© MAR 2026 | IRE Journals | Volume 9 Issue 9 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I9-1715572

Human Factors in Aircraft Maintenance Errors: Structured Review and Probability Modeling

NAZMUL HASAN ANIK CHAWDHURY[1], MAHABUB SULTAN[2]

[1] *Aerospace Engineering, Aviation and Aerospace University, Bangladesh*
[2] *School of Aeronautics, NWPU, China*

Abstract-Aircraft maintenance remains a high-consequence activity because latent deviations may survive task completion and only become visible during later operations, inspections, or abnormal conditions. This paper revises the original structured critical review by integrating a literature-informed predictive modeling demonstration based on fatigue, workload, documentation quality, training quality, handover quality, and experience. The review confirms that maintenance error is rarely the product of a single unsafe act. Instead, it emerges from the interaction of technician-level performance limits, local task conditions, procedural quality, supervisory choices, and broader organizational pressures. Fatigue and workload remain recurrent risk amplifiers, whereas documentation quality, competence development, communication quality, and organizational learning act as protective controls when they are operationally robust. To translate these mechanisms into a forward-looking safety tool, an illustrative logistic regression framework was fitted to a simulation-based dataset of 600 task-level observations derived from literature-consistent directional assumptions. The model achieved an area under the receiver operating characteristic curve of 0.777 on the hold-out set, with the strongest positive coefficients observed for workload and fatigue, while documentation quality emerged as the strongest protective predictor. Scenario analysis further showed a predicted error probability of 92.9% under high-fatigue, high-workload, poor-documentation conditions, compared with 1.4% under low-fatigue, better-documented, better-trained conditions. These results should not be interpreted as external validation of a deployable airline safety model, because the current exercise is simulation-based rather than trained on real organizational event records. Nevertheless, the combined review and model demonstrate how retrospective human-factors knowledge can be operationalized into prospective risk indicators inside maintenance safety management systems.

Index Terms- aircraft maintenance; human factors; maintenance error; aviation safety; HFACS-ME; fatigue; workload; documentation quality; logistic regression; safety management system.

I. INTRODUCTION

Aircraft maintenance is among the most safety-critical functions in aviation because it affects airworthiness, dispatch reliability, and system resilience directly. Unlike many flight-deck errors, maintenance-related deviations often remain latent until they combine with operating conditions, inspection failure, or later task interactions [1]-[3]. This delayed visibility complicates both detection and prevention.

The literature has consistently shown that maintenance error cannot be reduced to personal carelessness alone. Maintenance technicians work under shift pressure, task interruption, uneven environmental conditions, incomplete feedback, and strong dependence on manuals, technical orders, and handovers [1], [3], [4]. In such environments, error becomes a systems phenomenon expressed through individual actions rather than an individual defect isolated from its context.

Classical approaches such as MEDA and HFACS-ME made maintenance events more analyzable by structuring the investigation of unsafe acts, supervisory failures, and latent organizational conditions [5], [7], [13]. In parallel, research on fatigue, training, documentation, safety climate, leadership, and organizational learning showed that the same human-factor mechanisms that explain past maintenance failures can also inform prospective controls [6], [9]-[11], [14], [18], [21].

The central problem, however, is that these two strands are still often separated. Retrospective taxonomies explain what happened after an event, whereas proactive safety-management research addresses what should be monitored before a harmful event occurs. The present revision therefore extends the original review by adding a predictive modeling demonstration anchored in the paper's main mechanisms. The aim is not to claim operational validation from a synthetic dataset, but to show how literature-supported human-factor variables can be transformed into a prospective maintenance-risk indicator.

II. METHODOLOGY

A. Review design and scope

This manuscript retains the structure of a critical review rather than a meta-analysis because the underlying literature is heterogeneous in research design, analytical unit, outcome definition, and measurement logic [15]. The included works range from conceptual syntheses and accident analyses to taxonomy-based studies, survey-based behavioral research, and organizational safety-management investigations. Under such conditions, statistical pooling would imply a degree of comparability that the source literature does not fully support.

