

# A Systematic Review of AI-Based PHM Frameworks for Ship Systems and Equipment

Luis F. Mendoza<sup>1,2</sup>, Joan Suarez<sup>1,2</sup>, Edwin Puertas<sup>1</sup>, Edwin Paipa<sup>2</sup> and Juan Carlos Martinez-Santos<sup>1</sup>

<sup>1,2</sup> *Universidad Tecnológica de Bolívar, Cartagena, Bolívar, 130001, Colombia*  
luimendoza@utb.edu.co; jloaiza@utb.edu.co; epuerta@utb.edu.co; jcmartinezs@utb.edu.co

<sup>2</sup> *Corporación de Ciencia y Tecnología para el Desarrollo de la Industria Naval, Marítima y Fluvial, Cartagena, Bolívar, 130009, Colombia*  
lumendoza@cotecmar.com; jmsuarez@cotecmar.com; epaipa@cotecmar.com

## ABSTRACT

Maritime operations require high levels of availability and safety under changing cargo, operational, and environmental conditions. Therefore, anticipating degradation and failures in critical ship systems is an important operational requirement. Although predictive maintenance has advanced rapidly through data-driven approaches and artificial intelligence (AI), the available literature remains fragmented and does not consistently establish the relationships among the ship systems analyzed, the methodological approaches employed, and the conditions required for operational implementation.

This article presents a systematic literature review guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology to examine the development of AI-based predictive maintenance for ship systems from a Prognostics and Health Management (PHM) perspective. Scopus and Web of Science were searched, and predefined screening, eligibility, inclusion, and exclusion criteria were applied.

The findings show that research is strongly concentrated on propulsion systems, classified under Ship Work Breakdown Structure (SWBS) Group 200, and electric-plant systems, classified under SWBS Group 300, with most studies focusing on detection and diagnosis. Prognostic and prescriptive applications remain comparatively less developed. Methodologically, the literature is dominated by machine-learning and deep-learning approaches using operational and condition-monitoring variables associated with different operating regimes.

Overall, maritime PHM exhibits uneven levels of functional and implementation maturity. Important challenges remain in

model comparability, independent validation, and sustained onboard deployment, while opportunities persist for extending PHM to underrepresented ship systems and validating these approaches in real-world fleet operations.

## 1. INTRODUCTION

Asset maintenance is a key business function that supports operational continuity and effective lifecycle management of technical systems. Maintenance policies have evolved from corrective approaches, in which intervention occurs after failure, toward preventive strategies in which inspection and maintenance activities are scheduled in advance to reduce the likelihood of failure and control maintenance-related costs. However, strictly time-based preventive maintenance may result in unnecessary interventions while still allowing failures to occur between scheduled inspections. These limitations motivated the use of condition-monitoring techniques, in which operational signals and condition indicators are continuously or periodically assessed to detect degradation and emerging faults. The resulting condition information provides the basis for condition-based maintenance (CBM), in which maintenance decisions are adjusted according to the observed state of the asset (Ahmad & Kamaruddin, 2012; Jardine, Lin, & Banjevic, 2006).

In the maritime domain, maintenance should also be understood within the broader lifecycle-support framework of the vessel. Integrated Logistics Support (ILS) coordinates technical and management activities related to maintenance, supply support, technical data, support equipment, training, and sustaining engineering throughout the system life cycle (Casadiego Miranda et al., 2021). From this perspective, diagnostic, prognostic, and health-management activities form part of a broader supportability process in which asset-condition information can support not only maintenance execution but also availability management, lifecycle support, and logistics decision-making.

