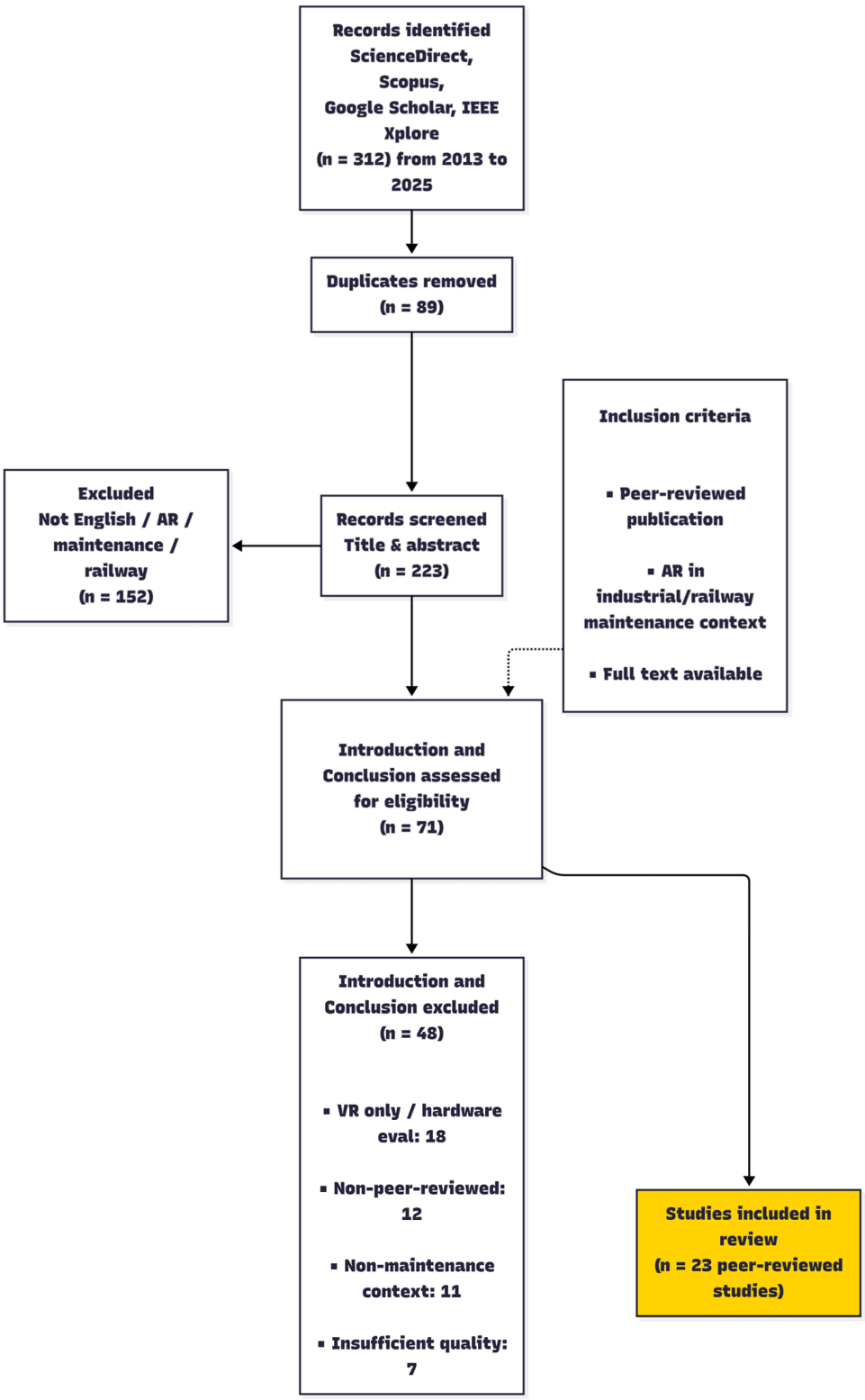


Figure 1. PRISMA Flow Diagram of the Systematic Literature Review on AR-Based Maintenance in Railway Systems (2013–2025).

Table 1. Key Thematic Trends in AR Adoption for Railway Maintenance.

Thematic Category	Article References	No. of Articles
Training and Knowledge Transfer	[11,12,31–34]	6
Real-Time and Remote Assistance	[9,10,35–39]	7
Usability, Human Factors and Challenges	[3–5,7–9,13,24,26,38–40]	12

Note: [39] and [9] appear in two thematic categories respectively, reflecting their cross-thematic contributions. The total count of unique references across the three categories is 23, with 2 references counted twice.

Training and Knowledge Transfer: AR has been extensively applied to training in railway maintenance, where its capacity to overlay interactive 3D content onto real equipment offers a compelling alternative to traditional instruction methods. Kim and Kim (2024) [11] developed AR-based training modules for air spring light maintenance on electric multiple units and reported high usability scores across technician groups, with measurable improvements in learning outcomes and operational readiness. Similarly, Cha et al. (2023) [31] demonstrated the effectiveness of AR-based educational content for light maintenance of railway vehicles, enabling personnel to interact with 3D models of block brakes and other components in a risk-free virtual environment. Kwon et al. (2021) [12] extended this approach to axle-mounted disc brake systems, designing AR training content specifically structured around the procedural complexity of the target operation.

These contributions collectively demonstrate that AR can significantly accelerate the transfer of procedural knowledge—a finding directly relevant to the knowledge retention challenge identified in the introduction. However, they share a common limitation: the training operations targeted are selected without explicit justification, leaving open the question of whether these particular operations represent the highest-value targets for AR investment.

Operational Efficiency and Real-Time Assistance: Beyond training, AR has demonstrated measurable impact on real-time maintenance efficiency. Garcia et al. (2024) [9] developed the Holorailway system for assembly operations in the railway industry, reporting a 47% reduction in global assembly time and an 88% decrease in errors—results that establish a compelling benchmark for AR-assisted procedural tasks. Scheffer et al. (2021, 2023) [10,36] developed a framework for AR-assisted decision-making in maintenance, proposing a visualize–guide–decide–track pipeline that structures the technician's interaction with physical equipment and digital overlays—a decision-support architecture that presupposes operation selection has already occurred, without addressing how that selection should be made. Jing et al. (2024) [38] investigated AR head-mounted display interface layouts for subway overhaul operations, demonstrating that well-designed AR interfaces significantly improve obstacle recognition and task completion in spatially constrained environments. The integration of AR with AI-driven decision support has also been explored by Malta, Farinha & Mendes (2023) [39], who evaluated user experience across industrial maintenance and support applications and found that AI-augmented AR substantially improves both task accuracy and technician confidence.

In parallel, the emergence of Digital Twin technology has opened new integration pathways for AR in railway maintenance. Bernal et al. (2024) [41] demonstrated that coupling AR interfaces with Digital Twin models of railway assets enables real-time synchronization between physical equipment states and digital overlays, significantly improving diagnostic accuracy in complex inspection tasks. This contribution, while outside the core corpus of the structured review, is cited here for its architectural relevance to the scalability criterion applied in the selection methodology proposed in this study—operations compatible with Digital Twin integration representing higher-value AR targets from a systems perspective.

Recent advances in deep learning-based visual inspection have introduced transformer-based architectures—including ViT [19], DINO [20], and SAM [42]—demonstrating strong performance in industrial anomaly detection and component recognition. These models represent the current frontier of learned visual inspection for industrial assets, though their operational deployment constraints in safety-critical AR contexts are discussed in relation to the tracking architecture adopted in the present system (Section 5).

These results are significant, but they reinforce the same pattern observed in the training literature: the operations chosen for AR integration are selected based on availability or perceived complexity, not on a systematic evaluation of their strategic value, scalability potential, or compatibility with the specific constraints of AR technology.

Challenges and Barriers to Adoption: Despite these promising results, the widespread integration of AR in railway maintenance remains limited by a set of persistent systemic barriers. Tretten, Illankoon and Candell (2021) [3] identify situation awareness as a critical challenge in digitalized maintenance environments, noting that the introduction of digital overlays can paradoxically degrade the operator's perception of the physical environment if not carefully designed. Mikušová et al. (2024) [13] highlight the persistent gap between laboratory demonstrations and field deployment, attributing it to insufficient integration with existing maintenance workflows and the absence of structured adoption roadmaps. At the organizational level, Kour et al. (2022) [24] document the human-centric barriers to AR adoption in railway asset management, including technician resistance, inadequate training infrastructure, and the difficulty of aligning AR tools with established maintenance cultures. More broadly, Rodríguez-Hernández et al. (2025) [7] conducted a comprehensive review of digitalization in railway maintenance through the lens of the International Union of Railways asset management framework, concluding that while digital technologies offer significant optimization potential, their successful implementation requires a far more integrated approach than is currently practiced.

2.3. Synthesis and Research Gap

The literature reviewed above converges on a clear conclusion: AR is technically capable of delivering significant value in railway maintenance, but its deployment remains opportunistic, fragmented, and insufficiently grounded in systematic selection criteria. Across training applications, real-time assistance systems, and efficiency-focused deployments, the choice of target operation is consistently treated as a pre-analytical given rather than as a research problem requiring structured resolution.

The consequences of this gap are concrete and measurable. Organizations deploying AR without a rigorous targeting methodology tend to invest in high-visibility but low-leverage applications—overlooking the operations where AR's specific capabilities in procedural guidance, error detection, and knowledge capture would generate the greatest systemic impact. Furthermore, deployments not designed with scalability in mind from the outset consistently remain isolated experiments, unable to provide the architectural foundation for a broader intelligent maintenance framework.

Table 2 makes this pattern explicit. Across the most cited contributions in AR-based railway maintenance, operation selection is consistently performed by convenience or industrial visibility, without explicit evaluation criteria and without any claim to generalizability. The difference is stark. Existing studies—including the most methodologically advanced contribution in this space [14], which identifies task-specific criteria relevant to AR deployment but stops short of operationalizing them into a selection procedure—leave the core decision unresolved. No study proposes a replicable, criteria-based methodology for AR operation selection in railway maintenance that combines quantitative multi-criteria evaluation with structured expert validation and explicit scalability assessment.

Table 2. Analytical Review of AR Applications in Railway and Industrial Maintenance.

Studies	Thematic Category	Explicit Criteria	Experts Validation	Replicable Framework	Limitation
[11,12,31–34]	Training & Knowledge Transfer	No	No	No	Not generalizable beyond single pilot
[3–5,7,8,13,24,26,39,40]	Usability & Human Factors	No	No	No	Not generalizable beyond single pilot
[10,35–37,39]	Real-Time Assistance	No	No	No	Not generalizable beyond single pilot
[9]	Real-Time Assistance/Usability	No	No	No	Not generalizable beyond single pilot
[38]	Assistance	No	No	No	No operational selection framework
[14]	Usability & Human Factors	Partial	No	No	Criteria not operationalized into methodology
This study	Multi-Criteria Decision-Making (MCDM)	Yes	Yes (n = 100)	Yes	Single-site validation; ordinal scoring resolution limited to 3 levels; AR architecture not yet tested beyond OP10

Beyond the absence of a structured selection methodology, the credibility of any expert-based validation depends critically on the rigor of panel recruitment and bias mitigation. Expert validation protocols in MCDM-based industrial studies vary significantly in terms of recruitment strategy, professional diversity, and consistency enforcement measures [43]—dimensions that are explicitly addressed in the validation design presented in Section 4.

Addressing this gap requires not only a structured selection methodology, but a deliberate choice of the decision-making method best suited to the specific constraints of railway maintenance contexts—where quantitative performance data is unavailable, expert knowledge is the primary input, and decision transparency is mandated by the regulatory framework. The selection of the Analytic Hierarchy Process (AHP) as the primary Multi-Criteria Decision-Making (MCDM) method in this study was not made by default. Table 3 presents a comparative analysis of five commonly applied MCDM methods in industrial maintenance and technology selection contexts—namely the Analytic Hierarchy Process (AHP), the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), the VlseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR), the Analytic Network Process (ANP), and the ELimination Et Choix Traduisant la REalité (ELECTRE)—evaluated against four requirements that are non-negotiable in the present study: expert tacit knowledge elicitation, consistency validation of expert judgements, interpretability for industrial practitioners, and scalability to large expert panels.

This comparison demonstrates that AHP is the only method among those reviewed that simultaneously satisfies all four requirements in the present context. This persistent absence of a structured, method-justified, expert-validated selection framework constitutes the central motivation and primary contribution of the present work. The five-step methodology developed here directly addresses this gap. It does so not as a secondary methodological consideration, but as the principal research contribution upon which all subsequent technological and industrial developments in this program are built.

Table 3. Comparative Analysis of MCDM Methods for AR Operation Selection in Industrial Maintenance Contexts.

Method	Documented Application Context	Strengths	Limitations	Expert Consistency Validation
AHP	Selection of maintenance actions in railway infrastructure [44]; predictive maintenance technique selection with 20 domain experts [45]; material and assembly method selection for railcar development [46]	Pairwise comparison structure elicits tacit expert knowledge; interpretable priority weights; built-in consistency ratio (CR) detects and filters incoherent judgments; validated in railway maintenance contexts	Judgment subjectivity increases with small expert panels; pairwise matrices may yield a CR > 0.10 requiring revision [28,47]; sensitivity to pairwise comparison scales	The CR mechanism provides a mathematically grounded acceptance threshold (CR < 0.10); applicable to both individual and group responses [27]
TOPSIS	Complex project selection in manufacturing engineering [48]; supplier evaluation and industrial performance ranking	Ranks alternatives according to their distance from the ideal and anti-ideal solutions; mathematically rigorous; objective when quantitative data are available	Requires quantitative numerical data for all criteria; no mechanism to elicit or validate expert judgments; sensitive to normalization methods and criterion weighting [49]; results vary depending on the data scale	No built-in consistency check; expert contributions are neither structured nor validated
VIKOR	Maintenance strategy selection for safety-critical assets combined with DEMATEL and ANP [50]	Identifies compromise solutions under conflicting criteria; robust when criteria cannot be fully compensated for each other	Does not provide a mechanism to improve decision-process consistency [51]; requires predefined weights; results are difficult to interpret without additional explanations for practitioners	No built-in consistency validation; relies on weights provided by external methods
ANP	Used with VIKOR for selecting maintenance strategies for critical assets [50]	Models interdependencies among criteria and clusters; extension of AHP toward non-hierarchical decision structures	Requires additional comparison matrices for each cluster; computational complexity increases nonlinearly with the number of criteria; impractical for large expert panels; requires strong methodological expertise from respondents	No built-in consistency control at the network level; CR applies only to individual cluster matrices
ELECTRE	Outranking-based selection in energy planning, logistics, and environmental assessment [52]	Handles non-compensatory criteria; suitable for ordinal preference structures; robust when certain criteria cannot substitute for one another	Outranking relations do not produce a single ranking list; results require significant interpretation; no expert consistency validation mechanism; difficult to communicate to non-specialist industrial decision-makers [52]	No built-in consistency validation; preference thresholds must be externally defined

3. Proposed Methodology

Grounded in the fundamental gap identified in Section 2.3—the systematic absence of principled operation selection methodologies across 25 reviewed studies—this study proposes a structured five-step decision procedure for the principled selection of an AR-compatible maintenance operation. The methodology draws on MCDM principles, a well-established approach for selecting the best alternative by simultaneously evaluating multiple criteria in engineering and industrial contexts [53], and on the Analytic Hierar-

chy Process (AHP), introduced by Saaty (1980) [27] and widely applied in maintenance engineering for its ability to derive consistent and transparent criterion weights through structured pairwise comparisons [54,55]. Together, MCDM and AHP provide a rigorous and replicable foundation for the proposed selection procedure.