The review was guided by three linked questions: which human and organizational factors are most consistently associated with maintenance error, what do current maintenance-error taxonomies explain well and where are their limits, and how can insights from retrospective classification be linked to proactive safety management in aviation maintenance organizations?

The evidence base prioritizes peer-reviewed literature directly relevant to maintenance work, maintenance deviations, safety-management practice, and technician working conditions, while retaining a small number of foundational sources that remain influential in aviation human-factors practice. This adjustment is methodologically more defensible than claiming a journal-only source base while simultaneously depending on foundational books, proceedings, and institutional materials.

B. Analytical logic

The review synthesizes the literature through three clusters: accident and event evidence, maintenance-error taxonomies and explanatory frameworks, and proactive safety-management research. The purpose of this structure is not merely descriptive. It allows the paper to trace whether the same mechanisms recur across different ways of studying maintenance safety.

This clustered reading repeatedly points to the same causal architecture. Fatigue, workload, documentation quality, communication quality, training adequacy, supervisory decisions, and organizational climate are not independent themes; they are connected variables inside a broader maintenance-safety system. That conclusion motivates the predictive extension reported in Section 5.

C. Predictive modeling extension

Because the reviewed literature does not provide a shared open dataset of task-level maintenance events with consistently coded error outcomes, the present modeling exercise is framed explicitly as an illustrative proof of concept rather than empirical validation. A synthetic dataset of 600 task-level observations was therefore used to demonstrate how literature-informed human-factor variables may be translated into a forward-looking maintenance-risk indicator.

The model inputs were fatigue score, workload score, documentation quality, training quality, handover quality, experience in years, and an interaction term between fatigue and workload. The binary outcome variable was maintenance error occurrence, coded as 1 for error and 0 for non-error. The data were split into 75% for training and 25% for testing, and the classifier was estimated using logistic regression with standardized predictors and balanced class weighting.

$$p_i = \frac{1}{1 + \exp(-\eta_i)}$$

$$\begin{aligned} \eta_i = \beta_0 + \beta_1\, Fatigue_i + \beta_2 Workload_i \\ + \beta_3 Documentation_i + \beta_4 Training_i \\ + \beta_5 Handover_i + \beta_6 Experience_i \\ + \beta_7\, (Fatigue_i \times Workload_i) \end{aligned}$$

In theoretical terms, positive coefficients indicate higher estimated error propensity, whereas negative coefficients indicate protective influence. Even so, coefficient signs for correlated predictors should be interpreted cautiously. When workload and fatigue are both strong drivers, interaction terms and secondary variables may reflect shared variance rather than a clean causal effect. This matters because a predictive model that appears numerically persuasive can still be conceptually misleading if coefficient interpretation ignores collinearity and the simulation origin of the data.

Table 1. Predictive model variables and their expected safety meaning.

Variable	Operational meaning	Expected effect
Fatigue score	Higher values indicate greater physical and psychological fatigue.	Increase error probability
Workload score	Higher values indicate greater task demand and pressure.	Increase error probability
Documentation quality	Higher values indicate clearer, more usable, and more consistent technical information.	Reduce error probability
Training quality	Higher values indicate stronger competence development and task preparation.	Reduce error probability
Handover quality	Higher values indicate better communication and shift-transfer discipline.	Reduce error probability
Experience years	Greater experience reflects accumulated task familiarity and error recognition.	Reduce error probability
Fatigue × workload	Interaction term used to test whether combined pressure changes risk beyond main effects.	Context dependent

III. THE HUMAN FACTORS CONTEXT OF AIRCRAFT MAINTENANCE

The earliest influential work in this field established that maintenance safety is shaped by the interaction of generic human limits with highly domain-specific working conditions [1], [4]. Technicians work with incomplete access to components, distributed teams, schedule pressure, variable environments, and delayed feedback about whether their actions ultimately succeed or fail. In contrast to cockpit operations, the person performing the task is often distant in time and space from the eventual operational outcome. That gap weakens error visibility and increases dependence on procedures, inspection, communication, and organizational memory.

