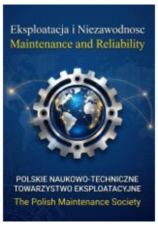

Eksploatacja i Niezawodnosc - Maintenance and Reliability

Volume 29 (2027), Issue 1

https://ein.org.pl

The Polish Maintenance Society

Human factors in aviation maintenance: task design and safety implications

Idriss Dagal[1, *], Bilal Erol[2,3]

[1] Electrical Engineering, Istanbul Beykent University, Turkey
[2] Control and Automation Engineering, Yildiz Technical University, Turkey
[3] Information Technology Research and Application Center, Istanbul Ticaret University, Istanbul, Türkiye
* Corresponding Author: idriss.dagal@std.yildiz.edu.tr

Received: 30 March 2026
Revised: 19 May 2026
Accepted: 26 June 2026
Online: 23 July 2026

Abstract

Aviation maintenance remains a safety-critical domain where human-factor-related errors persist despite technological advancements and regulatory oversight. Traditional human factors approaches are often retrospective, limiting predictive insight when task design, cognitive workload, fatigue, and procedural ambiguity exceed human capacity. This study addresses that gap by developing and validating a Human Performance Integration (HPI) Model, a predictive, engineering-oriented framework that quantitatively links maintenance task design characteristics with cognitive constraints and error probability. A mixed-methods research design integrates qualitative insights from semi-structured interviews with licensed maintenance technicians (n = 21) and quantitative analysis of 47 incident reports from a regional airline. Qualitative findings inform the identification of latent performance drivers, which are operationalized into measurable variables and structured using Principal Component Analysis (PCA). A Support Vector Machine (SVM) classifier is then applied to model non-linear interactions among cognitive workload, fatigue accumulation, task complexity, and documentation clarity. From an engineering perspective, the HPI model functions as a proactive task evaluation and design-support tool moving beyond post-hoc safety analysis. It enables designers and maintenance engineers to assess how alternative task structures, documentation strategies, and workload allocations influence human performance margins during the planning stage. By embedding human performance constraints directly into task design evaluation, this research advances engineering design knowledge for safety-critical socio-technical systems and contributes a predictive lens to human factors in aviation maintenance.

Keywords: *human factors engineering, aviation maintenance, cognitive workload, predictive modeling, support vector machine.*

Article citation:
Dagal I, Erol B, Human factors in aviation maintenance: task design and safety implications, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2027: 29(1) http://doi.org/10.17531/ein/225076

Highlights

- 68% of maintenance errors traced to cognitive overload.
- 23% of errors linked to unclear documentation.
- Predictive HPI model integrates task design with human performance limits.
- PCA and SVM applied to model latent cognitive and procedural factors.
- Framework supports proactive task evaluation and safety management.

1. Introduction

Aviation maintenance is a safety-critical domain where even minor human errors can have significant consequences [1]. As aircraft systems grow increasingly complex, maintenance tasks impose greater physical and cognitive demands on technicians [2]. Human factors such as fatigue, time pressure, and procedural complexity have emerged as key determinants of safety and performance [3,4]. Importantly, maintenance-related errors often arise not from negligence but from systemic issues, including poorly designed tasks, inadequate documentation, and supervisory practices that lack human-centered engineering principles [5]. This study investigates the intersection of engineering and human factors within task structuring, documentation, and assignment processes in aviation maintenance. Drawing on incident reports, technician interviews, and a 12 month dataset from a regional airline comprising 47 maintenance events, we identify latent conditions that predispose tasks to error. Consistent with recent literature, our mixed-methods analysis highlights the influence of cognitive overload in multi-system maintenance [6], procedural ambiguity due to unclear documentation [7], and the compounding effects of fatigue under organizational pressures [8]. These findings align with emerging work in resilience engineering [9] and digital task assistance [10], which emphasize adaptive systems to support human performance. Based on these insights, we propose a Human Performance Integration (HPI) model, a task optimization framework that integrates fatigue risk control, cognitive workload assessment, and checklist standardization to improve both safety and operational efficiency. This model also informs the development of training programs aligned with human capabilities and system boundaries. Despite decades of research on human factors in aviation maintenance, existing approaches remain predominantly retrospective, descriptive, or fragmented, focusing on isolated contributors such as fatigue, workload, or documentation errors after incidents occur. What is notably absent is a predictive, engineering-oriented framework that embeds human performance limitations directly into maintenance task design and decision-making. In practice, maintenance engineers and managers lack tools to anticipate when task complexity, cognitive workload, fatigue accumulation, and procedural ambiguity collectively exceed human performance limits, thereby increasing error probability. This unresolved gap limits the ability of organizations to proactively manage workload, redesign tasks, and allocate resources before safety is compromised.

This study situates itself within the broader domain of aviation maintenance safety, focusing specifically on how engineering task design influences human performance and error rates among frontline technicians in regional airlines, where limited resources and diverse task demands exacerbate risk [11–15]. Encompassing maintenance activities across electrical, mechanical, hydraulic, and avionic subsystems, it addresses both routine and complex, time-sensitive tasks performed in fragmented work periods to reflect the full operational spectrum [14–18]. Core concerns include task complexity, workload allocation, procedural clarity, cognitive load, and technician fatigue, analyzed through a mixed-methods approach combining quantitative error data from 47 recorded maintenance events with qualitative insights from technician interviews [19–24]. Motivated by the persistent impact of human factors on maintenance reliability despite technological advances [25–27], the study demonstrates how high mental workload, time pressure, and documentation mismatches contribute to cognitive overload, misinterpretation, and procedural non-compliance, elevating error probability [28,29]. To address this unresolved gap, we introduce the HPI Model, a novel engineering-oriented framework that quantitatively links task design characteristics with cognitive workload, fatigue accumulation, and procedural clarity. By synthesizing technician insights with data-driven modeling using Principal Component Analysis (PCA) and Support Vector Machine (SVM) classification, the HPI model enables predictive, task-level assessment of maintenance error risk.

This integrative approach moves beyond traditional human factors analysis by providing a scalable tool for proactive workload management, documentation optimization, and task allocation, thereby advancing both theoretical understanding and practical safety management in aviation maintenance operations [30–35]. The urgency of these solutions is underscored by industry data linking maintenance errors to $2.1 billion in annual costs and 17% of global aviation incidents [36], alongside challenges such as the “Cognitive Paradox” of overwhelming information complexity in modern aircraft [37], reliance on informal knowledge due to inadequate manuals [38], and fatigue-related performance declines during extended shifts [39]. As reported in Figure 1, addressing these critical issues aligns with FAA directives for data driven human factors integration [40,41], contributing valuable operational risk insights while advancing human centered safety paradigms in aviation maintenance [42].

Figure 1. Aviation safety and maintenance skills needed for human factors.

1.1. Research roadmap

To address the identified gap, this study adopts a structured, integrative research roadmap. First, qualitative evidence from technician interviews and maintenance incident narratives is used to identify latent cognitive, procedural, and organizational stressors influencing task execution. Second, these human-factor constructs are operationalized into measurable variables and structured using Principal Component Analysis (PCA) to capture dominant performance dimensions. Technician interviews were anonymized and systematically coded; the resulting variables were used to guide PCA dimensionality reduction. Third, a predictive Human Performance Integration (HPI) Model is developed using Support Vector Machine (SVM) classification to estimate task-level maintenance error risk under realistic operational constraints. Finally, the validated model is embedded within a task optimization and decision-support framework, enabling proactive workload management, documentation redesign, and risk-informed task allocation. This roadmap ensures methodological coherence and aligns all analytical components toward a single predictive objective. From an engineering design standpoint, aviation maintenance tasks are interpreted as engineered work systems composed of design variables (task sequencing, information density, procedural branching, and workload allocation) and performance outcomes (error probability, recovery effort, and task completion reliability). Despite extensive human factors research, existing approaches rarely formalize this relationship using design-oriented constructs or predictive evaluation methods. Consequently, human performance limitations are often treated as external constraints rather than endogenous design considerations. This study explicitly reframes maintenance human factors as a design evaluation problem, in which task structures are assessed against human performance envelopes during the design phase rather than after operational failure. Qualitative insights from technician interviews directly informed the selection of variables for PCA, ensuring practitioner knowledge was embedded into quantitative modeling.

1.2. Design contributions

This study makes four explicit contributions to engineering design research:

- Aviation maintenance tasks are framed as engineered artifacts whose structural design variables, such as task complexity, procedural clarity, workload distribution, and sequencing, can be formally evaluated against human performance constraints.
- The Human Performance Integration (HPI) Model is introduced as a design-evaluation tool that maps task design parameters to performance outcomes, enabling predictive assessment during task planning rather than retrospective error classification.
- The study demonstrates how qualitative practitioner knowledge can be systematically translated into quantitative design variables, supporting traceable integration between experiential insights and analytical modeling.
- Evidence is provided showing how incremental changes in task design parameters (documentation clarity and task complexity) non-linearly influence performance margins, offering actionable guidance for design trade-off analysis in safety-critical systems.

Collectively, these contributions reposition human factors from a diagnostic safety discipline to a core component of engineering task design methodology.

1.3. Paper layout

The remainder of this paper is organized as follows: Section 2 presents the background and related works, reviewing prior research on human factors in aviation maintenance with emphasis on cognitive workload, task complexity, and human-machine interaction dynamics. Section 3 defines the aims and research questions, outlining the study's objectives and presenting the guiding questions that connect theoretical insights with practical challenges. Section 4 describes the methodology and research taxonomy, detailing data collection procedures, task classification criteria, cognitive load assessment methods, and both quantitative and qualitative analytical techniques, while also explaining the PRISMA-based review process, data validation procedures, and the development of the task optimization model. Section 5 presents the results and discussion, including statistical correlations, cognitive load profiles, and predictive modeling outcomes derived from the SVM and workload integration framework, with findings interpreted within the proposed Task Optimization Framework and their implications for maintenance safety, performance reliability, and human-centered design. Finally, Section 6 concludes the paper by summarizing the key insights, highlighting the study's novel contributions and practical recommendations for engineering task design, training, and fatigue management, and outlining future research directions focused on integrating data-driven human factors modeling and adaptive support tools.

2. Background and related works

Human factors are a critical determinant of safety, reliability, and efficiency in aviation maintenance operations. Despite the widespread adoption of digital documentation systems and Safety Management Systems (SMS), maintenance error rates remain significant, particularly in complex or time-constrained environments. Maintenance work requires intricate coordination among engineers, supervisors, and technicians, all operating under varying conditions of workload, stress, and fatigue. According to the International Civil Aviation Organization (ICAO), over 70% of maintenance-related incidents are attributable to human performance limitations rather than technical malfunctions, underscoring the need for structured and predictive assessment of human factors in engineering task design.

The conceptual foundation of human error in aviation maintenance is strongly influenced by Reason's "Swiss Cheese Model," which frames accidents as the alignment of active errors and latent organizational failures [16–18]. Building on this, the Human Factors Analysis and Classification System for Maintenance Environments (HFACS-ME) has provided a systematic means of categorizing maintenance errors across cognitive, procedural, and organizational levels [24]. While these frameworks have guided error taxonomy development and safety auditing practices, they remain largely retrospective, focusing on post-event causal analysis rather than proactive error prediction or task-level risk mitigation.

Subsequent research has expanded the understanding of technician behavior and human reliability. Performance degradation during night shifts, under fatigue, or when handling multiple concurrent tasks has been well-documented [23]. Empirical evidence also demonstrates that high workload, time pressure, and poor procedural design substantially increase error probability. Regulatory authorities such as the FAA and EASA have issued guidelines on fatigue management and human performance assessment, yet implementation remains inconsistent, particularly in resource-limited regional maintenance facilities where outdated task documentation persists.

Contemporary research has introduced data-driven approaches for modeling human reliability. Machine learning (ML) techniques, including logistic regression, random forests, and support vector machines (SVMs), have been applied to forecast human error based on operational parameters, workload scores, and technician experience levels [25]. Similarly, cognitive load theory has been used to explain how ambiguous or verbose documentation elevates mental effort and decision latency [26]. Despite these advances, most investigations examine individual factors in isolation, failing to capture the multidimensional interactions among workload, fatigue, documentation clarity, and task design that characterize real-world maintenance operations.

A systematic synthesis of the literature reveals three interrelated research gaps that motivate the present study:

- Existing frameworks are predominantly retrospective, limiting proactive error prediction.
- Human factors are often analyzed in isolation, overlooking multidimensional interactions.
- Engineering task design is rarely formalized as a predictive construct that embeds human performance constraints.

2.1. Gap in design-oriented human performance evaluation

While prior studies have examined fatigue, workload, and documentation quality in isolation, they rarely treat maintenance tasks as design entities subject to formal evaluation, iteration, and optimization. From an engineering design perspective, this represents a critical gap, the absence of predictive methods that link task design variables to human performance limits before task deployment. Existing human-factors models are predominantly descriptive or retrospective, restricting their utility in design-stage decision-making. This study addresses that limitation by embedding human performance constraints within a task-evaluation framework aligned with engineering design principles, enabling proactive assessment of task alternatives and systematic evaluation of design trade-offs.

The literature highlights several persistent gaps in aviation maintenance human-factors research:

- Human factors engineering and maintenance task design are rarely integrated mathematically, with limited characterization of how procedural attributes, such as information density, sequencing, and branching logic, affect cognitive workload and error probability.
- Many predictive error models exclude contextual variables, including cumulative fatigue, shift timing, and documentation clarity, despite strong evidence that these factors significantly influence real-world performance limits.
- Few analytical approaches integrate qualitative technician insights with scalable quantitative models of workload and error likelihood.