Luis Mendoza et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.  
<https://doi.org/10.36001/IJPHM.2026.v17i2.4764>

The increasing availability of operational data and digital technologies has facilitated the implementation of CBM and enabled more advanced data-driven maintenance strategies. In particular, the growing volume and accessibility of condition-monitoring data have supported the development of models capable of estimating future asset condition and failure behavior, thereby extending condition-based approaches toward predictive maintenance (PdM) (Elmdoost-Gashti, Shafiee, & Bozorgi-Amiri, 2023). According to EN 13306, predictive maintenance is based on the assessment of asset condition and the forecasting of degradation in order to support maintenance decisions before functional failure occurs (*EN 13306:2017 Maintenance - Maintenance terminology*, 2017; Oldfield, McMonies, & Haig, 2014).

This broader PHM paradigm can be implemented through different reference models and architectures rather than through a single universally established workflow. Different standards and reference models organize PHM- and condition-based maintenance-related functions with different levels of scope and granularity. For example, the ISO 13374 series defines an open reference architecture for condition monitoring and diagnostics, covering the processing, communication, and presentation of condition information (*Condition Monitoring and Diagnostics of Machines: Data Processing, Communication and Presentation - Part 1: General Guidelines*, 2003). The Open System Architecture for Condition-Based Maintenance (OSA-CBM) operationalizes this functional view through interoperable processing blocks ranging from data acquisition and manipulation to state detection, health assessment, prognostic assessment, and advisory generation (MIMOSA, 2010). In parallel, IEEE Std 1856-2017 provides a normative framework for classifying PHM capabilities and planning their implementation throughout the life cycle of electronic systems (*IEEE Standard Framework for Prognostics and Health Management of Electronic Systems*, 2017). Therefore, PHM is treated in this study as an overarching paradigm encompassing multiple reference models, architectures, and implementations rather than as a unique prescribed workflow (Ellefsen, Æsøy, Ushakov, & Zhang, 2019; Soualhi, Lamraoui, Elyousfi, & Razik, 2022; Velasco-Gallego, Navas De Maya, Matutano Molina, Lazakis, & Cubo Mateo, 2023).

Despite these advances, implementing PHM under real shipboard operating conditions remains challenging. Major barriers include limited data availability and variable data quality (Noot, Birmele, & Rey, 2024; Lim, Oh, & Park, 2025), changes in operating regimes and environmental conditions, difficulties in defining representative degradation indicators, and limited model robustness across operating contexts (Elmdoost-Gashti et al., 2023; Ellefsen et al., 2019). These factors can substantially affect the practical effectiveness and transferability of CBM, PdM, and PHM solutions (Huang et al., 2024). Consequently, increasing attention has been given to methods capable of learning from limited, imbalanced,

and heterogeneous datasets (Velasco-Gallego et al., 2023; Kalafatelis et al., 2025; Benarfa, Dahmane, Brik, & Yagoubi, 2025).

Artificial intelligence has emerged as a promising means of addressing several of these PHM implementation challenges because it enables operational and condition-monitoring data to be used for detection, diagnosis, and prediction, particularly in complex systems for which purely physics-based modeling may be insufficient or prohibitively costly (Nguyen, Medjaher, & Tran, 2023; Ucar, Karakose, & Kırımca, 2024). In the maritime sector, the adoption of AI-based methods has accompanied the broader digital transformation of the industry and the growing interest in predictive and proactive maintenance strategies (Liang, Knutsen, Vanem, Æsøy, & Zhang, 2024). Accordingly, this article examines how artificial intelligence has been used to support PHM applications in shipboard systems, with particular attention to the systems investigated, the PHM functions addressed, the AI methods employed, and the reported data, validation, and deployment practices.

By examining these dimensions together, this review provides an integrated perspective on the current state of AI-based PHM research in shipboard systems. This approach enables the identification of research trends across ship systems, the assessment of PHM functional coverage, and the examination of how AI methods relate to the available data and reported validation practices. Furthermore, it helps identify gaps between research activity and the evidence supporting practical implementation, providing insights into current limitations and opportunities for future research.