Research on unsafe acts in maintenance reinforces this systems perspective. Skill-based errors, mistakes, and violations do occur, but they emerge inside local task conditions and wider organizational influences rather than in a vacuum [2], [3]. The practical implication is that maintenance safety cannot be improved reliably through exhortations to be more careful. It requires structural attention to staffing, scheduling, procedural design, supervision, and the quality of learning from weak signals.

This systems view also explains why retrospective blame cultures often fail. When organizations interpret a maintenance deviation only as an individual lapse, they may close the event without correcting the underlying conditions that made the lapse probable. The same task may then be repeated by another technician under nearly identical constraints. The literature therefore supports a move away from person-centered explanations toward integrated models that connect task-level performance limits with supervisory and organizational conditions.

IV. MAJOR HUMAN-FACTOR MECHANISMS

Table 2 summarizes the principal mechanisms retained from the original review and the main claims associated with each factor.

Table 2. Literature-supported mechanisms and their maintenance-safety implications.

Mechanism	Core finding	Implication

Fatigue and workload	Fatigue, workload, and sleepiness tend to rise together in maintenance settings [9], [10].	Fatigue management and workload control should be treated as active risk controls rather than wellness add-ons.
Documentation quality	Technical information quality affects comprehension and therefore maintenance reliability [6].	Documentation should be managed as an operational safety control, not merely an administrative requirement.
Training and competence	Competency-based task preparation improves adherence and error resistance [11].	Training quality matters most when tied to actual task demands, not only regulatory completion.
Communication and handover	Weak reporting and poor handover allow small deviations to persist across shifts [18].	Structured handover and reporting quality should be embedded in routine maintenance governance.
Safety climate and leadership	Leadership responses shape whether personnel speak up about unsafe conditions [18], [21].	A formal SMS is insufficient if day-to-day supervisory behavior discourages reporting or challenge.

A. *Fatigue and workload*

Fatigue is one of the most consistent mechanisms in aviation maintenance research. It degrades alertness, memory, checking discipline, communication quality, and willingness to challenge uncertain situations under time pressure [9], [10]. Importantly, fatigue is not merely an individual state. It is shaped by rostering, overtime, staffing, shift design, recovery opportunity, and how much degraded performance the organization informally tolerates.

Workload intensifies this problem because maintenance tasks are often time-constrained, interruption-prone, and procedurally dense. High workload does not simply add pressure; it competes with checking, cross-verification, documentation use, and coordination. That is why fatigue and workload should be read as interacting operational conditions rather than isolated risk items.

B. *Documentation and technical information quality*

Maintenance documentation is often misclassified as a paperwork issue. The reviewed literature suggests the opposite. Documentation is a direct human-factors variable because it shapes task interpretation, sequencing, comprehension, and procedural compliance [6]. Poor readability, fragmented instructions, ambiguous wording, outdated revisions, or weak availability at the point of work all create error pathways.

This point is more important than it first appears. Organizations frequently assume that procedural compliance solves risk, but compliance depends on whether the procedure is understandable and usable under real working conditions. A weak procedure can push technicians toward workaround behavior, memory substitution, and informal knowledge transfer, all of which increase the chance of deviation.

C. *Training and competence development*

Training is protective when it builds real competence rather than only checklist completion [11]. In maintenance settings, competence includes not just initial qualification but recurrent learning, task-specific familiarization, cross-shift coordination, error recognition, and the ability to identify degraded system states.

Poor training does not act alone; it amplifies the harm of fatigue, poor documentation, and time pressure because technicians have fewer cognitive and procedural resources available when work becomes demanding. For that reason, training quality should be interpreted as a resilience variable rather than a simple compliance variable.

D. *Communication, reporting, and organizational climate*

Maintenance work is distributed across technicians, shifts, specializations, and locations. Weak handover, incomplete reporting, or suppressive leadership responses allow small deviations to persist until they become operationally significant [18]. The literature indicates that a nominally positive safety culture cannot be judged only by its formal manuals or declared values. It must be judged by whether people actually report unsafe conditions and whether leadership responses reinforce or punish such behavior.

Communication problems are therefore rarely isolated. They often reflect wider issues such as hierarchy, blame, work pressure, understaffing, or fragmented coordination. A narrow corrective action focused only on reminding teams to communicate better may fail if these deeper conditions are left untouched.