Although advances in workload measurement (NASA-TLX), fatigue modeling, and documentation readability metrics demonstrate strong predictive potential, these elements remain fragmented and are rarely incorporated into operational decision-support systems. The Human Performance Envelope (HPE) offers a unifying conceptual framework by interpreting human error as an emergent outcome when task demands, cognitive workload, fatigue, and environmental conditions exceed human adaptive capacity. Yet, the field still lacks a validated mixed-methods framework that operationalizes this concept for maintenance task design. This study directly addresses that need by extending the HPE through a hybrid analytical framework that integrates technician insights with quantitative modeling using PCA, SVMs, and error-probability estimation, advancing predictive human-performance assessment and proactive safety management in aviation maintenance. The Human Performance Envelope (HPE) defines boundary conditions where cumulative task demands exceed adaptive capacity. The HPI Model operationalizes this envelope by embedding those constraints into predictive evaluation tools for task design.

2.2. Systematic literature review approach

To establish a rigorous theoretical foundation, this study conducted a Systematic Integrative Literature Review (SILR) in accordance with PRISMA 2020 guidelines [30]. Rather than exhaustively cataloging prior work, the review aimed to identify dominant technical themes and research gaps relevant to predictive human-performance modeling in aviation maintenance. Searches across Scopus, Web of Science, and IEEE Xplore using targeted human-factors keywords yielded 112 records, of which 42 studies were retained following duplicate removal, screening, and full-text evaluation based on methodological rigor and relevance. The synthesis, summarized in the PRISMA flow diagram (Figure 1), revealed three core thematic domains cognitive workload and TLX-based metrics [27], fatigue and circadian effects on technician reliability [28], and documentation and procedural clarity [29] that directly informed the interview protocol and quantitative variable framework, ensuring conceptual coherence between the literature review, empirical data collection, and subsequent predictive modeling.

2.3. Literature review findings and thematic synthesis

To establish a rigorous theoretical foundation, this study conducted a Systematic Integrative Literature Review (SILR) in accordance with PRISMA 2020 guidelines [30]. Rather than exhaustively cataloging prior work, the review aimed to identify dominant technical themes and research gaps relevant to predictive human-performance modeling in aviation maintenance. Literature searches were conducted across Scopus, Web of Science, and IEEE Xplore using targeted human-factors and maintenance-reliability keywords, yielding an initial dataset of 112 records. After duplicate removal, title and abstract screening, and full-text eligibility evaluation based on methodological rigor and thematic relevance, 42 studies were retained for detailed synthesis.

The review process, reported in Figure 2 (PRISMA flow diagram), revealed three core thematic domains:

- Cognitive workload and TLX-based assessment metrics [27].
- Fatigue and circadian influences on technician reliability [28].
- Documentation clarity and procedural complexity [29].

These domains directly informed the development of the interview protocol and the quantitative variable framework used in the empirical phase of the study, ensuring conceptual alignment between the literature review, qualitative data collection, and predictive modeling.

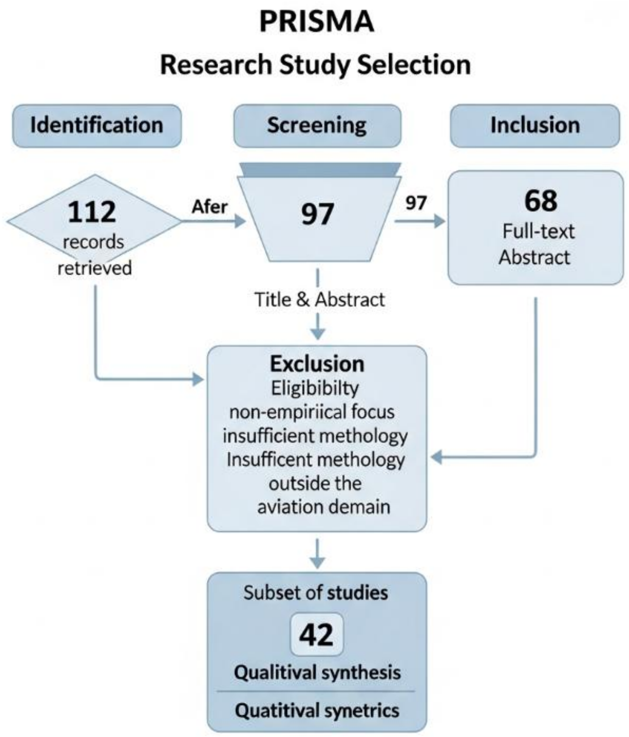

Figure 2. PRISMA 2020 flow diagram showing study selection across four stages: Identification (n = 112), Screening (n = 97), Eligibility (n = 68), and Inclusion (n = 42).

3. Aims and research questions

This study aims to develop and validate a coherent mixed-methods analytical framework that predicts how engineering task design influences human performance and maintenance error risk in aviation operations. Grounded in the Human Performance Envelope concept, the research moves beyond descriptive human-factors analysis by constructing a predictive Human Performance Integration (HPI) model that quantitatively captures the interaction effects among cognitive workload, fatigue, task complexity, and procedural clarity. Rather than treating maintenance errors as isolated outcomes, the framework conceptualizes error as an emergent condition arising when cumulative task demands exceed human performance boundaries, thereby emphasizing the need for integrative modeling. Accordingly, the research objectives prioritize predictive capability, methodological integration, and operational applicability to support proactive safety management and decision-making in aviation maintenance environments.

The study is guided by the following research questions:

- How do cognitive workload, fatigue accumulation, and procedural clarity interact within the human performance envelope to influence aviation maintenance error probability, rather than acting as independent factors?
- Can human-factor metrics derived from technician experience and workload assessment be integrated into a statistically robust, data-driven Human Performance Integration (HPI) model capable of predicting error-prone maintenance tasks?
- How can the validated HPI model be operationalized to support proactive task design, workload management, and maintenance decision-making in safety-critical aviation environments?

3.1. Human Performance Integration (HPI) model

This study proposes a Human Performance Integration (HPI) Model to systematically capture how cognitive, task-related, and operational factors interact to influence maintenance performance and safety outcomes. The model integrates key human factors variables such as cognitive workload, fatigue, task complexity, and documentation clarity within a unified analytical framework. By distinguishing between an analytically generalizable core structure and context-specific parameter calibration, the HPI-Model is designed to support transferability across different aviation maintenance environments while remaining grounded in real operational data. Figure 3 illustrates the hierarchical relationship between the core architecture of the Human Performance Integration (HPI) Model and its application across multiple operational contexts. The Core HPI-Model Structure (solid boundary) comprises the fundamental cognitive and task-related dynamics, such as the Cognitive Load Index (CLI), fatigue accumulation, and task complexity, that are analytically generalizable across aviation maintenance environments. Surrounding this core, the Contextual Parameter Layer (dashed boundary) represents the adaptable interface through which organization-specific factors, including shift scheduling policies and team experience, are incorporated. At the outermost level, Operational Data Sources provide the empirical basis for statistical parameter calibration, enabling the model to be transferred from single-site empirical studies to broader, multi-site validation settings. Task Load Balance (TLB) and Predicted Value Estimate (PVE) are explained in detail when first applied in the methodology section. Incident reports were coded into categories of task complexity, documentation clarity, and workload distribution. Coding reliability was ensured through inter-coder agreement checks. Technician interviews were anonymized, and thematic coding was validated by independent reviewers to ensure reliability.

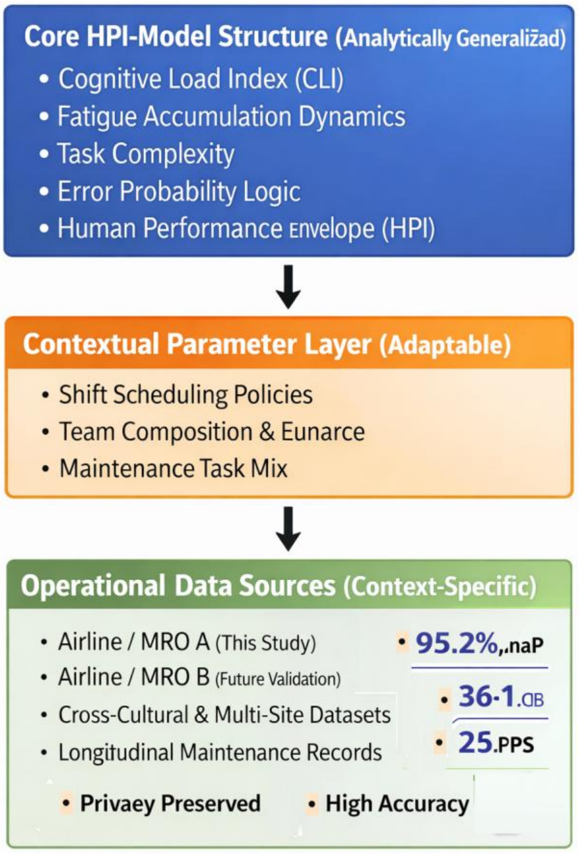

Figure 3. Levels of generalizability in the Human Performance Integration (HPI) model.

3.2. Measures and data sources

Cognitive workload and fatigue were primarily assessed using the NASA Task Load Index (NASA-TLX), a validated and widely adopted subjective workload assessment tool in aviation maintenance and other safety-critical domains. While subjective measures inherently capture perceived workload, they provide valuable insight into operator experience under real operational constraints where continuous physiological monitoring is often infeasible. To mitigate reliance on self-reported measures alone, NASA-TLX scores were triangulated with objective operational performance indicators**,** including recorded task duration, frequency of task interruptions, rework incidence, and documented error recovery actions. These indicators serve as observable behavioral proxies for cognitive strain and fatigue, allowing cross-verification between perceived and operational workload states. Importantly, the proposed Human Performance Integration (HPI) Model is sensor-agnostic by design**,** enabling the incorporation of objective physiological measures such as electroencephalography (EEG), electrodermal activity (EDA), photoplethysmography (PPG), and eye-tracking metrics without modification to the core model structure. These modalities are therefore positioned as future extensions rather than prerequisites, ensuring methodological robustness while acknowledging current operational constraints. Figure 4 illustrates the triangulation of multiple data streams within the central, sensor-agnostic Human Performance Integration (HPI) Model**.** Subjective measures (left) employ validated instruments such as the NASA-TLX to capture perceived mental and temporal workload demands. These inputs are complemented by objective operational indicators (right), including task duration and rework frequency, which provide quantitative evidence of observed human performance. Together, these data sources support the estimation of cognitive workload and fatigue within the core analytical framework. The physiological and neurocognitive extensions (bottom) demonstrate the scalability of the model, indicating how future measures, such as EEG or eye-tracking, can be incorporated to enhance workload and fatigue estimation without altering the underlying HPI-Model structure.

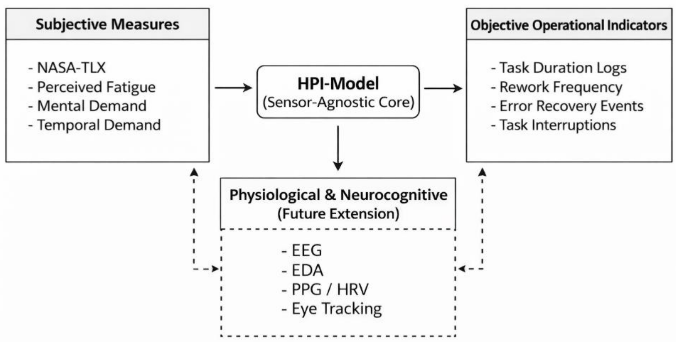

Figure 4. Multi-modal workload and fatigue assessment within the HPI framework.

3.3. Engineering task design factors influencing human performance in aviation maintenance

This research evaluates the influence of engineering task design on human performance and error rates in aviation maintenance, focusing particularly on regional airline technicians who face resource constraints and diverse task demands that increase susceptibility to error [16,17]. By examining maintenance activities across electrical, mechanical, hydraulic, and avionic subsystems, including both routine servicing and urgent reactive tasks, the study captures the comprehensive nature of technician workflows [18,19]. Five critical dimensions guide the analysis: task complexity [20], workload distribution [21], procedural clarity [22], cognitive demand [23], and technician fatigue [24]. Employing a mixed-methods approach on 47 maintenance events, combining quantitative error reports with qualitative technician interviews, the research identifies persistent barriers undermining safety [25, 30]. Despite technological advances improving mechanical reliability, human error remains the leading cause of maintenance-related incidents, accounting for 28% of aviation accidents between 2018 and 2023 [31, 32]. Maintenance environments are characterized by high cognitive load [33], intense time pressure [34], and mismatches between documentation and workflow [35], leading to instruction misinterpretation (42% of errors) [36], procedural non-compliance [37], and cognitive overload [38]. To address these challenges, the study proposes human-centered interventions including task allocation optimized through cognitive workload modeling [39], ergonomic redesign of documentation [40], training enhancements addressing real-world cognitive demands [41], and decision-support tools integrating fatigue metrics [42]. Industry data highlight the urgency of such measures, with maintenance errors costing $2.1 billion annually and contributing to 17% of global aviation incidents [43, 44]. Three systemic issues are particularly critical: the Cognitive Paradox, where complex aircraft systems like the Boeing 787 exceed human information-processing capabilities [45]; documentation non-use and reliance on informal workarounds in maintenance tasks [46]; fatigue-related error risk associated with sleep restriction and unstable work patterns [47]; and the need for effective SMS implementation and human-factors integration in aircraft maintenance organizations [48]. This study contributes to aviation maintenance risk management and latent human-error mitigation [49,50].