To achieve this objective, the review addresses the following research questions:

1. In which ship systems and equipment have PHM approaches and artificial intelligence models been applied, according to the literature?
2. What types of models and techniques are used in detection, diagnosis, prognosis, and prescription tasks?
3. What data management, validation, and deployment practices are reported in the implementation of PHM models on vessels?

This review is intended for researchers and practitioners in maintenance, reliability, and naval engineering. It provides a structured overview of shipboard PHM applications, the data-driven methods reported in the literature, and the main considerations affecting their operational implementation. The remainder of the article is organized as follows. Section 2 reviews the state of the art and establishes the research gap. Section 3 describes the systematic-review methodology. Section 4 presents the findings according to the three research questions: shipboard systems and equipment, PHM functions and AI methods, and implementation practices. Section 5 dis-

cusses the implications of the findings, and Section 6 presents the conclusions, limitations, and directions for future research.

## 2. STATE OF THE ART

Condition-Based Maintenance (CBM), Predictive Maintenance (PdM), and Prognostics and Health Management (PHM) are closely related concepts, but they should not be treated as interchangeable. CBM is a maintenance strategy in which intervention is triggered or adjusted according to evidence of the asset's current condition rather than solely according to predefined time or usage intervals. Its analytical support may therefore include condition monitoring, anomaly detection, and fault diagnosis without necessarily predicting how the asset condition will evolve in the future. PdM introduces an explicitly forward-looking capability by using observed condition and degradation information to estimate future asset health and the time at which maintenance may become necessary. Depending on the application, this prediction may be expressed as a degradation trajectory, failure probability, time-to-failure, future health state, or Remaining Useful Life (RUL) (Ahmad & Kamaruddin, 2012; Jardine et al., 2006; *EN 13306:2017 Maintenance - Maintenance terminology*, 2017).

Although PHM is commonly associated with reliability-centered maintenance contexts, PHM and reliability, availability, and maintainability (RAM) analyses address different methodological purposes. In this review, PHM is understood as the set of data-driven, model-based, or hybrid methods used to detect, diagnose, prognose, prescribe, or generate health-related information from monitored assets (Jardine et al., 2006; Lee et al., 2014; *Condition Monitoring and Diagnostics of Machines: Data Processing, Communication and Presentation - Part 1: General Guidelines*, 2003; International Organization for Standardization, 2015). Accordingly, the main focus of this article is on machine learning (ML) and deep learning (DL) models applied to PHM-related prediction, classification, and generative tasks.

By contrast, RAM analysis belongs to the broader reliability and maintainability framework and is mainly oriented toward estimating or predicting system-level performance metrics, such as reliability, availability, maintainability, mean time between failures (MTBF), mean time to repair (MTTR), downtime, failure rates, and other maintenance-performance metrics (International Electrotechnical Commission, 2014). Criticality analysis further supports this framework by identifying which assets, failure modes, or functions have the highest operational consequence (International Electrotechnical Commission, 2018).

PHM is broader than either maintenance strategy and is treated in this review as an engineering paradigm that integrates information acquisition and condition assessment with detection, diagnosis, prognosis, and health-management or advisory functions that support maintenance and operational decisions.

Operationally, condition monitoring or anomaly detection alone is therefore interpreted as a PHM-enabling Detection capability rather than as evidence of a complete PHM implementation. The transition toward PdM becomes explicit when future degradation, failure timing, or RUL is estimated, whereas stronger PHM integration is evidenced when condition, diagnostic, and/or prognostic information is connected to health assessment, maintenance prioritization, operational constraints, or recommended actions. RUL is consequently treated as a common prognostic output, but not as a mandatory requirement for every PHM implementation.