V. ILLUSTRATIVE PREDICTIVE MODELING RESULTS

The predictive extension was designed to test whether the mechanisms emphasized by the review could be translated into a usable risk-estimation framework. Because the dataset is simulation-based, the numerical outputs should be interpreted as a methodological demonstration rather than as operational performance evidence for a specific maintenance organization.

The dataset contained 600 observations, with an error prevalence of 28.5%. After a 75/25 train-test split, the hold-out performance yielded an accuracy of 0.700, precision of 0.485, recall of 0.767, F1-score of 0.595, and an area under the receiver operating characteristic curve of 0.777. The confusion matrix comprised 72 true negatives, 35 false positives, 10 false negatives, and 33 true positives. In practical terms, the model was better at identifying elevated-risk cases than at preserving very high precision, which is an intelligible trade-off for a conservative safety-screening tool.

Table 3. Hold-out predictive performance of the illustrative logistic regression model.

Metric	Value
Training observations	450
Test observations	150
Metric	**Value**
Positive-class prevalence	28.5%
Accuracy	0.700
Precision	0.485
Recall	0.767
F1-score	0.595
ROC-AUC	0.777
Confusion matrix	TN=72, FP=35, FN=10, TP=33

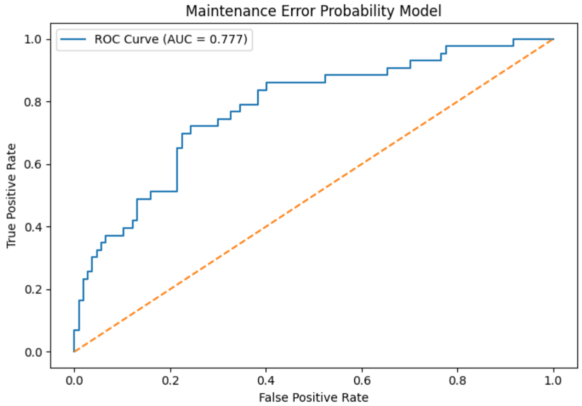

Figure 1. Receiver operating characteristic curve for the illustrative maintenance error probability model (AUC = 0.777).

A. *Coefficient interpretation*

Table 4 shows the fitted coefficients. Workload and fatigue were the two strongest positive predictors, with coefficients of 0.994 and 0.992, respectively. This is consistent with the review's argument that maintenance error probability rises when task demand and human performance degradation are simultaneously present.

Documentation quality was the strongest protective variable (-0.638), followed by experience (-0.276), training quality (-0.254), and handover quality (-0.173). This ranking is important. It suggests that the procedural information environment may be just as operationally influential as more visible staffing and fatigue concerns. In other words, a well-documented task can function as a compensating control, whereas poor documentation can magnify the effect of other adverse conditions.

The fatigue-workload interaction term was negative (-0.423), which should not be read too literally as evidence that combined pressure lowers risk. With correlated predictors and standardized estimation, interaction terms can reflect shared variance and coefficient competition rather than a clean causal signal. The safer interpretation is that the main effects of fatigue and workload dominate the model, while the interaction term requires more disciplined testing with real organizational data.

Table 4. Standardized logistic regression coefficients.

Predictor	Coefficient
workload score	0.994
fatigue score	0.992
handover quality	-0.173
training quality	-0.254
experience years	-0.276
fatigue × workload interaction	-0.423
documentation quality	-0.638

B. Scenario analysis

The scenario analysis provides a more intuitive interpretation of the model. A high-risk case defined by fatigue = 8, workload = 8, documentation quality = 3, training quality = 5, handover quality = 4, and experience = 2 years yielded a predicted error probability of 92.9%. By contrast, a lower-risk case defined by fatigue = 3, workload = 4, documentation quality = 8, training quality = 8, handover quality = 8, and experience = 10 years yielded a predicted error probability of 1.4%. The wide separation between these two scenarios is consistent with the review's central claim that maintenance error is generated by interacting conditions rather than by a single isolated factor.

Table 5. Illustrative scenario predictions.