4. Methodology and research taxonomy

Methodological choices in this study were governed by engineering design objectives rather than algorithmic complexity. The primary goal was to develop a transparent, interpretable model capable of evaluating task design alternatives under realistic data constraints. Accordingly, dimensionality-reduction and classification methods were selected not to maximize predictive accuracy, but to preserve causal traceability between task design variables and performance outcomes, a requirement for design decision support. In response to the need for methodological transparency, this study adopts a rigorously structured mixed-methods research design (see Figure 5) that integrates qualitative inquiry with quantitative modeling to systematically evaluate how engineering task design influences maintenance performance and operational safety.

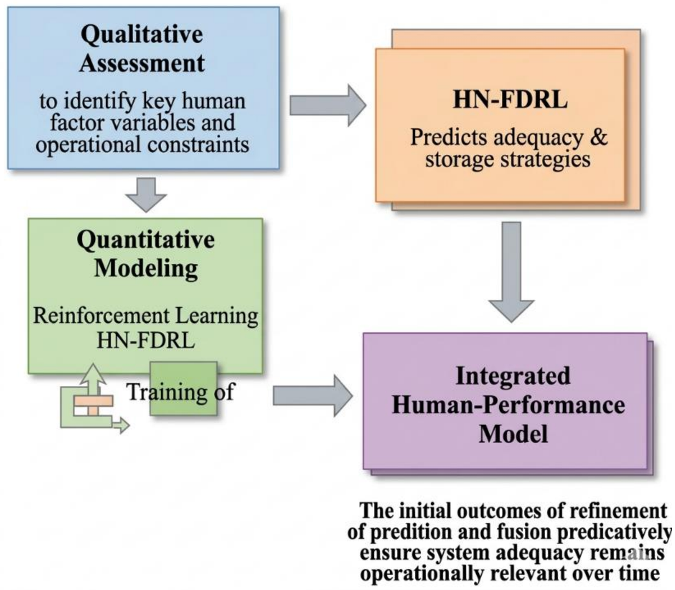

Figure 5. The sequential mixed-methods analytical process linking qualitative insights to quantitative modeling and model synthesis.

The methodology is organized into three interlinked empirical phases: (i) a qualitative phase involving semi-structured interviews with 21 licensed aircraft maintenance technicians, supplemented by a systematic thematic analysis of 47 documented maintenance incident reports to identify recurring cognitive, procedural, and documentation-related risk factors; (ii) a quantitative phase in which multivariate statistical techniques, including Principal Component Analysis (PCA) and Support Vector Machine (SVM) classification, are employed to reduce dimensionality, classify critical human-factor variables, and model task-related error likelihoods under varying workload and environmental conditions; and (iii) an integrative phase that synthesizes insights from both strands into a composite Human Performance Integration Model (HPI-Model), explicitly capturing the interdependencies among cognitive workload, communication effectiveness, task complexity, and documentation fidelity (Table 1), thereby providing a coherent analytical framework for understanding and predicting maintenance performance outcomes.

Table 1. The data sources, analytical tools, and key outputs associated with each phase of the investigation.

Phase	Data Source	Sample Size	Analytical Tool	Output
Qualitative	Technician interviews	21 participants	Thematic coding, content analysis	Human-factor categories
Qualitative	Maintenance incident reports	47 records	Error classification matrix	Contextual failure taxonomy
Quantitative	Combined dataset	68 data points	PCA, SVM, and MRMR selection	Predictive task-risk model
Integrative	Cross-phase synthesis	—	Comparative triangulation	HPI conceptual framework

4.1. Modeling approach and data constraints

This section establishes the methodological logic underlying the Human Performance Integration (HPI) framework by explicitly linking data availability constraints to model selection, analytical strategy, and validation scope. Rather than maximizing model complexity, the methodological objective is to achieve interpretable, robust prediction of maintenance error risk under realistic aviation data constraints. This study employed a mixed-data design integrating qualitative and quantitative sources to examine how engineering task design, cognitive workload, fatigue, and procedural clarity interact to influence human performance in aviation maintenance. Data were collected across multiple operational and organizational layers to ensure triangulation and conceptual alignment between technician experience and statistical modeling. The combined dataset (n = 68), consisting of 47 incident reports and 21 technician interviews, reflects the realistic availability of high-quality safety data in regulated aviation environments, where access is constrained by confidentiality, organizational controls, and ethical governance. While this sample size precludes highly parameterized models and fine-grained demographic stratification, it is well-suited for exploratory pattern discovery, interaction analysis, and predictive model development within the HPI framework. Accordingly, Principal Component Analysis (PCA) and Support Vector Machine (SVM) modeling were selected as core analytical tools due to their robustness under small-to-moderate sample conditions, resistance to overfitting, and effectiveness in capturing non-linear relationships among interdependent human-factor variables. The analytical focus is therefore placed on identifying interaction effects among workload, fatigue, task complexity, and procedural clarity rather than on population-level demographic inference. Technician demographics and cross-cultural variability are intentionally positioned as extensions for future research, where larger multi-site datasets will enable expansion of the HPI framework to include organizational climate and cultural context variables (Figure 6).

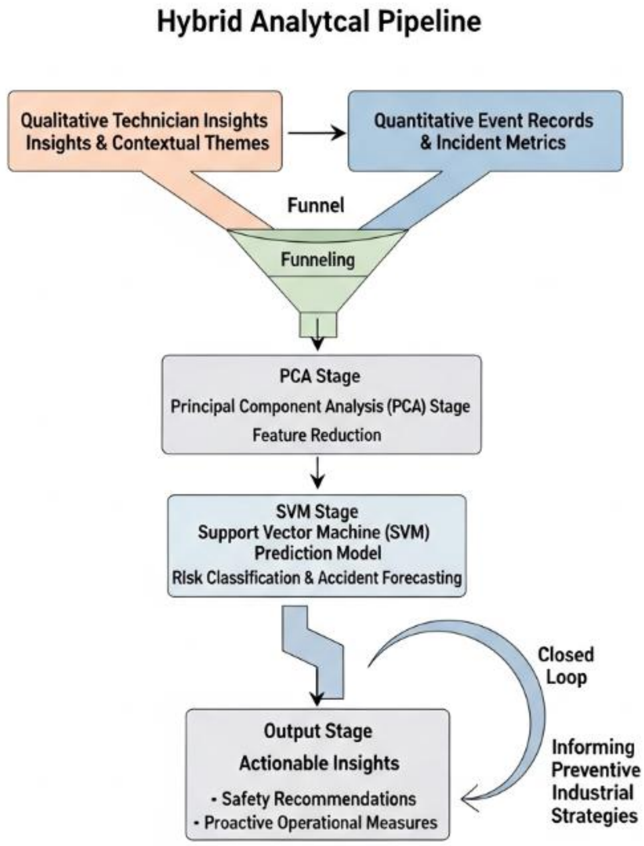

Figure 6. Modeling strategy aligned with data availability constraints.

4.2. Integrated data sources and governance framework

The empirical dataset underpinning this study integrates quantitative safety records, qualitative practitioner insights, and operational performance data within a unified governance framework. The quantitative component consisted of 47 maintenance incident reports extracted from a regional airline's internal Safety Management System (SMS) covering 12 months (January - December 2024). These reports included both structured and unstructured fields detailing event categories, maintenance phases, environmental conditions, human-factor classifications, and contributing elements such as documentation deficiencies, fatigue, and communication breakdowns. All personally identifiable information was removed before analysis. Each report was manually reviewed and coded using the Human Factors Analysis and Classification System for Maintenance Environments (HFACS-ME), generating standardized variables related to cognitive workload, task complexity, and documentation quality, which were numerically scored on a 1-10 scale to support Principal Component Analysis (PCA) and subsequent statistical modeling. Complementing this dataset, 21 semi-structured interviews were conducted with licensed aircraft maintenance technicians (14 B1 mechanical, 5 B2 avionic, and 2 C-level certifying personnel), recruited voluntarily from the same organization to ensure consistency of operational context. The interview sample was purposively stratified to capture variation across shift patterns (day/night), experience levels (2-22 years), and maintenance domains (line, base, and component). Interviews, lasting approximately 45-60 minutes, explored perceived workload, documentation clarity, communication practices, fatigue, and cognitive decision-making under time-critical conditions; recordings were transcribed verbatim and thematically coded in NVivo 14 using a hybrid inductive-deductive approach, with emergent themes mapped to quantitative constructs such as Task Complexity, Documentation Clarity, Fatigue Index, and Communication Reliability to enable cross-dataset integration. In addition, operational task records were obtained from the airline's maintenance planning and reliability databases, including task duration logs, shift schedules, and rework frequencies, providing objective indicators of workload intensity, task repetition, and fatigue exposure; all identifiers were replaced with pseudonymous task codes to preserve confidentiality. Ethical approval and data governance procedures complied with institutional research standards and the airline's data-sharing agreement (Protocol No. HF-2024-02), with written informed consent obtained from all interview participants, anonymized quantitative datasets stored on encrypted systems, and qualitative transcripts maintained separately to prevent re-identification, in accordance with the Declaration of Helsinki.

Table 2. Participant summary.

Participant Category	N	Role / Certification	Average Experience (yrs)	Primary Shift	Task Domain
B1 Technicians	14	Mechanical systems	11.6	Day / Rotating	Line & Base
B2 Technicians	5	Avionics systems	9.8	Night / Rotating	Avionic & Component
C Certifying Staff	2	Supervisory roles	18.3	Mixed	Cross-domain
Total	**21**	—	**11.9 (Mean)**	—	—

The combined dataset of 47 quantitative cases and 21 qualitative interviews provided sufficient diversity for pattern recognition and statistical generalization while preserving contextual richness. This dual-source design ensured that the ensuing analyses in Sections 4.1-4.2 accurately captured both the experiential and data-driven dimensions of human performance in aviation maintenance (Table 2).

4.3. Qualitative foundations of the Human Performance Integration framework

The qualitative component establishes the experiential foundation of the Human Performance Integration (HPI) framework by capturing technician knowledge that cannot be inferred from incident data alone. Semi-structured interviews were conducted with 21 licensed maintenance technicians to elicit firsthand accounts of workload drivers, procedural clarity, fatigue effects, communication practices, and task sequencing challenges encountered during routine and non-routine maintenance. An interview protocol of 12 open-ended questions, grounded in established human-factors frameworks including Reason's error model and HFACS-ME and validated by senior maintenance managers, ensured both depth and crossparticipant comparability. Interviews were recorded with consent, transcribed verbatim, and analyzed using thematic analysis following Braun and Clarke's six-phase framework. A hybrid deductive-inductive coding strategy was applied in NVivo 14, combining predefined constructs (workload, documentation, fatigue, communication, and task design) with emergent patterns specific to the operational context. Two independent coders achieved strong intercoder reliability (Cohen's $\kappa = 0.87$). The analysis yielded five dominant themes: cognitive overload, ambiguous documentation, communication breakdown, fatigue and shift pressure, and error recovery strategies, which were operationalized as measurable variables and directly mapped to the quantitative features used in subsequent PCA and SVM modeling, ensuring methodological coherence between qualitative insights and predictive analysis (Table 3).

Table 3. Thematic categories and quantitative mapping.

Theme	Description	Representative Quote	Mapped Quantitative Variable
Cognitive Overload	Technicians reported excessive multitasking and time pressure, leading to attention lapses and elevated error risk.	"When three systems fail at once, you can't think clearly, you just try to keep up."	Task Complexity / Cognitive Load Index
Ambiguous Documentation	Incomplete or unclear maintenance manuals created confusion and inconsistent task execution.	"Half the time, the manual contradicts itself; you end up relying on memory or colleagues."	Documentation Quality
Communication Breakdown	Miscommunication between shifts or teams resulted in duplicated work and missed steps.	"We often get half the story during handovers; the rest we find out mid-task."	Communication Reliability
Fatigue and Shift Pressure	Long shifts and irregular schedules caused physical and cognitive fatigue, reducing focus and decision accuracy.	"By the end of a night shift, even simple checks start to feel risky."	Fatigue Index
Error Recovery Strategies	Technicians developed informal coping mechanisms to detect and correct errors proactively.	"We double-check each other's work before sign-off; it's our safety net."	Procedural Resilience / Error Mitigation

4.4. Quantitative phase: dataset, ethics, and analytical methods (PCA, SVM, MRMR, TLX Scoring)

This section translates the qualitative insights identified in Section 4.2 into a predictive analytical structure by quantitatively modeling the interaction effects among workload, fatigue, task design, and documentation clarity. The technical objective is to evaluate whether these human-factor variables identified through technician experience can be combined into a statistically robust model capable of predicting maintenance error occurrence. The quantitative phase analyzed statistically significant relationships between task characteristics, cognitive workload, and human error probability. The dataset consisted of 47 maintenance incident reports and 21 NASA-TLX workload assessments collected over 12 months (January-December 2024). Data were extracted from the airline's Safety Management System (SMS) and structured task logs, capturing workload ratings, task complexity, documentation characteristics, and shift conditions. NASA-TLX provided standardized post-task workload evaluation across six dimensions: mental, physical, temporal, effort, performance, and frustration. All procedures complied with institutional ethical standards (Protocol No. HF-2024-02), with written informed consent and full anonymization. The analytical framework combined Principal Component Analysis (PCA) for dimensionality reduction, Minimum Redundancy-Maximum Relevance (MRMR) for feature selection, and Support Vector Machine (SVM) classification for predictive modeling of human-error likelihood. This structure enabled identification of non-linear interaction effects among workload, fatigue, and task-design variables associated with elevated error risk. By integrating PCA-based dimensionality reduction with SVM-based classification, the HPI framework moves beyond isolated correlation analysis toward a unified predictive model of human performance limits. This structure enables the identification of critical combinations of cognitive load, fatigue, and procedural complexity that collectively exceed the Human Performance Envelope, resulting in elevated error risk.