Step 1—Candidate Operation Identification and Redundancy Screening

The first step systematically identifies the full set of maintenance operations already addressed in the AR literature, for the purpose of excluding redundant targets from the selection pipeline. This screening is grounded in a structured literature review (detailed in Section 2) and serves a specific methodological function: ensuring that the proposed use case constitutes a novel AR contribution, not a replication of a documented deployment. Operations already extensively covered in peer-reviewed work (e.g., wheelset maintenance) are excluded at this stage.

Step 2—Maintenance Level Compatibility Analysis

Not all maintenance levels are equally compatible with AR technology. This step maps the maintenance levels defined by the AFNOR FD X 60-000 standard against a set of AR compatibility criteria—including operation visibility, procedural structure, cognitive load, and regulatory traceability. Operations belonging to levels that are either too simple (Level 1) or too complex and outsourced (Level 4) are excluded from further consideration.

Step 3—Multi-Criteria Filtering and Evaluation

Candidate operations identified in Steps 1 and 2 are evaluated through a structured multi-criteria decision matrix. The five evaluation criteria were selected based on a convergent analysis of the AR deployment literature and railway maintenance practice, and their relative weights were derived through AHP pairwise comparison.

Following Saaty's (1980) [27] pairwise comparison procedure, each criterion was compared against every other criterion on a 1–9 importance scale, producing the pairwise comparison matrix presented in Table 4. Priority weights were derived by normalizing the geometric mean of each row. The maximum eigenvalue λmax was calculated as 5.016, yielding a Consistency Index $CI = (\lambda max - n)/(n - 1) = (5.016 - 5)/4 = 0.004$. The Consistency Ratio was then computed as $CR = CI/RI = 0.004/1.12 = 0.0036$ (rounded to 0.004), where RI = 1.12 is the standard random index for n = 5 criteria [27]. This value is well below the accepted 0.10 threshold, confirming that the pairwise judgments are internally consistent and that the derived weights are methodologically valid.

Table 4. AHP Pairwise Comparison Matrix for Criteria Weighting.

Criteria	C1	C2	C3	C4	C5	Priority Weight
C1 Frequency	1	1/2	1/3	1	1	0.15
C2 Complexity	2	1	1/2	2	2	0.25
C3 Safety	3	2	1	3	3	0.30
C4 Traceability	1	1/2	1/3	1	1	0.15
C5 Scalability	1	1/2	1/3	1	1	0.15

The pairwise comparison matrix presented in Table 4 was constructed by the research team through a convergent synthesis of the AR deployment and railway maintenance literature, drawing on established priority structures from risk-based maintenance frameworks [24,56] and task-specific AR value analyses [14]. This author-derived approach is consistent with AHP applications in industrial contexts where quantitative performance data is unavailable and criterion weights must be grounded in domain

knowledge rather than empirical measurement [28]. The resulting weight structure was subsequently submitted for practitioner confirmation to a sub-panel of 13 senior railway maintenance professionals (≥11 years of experience) drawn from the broader expert cohort described in Step 4. Each member independently assessed whether the proposed priority ordering—Safety > Complexity > Frequency = Traceability = Scalability—was consistent with their operational experience. Consensus was reached without requiring revision of the initial weight structure, providing practitioner-grounded confirmation of the literature-derived priorities.

Each candidate operation is scored across all five criteria using a three-level ordinal scale (1–3). A three-level scale was selected rather than five- or ten-point alternatives to reflect the binary nature of the operational distinctions at stake: operations in this context are either clearly unsuitable, borderline, or clearly suitable for AR integration. This calibration minimizes inter-rater noise in a domain where scoring granularity does not improve decision quality, a choice supported by MCDM practice in industrial settings [53,55]. The five filtering dimensions are not arbitrarily defined: they constitute an operationalization of the classical cost–time–quality performance triangle [15] adapted to the specific constraints of AR deployment in regulated maintenance environments. Operational frequency and execution complexity translate the time dimension—identifying operations where AR guidance generates measurable intervention efficiency gains. Safety impact and traceability translate the quality dimension—prioritizing operations where error reduction and auditability deliver the highest systemic value. Scalability translates the cost dimension—ensuring that the selected operation constitutes a replicable entry point rather than an isolated pilot investment. The causal structure underlying this criterion selection is further detailed in Section 4, which maps the dominant failure modes of AR deployment in industrial maintenance onto the five filtering dimensions through an Ishikawa cause-and-effect decomposition:

(C1)—Operational Frequency (weight: 0.15): Operations performed more frequently generate more cumulative opportunities for AR-assisted error reduction and efficiency gains. Frequency is established as a standard prioritization factor in railway maintenance scheduling [56]. Scored as: Low (0–1 times/week = 1), Moderate (2–4 times/week = 2), Frequent (5–7 times/week = 3).

(C2)—Execution Complexity (weight: 0.25): Operations involving multiple simultaneous subtasks, cross-referenced data, and constrained physical environments represent the highest-value AR targets, as AR directly reduces cognitive load and omission errors in procedurally dense tasks [14]. Scored as: Low complexity (1), Moderate complexity (2), High complexity (3).

(C3)—Safety Impact (weight: 0.30): Given the safety-critical nature of high-speed railway maintenance, safety impact carries the highest weight in the matrix. Risk-based maintenance frameworks consistently assign maximum priority weight to safety consequences, as failures in safety-critical operations carry irreversible human and operational costs [24,56]. Scored as: Low safety impact (1), Moderate (2), High safety impact (3).

(C4)—Traceability and Documentation Requirements (weight: 0.15): Operations requiring detailed, standardized reporting are strong AR candidates, as AR can automate traceability—replacing manual form-filling with real-time digital documentation [10]. Scored as: No formal traceability required (1), Partial documentation (2), Full standardized traceability mandatory (3).

(C5)—Scalability toward AR and AI Integration (weight: 0.15): Operations that are procedurally structured and data-rich provide the strongest foundation for future integration with AI-based decision support and Reinforcement Learning [57,58]. AI-augmented AR has been shown to substantially improve both task accuracy and technician confidence

in industrial maintenance deployments [39], and the strategic relevance of scalability as a selection criterion is further supported by task-specific analysis of AR deployment value in procedurally complex operations [14]. Scored as: Low scalability potential (1), Moderate (2), High scalability potential (3).

Each candidate operation is scored across all five criteria on the 1–3 ordinal scale. A weighted composite score is computed as:

$$S = 0.15 \cdot C1 + 0.25 \cdot C2 + 0.30 \cdot C3 + 0.15 \cdot C4 + 0.15 \cdot C5$$

yielding a maximum possible score of 3.00. Operations are ranked accordingly, and the top three proceed to expert validation in Step 4. The robustness of this ranking to variations in criterion weighting is verified through a sensitivity analysis presented in Section 4.4, which tests six alternative weight configurations including equal weights across all criteria.

Step 4—Expert Validation and Benchmarking

The shortlisted operations are submitted to structured expert validation involving 100 railway maintenance professionals across five operational departments: rolling stock inspection, traction systems, electrical maintenance, preventive maintenance planning, and safety and quality assurance. Participants were recruited through purposive sampling coordinated with department heads, with inclusion criteria requiring a minimum of three years of active railway maintenance experience and direct operational familiarity with at least one Level 2 maintenance operation. Personnel with exclusively administrative roles were excluded. Participation was voluntary and responses were collected anonymously to minimize social desirability bias [43]. A nine-item questionnaire probes each operation across dimensions of complexity, AR benefit potential, error risk, scalability, and implementation feasibility. Responses were collected on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree) and analyzed using the Kruskal–Wallis H test to assess the independence of expert judgments across seniority levels—defined as junior (3–5 years), mid-level (6–10 years), and senior (>10 years of railway maintenance experience)—with a significance threshold of $p < 0.05$. This non-parametric approach was selected given the ordinal nature of Likert-scale data and the typical non-normality of expert response distributions in heterogeneous professional panels.

Step 5—Final Use Case Selection

The operation receiving the highest aggregated score across expert validation dimensions, confirmed by radar analysis against all shortlisted candidates, is selected as the pilot use case. The selection criterion prioritizes not only immediate AR suitability but also architectural scalability toward the broader AR and RL deployment framework of which this study is a part. It is important to note that the present study's contribution is delimited to the operation selection framework. The full technical implementation of the AR–DQL architecture—including the computer vision pipeline, Magic Leap 2 deployment constraints, mathematical DRL formulations governing state spaces, action bounds and reward functions, and deep learning evaluation metrics (mAP, IoU)—falls within the scope of the companion study [18].

Figure 2 visually illustrates this structured five-step methodological process, demonstrating the systematic and data-driven approach used to ensure the objective and well-founded selection of the AR use case.

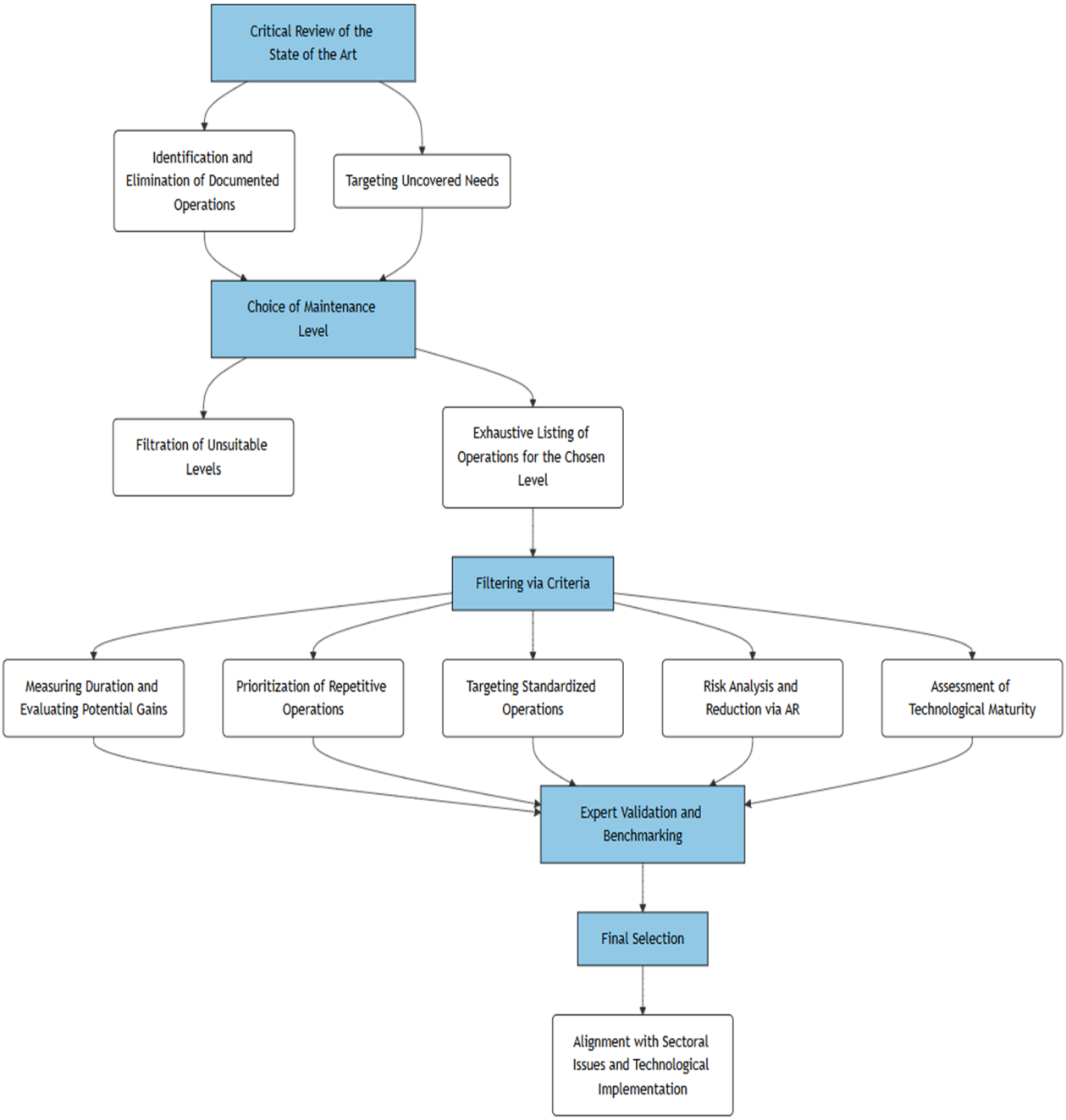


Figure 2. Structured Methodological Process for Selecting AR Use Case in Railway Maintenance.

4. Selection of the Target Maintenance Operation: Industrial Application and Validation

4.1. Industrial Context and Framework Definition

The selection methodology described in Section 3 is applied to a real industrial setting: a high-speed train maintenance facility operating in Morocco under the AFNOR FD X 60-000 standard. This partnership provided direct access to operational data, maintenance documentation, and domain expertise, constituting the empirical research environment within which the full five-step methodology is applied and validated. The facility operates a fleet of high-speed trainsets requiring structured preventive maintenance cycles across four classification levels. Level 2 operations—the focus of this study—are executed on a daily to weekly basis by teams of certified technicians across five operational departments.

As a preliminary analytical step and drawing on the four key factors for successful AR integration identified in prior work of this research program [8], Figure 3 maps the baseline preconditions that any candidate operation must satisfy before entering the selection pipeline. These preconditions—encompassing operational compatibility, safety compliance, knowledge transfer potential, and scalability toward digital integration—serve as qualitative filters applied prior to the quantitative scoring of Step 3, ensuring that only operations satisfying baseline feasibility requirements enter the multi-criteria evaluation.