Several authors agree that the scarcity of run-to-failure data and the limited availability of fault labels constitute central barriers to implementing prognosis under real operating conditions (Ellefsen et al., 2019). Similarly, insufficient formalization of data management and preprocessing practices has been reported, particularly regarding missing-value imputation and identification of operating states, which can introduce bias and limit external validity (Velasco-Gallego et al., 2023; Gharib & Kovács, 2023). In addition, model opacity and heterogeneous reporting practices may restrict the traceability required for operational adoption, especially when analytical decisions must be justified under safety-critical maritime conditions (Gharib & Kovács, 2023; Kalafatelis et al., 2025). Consistent with these limitations, previous evidence also indicates a marked concentration on detection and diagnosis, with comparatively lower coverage of advanced prognostic functions and maintenance-oriented decision support (Velasco-Gallego et al., 2023).

Previous maritime reviews have addressed complementary aspects of predictive maintenance and PHM, ranging from specific assets and PHM functions to data-driven methods, validation practices, and regulatory considerations. To clarify the positioning of the present study, Table 1 compares representative reviews according to their scope, search coverage, review size, PHM functions, AI methods, implementation considerations, and main analytical distinction.

The comparison in Table 1 shows that previous reviews provide complementary perspectives rather than addressing an identical analytical scope. P. Zhang et al. (2022) provide a broad process-oriented view of marine PHM from condition monitoring to maintenance decision-making, whereas Gharib and Kovács (2023) focus specifically on marine diesel-engine PHM. Velasco-Gallego et al. (2023) provide a systematic analysis of data-driven preprocessing, diagnosis, prognosis, and validation practices. Liang et al. (2024) extend the discussion toward classification-society requirements and technology assurance, while Kalafatelis et al. (2025) organize data-driven predictive-maintenance approaches according to maritime system components and emerging distributed-learning methods.

Where previous reviews reported both an explicit analytical corpus and functional frequencies, a quantitative comparison

Table 1. Comparison of representative maritime PHM and predictive-maintenance reviews.

| Study                         | Scope / systems                                                                       | Search coverage                                                                        | Review size                                       | PHM functions                                            | AI methods                                         | Implementation                                                              | Main distinction                                                                                                  |
|-------------------------------|---------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|---------------------------------------------------|----------------------------------------------------------|----------------------------------------------------|-----------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------|
| Zhang et al. (2022)           | Broad marine systems and equipment                                                    | Web of Science; mainly 2010–2020                                                       | >300 articles                                     | Monitoring, diagnosis, prognosis, maintenance decision   | ML-based diagnosis, health indicators, RUL methods | Data acquisition, preprocessing, storage, ship–shore transmission           | Process-oriented view of the complete marine PHM chain                                                            |
| Gharib and Kovács (2023)      | Marine diesel engines                                                                 | No explicit systematic search period reported                                          | Not explicitly reported                           | Monitoring, diagnosis, prognosis, maintenance decision   | Expert systems, statistical methods, ML, ANN, SVM  | Real-time monitoring and decision-support aspects                           | Asset-specific review focused on marine diesel-engine PHM                                                         |
| Velasco-Gallego et al. (2023) | Marine machinery and systems                                                          | IEEE Xplore, ACM DL, and Scopus; 2016–2022                                             | 88 primary studies                                | Detection, diagnosis, prognosis                          | ML, DL, preprocessing and validation methods       | Strong emphasis on preprocessing and validation                             | Systematic synthesis of data-driven diagnosis and prognosis                                                       |
| Liang et al. (2024)           | Broad maritime equipment                                                              | IEEE Xplore, Elsevier, Taylor & Francis, Wiley, and MDPI; 2017–2023                    | >200 articles                                     | Monitoring, diagnosis, prognosis, maintenance decision   | Physics-based, data-driven, ML and DL methods      | Classification rules, assurance, data management, onboard/shore processing  | Classification-society and technology-assurance perspective                                                       |
| Kalafatis et al. (2025)       | Propulsion, auxiliary, electrical, and hull systems                                   | No explicit systematic search period reported                                          | Not explicitly reported                           | Mainly detection, diagnosis, and prognosis               | ML, DL, federated and transfer learning            | Scalability, communication, privacy, and distributed learning               | Component-based review emphasizing distributed AI approaches                                                      |
| <b>This review</b>            | <b>Hull, propulsion, electric, command, and auxiliary systems mapped through SWBS</b> | <b>Scopus and Web of Science; no date restriction; eligible studies span 1993–2025</b> | <b>83 primary studies + 11 contextual reviews</b> | <b>Detection, Diagnosis, Prognosis, and Prescription</b> | <b>Comparative mapping of AI model families</b>    | <b>Data management, validation, deployment, and operational integration</b> | <b>Cross-dimensional synthesis linking SWBS systems, PHM functions, AI methods, and implementation conditions</b> |