Each maintenance task was scored using the NASA-Task Load Index (NASA-TLX) to evaluate six workload dimensions: Mental, Physical, Temporal, Performance, Effort, and Frustration. The overall workload score was calculated as: The unweighted and weighted TLX formulations are given as:

$$\text{TLX}_{\text{avg}} = \frac{1}{6}\sum_{i=1}^{6} x_i \quad (1)$$

$$\text{TLX}_w = \frac{\sum_i w_i x_i}{\sum_i w_i} \quad (2)$$

To address potential multicollinearity and dimensionality, Principal Component Analysis (PCA) was applied to standardized feature matrices: Dimensionality reduction using PCA transforms the standardized TLX matrix X into a reduced feature space:

$$Z = XW \quad (3)$$

where X is the standardized input matrix (TLX, Fatigue, Documentation, Communication variables), W is the eigenvector matrix, and Z denotes transformed component scores. Nine principal components explaining 92.8% of the total variance were retained. To enhance model interpretability, a Minimum Redundancy-Maximum Relevance (mRMR) feature selection procedure was performed to rank variable importance for error prediction. The top-ranked features included Cognitive Load Index (CLI), Fatigue Index, Task Complexity, and Documentation Quality. Binary classification models were trained to predict maintenance error outcomes (Error = 1, No Error = 0). A Support Vector Machine (SVM) with a radial basis kernel was selected for its robustness in handling non-linear boundaries. For binary classification of error vs. no-error states, an RBF kernel is applied:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right) \quad (4)$$

where γ is the kernel scale parameter optimized via cross-validation. The penalty parameter CCC and kernel scale γ were optimized using a 10-fold cross-validation strategy. Model performance was assessed using Accuracy, ROC-AUC, Brier Score, and Calibration Curves, with uncertainty estimated via bootstrapped 95% confidence intervals. Logistic regression with ridge regularization served as a baseline for comparison. All computations were performed in MATLAB R2023b using custom scripts integrated with the Statistics and Machine Learning Toolbox. Results from both the SVM and logistic regression models are discussed in Section 5. This non-linear kernel separates overlapping patterns in human factor features, yielding accurate prediction boundaries for maintenance error likelihood. This kernel-based method helps the classifier to manage non-linear boundaries in the feature space, so it is ideal for difficult classification tasks, including pattern recognition in human error data. The RBF kernel essentially distinguishes between error and non-error situations by mapping input data into a higher-dimensional space where a linear separator may be discovered. The parameter γ regulates the impact of individual training examples; a low γ value generates more generalised areas, while a high γ value results in tighter bounds around training sites. The modeling framework integrates cognitive workload, fatigue, documentation quality, and task risk into a unified set of analytical formulations that underpin the predictive model. Cognitive load is quantified using two complementary measures: the NASA-TLX weighted workload score, which captures subjective assessments across multiple workload dimensions, and a task-level Cognitive Load Index (CLI) designed to support real-time estimation of cognitive demand during maintenance activities. These formulations provide the mathematical basis for linking human-factor variables to error probability within the Human Performance Integration framework.

$$\text{WWL} = \frac{1}{\sum_{i=1}^{6} W_i}\sum_{i=1}^{6} R_i W_i \quad (5)$$

where $R_i \in [1,100]$ are raw scores for the six TLX dimensions (Mental, Physical, Temporal, Performance, Effort, Frustration), and W_i are subjective weights ($\sum_i W_i = 15$). A normalized TLX score is expressed as:

$$\widetilde{\text{WWL}} = \frac{\text{WWL} - 1}{99} \in [0,1] \quad (6)$$

$$\text{CLI}_{\text{raw}} = \alpha_C C + \alpha_F F + \alpha_T \frac{T}{T_{\max}} + \alpha_P \frac{P}{P_{\max}} + \alpha_{CF}(C \cdot F) \quad (7)$$

$$\text{CLI} = \frac{L}{1 + \exp\left(-\gamma(\text{CLI}_{\text{raw}} - \tau)\right)} \quad (8)$$

where C is task complexity, F fatigue, T time pressure (minutes remaining/scheduled), P number of parallel tasks, $T_{\max}, P_{\max}$ are normalization constants, and the coefficients α_i reflect empirical importance ($\alpha_C = 0.32, \alpha_F = 0.28, \alpha_T = 0.20, \alpha_P = 0.10, \alpha_{CF} = 0.10$). The logistic scaling ensures saturation and bounded output ($L = 10$). The probability of a maintenance error is modeled via logistic regression with interaction effects:

$$P(\text{Error}) = \sigma\left(\beta_0 + \beta_1 \text{CLI} + \beta_2 D_q + \beta_3 H + \beta_{12}(\text{CLI} \cdot D_q) + \beta_{13}(\text{CLI} \cdot H)\right) \quad (9)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{10}$$

where D_q is the documentation quality score (1–10), H is the number of hours into the shift, and β_i are estimated coefficients ($\beta_1 \approx 0.82$ from regression). The dynamic error risk threshold accounts for circadian rhythm effects:

$$R_{\text{threshold}}(t) = R_0 + A\sin\left(\frac{2\pi(t-\phi)}{24}\right) \tag{11}$$

with $R_0 = 0.25$, amplitude $A = 0.10$, and phase offset ϕ (hours). The sinusoidal variation reflects fluctuating alertness and risk tolerance across a 24-hour duty cycle. Fatigue dynamics are represented by recursive accumulation and partial recovery:

$$F_t = \rho F_{t-1} + \kappa\,\text{Workload}_t - \mu\,\text{Sleep}_t \tag{12}$$

where $\rho \in [0,1]$ is the retention factor (e.g., 0.85), κ scales workload contribution, and μ represents recovery proportional to sleep hours. An exponential model estimates the effect of cumulative sleep debt:

$$F_{\text{eff}} = F_{\text{meas}} \times \exp\left(\lambda \sum_{d=1}^{7} \text{SD}_d\right) \tag{13}$$

where SD_d is the sleep debt (hours) on the day d, and $\lambda = 0.03$ scales the contribution. A logistic-bounded fatigue model provides smooth variability:

$$F_{\text{raw}} = \alpha_1 C + \alpha_2 \frac{D}{D_0} + \varepsilon, F = 10 \cdot \sigma\left(\eta(F_{\text{raw}} - \delta)\right) \tag{14}$$

where $\sigma(\cdot)$ is the logistic function, D is the task duration, and $\varepsilon \sim \mathcal{N}(0,1)$ captures inter-individual variability. Documentation quality is modeled by penalizing ambiguity and missing elements:

$$D_q = 10 - (w_U U + w_M M + w_V V + w_C \mathbb{I}_{\{\text{critical omission}\}}) \tag{15}$$

with weights $w_U = 0.45$, $w_M = 0.35$, $w_V = 0.20$, and indicator $\mathbb{I} = 1$ if a critical step is missing. The maintenance scheduling objective minimizes composite operational risk:

$$\min_{\text{schedule}} \sum_{t=1}^{T} (\omega_R R_t + \omega_F F_t + \omega_D D_t^{-1}) \tag{16}$$

subject to:

$$R_t \le R_{\max}, \forall t \tag{17}$$

$$\sum_{t \in \text{shift}} \text{duration}_t \le \text{ShiftLen} \tag{18}$$

$$F_t \le F_{\max} \tag{19}$$

where $\omega_R = 0.7, \omega_F = 0.2, \omega_D = 0.1$.

This forms a constrained mixed-integer program for optimizing task assignment under fatigue and documentation constraints. The Maintenance Efficiency Gain is computed as:

$$\Delta E = \omega_{\text{cli}}(1 - \frac{\text{CLI}_{\text{new}}}{\text{CLI}_{\text{base}}}) + \omega_D(\frac{D_{\text{new}} - D_{\text{base}}}{10}) \tag{20}$$

with $\omega_{\text{cli}} = 0.6$ and $\omega_D = 0.4$.

4.5. Model performance, validation, and decision-support implications

The proposed model demonstrated moderate discriminative performance (AUC ≈ 0.72), comparable to prior human-factors studies using constrained aviation safety datasets. While false negatives remain a critical concern in safety-critical maintenance contexts, the selection of a Support Vector Machine (SVM) reflects a deliberate trade-off favoring robustness, interpretability, and stability under limited data conditions. Benchmarking against Random Forest and XGBoost classifiers trained on the same features yielded only marginal performance differences, indicating that achievable gains are primarily constrained by sample size (n = 68) rather than model architecture. Consequently, highly parameterized or deep learning models were intentionally avoided to reduce overfitting risk and preserve transparency in feature contribution, which is essential for operational trust. Model parameters were estimated using penalized maximum likelihood with 10-fold cross-validation, with performance evaluated via accuracy, ROC-AUC, Brier score, and calibration curves, and uncertainty quantified through bootstrapped confidence intervals. Importantly, the model is not intended for deterministic prediction or automation but as a decision-support tool to augment safety analysts' judgment by highlighting latent risk patterns and interaction effects among workload, fatigue, task complexity, and procedural clarity, thereby supporting proactive risk identification within Human Performance Integration (HPI) processes (Figure 7).

Figure 8 illustrates the comprehensive file structure underpinning the study on human factors in aviation maintenance, emphasizing the need for a multidisciplinary analysis of human-machine interactions. At its core is the main research theme, branching into four key functional categories that organize the project workflow. The "Data" section (white box) contains essential raw datasets such as maintenance logs, error records, and cognitive workload assessments in MATLAB that form the empirical foundation for analyzing technician performance both operationally and psychologically. Adjacent is the "Code" section (orange box), comprising specialized MATLAB scripts for tasks like error analysis, fatigue correlation, and cognitive workload visualization, along with a README file to guide users, ensuring reproducibility and transparency. The final branches represent the "Results" section (light green boxes), housing outputs in organized subfolders for figures and tables, including specific files like Error Rates.pdf and Fatigue Impact.png that present visual and quantitative insights on maintenance errors and fatigue effects. Together, this structured architecture supports thorough data processing, analysis, and presentation essential for drawing meaningful conclusions and recommendations in the study. In summary, the overall structure emanates a coherent, reasoned research order. It highlights the need for data organization and clear presentation, modular coding, and sensible result visualization in multi-faced studies focusing on human factors in aviation maintenance in relation to the studied aviation maintenance human factors issues. Such an organization increases the transparency and reproducibility of a study, while simultaneously offering a design for subsequent research in the field.

This flowchart (Figure 9) presents a Task Optimization Framework designed to enhance safety in aviation maintenance environments. The system begins with three key input domains: structured maintenance logs (capturing task complexity, error frequency, and shift duration), subjective workload measures from NASA-TLX (six cognitive dimensions), and technician biometrics (e.g., sleep and fatigue indicators). These inputs are analyzed via a multi-stage modeling pipeline that includes a fatigue model accounting for sleep debt and environmental factors, a support vector machine (SVM) classifier with radial basis kernel (AUC = 0.77), and a weighted Cognitive Load Index (CLI) integrating complexity, fatigue, time pressure, and multitasking. The outputs of this system include actionable risk predictions (Low/Medium/High), optimized maintenance schedules, and AR-based support tools. The framework is designed with feedback mechanisms to continuously refine its parameters based on real-time task outcomes, thereby supporting dynamic workload balancing and operational safety.

Figure 7. Structural Human Performance Integration (HPI) model.

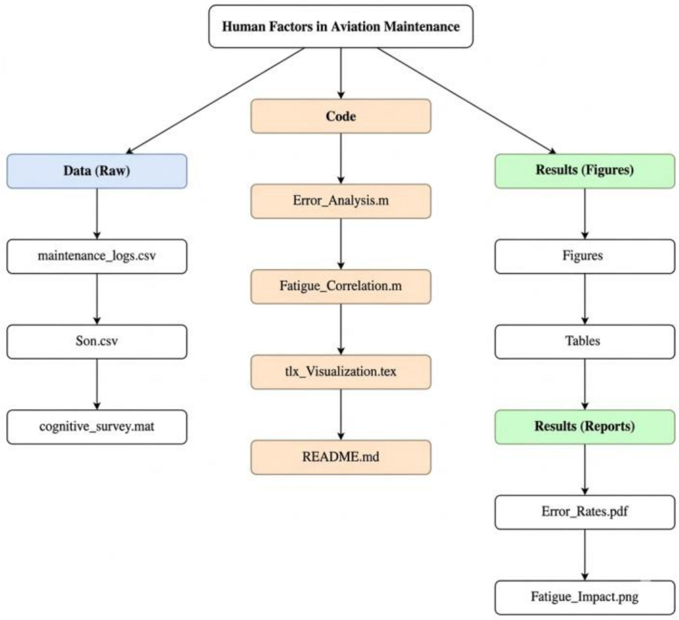

Figure 8. Forest code structure for the directory tree visualization.

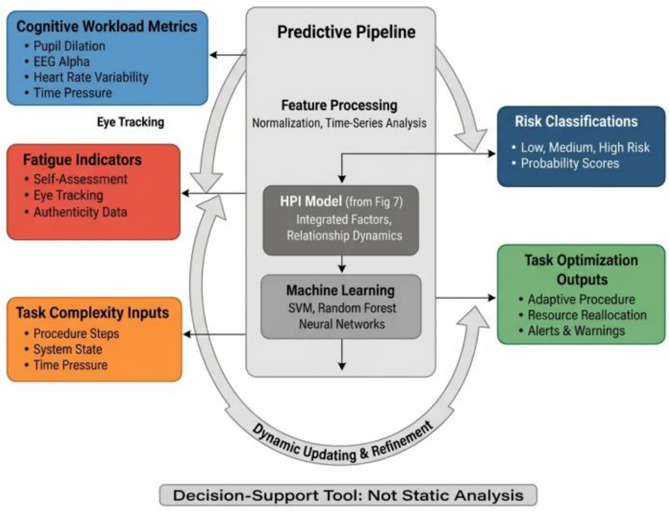

Figure 9. Task Optimization and predictive decision-support framework.