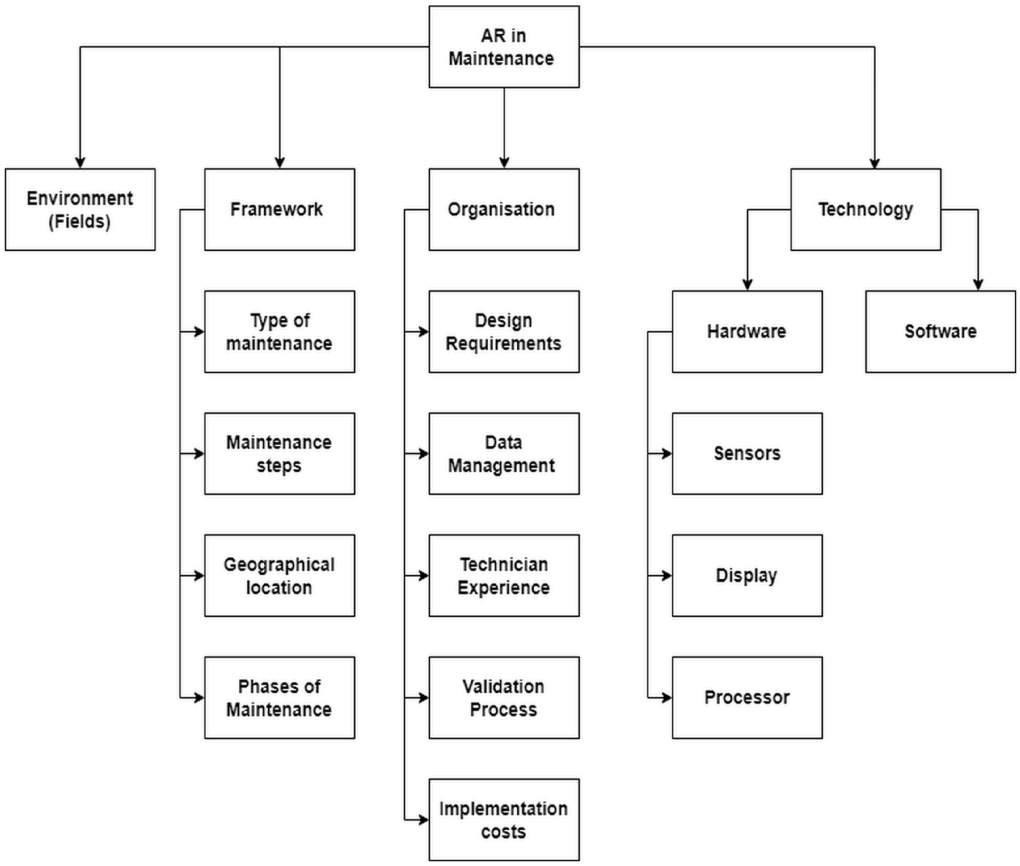


Figure 3. Key factors for successful AR Integration in Maintenance Processes [8].

To systematically identify the root causes of inefficient or slow maintenance that AR could address, an Ishikawa cause-and-effect diagram (Figure 4) was constructed around the central problem statement: "inefficient or slow maintenance execution with AR potential." The Ishikawa diagram is a recognized quality management tool for structured root cause analysis in industrial settings [59], and its application here serves to translate operational pain points into structured AR integration requirements across six causal dimensions: Methods, Machines, Manpower, Measurement, Materials, and Environment.

Notably, the six root causes identified through this analysis do not merely describe the problem—they directly inform and justify the five evaluation criteria applied in Step 3. Specifically, the Manpower and Measurement dimensions motivate the inclusion of Complexity (C2) and Safety Impact (C3); the Methods dimension grounds the Traceability criterion (C4); and the Environment dimension reinforces the Frequency criterion (C1), reflecting the operational regularity with which constrained-environment tasks are executed.

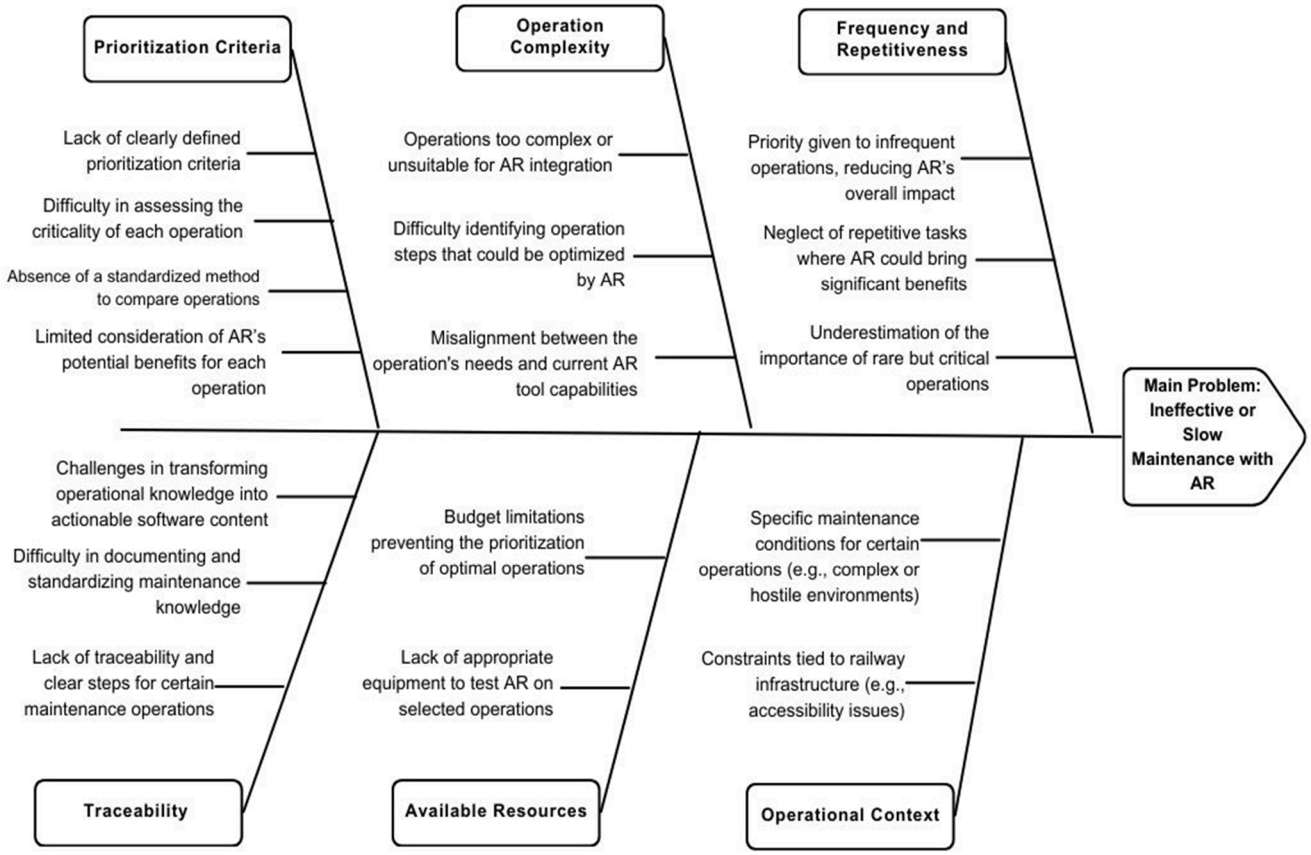


Figure 4. Ishikawa Diagram for Selecting a Maintenance Operation Assisted by AR.

4.2. Step 1 Applied—Redundancy Screening and Prerequisite Filtering

Prior to multi-criteria scoring, two categories of operations were excluded from the candidate pool through the redundancy screening and prerequisite filtering procedure defined in Step 1.

First, wheelset-related operations were excluded because this category has been extensively addressed in the AR railway literature [11,12], and targeting an already-documented operation would not constitute a novel AR contribution—consistent with the redundancy elimination objective of Step 1.

Second, operations lacking detailed standardized maintenance plans were excluded, as the existence of a documented procedural baseline is a non-negotiable prerequisite for AR-guided overlay: without a formalized procedure, AR cannot deliver consistent step-by-step guidance or automate traceability reporting [10]. All remaining operations entering Step 3 satisfy this prerequisite.

4.3. Step 2 Applied—Maintenance Level Compatibility Analysis

The facility operates a four-level maintenance classification in accordance with the AFNOR FD X 60-000 standard, ranging from simple operator-level interventions (Level 1) to complex overhaul operations (Level 4). Table 5 presents a structured comparison of each maintenance level against AR compatibility criteria—including procedural structure, operator qualification, intervention frequency, and regulatory traceability requirements.

The analysis confirms that Level 2 maintenance represents the optimal entry point for AR integration. Level 2 operations are characterized by moderate complexity, structured procedural steps, frequent execution, and mandatory standardized documentation—conditions that simultaneously maximize AR's guidance value, traceability automation potential, and organizational return on investment. This conclusion is consistent with

findings from the broader AR maintenance literature: Kour et al. (2022) [24] identify structured procedural compatibility and organizational accessibility as the two primary enablers of successful AR deployment in asset-intensive industries, both of which are uniquely satisfied at Level 2 in the AFNOR classification.

Table 5. Comparison of Adaptability of the Four Maintenance Levels with AR.

Criterion	Level 1	Level 2	Level 3	Level 4
Description	Simple tasks: inspections, cleaning, adjustments.	Technical interventions requiring tools and specific diagnostics.	Heavy repairs or replacements requiring a specialized workshop.	Complete revisions and complex overhauls with disassembly.
AR Integration Feasibility	Not feasible: tasks are too simple to justify AR deployment overhead.	Highly feasible: structured procedural steps are directly compatible with AR overlay guidance.	Partially feasible: heavy equipment and unstructured environments limit current AR capabilities.	Not feasible: extreme variability and specialized resources exceed current AR technological reach.
AR Compatibility	Not useful for simple tasks, little added value.	Ideal: AR provides visual overlays for diagnostics and detailed instructions.	Difficult: requires heavy equipment, incompatible with current AR tools.	Low: complex processes requiring specialized human and technological resources.
Intervention Cost	Low	Medium	High	Very high
Impact on Performance	Limited	Significant: directly improves reliability and avoids breakdowns requiring heavy interventions.	Moderate: important impact but comes too late to avoid critical degradation.	Indirect: long-term impact, little immediate effect on daily performance.
Organizational Accessibility	Full internal access: executable by non-specialized operators without AR support infrastructure.	Full internal access: performed by trained technicians within a structured organizational framework, compatible with AR deployment.	Restricted internal access: requires specialized workshop and external coordination, limiting AR integration scope.	Outsourced: Level 4 equivalent tasks handled externally, outside the organization's AR deployment authority.
AR Return on Investment (ROI)	Too low	Excellent: quick ROI due to structured procedures, reduced error risk, and direct impact on train availability.	Moderate: justified only for specific and critical tasks.	Low: high deployment cost for marginal operational benefits.
Regulatory Compliance Burden	Minimal: no formal certification or traceability requirements, AR adds no compliance value.	Moderate: standardized documentation required; AR can automate traceability and reduce compliance risk.	High: interventions require formal certification and workshop-level controls incompatible with AR-only solutions.	Very high: regulatory oversight requires multi-party validation; AR cannot substitute for required human expertise.

Level 1 operations are excluded because their simplicity does not justify the deployment overhead—AR adds no measurable guidance value where procedures require minimal cognitive engagement. Levels 3 and 4 are excluded for the opposite reason: their complexity, specialized resource requirements, and outsourced execution place them beyond the current reach of AR-only solutions, a limitation documented by Mikušová et al. (2024) [13] in their analysis of the gap between AR laboratory demonstrations and operational field deployment. This bidirectional exclusion logic—eliminating both the too-simple and the too-complex—is a distinctive feature of the proposed methodology that distinguishes it from prior work, which applies AR to operations selected without explicit compatibility screening.

4.4. Step 3 Applied—Candidate Operation Listing and Initial Filtering

Following the exclusions performed in Section 4.2, the remaining Level 2 preventive maintenance operations were systematically listed and evaluated through the multi-criteria decision matrix defined in Section 3. Table 6 summarizes these operations along with their estimated execution duration and weekly frequency, anonymized in accordance with the confidentiality requirements of the industrial partner. Duration is classified as Low (<2 h), Moderate (2–5 h), or High (>5 h); frequency as Low (0–1 times/week), Moderate (2–4 times/week), or Frequent (5–7 times/week).

Table 6. Multi-Criteria Decision Matrix—Scoring Results.

Op	C1 (15%)	C2 (25%)	C3 (30%)	C4 (15%)	C5 (15%)	Score /3.00	Justification
OP10	3 (Frequent)	3 (High)	3 (Critical)	3 (Full mandatory)	2 (Moderate-High)	2.85	C1: 7–9 executions/week, highest frequency among all Level 2 candidates. C2: 5 sequential sub-steps requiring simultaneous temperature reading, oil level cross-referencing, tolerance table consultation, anomaly detection, and mandatory form completion in a low-light constrained environment. C3: Direct linkage to the Main Transformer; immediate safety and service consequence. C4: Full standardized traceability mandatory under AFNOR FD X 60-000. C5: Highly structured and data-rich; strong AI integration potential, partially constrained by current sensor availability.
ATS0	2 (Moderate)	2 (Moderate)	2 (Moderate)	2 (Partial)	2 (Moderate)	2.00	C1: 5–7 executions/week; frequency approaches the upper moderate bound but is driven by fleet rotation scheduling rather than an intrinsic daily requirement. C2: Multi-step procedure with moderate cognitive load. C3: Moderate impact on availability. C4: Partial standardized documentation required. C5: Moderately structured; compatible with AR extension.
OP37	2 (Moderate)	2 (Moderate)	1 (Low-Moderate)	2 (Partial)	1 (Low)	1.70	C1: 2 executions/week; moderate frequency despite long execution duration. C2: Multi-step procedure with moderate complexity. C3: Low-to-moderate safety impact. C4: Partial documentation required. C5: Limited scalability toward AI integration.
Op D	1 (Low)	1 (Low)	2 (Moderate)	2 (Partial)	1 (Low)	1.55	—
Op E	2 (Moderate)	2 (Moderate-High)	1 (Low)	1 (Partial)	1 (Low)	1.45	—
Op B	2 (Moderate)	1 (Low)	1 (Low-Moderate)	2 (Partial)	1 (Low)	1.30	—
Op G	1 (Low)	2 (Moderate-High)	1 (Low)	1 (Partial)	1 (Low)	1.25	—
Op I	1 (Low)	2 (Moderate-High)	1 (Low)	1 (Partial)	1 (Low)	1.25	—
Op A	1 (Low)	2 (Moderate)	1 (Low)	1 (Partial)	1 (Low)	1.25	—
Op F	1 (Low)	1 (Low-Moderate)	1 (Low)	2 (Partial)	1 (Low)	1.15	—
Op H	1 (Low)	1 (Low)	1 (Low-)	2 (Partial)	1 (Low)	1.15	—
Op C/J	1 (Low)	1 (Low)	1 (Low)	1 (Partial)	1 (Low)	1.00	—

A sensitivity analysis was conducted by systematically varying the weights of the two highest-weighted criteria—Safety (C3) and Complexity (C2)—by ±20% of their baseline values, with the resulting delta redistributed equally across the remaining criteria to maintain a unit sum. An additional scenario with equal weights across all five criteria was also tested. Table 7 presents the six weight configurations and their resulting composite scores. Across all six scenarios, OP10 retained the highest composite score—ranging from 2.800 to 2.865—with a minimum margin of 0.800 points over the second-ranked operation (ATS0, score = 2.000 in all scenarios). The top-three ranking remained stable across all configurations (OP10 > ATS0 > OP37), confirming that the selection outcome is robust to moderate and extreme variations in criterion weighting.

Table 7. Sensitivity Analysis—Composite Scores Under Six Weight Configurations.