Scenario	Condition summary	Predicted error
High-risk profile	High fatigue and workload, weak documentation, lower experience, weaker handover	92.9%
Lower-risk profile	Lower fatigue and workload, stronger documentation, stronger training, better handover	1.4%

VI. DISCUSSION

The combined review and predictive demonstration strengthen the original argument in two ways. First, they show that the most stable mechanisms in the maintenance human-factors literature can be converted into a prospective analytical structure. Second, they make clear that not all factors play equivalent roles once translated into a predictive framework. In the present results, workload, fatigue, and documentation quality dominate the directional pattern of risk.

This is not a trivial outcome. Maintenance organizations often treat fatigue as a recognized hazard but documentation quality as a secondary administrative concern. The model suggests that this hierarchy may be misleading. Documentation quality emerged as the strongest protective variable in the fitted results, which is compatible with the review's interpretation of documentation as an operational control that shapes comprehension, sequencing, and procedural adherence rather than as passive text [6].

At the same time, the model should not be oversold. A simulation-based logistic regression can demonstrate feasibility, but it cannot establish external validity, transportability across fleets, or the real frequency of maintenance error under specific organizational conditions. An apparently good AUC does not rescue a model from weak data provenance. For that reason, the present results are best read as a bridge between theory and implementation, not as a finished predictive product.

VII. PRACTICAL IMPLICATIONS FOR MAINTENANCE SAFETY MANAGEMENT

The review and model together suggest five practical priorities for maintenance organizations. First, fatigue management should be integrated with workload design rather than handled as a standalone awareness issue. Second, technical documentation should be governed as a living safety control, with explicit attention to readability, availability, revision discipline, and use at the point of work. Third, training should be evaluated against real task competence and error resistance, not only against completion records. Fourth, handover quality and reporting behavior should be treated as measurable operational indicators. Fifth, predictive dashboards inside safety management systems should be designed to support early warning and supervisory attention rather than to automate blame assignment.

There is also an important governance lesson. Retrospective frameworks such as MEDA and HFACS-ME should not be discarded. They remain useful for structured investigation. The stronger position is to connect them with prospective indicators so that organizations learn not only from what has already failed, but also from conditions that make failure more likely.

VIII. LIMITATIONS AND FUTURE RESEARCH

Three limitations should be emphasized. First, the predictive dataset used here is synthetic and therefore does not represent validated airline or maintenance-organization event records. Second, the current model uses a deliberately small variable set in order to keep the demonstration interpretable. Real maintenance risk also depends on shift timing, staffing adequacy, task complexity, environmental conditions, inspection quality, tooling, and supervisory pressure. Third, the fitted coefficients are sensitive to simulated assumptions and predictor correlation.

Future research should therefore move from proof of concept to empirical validation. The most useful next step would be a real task-level dataset that combines coded maintenance outcomes with human-factor exposures such as fatigue, workload, shift length, documentation use, training history, and handover quality. Once such data exist, multilevel logistic or Bayesian models would likely be more appropriate than a simple single-level classifier because maintenance work is nested within technicians, shifts, teams, fleets, and organizations.

IX. CONCLUSION

This revised paper argues that the most important unresolved issue in aircraft maintenance human factors is no longer whether maintenance error matters, but how established human-factor knowledge can be converted into proactive safety controls. The review confirms that fatigue, workload, documentation quality, training, communication, leadership, and organizational climate are recurring mechanisms across accident analysis, error classification, and safety-management research.

The illustrative predictive model extends that argument by showing that these mechanisms can be operationalized into a maintenance error probability framework. In the present results, workload and fatigue behaved as the strongest positive risk drivers, whereas documentation quality behaved as the strongest protective factor. The model achieved a hold-out ROC-AUC of 0.777 and produced scenario predictions ranging from 1.4% to 92.9%, demonstrating clear conceptual sensitivity to changing maintenance conditions.

The deeper lesson is not that a synthetic logistic model should be deployed directly. It is that aviation maintenance now has a plausible path from retrospective explanation to prospective monitoring. That path becomes scientifically credible only when organizations collect real, high-quality task-level data and use predictive tools as supplements to professional judgment, reporting culture, and structured investigation rather than as replacements for them.