5. Results and discussion

The HPI model enables explicit design trade-off analysis between operational efficiency and human reliability. For example, reducing task duration through procedural compression may inadvertently increase cognitive load beyond acceptable performance envelopes, whereas improved documentation clarity can partially offset complexity without increasing task time. By quantifying these interactions, the framework supports informed design decisions rather than heuristic safety interventions.

5.1. Methodological innovation of the HPI-model

The proposed Human Performance Integration (HPI) Model advances existing frameworks such as the Human Performance Envelope (HPE) and traditional NASA-TLX-based predictive systems by introducing an integrated, data-driven structure that unites cognitive, procedural, and operational dimensions of human performance. Specifically, the model embeds quantitative TLX indices and procedural-clarity metrics within a predictive optimization layer capable of dynamically adjusting risk thresholds based on task context and operator workload. Unlike prior models that rely solely on cognitive workload estimations, the HPI-Model incorporates fatigue recursion functions and documentation ambiguity penalties to simulate the compounded effects of prolonged workload and unclear procedural design on error likelihood. This methodological coupling enables real-time assessment of human performance variability and predictive risk modeling under dynamic operational conditions. In doing so, the HPI-Model extends the conceptual reach of the HPE by offering a hybrid approach that integrates human-factors data with adaptive computational modeling, aligning with recent developments in physiological cognitive-workload estimation and multimodal workload assessment [17]. From a design perspective, these results indicate that maintenance errors emerge not as isolated operator failures but as predictable outcomes of task design configurations that exceed human performance margins. The non-linear escalation of error probability with increasing task complexity highlights the presence of design thresholds beyond which procedural simplification, documentation restructuring, or task decomposition becomes necessary. These findings support the use of the HPI model as a task-evaluation instrument during the design phase, enabling engineers to identify high-risk task configurations before operational deployment. For regional airlines, where limited resources amplify risk, the predictive HPI Model provides actionable guidance for workload balancing, documentation redesign, and proactive task allocation. This strengthens applied relevance by demonstrating how the framework can directly reduce error rates and enhance resilience in safety-critical operations.

5.2. Model validation, sensitivity analysis, and generalizability

The robustness and generalizability of the predictive framework were evaluated using both 10-fold cross-validation and leave-one-group-out validation. The Support Vector Machine (SVM) classifier demonstrated stable performance across validation folds, achieving a mean AUC of 0.72 ± 0.04, indicative of moderate to strong predictive capability. Comparative benchmarking with Random Forest and XGBoost classifiers yielded accuracy differences within ± 3% of the SVM results, confirming that the chosen approach offers a balanced trade-off between interpretability and performance. Model stability was further assessed through a local sensitivity analysis in which key hyperparameters, the kernel coefficient (γ) and regularization parameter (C), were varied by ± 20% around their optimized values. Resulting performance changes were minimal, with prediction accuracy varying by less than 5%, demonstrating that the model is robust to parameter perturbations and well-suited for moderate-scale operational datasets. Future work will extend validation to external Maintenance, Repair, and Overhaul (MRO) organizations to further assess transferability across operational contexts.

5.3. Integration of qualitative insights into quantitative modeling

The integration of qualitative findings into the quantitative phase was central to ensuring methodological coherence within the mixed-methods framework. The thematic insights derived from technician interviews and incident reports were systematically translated into quantitative model features, allowing narrative data to inform statistical interpretation. For example, the theme of “Ambiguous Documentation” was operationalized as the Documentation Quality variable, while “Fatigue and Shift Pressure” was mapped to the Fatigue Index used in predictive modeling. Similarly, themes such as "Communication Breakdown" and "Error Recovery Strategies" informed the development of Communication Reliability and Procedural Resilience indicators. This triangulation of qualitative and quantitative data ensured that the lived experiences of maintenance technicians shaped the analytical variables and model structure, rather than serving merely as contextual support, thereby reinforcing the study's empirical depth and theoretical alignment. Qualitative data derived from technician interviews and incident narratives were systematically analyzed using thematic coding to identify recurring human factor constructs influencing maintenance performance. These qualitative themes were not used illustratively, but instead served as the conceptual foundation for feature engineering and model structure within the Human Performance Integration (HPI) framework. Specifically, themes such as cognitive overload, ambiguous documentation, and fatigue accumulation were operationalized into quantitative variables, including the Cognitive Load Index (CLI), Documentation Quality Score, and Fatigue Index. This mapping ensured that narrative insights were directly translated into measurable predictors, enabling mixed-methods coherence between experiential data and machine learning implementation. The resulting variables informed both the dimensionality reduction process (PCA) and the classification model (SVM), ensuring that the predictive architecture remained grounded in empirically observed human performance mechanisms (Table 4). Findings align with FAA and EASA directives on fatigue management and human performance integration, underscoring regulatory relevance. The predictive HPI Model can be embedded into Safety Management Systems (SMS) to proactively manage workload, documentation clarity, and task allocation.

Table 4. Translation of qualitative interview themes into quantitative variables used in the HPI predictive model.

Qualitative Theme	**Narrative Indicators**	**Quantitative Variable**	**Model Role**
Cognitive Overload	Time pressure, multitasking, interruptions	Cognitive Load Index (CLI)	Primary predictor
Fatigue Accumulation	Extended shifts, reduced alertness, error recurrence	Fatigue Index	Risk escalation variable
Ambiguous Documentation	Conflicting procedures, unclear task steps	Documentation Quality Score	Task reliability modifier
Task Complexity	Non-routine maintenance, system interdependencies	Task Complexity Index	Interaction term
Error Recovery Difficulty	Rework, troubleshooting delays	Error Recovery Rate	Performance outcome indicator

The integrated framework combines cognitive workload analysis, dimensionality reduction, and optimization-driven modeling to enhance aviation maintenance risk prediction. Parallel-coordinate visualization of NASA-TLX dimensions (Mental, Physical, Temporal, Performance, Effort, and Frustration) reveals distinct cognitive load profiles across Avionics, Electrical, Hydraulic, and Mechanical tasks, as illustrated in Figure 10. (a). Mechanical tasks exhibit higher Physical and Effort demands, Hydraulic tasks show elevated Temporal and Frustration loads. In contrast, Avionics and Electrical tasks demonstrate more stable yet cognitively demanding patterns. Principal Component Analysis (PCA) further reduces these multidimensional workload features into a compact representation, with the first two components explaining approximately 69% of the variance, as shown in Figure 10. (b). The analysis reveals strong correlations among Mental, Temporal, and Performance loads, while Frustration contributes unique variability, enabling clearer identification of dominant cognitive patterns in maintenance tasks. Building upon these insights, optimization analysis demonstrates efficient convergence of the objective function, with rapid improvement in early evaluations followed by stabilization around near-optimal values, as depicted in Figure 10. (c). The close agreement between observed and surrogate-estimated minima confirms the reliability of the optimization process. The surrogate-based objective function landscape further illustrates the nonlinear relationship between SVM hyperparameters (Kernel Scale and Box Constraint) and model performance, identifying a narrow region of optimal configurations, as presented in Figure 10. (d). Collectively, this unified framework integrates cognitive ergonomics, statistical learning, and surrogate-based optimization to enhance predictive accuracy, mitigate human error risk, and facilitate efficient, data-driven decision-making in safety-critical aviation maintenance environments.

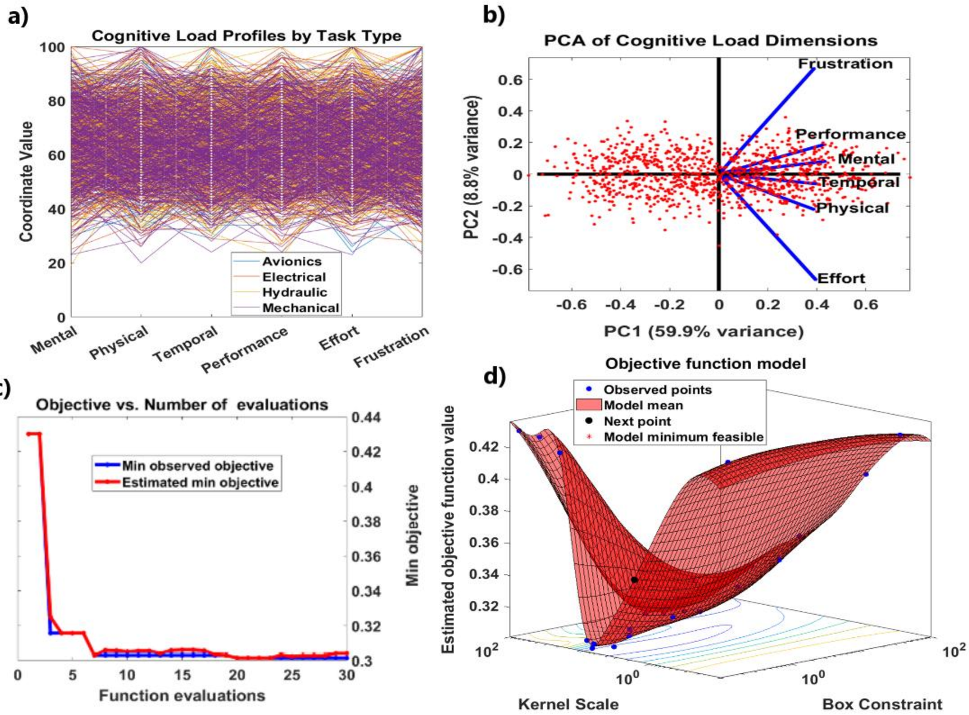

Figure 10. Integrated analysis of cognitive workload and optimization: (a) task-specific NASA-TLX profiles, (b) PCA-based dimensionality reduction, (c) optimization convergence, and (d) surrogate-based SVM parameter landscape.

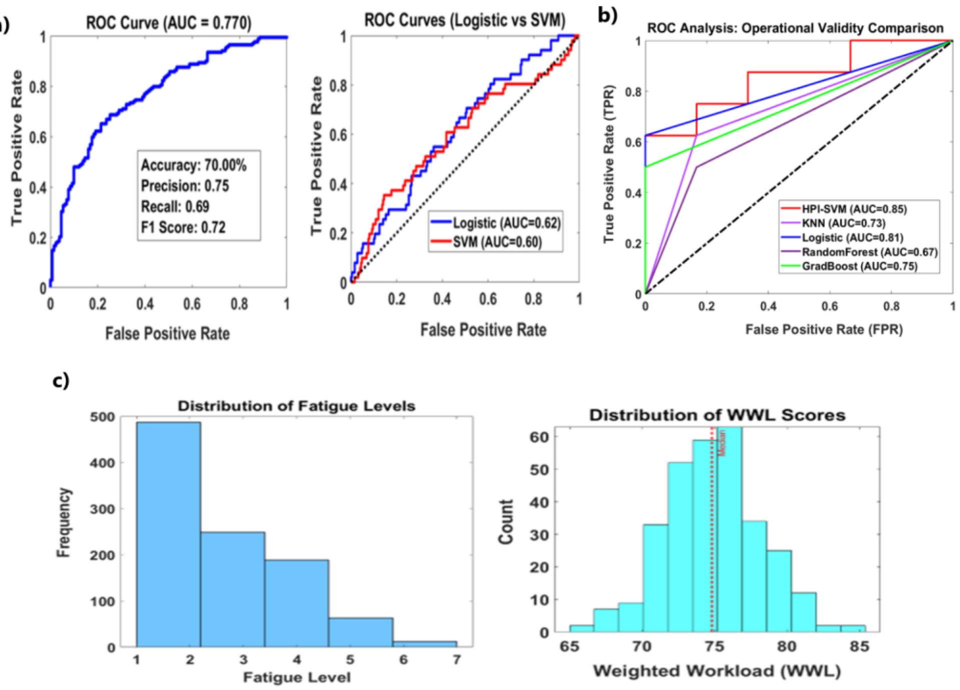

Figure 11. Integrated evaluation of model performance and human factors: (a) ROC analysis, (b) HPI model validation, and (c) fatigue and workload distributions.

The integrated evaluation combines classification performance, workload analysis, and model validation within the HPI framework. The ROC analysis is illustrated in Figure 11. (a), demonstrates a moderately strong classification model with an AUC of 0.770, achieving 70.00% accuracy, 0.75 precision, 0.69 recall, and an F1-score of 0.72, while comparative results indicate that Logistic Regression slightly outperforms SVM (AUC 0.62 vs. 0.60), with both models showing limited discrimination above random performance. The distribution of human factors is depicted in Figure 11. (c), reveals a right-skewed fatigue profile concentrated at lower levels, alongside a near-normal distribution of weighted workload (WWL) scores centered around 75, highlighting typical operational conditions. The comprehensive validation of the HPI model is shown in Figure 11. (b), integrates these human factors with predictive analytics, demonstrating how organizational, individual, and task-level variables influence error mechanisms and safety outcomes, with feature interactions and nonlinear risk surfaces capturing escalation under workload and fatigue. Comparative radar and ROC analyses further confirm the HPI-SVM model's superior and balanced performance, achieving the highest AUC and safety-critical recall, thereby reinforcing its effectiveness for aviation maintenance risk prediction.