Operation	Baseline	S1 Safety +20%	S2 Safety −20%	S3 Complexity +20%	S4 Complexity −20%	S5 Equal Weights
OP10	2.850	2.865	2.835	2.862	2.838	2.800
ATS0	2.000	2.000	2.000	2.000	2.000	2.000
OP37	1.700	1.505	1.595	1.575	1.525	1.600
Op D	1.55	1.495	1.405	1.425	1.475	1.400
Op E	1.45	1.370	1.430	1.438	1.363	1.400
Op B	1.30	1.270	1.330	1.275	1.325	1.400
Op G	1.25	1.235	1.265	1.300	1.200	1.200
Op I	1.25	1.235	1.265	1.300	1.200	1.200
Op A	1.25	1.235	1.265	1.300	1.200	1.200
Op F	1.15	1.135	1.165	1.137	1.163	1.200
Op H	1.15	1.135	1.165	1.137	1.163	1.200
Op C/J	1.00	1.000	1.000	1.000	1.000	1.000

The complete ranking of all twelve candidate operations remained invariant across all six weight configurations. Minor score fluctuations are observed in the lower-ranked operations under certain scenarios—most notably under equal weights (S5), where Operations B, F, and H converge toward 1.400 and 1.200 respectively—but these variations do not produce any rank reversal at any position in the hierarchy. The three operations shortlisted for expert validation (OP10, ATS0, OP37) are separated from the remaining candidates by a decisive score gap that no tested weight configuration can bridge, confirming that the shortlist is methodologically stable and independent of the specific AHP weight assumptions.

The robustness of the selection outcome to scoring scale granularity follows directly from the score gap documented above. The 0.850-point margin separating OP10 (2.850) from the second-ranked candidate ATS0 (2.000) represents 28% of the total three-point scale range. For a finer-grained five-point scale to alter this ranking, expert scoring patterns across all five criteria simultaneously would need to reverse their direction—a scenario that neither the six weight configurations tested here nor the convergent expert validation reported in Section 4.5 (OP10 mean preference: 54.0% vs. 28.8% for ATS0 across nine questions) renders plausible. A five-point Likert scale would improve scoring resolution among lower-ranked candidates where margins are smaller, but would not bridge a gap of this structural magnitude at the top of the ranking.

4.5. Step 4 Applied—Expert Validation and Benchmarking

The three operations scoring highest in the multi-criteria matrix—OP10 (2.850), ATS0 (2.000), and OP37 (1.7)—were selected for expert validation. The decision to submit exactly

three candidates, rather than a broader or narrower shortlist, reflects a deliberate methodological trade-off: fewer than three candidates would reduce the discriminatory power of expert comparison, while more than three would dilute expert attention across operationally heterogeneous options and increase respondent fatigue in a nine-item questionnaire format. A shortlist of three also ensures that the score gap separating these operations from the remaining candidates—minimum 0.150 points below OP37—is preserved as a meaningful selection threshold.

These three operations were subsequently submitted to structured expert validation involving 100 railway maintenance professionals, distributed across five operational departments: Planning and Methods (n = 20), Electrical (n = 25), Mechanical (n = 25), Industrial Coordination (n = 15), and Engineering (n = 15).

Three explicit measures were implemented to ensure recruitment neutrality and minimize response bias [43]: (1) Participants were approached exclusively through department heads using pre-established staff lists, precluding voluntary self-selection and the participation bias it introduces. (2) Full anonymity of responses was guaranteed throughout the data collection process, minimizing social desirability effects on expert judgments. (3) No participant was informed of the preliminary MCDM ranking prior to completing the questionnaire, eliminating anchoring effects that could have artificially inflated consensus around the top-ranked candidate.

Participants were recruited through purposive sampling with inclusion criteria requiring a minimum of three years of active railway maintenance experience and direct operational familiarity with at least one Level 2 maintenance operation. Personnel with exclusively administrative roles were excluded.

The sample encompasses three experience brackets—3–5 years (n = 50), 6–10 years (n = 37), and 11–15 years (n = 13)—as illustrated in Figure 5, ensuring representation across seniority levels and robustness against individual expertise bias, a methodological requirement established in expert validation frameworks for industrial AR deployment [14].

The nine-item questionnaire probed the three shortlisted operations across the same five dimensions as the decision matrix: complexity, AR benefit potential, error risk, scalability, and implementation feasibility. This alignment ensures full methodological consistency between the quantitative and qualitative phases of the selection. The specific questions were:

1. Which operation has the highest complexity and would most benefit from AR assistance?
2. Which operation could benefit most from 3D modeling of maintenance steps?
3. Which operation would gain the most time savings from AR-guided execution?
4. Which operation has the highest risk of human error addressable by AR?
5. Which operation offers the best scalability for future AR applications?
6. Which would be easiest to integrate into existing processes with minimal modification?
7. Which operation would most benefit from improved operator safety through AR?
8. Considering implementation resources, which operation is the most realistic starting point?
9. Overall, which operation presents the best conditions to serve as an AR use case?

Prior to interpreting the directional results, three statistical tests were conducted to establish the validity, significance, and robustness of the findings.

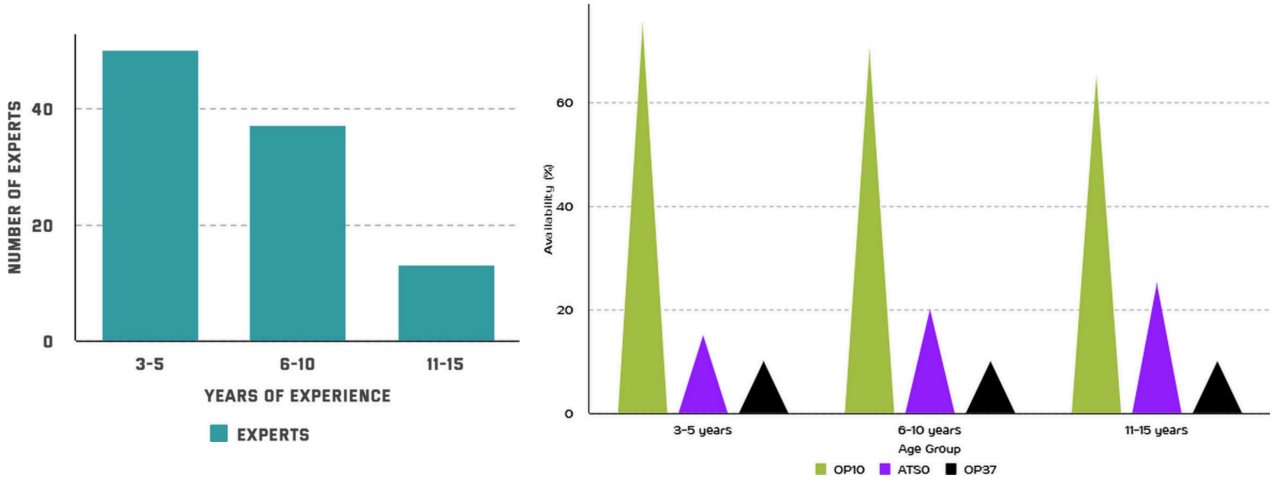
First, a Kruskal–Wallis test was conducted to determine whether the distribution of votes across the three candidate operations differed significantly across the nine questionnaire items. The test yielded $H = 18.479$ ($p < 0.001$), confirming that the observed differences in preference between OP10, OP37, and ATS0 are statistically significant and cannot be attributed to random variation.

Second, a Kruskal–Wallis test was applied to assess whether expert experience level—defined as three brackets: 3–5 years (n = 50), 6–10 years (n = 37), and 11–15 years (n = 13)—

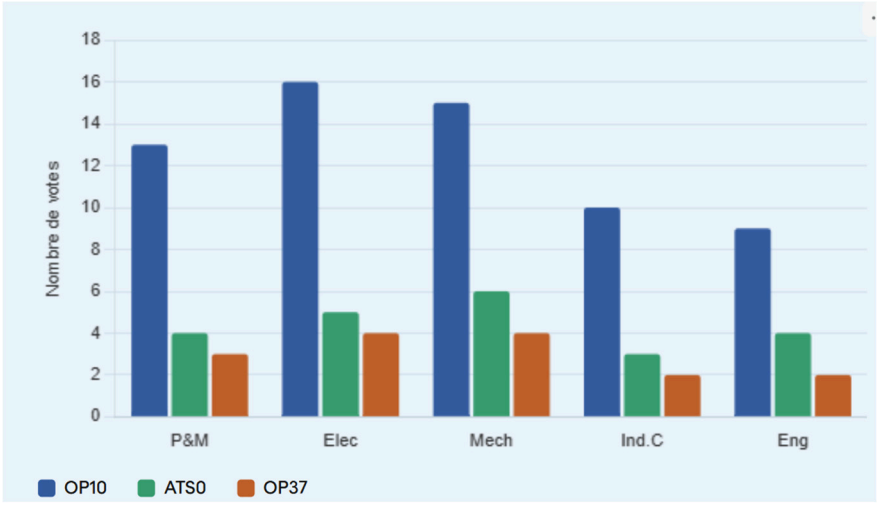
influenced preference for OP10. The test yielded no statistically significant difference in OP10 preference across experience groups ($p > 0.05$), indicating that OP10 selection is consistent across different levels of professional experience. Median OP10 selections were comparable across groups (3–5 years: 5.0/9; 6–10 years: 5.0/9; 11–15 years: 4.5/9). This confirms that convergence on OP10 is robust and independent of experience level, supporting the validity of the expert consensus.

Figure 5a illustrates the distribution of respondents by experience level across the three seniority brackets. Figure 5c presents the distribution of operation evaluations across all nine criteria disaggregated by technician experience level. To further assess whether departmental affiliation introduced systematic variation, vote distributions for OP10 were disaggregated across the five operational departments for Q9, as illustrated in Figure 5b. The quantitative stability of these results across both seniority and departmental subgroups is synthesized following the three statistical tests reported here.

Third, Kendall's coefficient of concordance (W) was computed, yielding W = 0.037 ($\chi^2 = 29.949$, $p < 0.001$). Given the compositional diversity of the panel, this value requires careful interpretation, which is addressed in detail in Section 6.1.



(a) (c)



(b)

Figure 5. Expert survey results: (**a**) distribution of years of experience of surveyed technicians and engineers; (**b**) distribution of expert votes by operational department for Q9 (n = 100); (**c**) distribution of operation evaluations based on technician and engineer experience for each criterion.

The stability of the final selection score is supported by evidence operating at three complementary levels. At the quantitative level, the sensitivity analysis reported in Section 4.4 demonstrates that OP10's composite score of 2.850 remains the highest across all six weight configurations tested, with a minimum margin of 0.800 points over the second-ranked candidate—a gap that no tested variation in criterion weighting can bridge. At the individual level, the Mann–Whitney result (U = 1283.5, p = 0.230) confirms that seniority-based subgroup variance does not influence the directional preference for OP10, with median selection rates of 5.0/9 and 4.5/9 for junior and senior experts respectively—a difference in practical insignificance. At the departmental level Table 8), the maximum inter-group variance in OP10 preference across the five operational departments is 5 percentage points (range: 60–65% on Q9), confirming that localized professional perspectives do not destabilize the aggregate selection outcome (Figure 6). Taken together, these three convergent stability indicators confirm that OP10 is the optimal AR use case, a conclusion robust to both quantitative weight assumptions and qualitative subgroup variation.

Table 8. Expert Vote Distribution per Question (n = 100).

Question	Dimension	OP10	OP37	ATS0
Q1—Complexity & AR benefit	Complexity	63%	28%	9%
Q2—3D modeling benefit	Complexity	57%	19%	24%
Q3—Time savings potential	Efficiency	42%	48%	10%
Q4—Human error risk	Safety	58%	24%	18%
Q5—Scalability for future AR	Scalability	48%	27%	25%
Q6—Ease of integration	Feasibility	37%	37%	26%
Q7—Operator safety improvement	Safety	67%	21%	12%
Q8—Realistic starting point	Feasibility	44%	34%	22%
Q9—Overall best AR use case	Global	70%	21%	9%
Mean across all questions	—	54.0%	28.8%	17.2%

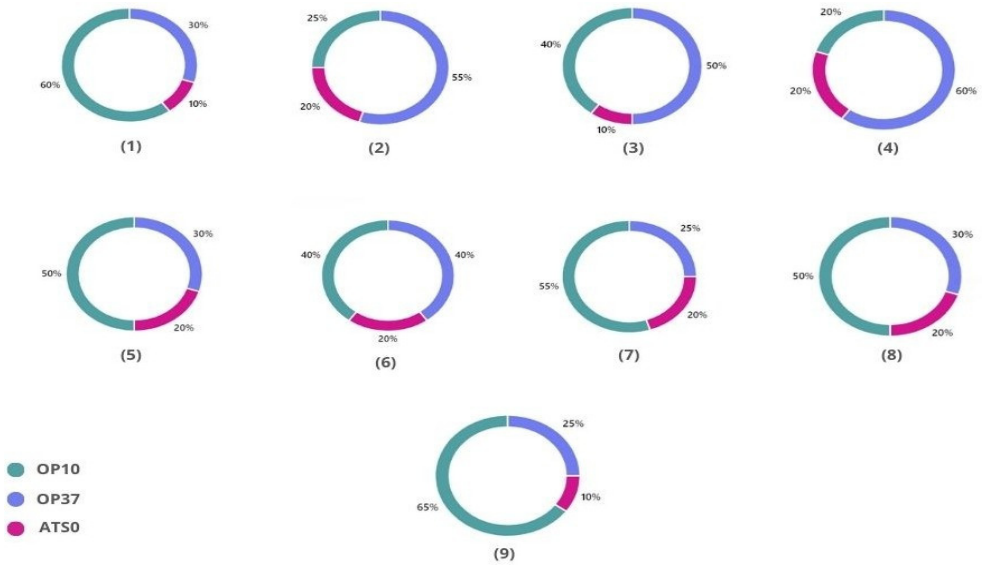


Figure 6. Percentage of responses per operation for each question.

Two items warrant specific discussion. On Q3 (time savings potential), OP37 received a slight plurality over OP10 (48% vs. 42%), reflecting expert perception that OP37's longer

baseline execution time offers proportionally greater scope for AR-driven efficiency gains. However, this item-level result does not alter the overall ranking: OP10 dominates on the seven remaining questions, and the Q3 divergence is consistent with the lower safety and complexity scores assigned to OP37 in the quantitative matrix—where time savings alone, absent safety and complexity justification, are an insufficient basis for use case selection. On Q6 (ease of integration), OP10 and OP37 tied at 37%, reflecting comparable implementation feasibility. This tie is expected given that both operations share similar procedural documentation maturity within the facility.

4.6. Step 5 Applied—Final Selection and Use Case Characterization

OP10—the insulating oil level verification of the Main Transformer—was selected as the most suitable pilot use case for AR integration. The radar diagram in Figure 7 confirms OP10's systematic dominance across all five evaluation dimensions relative to ATS0 and OP37.