ACKNOWLEDGMENT

The authors acknowledge the use of computational tools, including Python, for the development of the illustrative predictive modeling presented in this study. The authors also express their appreciation for the academic guidance and resources that supported the completion of this work.

REFERENCES

[1] Reason, J. (1997). Managing the Risks of Organizational Accidents. Ashgate.

[2] Hobbs, A., & Williamson, A. (2003). Associations between errors and contributing factors in aircraft maintenance. Human Factors, 45(2), 186-201.

[3] Johnson, W. B., & Watson, J. T. (2010). Human Factors in Aviation Maintenance. In G. Salvendy (Ed.), Handbook of Human Factors and Ergonomics (4th ed., pp. 1133-1156). John Wiley & Sons.

[4] Taylor, J. C., & Christensen, J. M. (1998). Airline maintenance resource management: A guide for the trainer and practitioner. Federal Aviation Administration.

[5] Rankin, W. L., & Allen, J. (1996). Boeing's Maintenance Error Decision Aid (MEDA). In Proceedings of the Human Factors and

Ergonomics Society Annual Meeting (Vol. 40, No. 19, pp. 885-889). SAGE Publications.

[6] Hsia, P. Y. (2007). The effects of technical document readability on user performance. Journal of Technical Writing and Communication, 37(3), 275-294.

[7] Wiegmann, D. A., & Shappell, S. A. (2003). A human error approach to aviation accident analysis: The human factors analysis and classification system. Ashgate.

[8] Krulak, D. C. (2004). Human factors in maintenance: impact on aircraft mishap frequency and severity. Naval Postgraduate School.

[9] Wang, M. J., & Chuang, C. C. (2014). A study of human factors in aviation maintenance. Journal of Air Transport Management, 37, 1-7.

[10] da Silva, J. C., et al. (2024). Human fatigue in the aircraft maintenance environment. Safety Science, 170, 106224.

[11] Walter, D. (2015). Competency-based on-the-job training for aircraft maintenance engineers. Journal of Aviation/Aerospace Education & Research, 24(2), 1.

[12] Patankar, M. S., & Taylor, J. C. (2004). Risk management and error reduction in aviation maintenance. Ashgate.

[13] Illankoon, P., Tretten, P., & Kumar, U. (2019). A prospective study of maintenance deviations using HFACS-ME. International Journal of Industrial Ergonomics, 74, 102857.

[14] McDonald, N., Corrigan, S., Daly, C., & Cromie, S. (2000). Safety management systems and safety culture in aircraft maintenance organisations. Safety Science, 34(1-3), 151-176.

[15] Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, 333-339.

[16] Marais, K., & Robichaud, M. (2012). Analysis of aircraft maintenance-related accidents and incidents. Journal of Aviation Technology and Engineering, 1(2), 4.

[17] Hobbs, A. (2008). An overview of human factors in aviation maintenance. Part-66 Certifying Staff, 1-14.

[18] Aktas, E., & Kagnicioglu, D. (2021). The effect of safety leadership and safety climate on safety behavior in aircraft maintenance. Journal of Air Transport Management, 92, 102021.

[19] Parker, S. K., & Axtell, C. M. (2001). Seeing another viewpoint: Antecedents and outcomes of employee perspective taking. Academy of Management Journal, 44(6), 1085-1100.

[20] Profit, A. D., & Denison, D. R. (2014). Organizational culture and safety in aviation maintenance. Journal of Aviation/Aerospace Education & Research, 23(2), 1.

[21] Zohar, D. (2010). Thirty years of safety climate research: Reflections and future directions. Accident Analysis & Prevention, 42(5), 1517-1522.

[22] Okine, E. A. (2025). Evolution of human factors research in aviation safety. Transportation Research Interdisciplinary Perspectives.

[23] Georgiou, A. M. (2009). The Effect of Human Factors in Aviation Maintenance Safety. 2009 International Symposium on Aviation Psychology.

[24] Marcus, J. H., & Rosekind, M. R. (2017). Fatigue in aviation: a big data approach. NASA Ames Research Center.