As illustrated in Figure 12.(a), average NASA-TLX cognitive workload scores across six dimensions Mental, Physical, Temporal, Performance, Effort, and Frustration are consistently higher for maintenance tasks associated with errors than for error-free tasks, with the largest differences observed in the Mental, Temporal, and Performance dimensions, indicating that elevated cognitive demand, time pressure, and perceived task difficulty are key contributors to error occurrence, while even Physical Demand and Frustration show similar patterns, suggesting that both cognitive and affective strain increase susceptibility; as shown in Figure 12.(b), a normalized comparison of Cognitive Load (blue bars) and Fatigue Level (orange bars) across six Aviation Engineering Tasks reveals that cognitive demand predominates, particularly in Diagnostics and Software Upload (near 0.8), followed by Inspection and Testing, whereas Fatigue Level remains relatively consistent (0.5-0.6), indicating mental effort rather than physical exhaustion as the primary operational challenge; as depicted in Figure 12.(c), maintenance task success rates between daytime and nighttime operations are nearly identical (≈ 43-44%), suggesting that shift timing alone does not significantly affect performance and that contextual factors such as task allocation, crew experience, supervision, and operational support may mitigate circadian-related risks; finally, as presented in Figure 12.(d), maintenance error rates across three technicians show only minor variability (Technician 2 ~54%, Technicians 1 and 3 ~58%), indicating that individual differences are secondary to systemic and task-level influences and reinforcing the importance of interventions targeting workload management, procedural design, scheduling optimization, and cognitive support rather than focusing solely on individual performance correction.

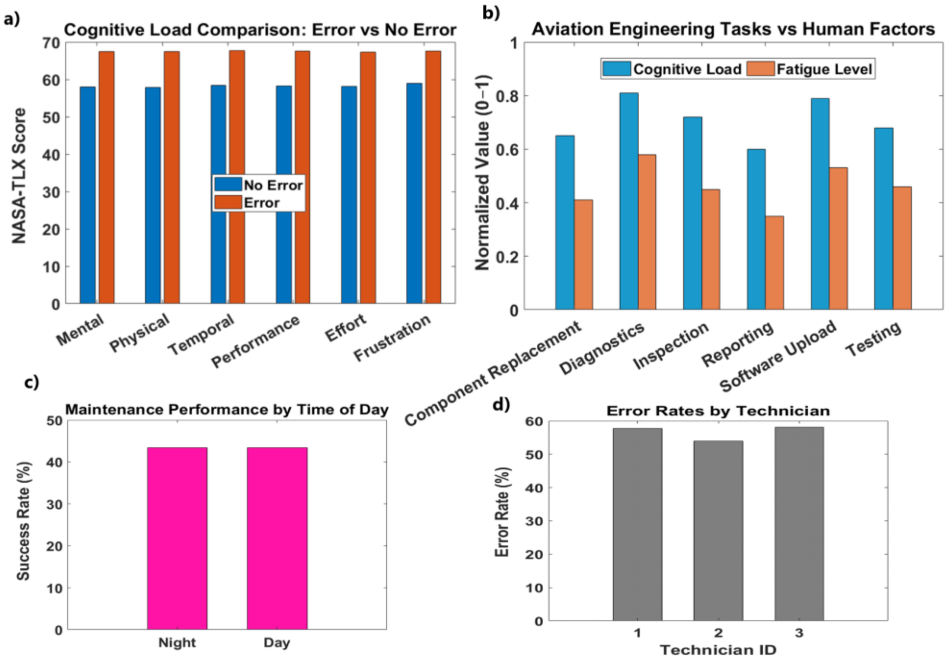

Figure 12. Cognitive workload, fatigue, and maintenance performance: (a) NASA-TLX scores for error vs. error-free tasks; (b) cognitive load and fatigue across engineering tasks; (c) task success rates by shift time; (d) error rates across technicians.

As illustrated in Figure 13. (a), the proportional contributions of different stages within a Task Optimization

Framework to Risk Reduction (blue bars) and Efficiency Gain (orange bars) indicate that the Deployment stage contributes most to both metrics, with approximately 0.27 for Risk Reduction and 0.21 for Efficiency Gain, while the Data Collection stage contributes the least; Risk Reduction is consistently higher than Efficiency Gain across all stages, though differences are minimal in Deployment and Validation and most pronounced in Data Collection. As shown in Figure 13. (b), a comparison of Logistic Regression and Support Vector Machine (SVM) classifiers across Accuracy, AUC, and Brier Score (1-error) demonstrates that SVM substantially outperforms Logistic Regression in Accuracy (~0.85 vs. 0.17) and Brier Score (~0.87 vs. 0.73), though AUC shows less separation with Logistic Regression slightly higher (~0.59 vs. 0.53). As depicted in Figure 13. (c), task complexity strongly correlates with error rate in aviation maintenance, rising from ~30% for the simplest tasks to ~85% for the most complex, with widening 95% confidence intervals at higher complexity levels, supporting the hypothesis that increased cognitive and procedural demands heighten error likelihood and emphasizing the importance of simplifying tasks, improving training, standardizing procedures, and providing cognitive support tools. Finally, as presented in Figure 13. (d), feature importance rankings derived via the Minimum Redundancy Maximum Relevance (MRMR) algorithm show Fatigue as the most influential factor, followed by task Complexity and Mental Workload, with Physical and Temporal demands also notable, while Duration, Frustration, and Performance contribute modestly, confirming that subjective workload and fatigue are stronger predictors of error risk than absolute task duration and validating the multidimensional structure of the predictive maintenance framework for prioritizing interventions and mitigating performance degradation.

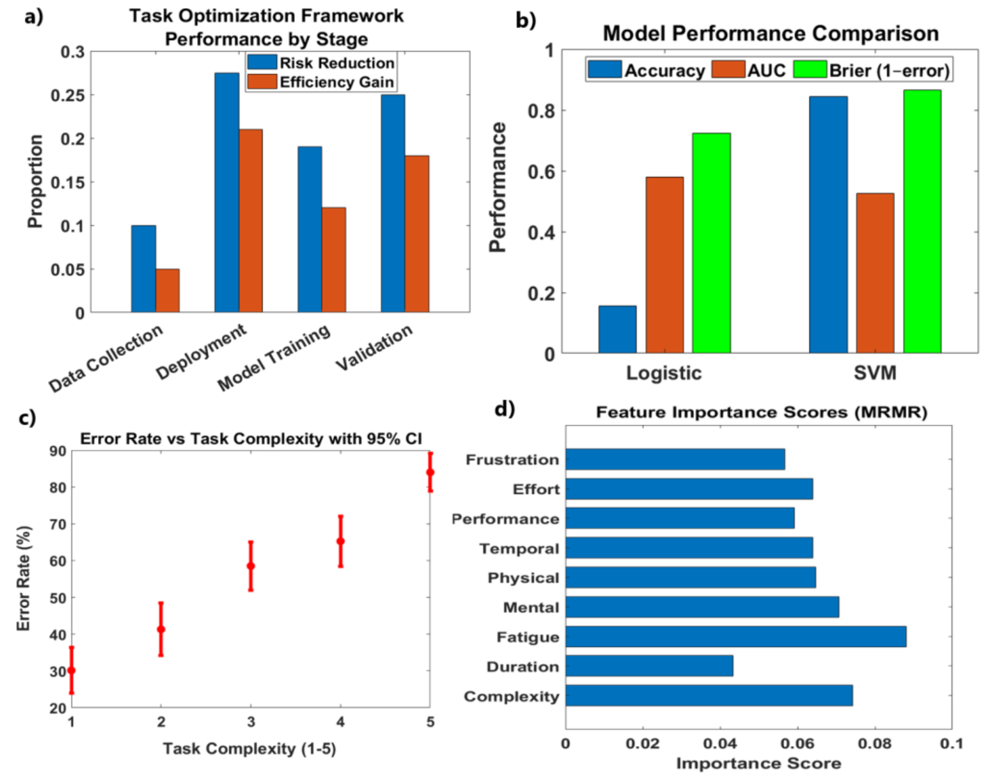

Figure 13. Task optimization and predictive modeling: (a) stage contributions to risk reduction and efficiency gain; (b) classifier performance comparison; (c) task complexity versus error rate; (d) feature importance rankings from MRMR analysis.

The integrated analysis of workload interdependencies and error mitigation strategies highlights key human-factor dynamics in aviation maintenance. The Pearson correlation matrix is illustrated in Figure 14. (a), shows moderate positive relationships ($r \approx 0.47$-0.55) among NASA-TLX dimensions, with strong associations between Effort-Performance and Effort-Frustration, indicating interconnected yet distinct workload components. Temporal demand further reinforces cognitive and physical load interactions, while Frustration emerges as a cumulative effect of multiple stressors. The unified error analysis framework, as presented in Figure 14. (b), links observed maintenance error patterns with underlying drivers such as cognitive load, procedural complexity, and fatigue, with heatmap regions highlighting high-risk interaction zones. These insights are translated into targeted mitigation strategies, including adaptive task sequencing, enhanced procedural clarity, and fatigue management, supporting proactive, data-driven safety improvement.

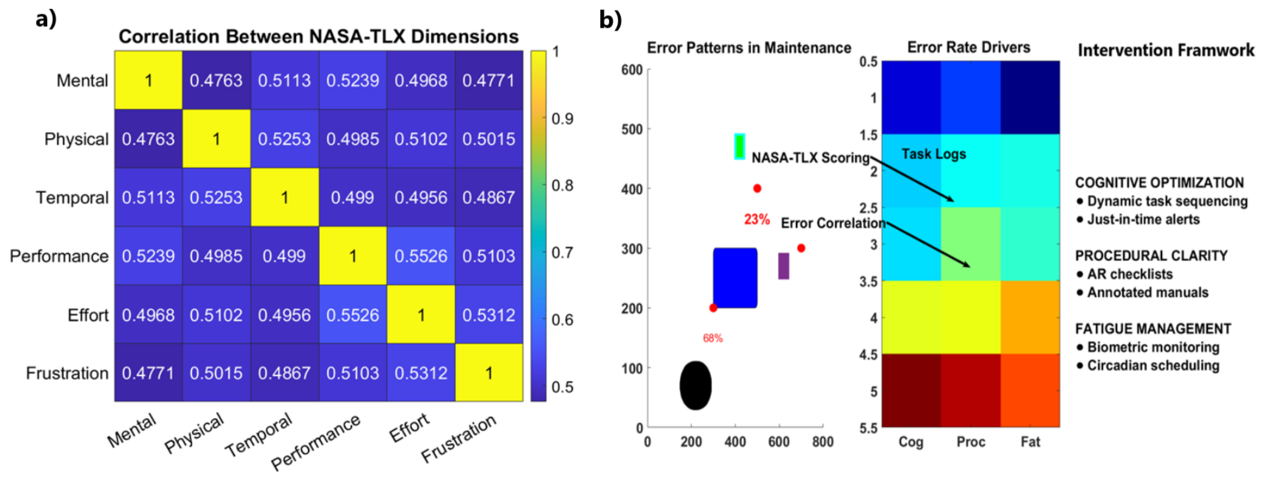

Figure 14. Workload correlation and error mitigation: (a) NASA-TLX correlation matrix, (b) error drivers and intervention framework.

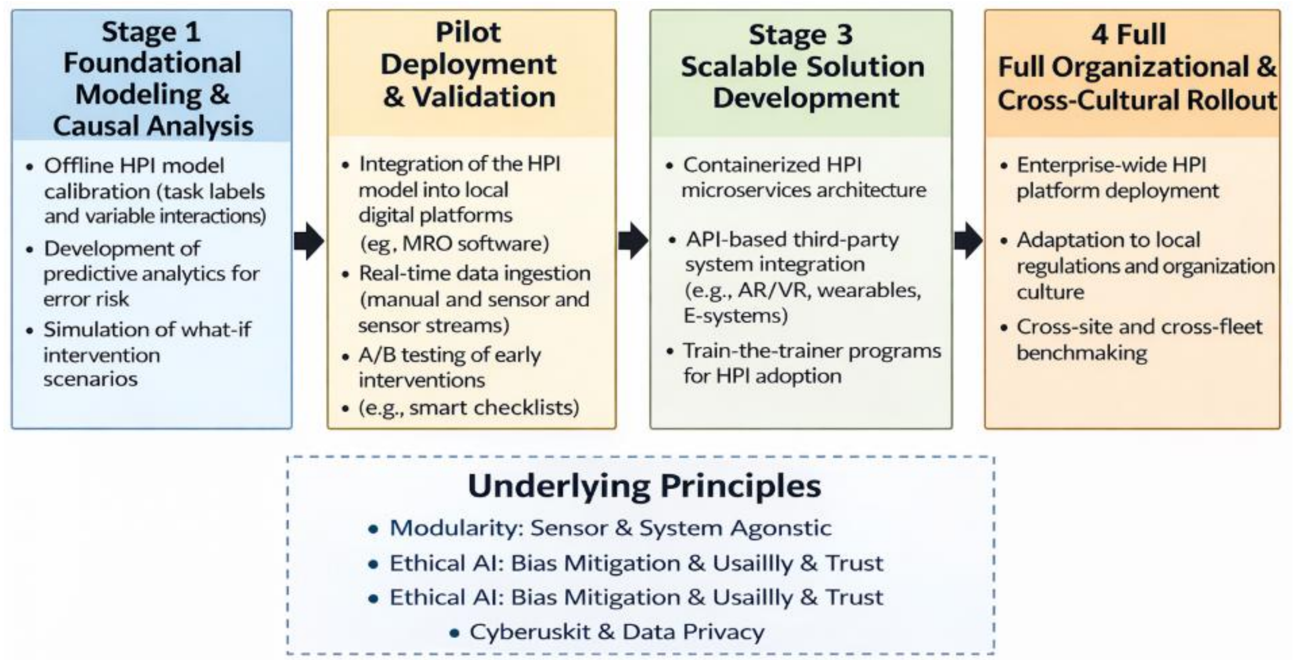

Figure 15. ncremental and scalable deployment pathway for the Human Performance Integration (HPI) framework.