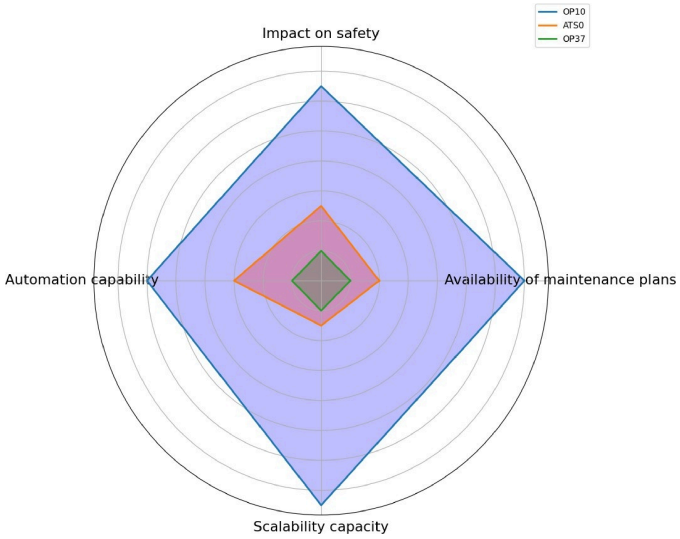


Figure 7. Radar Diagram—Comparative Analysis of Criteria for OP10, ATS0, and OP37.

OP10 involves five sequential steps requiring meticulous technical attention and simultaneous multitasking across measurement, cross-referencing, and standardized reporting tasks:

i. Preparation for Verification—If the Main Transformer (MT) temperature differs from ambient, pumps must operate for at least 5 min to homogenize oil temperature, preventing measurement errors from thermal gradients.
ii. Temperature Measurement—A laser-guided thermometer is used at the MT center-of-gravity marker (Figure 8a) to obtain precise, non-intrusive temperature readings.
iii. Oil Level Verification—The oil level is read on a graduated scale (Figure 8b) and cross-referenced against the measured temperature within a ± 7 °C tolerance. Anomalies trigger mandatory visual inspection for leaks.
iv. Visual Inspection—Conducted at the silica gel breather (Figure 8c), this step evaluates desiccant bead condition, seal and joint integrity, and signs of wear or leakage.
v. Standardized Reporting—The technician completes a mandatory traceability form (Figure 8d) documenting all measurements, observations, and findings—a regulatory requirement that AR can automate in real time, replacing manual form-filling with digitally generated, timestamped records.

With OP10 fully characterized across its procedural, safety, and traceability dimensions, the following section defines the technological architecture required to implement this selection.



(**a**)



(**b**)



(**c**)

SIANA - - RGVM PRODUCT SELF-CHECK SHEET

Level 2 Preventive Maintenance type OP10 following ===> V T6 114-J4K

Date: **Track:** **Trains:**

Operators(s): **Laser-sighted thermometer number:**

Thermometer use-by date:

Examination	M1	M2	Di N°	Observation
Oil level check				
Silica gel check				
Main transformer N°				

Oil level check:				Di N°	Observation
M1		M2			
T° taken with the laser sighted thermometer	°C	T° taken with the laser sighted thermometer	°C		
Oil level reading on the strip	°C	Oil level reading on the strip	°C		
Absolute value differences between the two T° readings	°C	Absolute value differences between the two T° readings	°C		

THE OIL LEVEL MUST CORRESPOND TO THE TEMPERATURE READ BY THE LASER-SIGHTED THERMOMETER WITH A TOLERANCE OF ±7 °C

Silica Gel Control		Di N°	Observation
M1	M2		
DESICCANT BEAD COLOR	DESICCANT BEAD COLOR		

ORANGE COLOR = DRY => LEAVE IN SERVICE

GREEN OR WHITE COLOR = WET => REPLACE DESICCANT BEADS

BLACK COLOR + OIL => REPLACE DESICCANT BEADS

VALIDATION OF OPERATIONAL CONFORMITY

LAST AND FIRST NAME, Signature (S) Operator(s):	Manager validation	
	Name	
	Position	
	Date	
	Signature	

(**d**)

Figure 8. OP10 operational components: (**a**) MT center-of-gravity marker for laser temperature measurement; (**b**) oil level reading on graduated scale; (**c**) silica gel breather of the main transformer; (**d**) standardized reporting form.

5. Technological Stack and Experimental Validation

This section serves a delimited and explicit function within the present article. Sections 5.1–5.3 constitute original contributions of this work: the implementation framework, software stack selection, and hardware selection were performed as direct outputs of the operation selection methodology developed in Sections 3 and 4—demonstrating that the principled identification of OP10 as the optimal AR target conditions not only the decision to deploy AR, but the specific technological architecture required to do so effectively. The experimental results summarized in Section 5.4, however, are drawn from a companion publication [18], which provides the complete experimental design, training protocol, hardware calibration procedure, and inferential statistical analysis underlying these figures. Their inclusion here is strictly limited to establishing the empirical alignment between the selection methodology and the deployment outcomes it enables. Readers seeking the full experimental detail are directed to [18].

5.1. Implementation Framework

With OP10 identified and validated as the optimal AR target through the five-step methodology, this section defines the technological architecture required to implement this selection—transitioning from the evidence-based choice of Section 4 to its concrete industrial implementation.

To ensure a structured and traceable transition from the existing manual procedure to an AR-assisted workflow, a detailed implementation framework was developed prior to any software or hardware deployment. Rather than modeling the Main Transformer in its entirety, three components were selected for 3D reconstruction—the silica gel breather, the oil gauge display, and the oil basin marker (Figure 9)—chosen for their direct relevance to the inspection steps and their geometric stability under AR tracking conditions. These components were modeled at realistic scale to minimize overlay errors and ensure measurement precision consistent with the ±7 °C tolerance required by the OP10 procedure. Figure 10 maps the reasoning and sequencing logic for integrating AR into each step of the OP10 maintenance operation, defining the successive decision points, system interactions, and operator feedback loops that the AR application must support.

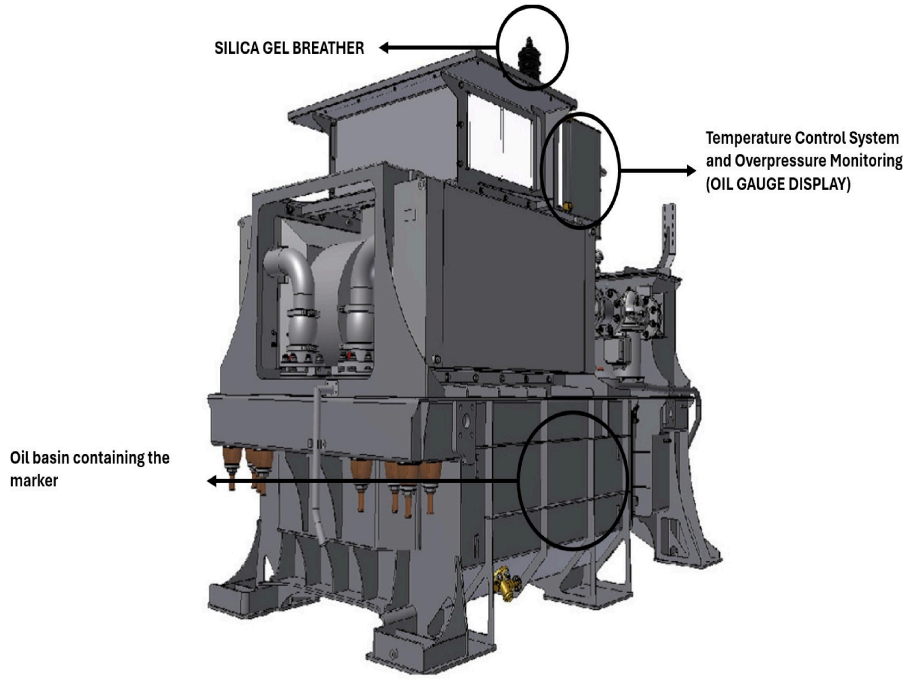


Figure 9. Components of the MT Modeled for Accurate Oil Level Measurement.

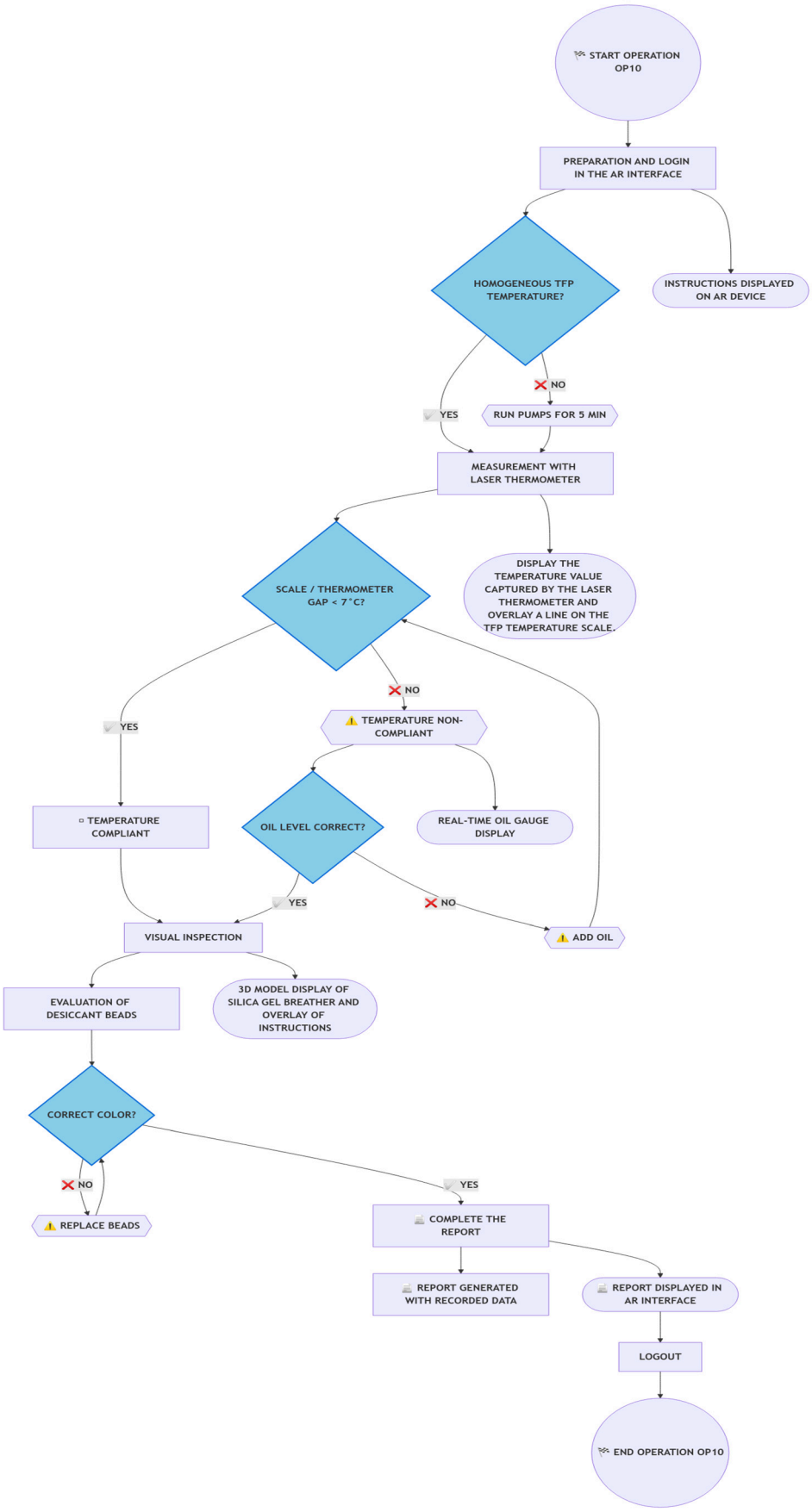


Figure 10. Implementation Framework and Reasoning for Integrating AR into OP10 Maintenance Steps.

The functional requirements derived from this framework fall into three categories. First, interactive display: the AR interface must deliver step-by-step procedural guidance and real-time display of critical values—temperature readings, oil level thresholds, and silica gel color states—directly in the operator's field of view. Second, 3D visual assistance: marker-based overlays on the oil gauge and a 3D model of the silica gel basin must enable precise visual inspection without requiring the operator to consult paper documentation. Third, automated reporting: voice or gesture-based input must allow the operator to record observations in real time, with automatic consolidation into the standardized OP10 reporting form, eliminating the need for a second agent to handle documentation during the operation.

The overall software development flow, from sensor input through application and scene layers to display output, is illustrated in Figure 11 and served as the guiding architecture for the subsequent prototype development.

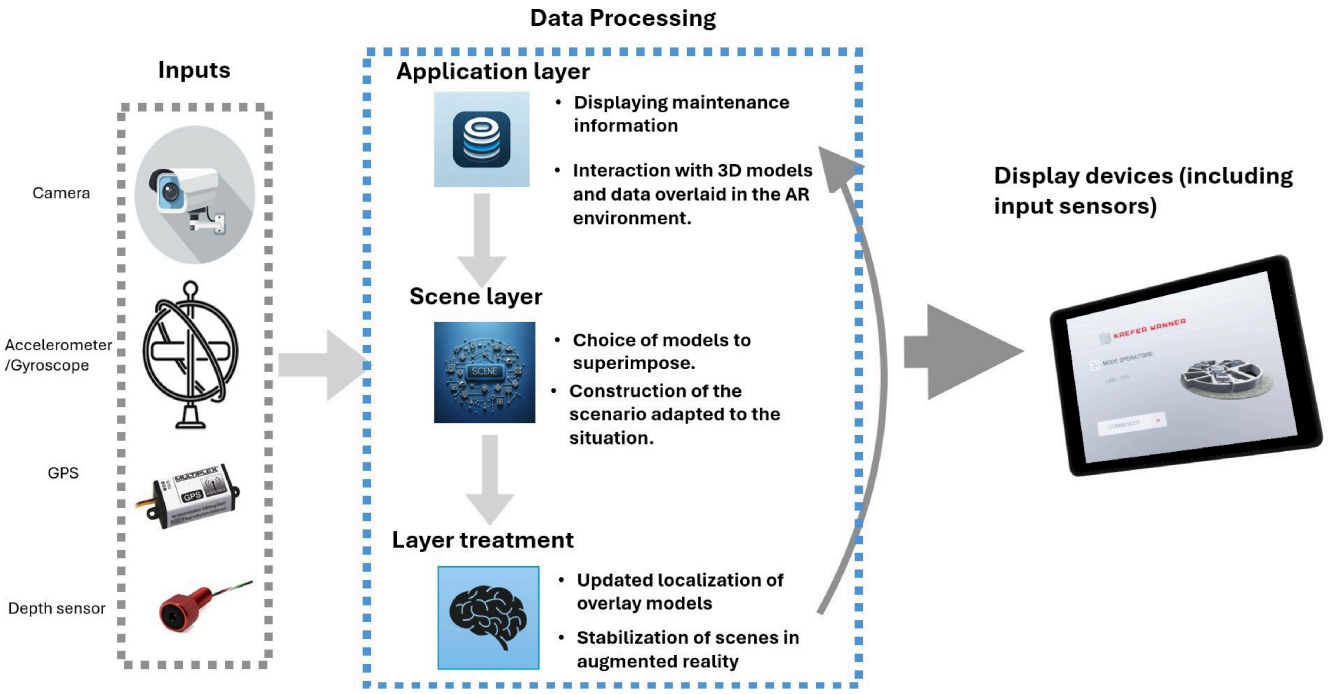


Figure 11. Overall Flow Diagram for Software Development Process.

5.2. Software Stack Selection

The software stack was selected through a systematic evaluation of existing AR solutions across three functional categories recognized by the extended reality (XR) community: AR Content Management Systems (CMS AR), AR Development Software Development Kits (SDK AR), and Game Engines for AR. The selection process was anchored in the specific functional requirements of OP10—prioritizing AR instruction display capability, 3D object interaction, and report automation—rather than on general literature recommendations, ensuring direct alignment between technological capability and operational need.

Candidate solutions were evaluated across two complementary dimensions. Technical capability was assessed against the three functional requirements defined in Section 5.1, drawing on vendor documentation and peer-reviewed evaluations of XR deployment in industrial maintenance [40,60]. Adoption reliability was assessed using user ratings and review volumes aggregated on the G2 platform for 2022–2025, which provides a large-sample proxy for real-world deployment maturity and cross-industry validation—a criterion particularly relevant when selecting foundational infrastructure for an industrial pilot (Table 9). These two dimensions are treated as complementary:

technical capability determines functional eligibility, while adoption reliability provides secondary confirmation that the selected solutions have demonstrated stability beyond controlled research environments.

Table 9. Comparison of AR Solutions Based on Key Criteria.

Category	Product	Version	Rating/Review Count	AR Instruction Display	3D Object Interaction	Report Automation/ CMS
CMS AR	echo3D	https://www.echo3d.com/	4.4/583	Strong	Medium	Weak
CMS AR	ArborXR	2025.47.0	4.9/564	Strong	Weak	Strong
CMS AR	Zapworks	https://zap.works/	4.5/530	Strong	Strong	Weak
CMS AR	Strapi	v5	4.7/564	Strong	Medium	Medium
SDK AR	ARKit	5.x	4.3/574	Strong	Strong	Weak
SDK AR	ARCore	1.48	4.5/580	Strong	Strong	Weak
SDK AR	Vuforia Engine	11	4.3/554	Strong	Strong	Weak
SDK AR	ARToolKit	1.1.9	4.9/518	Strong	Strong	Weak
SDK AR	OpenXR	1.1	4.6/544	Strong	Strong	Medium
Game Engine	Unity	6 (6000.3 LTS)	4.5/358	Strong	Strong	Weak

From this evaluation, three complementary solutions were selected to form an integrated and scalable stack. ArborXR (CMS AR) enables centralized management and deployment of AR content across heterogeneous headset fleets, ensuring consistent delivery of procedural instructions across devices and maintenance sites. Although ArborXR offers limited native 3D object manipulation capability—a function handled within the development environment rather than the CMS layer—its centralized content management and deployment tracking features are precisely those required at the operational management level of the OP10 deployment. Vuforia Engine (SDK AR) provides hybrid tracking capabilities that combine ARKit/ARCore-based SLAM for environmental mapping with proprietary millimeter-accurate image and 3D target detection—a combination that neither pure SLAM nor marker-only solutions can match, and that is particularly suited to the constrained, low-contrast visual environments typical of train engine compartments [40]. Unity (AR Game Engine) serves as the development and rendering platform, enabling high-fidelity real-time 3D visualization and seamless integration with Vuforia's tracking outputs [60].

5.3. Hardware Selection

Hardware selection was governed by four operational constraints specific to the OP10 context: compatibility with both Vuforia Engine and ArborXR, capacity to function in low-light engine compartments, ergonomic suitability for prolonged hands-free use, and versatility for future AR use cases within the same maintenance program. A systematic comparison of devices supported by both software solutions was conducted (Table 10), from which two candidates emerged as technically qualified: the RealWear HMT-1/HMT-1Z1 and the Magic Leap 2.

The RealWear HMT-1/HMT-1Z1 offers robust hands-free operation through voice command interaction, a modular mounting system compatible with standard safety helmets, and an attachable LED lamp that directly addresses the low-light constraint of OP10. Its approximately 380 g weight and head-mounted configuration minimize operator fatigue during multi-step inspections. The Magic Leap 2, while lacking integrated lighting, delivers superior immersive AR visualization through advanced 3D rendering and direct Unity integration, with a visor weight of approximately 260 g better suited to extended precision

tasks. The low-light constraint is addressed by attaching a lamp to the operator's existing personal protective equipment—a solution that preserves both visibility and ergonomic comfort without requiring additional hardware.

Table 10. Devices Supported by Vuforia Engine and ArborXR.

Brand	Devices Compatible with Vuforia	Devices Compatible with ArborXR	Manufacturer	City	Country
Apple	iPhone (ARKit), iPad	Apple Vision Pro	Apple Inc.	Cupertino, CA	USA
Samsung	Galaxy S, Tab, A (ARCore)	—	Samsung Electronics Co., Ltd.	Suwon	Republic of Korea
Google	Pixel 6/7/8, Pixel Tablet	—	Google LLC	Mountain View, CA	USA
Microsoft	HoloLens 2	—	Microsoft Corporation	Redmond, WA	USA
Magic Leap	Magic Leap 2	Magic Leap 2	Magic Leap, Inc.	Plantation, FL	USA
RealWear	HMT-1, HMT-1Z1	HMT-1, HMT-1Z1, Navigator 500	RealWear, Inc.	Vancouver, WA	USA
Meta	—	Quest 3, Quest Pro, Quest 2	Meta Platforms, Inc.	Menlo Park, CA	USA
Pico VR	—	Neo 3 Pro, PICO 4 Enterprise	PICO Technology (ByteDance subsidiary)	Beijing	China
HTC VIVE	—	VIVE Focus 3, XR Elite	HTC Corporation	New Taipei City	Taiwan
Vuzix	—	M4000, Blade	Vuzix Corporation	Rochester, NY	USA

On the basis of this analysis, the Magic Leap 2 was selected as the primary deployment device for the OP10 use case, with the RealWear HMT-1/HMT-1Z1 retained as a validated alternative for environments where voice-command interaction and integrated lighting take priority over visualization fidelity. Known operational constraints of the Magic Leap 2 were explicitly evaluated prior to selection. Its 70° diagonal FOV (45° H × 55° V) [61]—wider than the HoloLens 2's 52° diagonal (43° H × 29° V) [62]—is particularly relevant to the spatial configuration of OP10, where the three inspection zones (silica gel breather, oil gauge, oil basin marker) are distributed across a 65° angular spread within the engine compartment; the wider FOV reduces the number of head reorientations required during a 30 min inspection session, minimizing procedural interruptions. The absence of integrated illumination is addressed by attaching a standard frontal lamp to the operator's existing PPE—a solution already standard practice at the facility and validated as non-interfering with Vuforia tracking under the low-contrast lighting conditions of the compartment. At 260 g, the Magic Leap 2 visor is less than half the weight of the HoloLens 2 (566 g) [62], a decisive ergonomic advantage for extended precision tasks in a constrained posture. These constraint-aware considerations, combined with the device's native Unity integration and ArborXR compatibility, grounded the final hardware selection in operational reality rather than benchmark performance alone.

5.4. Experimental Validation—Summary of Results from [18]

The technological stack defined in Sections 5.1–5.3 was subsequently integrated with an IoT sensor layer and a Deep Q-Learning (DQL) optimization module to form a complete hybrid AR-RL framework. The full experimental design, validation protocol, and inferential statistical analysis underlying the results summarized below are reported in the companion publication [18]; the present section is strictly limited to the performance outcomes that establish the operational consequences of the selection decision developed in Sections 3 and 4.

The integrated system was validated over a three-month period under simulated and semi-operational conditions representative of the Moroccan high-speed rail maintenance environment. Across all key performance indicators, results were significant and consistent, as summarized in Table 11 and Figure 12.

Table 11. Summary of Experimental Validation Results for AR-Assisted OP10—Three-Month Evaluation Period.

Performance Indicator	Baseline (Manual Procedure)	AR-Assisted Result	Improvement Range	Measurement Condition
Intervention time per execution	~45–60 min (estimated)	~28–38 min	−34% to −47%	Simulated + semi-operational
Human error rate per execution	Reference = 100%	30–45% of baseline	−55% to −70%	Simulated + semi-operational
Failure-related costs (OP10-linked)	Reference = 100%	58–72% of baseline	−28% to −42%	Semi-operational
Additional efficiency gain (RL layer)	—	+15–20% vs. AR-only	+15% to +20%	Simulated
Evaluation duration	—	3 months	—	Moroccan HSR facility
Conditions	—	Simulated + semi-operational	—	Pre-deployment validation

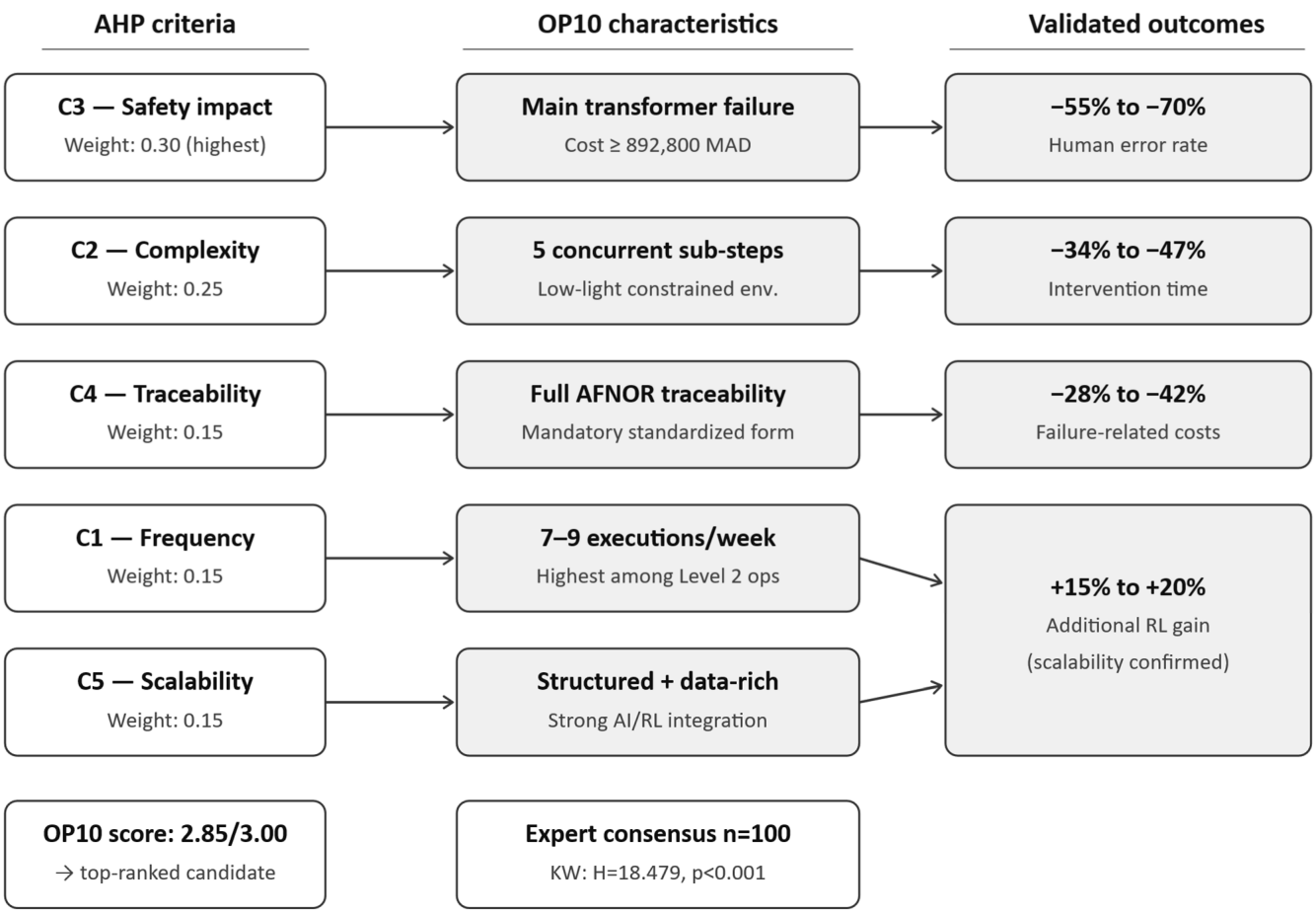


Figure 12. From Criteria to Outcomes: AHP-Driven Selection of OP10 and Experimental Validation of AR Integration Performance.

These results validate the upstream selection decision in two respects. First, the 34–47% reduction in intervention time is directly attributable to the procedural density of OP10—the precise characteristic that drove its maximum C2 score in the multi-criteria matrix. Second, the 55–70% reduction in human error rates reflects the measurable impact of AR guidance on the two failure modes prioritized by the C3 Safety Impact criterion: omission errors and measurement misinterpretation. The 28–42% decline in failure-related costs confirms the return on investment anticipated by the MCDM scoring and validates the economic rationale for targeting safety-critical, high-traceability operations as primary AR candidates.

Taken together, these results confirm that OP10, as identified through the five-step decision framework, proved not only technically compatible with AR integration but capable of generating substantial operational gains when augmented with adaptive AI—an outcome that would not have been predictable without the systematic characterization performed in Sections 3 and 4.

6. Discussion

6.1. Interpretation of Selection Results

The convergence of the multi-criteria decision matrix and the expert validation on OP10 does not reflect a circularity in the methodology but rather a genuinely distinctive operational profile that sets this operation apart from all other Level 2 candidates along the dimensions most critical to AR deployment success.

The composite score of 2.85/3.00 achieved by OP10 carries concrete industrial significance: a score at this level indicates that the operation approaches the theoretical maximum of AR suitability across all weighted criteria simultaneously—not on one or two isolated dimensions, but across the full evaluation space. The score gap between OP10 (2.85) and the second-ranked candidate ATS0 (2.00) amounts to 0.85 points on a three-point scale, compared to only 0.30 points separating ATS0 from OP37 (2.00 vs. 1.70). This non-linear differentiation at the top of the ranking confirms that OP10's superiority is genuine rather than marginal.

Three factors explain this dominance, each aligned with the weighting scheme defined in Section 3. First, OP10 concerns the Main Transformer, the most critical component of the high-speed trainset. Its failure has immediate safety implications and a replacement cost of at least 892,800 MAD (about 86,848 USD). This explains OP10's maximum score on C3, the most heavily weighted criterion (0.30), and makes safety the main driver of its overall advantage. Second, OP10 has the highest cognitive demand among all Level 2 operations. It matches the conditions under which AR is most valuable: visually guided, multi-step, and cognitively demanding tasks in which omission errors are both likely and consequential [14]. This justifies its top score on C2 (0.25). Third, OP10 is the only candidate subject to full standardized traceability under AFNOR FD X 60-000. This regulatory requirement becomes a deployment advantage because AR can automate real-time documentation while reducing compliance risk, which supports its maximum score on C4.