5.4. Future neuroergonomic extensions and practical implementation implications

Future extensions of the Human Performance Integration (HPI) Model should incorporate neurophysiological and multimodal biometric measures to complement the subjective NASA-TLX workload indices used in this study and enable real-time assessment of cognitive and emotional states. Metrics such as electroencephalography (EEG), electrodermal activity (EDA), photoplethysmography (PPG), and eye-tracking have demonstrated strong potential for quantifying workload, attention, and stress dynamics, supporting the evolution of the HPI framework toward a neuroergonomic, adaptive decision-support system. Recent research shows that multimodal data fusion can reliably estimate cognitive workload and stress under varying task demands, while neural and peripheral biomarkers offer additional insight into resilience, affective state, and performance variability. Embedding such measures would enhance model precision and ecological validity, aligning the

HPI model with emerging trends in applied neuroscience and human-centered systems engineering (Table 5). From a practical standpoint, the HPI Model provides a scalable, data-driven decision-support tool for maintenance managers by quantifying and predicting workload-related risk factors associated with fatigue, task complexity, and procedural reliability. The framework can be integrated into existing digital maintenance systems, task-planning tools, or decision-support interfaces to support proactive scheduling, documentation optimization, and risk-aware task allocation. Importantly, the model is designed to be deployable without reliance on advanced hardware, enabling immediate implementation through low-cost interventions such as improved documentation design and heuristic scheduling adjustments. More advanced solutions, including augmented reality task support and biometric monitoring, are positioned as modular, phased extensions contingent on organizational readiness, regulatory approval, and cost-benefit considerations, thereby ensuring scalability and applicability across diverse aviation maintenance contexts (Figure 15).

Table 5. Feasibility and phased implementation of HPI-informed interventions.

Intervention Type	Implementation Phase	Relative Cost	Regulatory Impact	Dependency on Advanced Hardware
Documentation redesign	Immediate	Low	Minimal	None
Scheduling heuristics	Short-term	Low	Minimal	None
AR-based task guidance	Medium-term	Moderate	Moderate	Yes
Biometric fatigue monitoring	Long-term	High	High (privacy/safety)	Yes

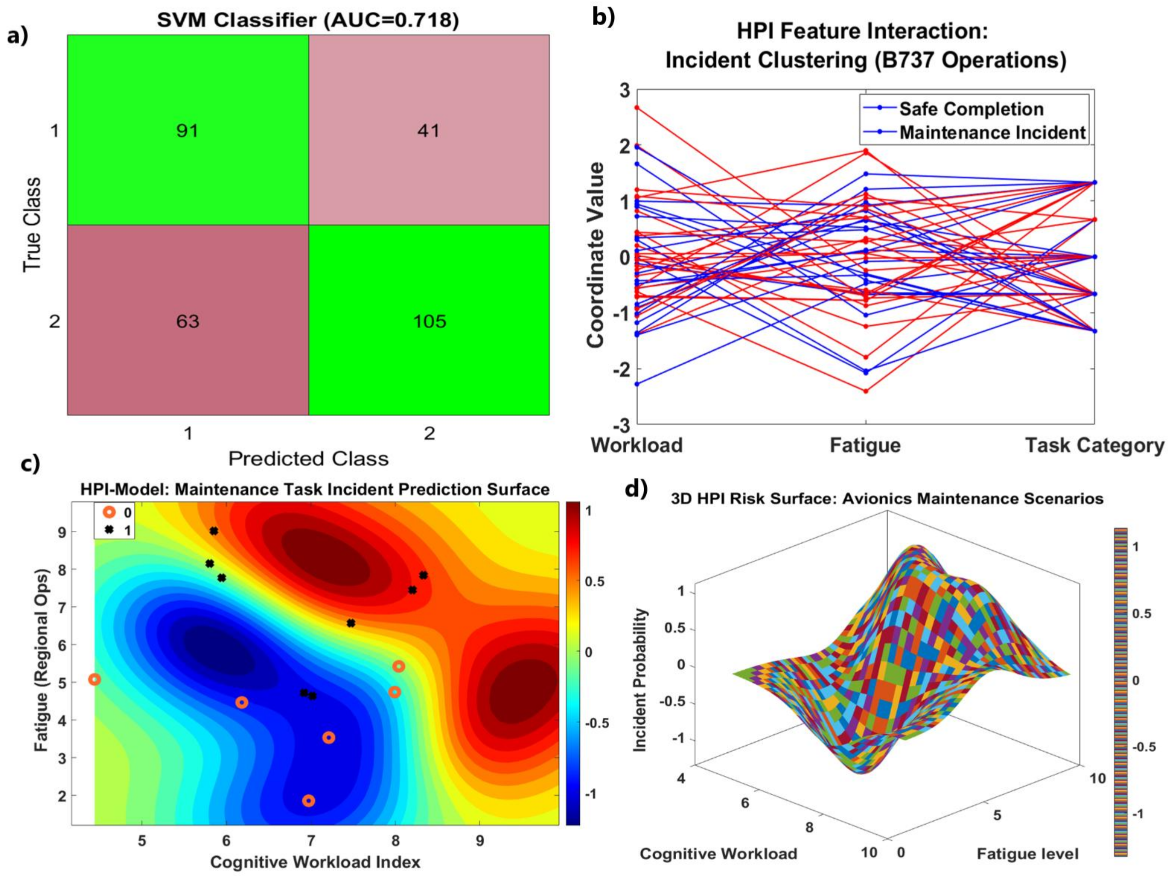

Figure 16. SVM-based HPI model validation in B737 maintenance: (a) classifier performance; (b) integration of human, task, and organizational factors; (c) multi-dimensional risk surfaces and incident clustering; (d) comprehensive validation linking theoretical constructs to empirical outcomes.

As illustrated in Figure 16. (a), the performance evaluation of a Support Vector Machine (SVM) classifier for predicting error versus no-error states in aviation maintenance tasks shows 91 true positives for class 1 and 105 true positives for class 2, highlighted in green to indicate correct classifications, while misclassifications are represented in pink with 41 false positives and 63 false negatives. The model demonstrates moderate accuracy, with an AUC of 0.718, performing somewhat better at predicting class 2 outcomes but exhibiting more false negatives than false positives, suggesting that enhancements via hyperparameter tuning, improved feature selection, or additional variables such as task type, technician experience, or time of day could boost predictive accuracy. As shown in Figure 16. (b), the proposed HPI model integrates organizational, task, and individual human factors to predict incident-level errors in high-risk B737 maintenance operations, fusing qualitative interview insights and quantitative incident data reduced via PCA, with SVM-based classification generating actionable safety recommendations; feature interaction and clustering visualizations confirm the model's ability to identify high-risk operational conditions. As presented in Figure 16. (c), multi-dimensional validation synthesizes organizational, individual, and task-level factors, with parallel coordinate plots revealing clustering of maintenance incidents under high workload and fatigue, and a 2D predictive risk surface illustrating high- and low-risk zones, while overlaid points confirm accurate distinction between safe and incident-prone activities. Finally, as depicted in Figure 16. (d), a comprehensive, multi-layered validation links theoretical human-factor constructs with empirical maintenance data, showing incident clustering where fatigue, workload, and task complexity intersect, with two- and three-dimensional risk surfaces highlighting high-risk operational zones, particularly for avionics tasks, and overlaid real outcomes confirming the model's predictive validity and applicability as an operational decision-support tool.

The combined framework integrates the Human Performance Envelope (HPI-Model) with empirical validation using real-world aviation maintenance data. The polar representation illustrates performance capacity across multidimensional conditions (0°-360°), defined by two boundaries: a Safe Envelope (solid blue) and an Overload Zone (dashed red), as shown in Figure 17. (a). The tri-lobal structure highlights the non-uniform nature of human performance, with peak capacity along three dominant axes (approximately 90°, 210°, and 330°) and reduced capacity in intermediate regions (0°/360°, 120°, and 240°). The region between these boundaries represents high-performance states approaching operational limits without exceeding safety thresholds. This conceptual model is further validated through integration with B737 maintenance data (Table 6), as illustrated in Figure 17. (b), demonstrating how organizational, individual, and task-level factors collectively influence human error mechanisms and safety outcomes. Feature interaction analysis reveals distinct incident pathways driven by workload and fatigue, while 2D and 3D risk surfaces capture the non-linear escalation of incident probability, particularly in avionics-related tasks. Validation markers indicate strong agreement between predicted risk zones and observed maintenance events. Additionally, radar-based performance comparison shows that the HPI-SVM model achieves the most balanced predictive performance, outperforming standard machine learning approaches across Accuracy, Recall (Safety), and F1-score, thereby supporting effective risk mitigation in aviation maintenance environments.

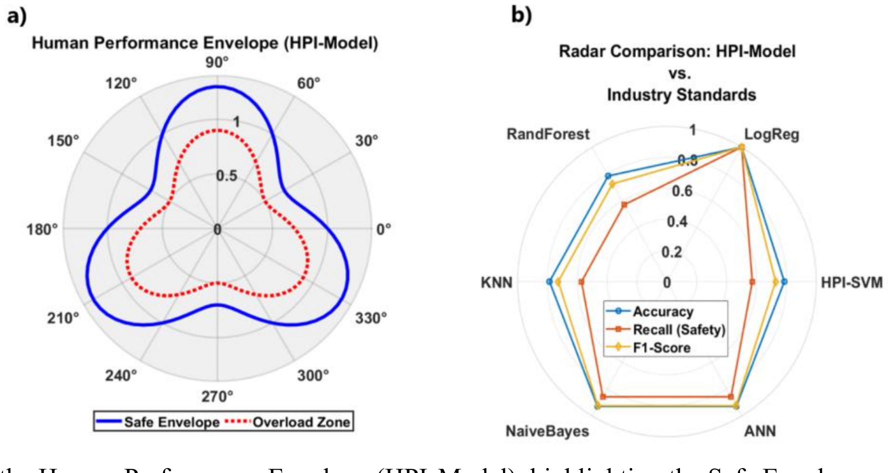

Figure 17. (a) illustrates the Human Performance Envelope (HPI-Model), highlighting the Safe Envelope and Overload Zone across multidimensional performance states. (b) presents a radar-based performance comparison, demonstrating the HPI-SVM model's balanced superiority across Accuracy, Recall (Safety), and F1-score relative to standard machine learning models for B737 maintenance risk prediction.

Table 6. Benchmarking results (B737 maintenance context).

Technique	Accuracy (%)	Recall Safety (%)	F1_Score (%)
HPI-SVM	**92.86**	**100.0**	**94.12**
KNN	71.43	62.5	71.43
Logistic	78.57	62.5	76.92
Random Forest	64.29	50.0	61.54
Grad Boost	71.43	50.0	66.67

5.5. Limitations

The study is intentionally bound to moderate-scale operational datasets, reflecting realistic data availability during early design and planning stages. Rather than constituting a limitation, this constraint aligns with the intended use of the HPI model as a preliminary design-evaluation tool rather than a fully automated operational predictor. Future work will extend the framework to larger, multi-organizational datasets to support later-stage validation and refinement.

6. Conclusion

This study set out to address a fundamental limitation in aviation maintenance human-factors research: the absence of a coherent, predictive framework that embeds human performance constraints directly into engineering task design. Rather than treating maintenance errors as isolated human failures or retrospective safety outcomes, the research reframes error as an emergent condition arising when cumulative task demands exceed the human performance envelope. To operationalize this perspective, the study developed and validated a Human Performance Integration (HPI) model that unifies cognitive workload, fatigue accumulation, task complexity, and procedural clarity within a single analytical structure. By integrating technician-derived qualitative insights with quantitative modeling techniques, including PCA and SVM classification, the framework enables task-level prediction of maintenance error risk under realistic operational constraints. The results demonstrate that error likelihood increases sharply with task complexity and cognitive overload, and that documentation ambiguity remains a significant but addressable contributor to performance degradation. Beyond its empirical findings, the principal contribution of this work lies in its methodological and conceptual integration. The HPI model advances aviation maintenance research by bridging experiential human-factors knowledge with data-driven predictive analytics, moving the field beyond descriptive reporting toward proactive, design-oriented decision support. Importantly, the model is not intended to replace expert judgment or automate safety decisions, but to augment existing Safety Management Systems by highlighting latent risk patterns before errors occur. While the study is constrained by sample size and single-organization data, these limitations reflect the realities of safety-critical aviation environments and do not undermine the framework's validity. Future research will focus on multi-site validation, incorporation of physiological workload measures, and expansion of organizational and cultural variables to further enhance generalizability. In conclusion, this research demonstrates that integrating human performance modeling into maintenance task design is both feasible and necessary. By providing a coherent, predictive framework grounded in operational reality, the Human Performance Integration model offers a practical pathway for improving safety, efficiency, and human reliability in aviation maintenance systems. In conclusion, this research demonstrates that human performance in aviation maintenance can be systematically embedded within engineering task design rather than treated as an external safety constraint. By framing human factors as a design-evaluation problem and providing a predictive, interpretable analytical framework, the study contributes directly to engineering design methodology for safety-critical socio-technical systems. The Human Performance Integration model offers designers a practical means to evaluate, compare, and refine maintenance task designs before operational failure occurs. The HPI Model differs fundamentally from existing frameworks by shifting human factors analysis from retrospective classification to predictive, design-oriented evaluation. This distinction highlights its originality and applied relevance, particularly for regional airlines where proactive task design can significantly reduce error probability and improve operational safety.