The expert validation results reinforce these structural findings. The Kruskal–Wallis result ($H = 18.479$, $p < 0.001$) formally confirms that OP10's dominance across the nine questionnaire items is statistically significant and cannot be attributed to distributional noise. The seniority-independence result (Mann–Whitney U, $p = 0.230$) further strengthens this interpretation by confirming that the consensus on OP10 is not driven by any particular experience group—a critical robustness property in industrial expert validation [14].

The low Kendall's W (0.037) reported in Section 4.5 warrants explicit interpretation and should not be read as evidence of expert disagreement. Kendall's W measures global rank concordance—whether all experts agree on the complete ordering of all three candi-

dates simultaneously across all nine questions. In a heterogeneous panel spanning five departments with functionally distinct operational priorities, low global concordance is both expected and well-documented: Schmidt (1997) [63] reports W values between 0.01 and 0.15 in multi-professional panels, with a median near 0.05; Rowe and Wright (1999) [64] show that panels combining operational and strategic roles structurally produce W below 0.10 regardless of directional consensus quality; Hasson and Keeney (2011) confirm that in panels exceeding 50 multi-disciplinary experts [65], $W < 0.05$ is common and does not contradict a strong directional signal; and Kalaian and Kasim (2012) report W values between 0.02 and 0.08 in multi-departmental industrial validation studies [66]. The $W = 0.037$ reported here falls squarely within this documented range and is therefore expected rather than anomalous.

The methodological distinction between global rank concordance and directional preference convergence is critical here. Kendall's W is inherently sensitive to role-specific weighting of secondary criteria—a structural property of any compositionally diverse panel. The chi-square statistic ($\chi^2 = 29.949$, $p < 0.001$) operates on a different inferential level: it confirms that the distribution of votes across OP10, ATS0, and OP37 is non-random, and that OP10's systematic plurality across all nine questions constitutes a genuine and reproducible directional signal. Low W and significant H are not contradictory—they reflect departmental heterogeneity in full rank ordering coexisting with unambiguous directional convergence on OP10 [63], a pattern entirely consistent with the panel composition documented in Section 4.5.

The methodological distinction between global rank concordance and directional preference convergence is critical here. Kendall's W measures consistency of complete rankings across all items simultaneously—a metric inherently sensitive to role-specific weighting of secondary criteria. The chi-square statistic ($\chi^2 = 29.949$, $p < 0.001$) operates on a different inferential level: it confirms that the distribution of votes across OP10, ATS0, and OP37 is non-random, and that OP10's systematic plurality across all nine questions constitutes a genuine and reproducible directional signal embedded within, rather than obscured by, the heterogeneity of the panel.

To further assess whether this directional convergence was consistent across departmental subgroups—rather than driven by any single department—the proportion of OP10 selections was examined separately for each of the five operational departments. OP10 received a plurality of votes across all five departments, confirming that the observed consensus is not an artifact of subgroup composition but reflects a consistently held preference across operationally distinct roles. Formal within-department concordance analysis using Kendall's W was not performed due to subgroup sizes ranging from $n = 13$ to $n = 25$, which fall below the threshold recommended for stable W estimation in small stratified samples [63]. However, the convergence of the global Kruskal–Wallis result, the cross-departmental plurality pattern, and the seniority-independence validation (Mann–Whitney U, $p = 0.230$) together constitute a multi-layered robustness argument that is not dependent on any single statistical indicator.

6.2. Confrontation with the Literature

The AR deployment results—a 34 to 47% reduction in intervention time and a 55 to 70% decrease in human error rates—position this study within and beyond the current performance benchmarks of the AR railway maintenance literature.

The most directly comparable result is that of Garcia et al. (2024) [9], who reported a 47% reduction in assembly time and an 88% decrease in errors using the Holorailway AR system for railway assembly operations. The intervention time reduction achieved here (34–47%) converges with Garcia et al.'s upper bound, confirming that AR-guided proce-

dural overlay delivers comparable efficiency gains across different operation types and railway contexts. The error reduction range reported here (55–70%), however, is substantially lower than Garcia et al.'s 88%. This divergence is entirely expected: assembly tasks involve discrete binary verification steps where AR guidance can virtually eliminate errors, whereas OP10's inspection and measurement tasks involve continuous variables—oil levels, temperature readings, tolerance cross-referencing—where human judgment cannot be fully displaced by AR overlay, preserving a residual error margin that is operationally irreducible under current AR capabilities.

Beyond this performance comparison, Garcia et al.'s operation was selected without systematic justification, whereas OP10 was selected through a principled five-step methodology. This distinction is not incidental: it directly conditions the generalizability of the deployment architecture. A deployment anchored in a systematically selected operation—one that satisfies explicit criteria for procedural structure, safety impact, and traceability compatibility—provides a replicable blueprint for scaling AR to other operations. A deployment anchored in a convenient or visible operation provides only an isolated data point.

Compared to the visualize–guide–decide–track pipeline of Scheffer et al. (2021) [10], the present framework contributes one structural element that pipeline presupposes but does not address: a systematic pre-deployment operation selection methodology. This directly addresses the gap identified in the literature review of Section 2—that prior AR deployments treat operation selection as pre-analytical—and establishes a rigorous entry point from which the guidance, decision, and tracking layers of any AR pipeline can be built on operationally justified foundations. This positioning is consistent with the broader trajectory identified by Rodríguez-Hernández et al. (2025) [7], who conclude that successful AR integration in railway organizations requires a far more structured and methodology-driven approach than is currently practiced.

6.3. Limitations

Three categories of limitations qualify the findings of this study.

Mono-site industrial context. The full five-step methodology was applied to a single high-speed train maintenance facility operating under Moroccan regulatory and organizational conditions. While the AFNOR FD X 60-000 standard is internationally recognized and the facility's operational characteristics are broadly representative of high-speed rail maintenance environments, the specific scores assigned to each operation in the multi-criteria matrix reflect this particular facility's maintenance culture, fleet composition, and regulatory interpretation. Replication in facilities operating under different standards—EN 13306 in European contexts [67], or JIS E in Japanese contexts—may yield different candidate operation rankings, and the methodology should be re-applied rather than the specific scores transferred.

Ordinal scoring resolution. The three-level ordinal scale (1–3) used in Step 4 was selected to minimize inter-rater variability, consistent with MCDM practice in operationally constrained contexts [53,55]. However, this scale limits discriminatory precision between operations that differ in degree rather than in kind within a single scoring level. A finer-grained scale (1–5 or 1–10) would increase resolution at the cost of higher inter-rater variance and reduced consensus reliability—a trade-off that future applications should calibrate to the specific decision context.

Single-operation AR validation scope. The AR architecture developed and validated in this study is specific to OP10 and has not yet been tested on other Level 2 operations. While the technological stack—ArborXR, Vuforia Engine, Unity, and Magic Leap 2—was selected with cross-operation scalability as an explicit criterion, the generalizability of the AR

interface design, tracking configuration, and reporting automation to other maintenance operations remains to be empirically confirmed. Future work should extend the AR deployment to the next-ranked candidates (ATS0 and OP37) to assess both the portability of the architecture and the predictive validity of the selection methodology.

6.4. Generalizability and Conditions for Replication

The five-step decision methodology is designed to be replicable across maintenance contexts beyond the specific industrial setting in which it was applied. Its generalizability rests on three inherent properties independent of the railway sector: the use of a standard MCDM framework adaptable to any criterion set, the AHP-derived weighting procedure transparent and re-applicable to different operational priorities, and the expert validation protocol scalable to any organizational size and departmental structure.

Three conditions must be satisfied for successful replication in a new industrial context. First, a formalized maintenance classification standard must be in place—equivalent to AFNOR FD X 60-000 or any recognized national or international norm—to enable the maintenance level compatibility analysis of Step 2. Second, detailed standardized maintenance plans must exist for candidate operations, as AR-guided overlay requires a documented procedural baseline from which to construct the guidance sequence. Organizations operating primarily on tacit knowledge will need to invest in procedural documentation before AR deployment becomes feasible. Third, access to a representative sample of domain experts is required for Step 5—ideally spanning multiple operational departments and experience levels, as the cross-sectional robustness demonstrated in this study (Mann–Whitney $p = 0.230$) is a function of the diversity of the validation panel.

Beyond the railway sector, the methodology is directly applicable to any asset-intensive industry where maintenance operations are regulated, procedurally structured, and safety-critical—including aerospace MRO, nuclear plant maintenance, oil and gas inspection, and advanced manufacturing. Replication across these sectors would follow the same five-step sequence validated here—beginning with a sector-specific redundancy screening and maintenance level compatibility analysis, followed by multi-criteria scoring against the target organization's operational priorities, and concluding with expert validation composed to reflect its departmental structure. The railway application reported in this study constitutes the first validated node of this broader deployment trajectory. In each of these contexts, the fundamental research gap addressed by this study—the absence of a criterion-based methodology for selecting which operation to target first for AR integration—is equally present, and the five-step framework provides a directly transferable resolution.

From a technological evolution perspective, the deterministic tracking architecture adopted in this study—grounded in Vuforia Engine's geometry-anchored spatial registration against the pre-existing 3D model of the Main Transformer—was explicitly conditioned by the availability of that model. This architectural prerequisite, while fully satisfied for OP10, will not be universally present across candidate operations identified by future applications of the five-step framework. For operations where no validated 3D model exists, the deployment architecture will necessarily evolve toward foundation vision models: SAM-based segmentation for dynamic component boundary detection and ViT-based markerless recognition for geometry-independent spatial registration. Integration of these approaches is planned for the next-ranked candidates ATS0 and OP37 in the near term (12–18 months), and their validation under safety-critical operational conditions represents the primary technological development horizon of this research program over the following 12–24 months. This trajectory positions the present study not as a terminal deployment, but as the first validated node in a scalable AR integration framework whose architectural evolution is directly driven by the operation selection methodology developed here.

7. Conclusions

This study addressed a question that the AR maintenance literature has consistently overlooked: not how to deploy AR in railway maintenance, but which operation should be targeted first, and according to what criteria. By treating operation selection as a research problem rather than a preliminary convenience, this work contributes three original elements to the field.

First, a structured five-step MCDM-based methodology that determines which maintenance operation to target before any technological deployment begins—a decision step that prior work consistently treated as pre-analytical. Second, its validated application to a real high-speed railway facility in Morocco, converging through quantitative scoring and expert consensus (n = 100) on OP10 as the optimal AR candidate. Third, a preliminary empirical validation demonstrating that this principled selection directly conditions deployment outcomes: a 34–47% reduction in intervention time, a 55–70% decrease in human error rates, and a 28–42% decline in failure-related costs—gains traceable to the operational characteristics that drove OP10's selection in the first place.

These results carry significance at three levels, each grounded in the quantitative evidence produced by this study. At the operational level, a 34 to 47% reduction in intervention time for OP10—an operation executed on a frequent weekly basis—applied across a fleet of 50 high-speed trainsets implies a recovery of approximately 170 to 235 technician-hours per year for this single operation alone, with direct consequences for fleet availability, maintenance scheduling density, and labor cost structure. A 55 to 70% reduction in human error rates on an operation directly linked to the Main Transformer—whose replacement costs a minimum of 892,800 MAD (86,848 USD)—translates into a proportional reduction in unplanned interventions, fleet immobilization events, and regulatory non-compliance risk. These figures confirm that the return on AR investment is not merely operational but financial, with a payback horizon that can be modeled against the capital cost of the Magic Leap 2 deployment architecture using the performance ranges reported here.

At the methodological level, the five-step framework validated in this study is replicable across maintenance levels, train types, regulatory environments, and industrial sectors—providing a generalizable template for AR deployment prioritization that extends beyond the railway context. The AHP-derived weighting procedure is transparent and re-applicable to different operational priority structures, and the expert validation protocol is scalable to any organizational size and departmental configuration.

At the scientific level, this work establishes operation selection as a legitimate and necessary research step in the AR deployment pipeline—one that conditions the scalability, transferability, and return on investment of any subsequent technological development. AR-automated traceability further creates a structured digital record of expert decision-making during each maintenance execution, progressively building an operational knowledge base that reduces organizational dependence on tacit individual expertise and mitigates the skill gap risk created by technician retirement.

As the third component of a multi-article research program combining AR and Reinforcement Learning for intelligent railway maintenance, this work provides the critical bridge between the strategic findings of prior contributions and their concrete industrial application.

Future work will extend this methodology along four directions. First, the five-step framework will be applied to additional Level 2 operations at the same facility, with the objective of progressively building a prioritized AR integration roadmap across the full preventive maintenance cycle—moving from a single validated use case toward a systematic deployment architecture. Second, longitudinal validation under fully operational conditions—involving live trainsets, real traceability requirements, and uncontrolled en-

vironmental variability—will be conducted to confirm the stability of the performance gains reported here and to refine the RL policy under real-world feedback dynamics. Third, organizational dimensions of deployment—including technician onboarding protocols, change management strategies, data governance infrastructure, and cost modeling across multi-operation AR programs—will be systematically addressed in subsequent contributions, completing the adoption framework initiated by the systematic review that opened this research program. Fourth, the instrumentation of high-speed trainsets under fully operational conditions will generate the annotated component datasets required to integrate transformer-based vision models such as ViT as a geometry-agnostic complement to the deterministic tracking layer validated in this study. Whether AR, combined with adaptive AI, can progressively automate not only task guidance but also the continuous optimization of maintenance decisions themselves remains a central open question for future work.

Author Contributions: Conceptualization, T.O.; Methodology, T.O.; Software, T.O.; Validation, T.O.; Formal analysis, T.O. and H.E.K.; Investigation, T.O.; Resources, T.O.; Data curation, T.O.; Writing—original draft, T.O.; Writing—review and editing, A.A. and S.M.; Visualization, A.A. and S.M.; Supervision, A.A. and S.M.; Project administration, S.M. and H.E.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Author Sedki Mohamed is employed by Office National des Chemins de Fer and Author Hicham El Kimi is employed by Société Marocaine de Maintenance des Rames à Grande Vitesse. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Park, K.J.; Jung, J.D. A study for the development of the reliability/availability management system of the urban transit vehicles. *J. Korean Soc. Railw.* **2013**, *16*, 163–168. [CrossRef]
2. Bianchi, G.; Fanelli, C.; Freddi, F.; Giuliani, F.; La Placa, A. Systematic review railway infrastructure monitoring: From classic techniques to predictive maintenance. *Adv. Mech. Eng.* **2025**, *17*, 16878132241285631. [CrossRef]
3. Tretten, P.; Illankoon, P.; Candell, O. Digitalization of Railway Maintenance: A Situation Awareness Perspective. In *Advances in Human Factors and System Interactions, AHFE 2021*; Lecture Notes in Networks and Systems; Springer: Cham, Switzerland, 2021; Volume 265, pp. 202–209.
4. Runji, J.M.; Lee, Y.-J.; Chu, C.-H. Systematic Literature Review on Augmented Reality-Based Maintenance Applications in Manufacturing Centered on Operator Needs. *Int. J. Precis. Eng. Manuf.-Green Technol.* **2022**, *10*, 1491–1510. [CrossRef]
5. del Amo, I.F.; Erkoyuncu, J.A.; Roy, R.; Palmarini, R.; Onoufriou, D. A Systematic Review of Augmented Reality Content-Related Techniques for Knowledge Transfer in Maintenance Applications. *Comput. Ind.* **2018**, *103*, 47–71. [CrossRef]
6. Di Natali, C.; Mattila, J.; Kolu, A.; De Vito, P.; Gauttier, S.; Morata, M.; Garcia, M.; Caldwell, D. Smart tools for railway inspection and maintenance work, performance and safety improvement. In Proceedings of the Transportation Research Arena TRA 2022, Lisbon, Portugal, 27–30 April 2022; pp. 1–8.
7. Rodríguez-Hernández, M.; Crespo-Márquez, A.; Sánchez-Herguedas, A.; González-Prida, V. Digitalization as an Enabler in Railway Maintenance: A Review from the International Union of Railways Asset Management Framework Perspective. *Infrastructures* **2025**, *10*, 96. [CrossRef]
8. Oumaima, T.; Abdeslam, A.; Mohamed, S.; Hicham, E.K. A systematic review of different adoption strategies for augmented reality in maintenance: Bridging theory and practice. *Discov. Appl. Sci.* **2025**, *8*, 18. [CrossRef]
9. Garcia, C.; Ortega, M.; Ivorra, E.; Contero, M.; Mora, P.; Alcañiz, M.L. Holorailway: An Augmented Reality System to Support Assembly Operations in the Railway Industry. *Adv. Manuf.* **2024**, *12*, 764–783. [CrossRef]

10. Scheffer, S.; Martinetti, A.; Damgrave, R.; Thiede, S.; van Dongen, L. How to Make Augmented Reality a Tool for Railway Maintenance Operations: Operator 4.0 Perspective. *Appl. Sci.* **2021**, *11*, 2656. [CrossRef]
11. Kim, K.-S.; Kim, C.-S. Development and Usability Evaluation of Augmented Reality Content for Light Maintenance Training of Air Spring for Electric Multiple Unit. *Appl. Sci.* **2024**, *14*, 7702. [CrossRef]
12. Kwon, H.-J.; Lee, S.-I.; Park, J.-H.; Kim, C.-S. Design of Augmented Reality Training Content for Railway Vehicle Maintenance Focusing on the Axle-Mounted Disc Brake System. *Appl. Sci.* **2021**, *11*, 9090. [CrossRef]
13. Mikušová, N.; Neradilová, H.; Hlatká, M.; Fedorko, G.; Molnár, V.; Kráľ, J. Use of Augmented Reality in Railway Transport. *Transp. Res. Procedia* **2024**, *77*, 253–259. [CrossRef]
14. Bock, L.; Bohné, T.; Tadeja, S.K. Decision Support for Augmented Reality-Based Assistance Systems Deployment in Industrial Settings. *Multimed. Tools Appl.* **2024**, *84*, 23617–23641. [CrossRef] [PubMed]
15. Atkinson, R. Project management: Cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria. *Int. J. Proj. Manag.* **1999**, *17*, 337–342. [CrossRef]
16. Elia, V.; Gnoni, M.G.; Lanzilotto, A. Evaluating the application of augmented reality devices in manufacturing from a process point of view: An AHP based model. *Expert Syst. Appl.* **2016**, *63*, 187–197. [CrossRef]
17. *AFNOR FD X 60-000*; Maintenance—La Fonction Maintenance. AFNOR: Saint-Denis, France, 2002.
18. Tayyarat, O.; Ahmadi, A.; Sedki, M.; El Kimi, H. Enhancing High-Speed Train Maintenance through Augmented Reality and Deep Reinforcement Learning: A Hybrid Optimization Approach. In *Proceedings of the 2025 International Conference on Circuit, Systems and Communication (ICCSC), Fez, Morocco, 19–20 June 2025*; IEEE: Piscataway, NJ, USA, 2025. [CrossRef]
19. Bonfiglioli, L.; Buda, M.; Caprino, G.; Ferretti, G.; Ficarra, V.; Grana, C. Vision Transformer in Industrial Visual Inspection. *Appl. Sci.* **2022**, *12*, 11981. [CrossRef]
20. Yang, Q.; Guo, R. An Unsupervised Method for Industrial Image Anomaly Detection with Vision Transformer-Based Autoencoder. *Sensors* **2024**, *24*, 2440. [CrossRef] [PubMed]
21. Alatawi, H.; Albalawi, N.; Shahata, G.; Aljohani, K.; Alhakamy, A.; Tuceryan, M. Augmented Reality-Assisted Deep Reinforcement Learning-Based Model towards Industrial Training and Maintenance. *Sensors* **2023**, *23*, 6024. [CrossRef] [PubMed]
22. Sulistiyono, M.; Hasyim, J.W.; Bernadhed, B.; Liantoni, F.; Sidauruk, A. Comparative study of marker-based and markerless tracking in augmented reality under variable environmental conditions. *J. Soft Comput. Explor.* **2024**, *5*, 413–422. [CrossRef]
23. PTC Inc. Vuforia Engine—Model Targets from CAD. Vuforia Developer Library. 2024. Available online: https://developer.vuforia.com/library/vuforia-engine/images-and-objects/model-targets/model-targets/ (accessed on 2 April 2026).
24. Kour, R.; Castaño, M.; Karim, R.; Patwardhan, A.; Kumar, M.; Granström, R. A Human-Centric Model for Sustainable Asset Management in Railway: A Case Study. *Sustainability* **2022**, *14*, 936. [CrossRef]
25. Davis, F.D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **1989**, *13*, 319–340. [CrossRef] [PubMed]
26. Bahaei, S.; Gallina, B.; Vidović, M. A case study for risk assessment in AR-equipped sociotechnical systems. *J. Syst. Archit.* **2021**, *119*, 102250. [CrossRef]
27. Saaty, T.L. *The Analytic Hierarchy Process*; McGraw-Hill: New York, NY, USA, 1980.
28. Liberatore, M.J.; Nydick, R.L. The analytic hierarchy process in medical and health care decision making: A literature review. *Eur. J. Oper. Res.* **2008**, *189*, 194–207. [CrossRef]
29. Kagermann, H.; Wahlster, W.; Helbig, J. *Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0: Securing the Future of German Manufacturing Industry*; Acatech–National Academy of Science and Engineering: Frankfurt, Germany, 2013.
30. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef] [PubMed]
31. Cha, J.H.; Kim, K.S.; Kim, C.S.; Kwon, H.J. Development of Educational Content Based on Augmented Reality for Light Maintenance of Railway Vehicle. *Int. J. Future Comput. Commun.* **2023**, *12*, 54–57. [CrossRef]
32. Kwon, H.-J.; Song, J.-H.; Kim, K.-S.; Kim, C.-S. Methodology for Digitalizing Railway Vehicle Maintenance Training Using Augmented Reality. *Informatics* **2025**, *12*, 101. [CrossRef]
33. Kang, G.H.; Kwon, H.J.; Chung, I.S.; Kim, C.S. A Study on the Development of Augmented Reality Contents for Air Compressor of Railway Vehicles. In Proceedings of the 2023 Prognostics and Health Management Conference (PHM), Paris, France, 31 May–2 June 2023; pp. 59–63. [CrossRef]
34. Kang, G.H.; Kwon, H.J.; Chung, I.S.; Kim, C.S. Development and Usability Evaluation of Mobile Augmented Reality Contents for Railway Vehicle Maintenance Training: Air Compressor Case. *Adv. Sci. Technol. Eng. Syst. J.* **2024**, *9*, 91–103. [CrossRef]
35. Mourtzis, D.; Siatras, V.; Angelopoulos, J. Real-Time Remote Maintenance Support Based on Augmented Reality (AR). *Appl. Sci.* **2020**, *10*, 1855. [CrossRef]
36. Scheffer, S.S.; Martinetti, A.A.; Damgrave, R.R.; van Dongen, L.L. Supporting Maintenance Operators Using Augmented Reality Decision-Making: Visualize, Guide, Decide & Track. *Procedia CIRP* **2023**, *119*, 782–787. [CrossRef]

37. Wang, J.; Qi, Y. A Multi-User Collaborative AR System for Industrial Applications. *Sensors* **2022**, *22*, 1319. [CrossRef] [PubMed]
38. Jing, C.; Cheng, X.; Wang, X.; Yu, D.; Yao, X. Effects of AR-HMD Interactive Interface Layout on Usability and Obstacle Avoidance in Subway Overhauls. *Int. J. Hum.-Comput. Interact.* **2025**, *41*, 5817–5830. [CrossRef]
39. Malta, A.; Farinha, T.; Mendes, M. Augmented Reality in Maintenance—History and Perspectives. *J. Imaging* **2023**, *9*, 142. [CrossRef] [PubMed]
40. Porsch, R.; Lehmann, M.A. Evaluation of Marker-Based AR-Tracking with Vuforia in the Context of Rail Vehicle Maintenance. In *Lecture Notes in Networks and Systems*; Springer Nature Switzerland: Cham, Switzerland, 2023; pp. 65–71.
41. Bernal, E.; Wu, Q.; Spiryagin, M.; Cole, C. Augmented Digital Twin for Railway Systems. *Veh. Syst. Dyn.* **2024**, *62*, 67–83. [CrossRef]
42. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment Anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2–3 October 2023. [CrossRef]
43. Guillén-Mena, V.; Quesada-Molina, F.; Astudillo-Cordero, S.; Lema, M.; Ortiz-Fernández, J. Lessons learned from a study based on the AHP method for the assessment of sustainability in neighborhoods. *MethodsX* **2023**, *11*, 102440. [CrossRef] [PubMed]
44. Nyström, B.; Söderholm, P. Selection of maintenance actions using the analytic hierarchy process (AHP): Decision-making in railway infrastructure. *Struct. Infrastruct. Eng.* **2010**, *6*, 467–479. [CrossRef]
45. Prommachan, W.; Srisawat, S.; Tongsiri, C. Selection Criteria for Evaluating Predictive Maintenance Techniques for Rotating Machinery using the Analytic Hierarchical Process (AHP). *Eng. Technol. Appl. Sci. Res.* **2024**, *14*, 12851–12856. [CrossRef]
46. Daniyan, I.; Mpofu, K.; Adeodu, A.; Tlhabadira, I. The use of Analytical Hierarchy Process (AHP) decision model for materials and assembly method selection during railcar development. *Cogent. Eng.* **2020**, *7*, 1833433. [CrossRef]
47. Wedley, W.C. Consistency prediction for incomplete AHP matrices. *Math. Comput. Model.* **1993**, *17*, 151–161. [CrossRef]
48. Shamsuzzoha, A.; Piya, S.; Shamsuzzaman, M. Application of fuzzy TOPSIS framework for selecting complex project in a case company. *J. Glob. Oper. Strateg. Sourc.* **2021**, *14*, 528–566. [CrossRef]
49. Vavrek, R.; Becica, J. Similarity of TOPSIS results based on criterion variability: Case study on public economic. *PLoS ONE* **2022**, *17*, e0271951. [CrossRef] [PubMed]
50. Arjomandi, M.A.; Dinmohammadi, F.; Mosallanezhad, B.; Shafiee, M. A fuzzy DEMATEL-ANP-VIKOR analytical model for maintenance strategy selection of safety critical assets. *Adv. Mech. Eng.* **2021**, *13*, 1687814021994965. [CrossRef]
51. Opricovic, S.; Tzeng, G.H. Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. *Eur. J. Oper. Res.* **2004**, *156*, 445–455. [CrossRef]
52. Guitouni, A.; Martel, J.M. Tentative guidelines to help choosing an appropriate MCDM method. *Eur. J. Oper. Res.* **1998**, *109*, 501–521. [CrossRef]
53. Taherdoost, H.; Madanchian, M. Multi-Criteria Decision Making (MCDM) Methods and Concepts. *Encyclopedia* **2023**, *3*, 77–87. [CrossRef]
54. Emrouznejad, A.; Marra, M. The state of the art development of AHP (1979–2017): A literature review with a social network analysis. *Int. J. Prod. Res.* **2017**, *55*, 6653–6675. [CrossRef]
55. Ding, S.H.; Kamaruddin, S. Maintenance policy optimization—Literature review and directions. *Int. J. Adv. Manuf. Technol.* **2015**, *76*, 1263–1283. [CrossRef]
56. Sasidharan, M.; Burrow, M.P.N.; Ghataora, G.S.; Marathu, R. A Risk-Informed Decision Support Tool for the Strategic Asset Management of Railway Track Infrastructure. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2022**, *236*, 412–425. [CrossRef]
57. Retzlaff, C.O.; Das, S.; Wayllace, C.; Mousavi, P.; Afshari, M.; Yang, T.; Saranti, A.; Angerschmid, A.; Taylor, M.E.; Holzinger, A. Human-in-the-Loop Reinforcement Learning: A Survey and Position on Requirements, Challenges, and Opportunities. *J. Artif. Intell. Res.* **2024**, *79*, 359–415. [CrossRef]
58. Xu, Z.; Zhang, Q. Predictive maintenance optimization for industrial equipment via reliable prognosis and risk-aware reinforcement learning. *Complex Intell. Syst.* **2025**, *12*, 11. [CrossRef] [PubMed]
59. Tague, N.R. *The Quality Toolbox*, 2nd ed.; ASQ Quality Press: Milwaukee, WI, USA, 2005.
60. Jeong, H.; Choi, K.-S.; Son, Y.; Yang, Y. *ADA: Augmented-Reality Development Adaptor for Supporting AR Glass*; IEEE: Piscataway, NJ, USA, 2023; pp. 1513–1516.
61. Magic Leap. Field of View (FOV)—Magic Leap 2. Magic Leap Developer Documentation. 2023. Available online: https://developer-docs.magicleap.cloud/docs/device/hardware/fov/ (accessed on 2 April 2026).
62. Microsoft. HoloLens 2 Hardware. Microsoft Learn Documentation. Available online: https://learn.microsoft.com/en-us/hololens/hololens2-hardware (accessed on 2 April 2026).
63. Schmidt, R.C. Managing Delphi Surveys Using Nonparametric Statistical Techniques. *Decis. Sci.* **1997**, *28*, 763–774. [CrossRef]
64. Rowe, G.; Wright, G. The Delphi technique as a forecasting tool: Issues and analysis. *Int. J. Forecast.* **1999**, *15*, 353–375. [CrossRef]
65. Hasson, F.; Keeney, S. Enhancing rigour in the Delphi technique research. *Technol. Forecast. Soc. Change* **2011**, *78*, 1695–1704. [CrossRef]

66. Kalaian, S.A.; Kasim, R.M. Terminating Sequential Delphi Survey Data Collection. *Pract. Assess. Res. Eval.* **2012**, *17*, 5. [CrossRef]
67. *EN 13306:2017*; Maintenance—Maintenance Terminology. European Committee for Standardization: Brussels, Belgium, 2017.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.