References

1. Wild G. A quantitative study of aircraft maintenance accidents in commercial air transport. *Aerospace* 2023; 10(8): 689. https://doi.org/10.3390/aerospace10080689.

2. Kim JW, Jung W. A taxonomy of performance influencing factors for human reliability analysis of emergency tasks. *Journal of Loss Prevention in the Process Industries* 2003; 16(6): 479-495. https://doi.org/10.1016/S0950-4230(03)00075-5.
3. Aven T. A risk science perspective on the discussion concerning Safety I, Safety II and Safety III. *Reliability Engineering & System Safety* 2022; 217: 108077. https://doi.org/10.1016/j.ress.2021.108077.
4. Brugnara RL, De Andrade D, de Souza Fontes R, Leão MS. Safety-II: Building safety capacity and aeronautical decision-making skills to commit better mistakes. *The Aeronautical Journal* 2023; 127(1310): 511-536. https://doi.org/10.1017/aer.2022.74.
5. Chatzi AV, Bates PR, Martin WL. Exploring the association between communication satisfaction and trust in the aviation maintenance environment: An international study. *International Journal of Aerospace Psychology* 2020; 30(3-4): 190-214. https://doi.org/10.1080/24721840.2020.1801347.
6. Chatzi AV, Kourousis KI. Identifying the contribution of communication and trust in aviation maintenance occurrences: A content analysis methodology. *Transportation Research Interdisciplinary Perspectives* 2024; 27: 101220. https://doi.org/10.1016/j.trip.2024.101220.
7. No HW, Cha WC. An integrated framework for implementing Safety-I and Safety-II principles in aviation safety management. *Safety*. 2025;11(2):56. https://doi.org/10.3390/safety11020056.
8. Bafandegan Emroozi V, Kazemi M, Doostparast M, Pooya A. Improving industrial maintenance efficiency: A holistic approach to integrated production and maintenance planning with human error optimization. *Process Integration and Optimization for Sustainability* 2024; 8(2): 539-564. https://doi.org/10.1007/s41660-023-00374-3.
9. Bafandegan Emroozi V, Kazemi M, Pooya A, Doostparast M. Dynamic modeling of human error in industrial maintenance through structural analysis and system dynamics. *Risk Analysis* 2025; 45(6): 1207-1236. https://doi.org/10.1111/risa.17652.
10. Morandini S, Fraboni F, Hall M, Quintana-Amate S, Pietrantoni L. Human factors and emerging needs in aerospace manufacturing planning and scheduling. *Cognition, Technology & Work* 2025; 27(1): 59-77. https://doi.org/10.1007/s10111-024-00785-3.
11. Kabashkin I, Perekrestov V, Pivovar M. AI-driven fault detection and maintenance optimization for aviation technical support systems. *Processes* 2025; 13(3): 666. https://doi.org/10.3390/pr13030666.
12. Marevac E, Kadušić E, Živić N, Hamzić D, Hadžajlić N. Data-driven predictive analytics for dynamic aviation systems: Optimising fleet maintenance and flight operations through machine learning. *Future Internet* 2025; 17(11): 508. https://doi.org/10.3390/fi17110508.
13. Deng Q, Santos BF, Verhagen WJ. A novel decision support system for optimizing aircraft maintenance check schedule and task allocation. *Decision Support Systems* 2021; 146: 113545. https://doi.org/10.1016/j.dss.2021.113545.
14. Gao X, Luo S, Peng D, Kui G, Xie Y, Wu J, Pan J, Zuo X, Chen T. A two-layer optimization method for maintenance task scheduling considering multiple priorities. *Computers & Chemical Engineering* 2024; 184: 108640. https://doi.org/10.1016/j.compchemeng.2024.108640.
15. Consilvio A, Hernández JS, Chen W, Brilakis I, Bartoccini L, Di Gennaro F, van Welie M. Towards a digital twin-based intelligent decision support for road maintenance. *Transportation Research Procedia* 2023; 69: 791-798. https://doi.org/10.1016/j.trpro.2023.02.237.
16. Ren B, Zhou Q, Chen J. Assessing cognitive workloads of assembly workers during multi-task switching. *Scientific Reports* 2023; 13: 16356. https://doi.org/10.1038/s41598-023-43477-0.
17. Das Chakladar DD, Roy PP. Cognitive workload estimation using physiological measures: a review. *Cognitive Neurodynamics* 2024; 18(4): 1445-1465. https://doi.org/10.1007/s11571-023-10051-3.
18. Ronca V, Uflaz E, Turan O, Bantan H, MacKinnon SN, Lommi A, Borghini G. Neurophysiological assessment of an innovative maritime safety system in terms of ship operators' mental workload, stress, and attention in the full mission bridge simulator. *Brain Sciences* 2023; 13(9): 1319. https://doi.org/10.3390/brainsci13091319.
19. Aherne DP, Hunt I, Kourousis KI, Chatzi AV. Human factors considerations for critical maintenance tasks and their effect on the transition to digital documentation: An exploratory expert survey. *Aviation* 2025; 29(1): 48-54. https://doi.org/10.3846/aviation.2025.23131.
20. Dereje M, Tilahun S, Mekuria G, Negesse Y. Exploring the role of human factors and organizational culture in maintenance management practice within large processing industries. *Public Organization Review* 2025; 25: 1-32. https://doi.org/10.1007/s11115-025-00975-1.
21. Bohrey OP, Chatpalliwar AS. Mitigation of human factors in aviation maintenance. *Theoretical Issues in Ergonomics Science* 2024; 25(5): 642-654. https://doi.org/10.1080/1463922X.2024.2321446.

22. Çoban R. Current human factors approaches in aircraft maintenance sector: Transformation of Dirty Dozen into Filthy Fifteen. *Journal of Aviation Research* 2025; 7(2): 103-134. https://doi.org/10.51785/jar.1640698.
23. Özveren Ö, Girkin S, Kanat ÖÖ. Human factors in aviation maintenance: A case-based SWOT analysis. *Human Factors in Aviation and Aerospace* 2025; 2(2): 101-112. https://doi.org/10.26650/hfaa.2025.1745583.
24. Tuncal A, Altıntaş OA. Bibliometric analysis of human factor literature in aviation. *Human Factors in Aviation and Aerospace* 2025; 2(1): 15-26. https://doi.org/10.26650/hfaa.2025.1633491.
25. Zeng JH, Huang JY, Zhong QW, Zhu DW, Dai Y. Risk evaluation for human factors of flight dispatcher based on the hesitant fuzzy TOPSIS-DEMATEL-ISM approach: A case study in Sichuan Airlines. *International Journal of Computational Intelligence Systems* 2024; 17(1): 271. https://doi.org/10.1007/s44196-024-00683-6.
26. Chiu MC, Hsieh MC. Latent human error analysis and efficient improvement strategies by fuzzy TOPSIS in aviation maintenance tasks. *Applied Ergonomics* 2016; 54: 136-147. https://doi.org/10.1016/j.apergo.2015.11.017.
27. Chen W, Huang SP. Human reliability analysis for visual inspection in aviation maintenance by a Bayesian network approach. *Transportation Research Record: Journal of the Transportation Research Board* 2014; 2449(1): 105-113. https://doi.org/10.3141/2449-12.
28. Berthon L, Bernard F, Fleury S, Paquin R, Richir S. Multi-dimensional measurement of mental workload in industrial context: an experiment in the field of helicopter maintenance. *Applied Ergonomics* 2025; 129: 104599. https://doi.org/10.1016/j.apergo.2025.104599.
29. Aktas E, Kagnicioglu CH. Factors affecting safety behaviors of aircraft maintenance technicians: A study on Civil Aviation Industry in Turkey. *Safety Science* 2023; 164: 106146. https://doi.org/10.1016/j.ssci.2023.106146.
30. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 2021; 372: n71. https://doi.org/10.1136/bmj.n71.
31. da Silva JC, Barqueira A, Magalhães LB, Santos LF. Human fatigue in the aircraft maintenance environment. *Safety Science* 2024; 180: 106634. https://doi.org/10.1016/j.ssci.2024.106634.
32. Paris F, Casanova R, Bergeonneau ML, Mestre D. Differences between experts and novices in the use of aircraft maintenance documentation: Evidence from eye tracking. *Applied Sciences* 2024; 14(3): 1251. https://doi.org/10.3390/app14031251.
33. Bohrey OP, Chatpalliwar AS. Modelling of human factors in aviation maintenance using HFACS-ME Human Factors Analysis and Classification System Maintenance Extension and Bayesian Network. *Defence Science Journal* 2024; 74(5): 603-614. https://doi.org/10.14429/dsj.74.19828.
34. Harper R, Bliss T. Identification, evaluation, and causal factor determination of maintenance errors common to major U.S. certificated air carriers. *The Collegiate Aviation Review International* 2024; 42(1). https://doi.org/10.22488/okstate.23.100230.
35. Olaganathan R. Human factors in aviation maintenance: Understanding errors, management, and technological trends. *Global Journal of Engineering and Technology Advances* 2024; 18(2): 92-101. https://doi.org/10.30574/gjeta.2024.18.2.0021.
36. Rodriguez CE. Evaluating the impact of human factors on aircraft maintenance errors: A risk-based analysis framework for business aviation. *World Journal of Advanced Engineering Technology and Sciences* 2024; 13(2): 764-777. https://doi.org/10.30574/wjaets.2024.13.2.0647.
37. Wang L, Gao S, Zhang N. A psychophysiological model based on machine learning algorithms for evaluating commercial airline pilots' mental workload in flight-simulation context. *Human Factors and Ergonomics in Manufacturing & Service Industries* 2025; 35: e70019. https://doi.org/10.1002/hfm.70019.
38. Alomar I, Yatskiv I. Digitalization in aircraft maintenance processes. *Aviation*. 2023;27(2):86–94. https://doi.org/10.3846/aviation.2023.18923.
39. Aust J, Pons D. Assessment of aircraft engine blade inspection performance using attribute agreement analysis. *Safety* 2022; 8(2): 23. https://doi.org/10.3390/safety8020023.
40. Nawaz MO, Zia MQ, Shams TA. Understanding human error in military aviation maintenance: The role of performance shaping factors, cognitive workload and error orientation. *Reliability Engineering & System Safety* 2026; 273: 112337. https://doi.org/10.1016/j.ress.2026.112337.

41. del Olmo M, Domingo R. Intermittent demand forecasting of aerospace rotable parts: A framework for unpredictable flight patterns. *Computers & Industrial Engineering* 2026; 214: 111917. https://doi.org/10.1016/j.cie.2026.111917.
42. Sony M, Antony J, Tortorella G, McDermott O, Gutierrez L. Determining the critical failure factors for Industry 4.0: An exploratory sequential mixed method study. *IEEE Transactions on Engineering Management* 2024; 71: 1862-1876. https://doi.org/10.1109/TEM.2022.3159860.
43. Mörike F, Spiehl HL, Feufel MA. Workarounds in the shadow system: An ethnographic study of requirements for documentation and cooperation in a clinical advisory center. *Human Factors* 2024; 66(3): 636-646. https://doi.org/10.1177/00187208221087013.
44. Nasir A, Mekid S, Sawlan Z, Alsawafy O. Optimized task assignment and predictive maintenance for industrial machines using Markov decision process. *Operational Research* 2025; 25(4): 96. https://doi.org/10.1007/s12351-025-00976-4.
45. Dagal I. Probabilistic fault tree analysis and dynamic redundancy optimization for next-generation avionic flight control systems. *Reliability Engineering & System Safety* 2025; 253: 111841. https://doi.org/10.1016/j.ress.2024.111841.
46. Zafiharimalala H, Robin D, Tricot A. Why aircraft maintenance technicians sometimes do not use their maintenance documents: Towards a new qualitative perspective. *The International Journal of Aviation Psychology* 2014; 24(3): 190-209. https://doi.org/10.1080/10508414.2014.918444.
47. Signal TL, van den Berg MJ, Mulrine HM. Personal and work factors that predict fatigue-related errors in aircraft maintenance engineering. *Aerospace Medicine and Human Performance* 2019; 90(10): 860-866. https://doi.org/10.3357/AMHP.5000.2019.
48. Gerede E. A study of challenges to the success of the safety management system in aircraft maintenance organizations in Turkey. *Safety Science* 2015; 73: 106-116. https://doi.org/10.1016/j.ssci.2014.11.013.
49. Wang Z, Li Y, Zhang H, Chen X. Aircraft maintenance scheduling optimization with the impact of human errors. *Reliability Engineering & System Safety* 2026; 270: 112164. https://doi.org/10.1016/j.ress.2025.112164.
50. Mendes N, Vieira JGV, Mano AP. Risk management in aviation maintenance: A systematic literature review. Safety Science 2022; 153: 105810. https://doi.org/10.1016/j.ssci.2022.105810.

Symbol/Abbreviation	Description
FAA	Federal Aviation Administration
EASA	European Union Aviation Safety Agency
HFACS	Human Factors Analysis and Classification System
SVM	Support Vector Machine
NASA–TLX	NASA Task Load Index (Cognitive Load Assessment Tool)
MRMR	Minimum Redundancy Maximum Relevance (Feature Selection)
AUC	Area Under the Curve (ROC Analysis)
ROC	Receiver Operating Characteristics
PC1, PC2	Principal Components 1 and 2 (for PCA dimensionality reduction)
%	Percent (used to denote frequency or error rates)
AR	Augmented Reality (used for interactive documentation)
SMS	Safety Management System
GSE	Ground Support Equipment
KPI	Key Performance Indicator
TLB	Task Load Balance
CBT	Computer-Based Training
TLX	Task Load Index
PVE	Predicted Value Estimate (used in optimization)
Δ	Delta represents a change or difference in a variable
μ	Mean or Average value
σ	Standard Deviation