was also possible. Velasco-Gallego et al. (2023), for example, analyzed 88 primary studies published between 2016 and 2022 and reported that approximately 3% addressed fault prognosis. In the present 83-study analytical sample, 21 studies (25.3%) include an explicit prognostic component, comprising 15 studies classified as Prognosis, three as Diagnosis + Prognosis, two as Detection + Prognosis, and one as Diagnosis + Prognosis + Prescription.

Nevertheless, this difference should not be interpreted directly as evidence of temporal progression in maritime PHM. The two reviews differ in database coverage, eligibility criteria, system scope, and operational definitions used to classify PHM functions. Broader reviews such as Zhang et al. (2022), which retained more than 300 articles mainly from 2010–2020, and Liang et al. (2024), which reviewed more than 200 articles focused on 2017–2023, provide broader review corpora but do not employ the same study-level functional normalization adopted in the present review. Accordingly, these comparisons are used to contextualize differences in literature coverage and PHM functional emphasis rather than as a meta-analytic comparison of equivalent populations.

These contributions establish important foundations for maritime PHM; however, among the representative reviews compared in Table 1, no single study simultaneously structures the evidence according to a formal ship-system decomposition, normalizes the reported PHM functions for quantitative comparison, maps the AI model families used for those func-

tions, and systematically examines data-management, validation, and deployment practices. The present review addresses this intersection by combining SWBS-based system mapping with the review-specific Detection–Diagnosis–Prognosis–Prescription normalization and an implementation-oriented analysis covering data practices, validation, and operational integration.

Accordingly, the contribution of this review does not lie in proposing a new systematic-review procedure, since evidence identification, screening, and selection follow established PRISMA 2020 guidelines. Instead, its contribution lies in the cross-dimensional synthesis of the AI–PHM–maritime literature through four connected analytical dimensions: (i) ship systems and equipment using the SWBS structure, (ii) normalized PHM functions, (iii) AI model families, and (iv) implementation practices encompassing data management, validation, and deployment. This integrated view enables system-level concentration, PHM functional coverage, methodological choices, evidential strength, and operational implementation gaps to be examined within the same evidence base.

More specifically, the integrated framework distinguishes three issues that would otherwise be conflated: the number of studies addressing a ship-system domain, the PHM functions addressed within that domain, and the strength of the reported implementation evidence. For example, a higher share of downstream PHM functions does not necessarily imply inde-

pendent validation or sustained deployment. Thus, the resulting cross-dimensional comparisons are intended to identify mismatches between research activity, functional scope, and evidential maturity, rather than to rank equipment groups by publication count alone.

### 3. METHODOLOGY

This systematic review was conducted in accordance with the PRISMA 2020 guidelines (Page et al., 2021) to identify, select, and synthesize evidence on the application of artificial intelligence (AI) and machine learning models in PHM-oriented applications on ships.

Following the PRISMA 2020 framework, the review process is described in terms of: (i) information sources, (ii) the database-specific search strategy based on predefined keyword groups and Boolean operators, and (iii) study screening, eligibility assessment, data extraction, and coding.

#### 3.1. Sources of Information

We used Scopus and Web of Science (WoS) to identify relevant scientific literature. We selected these databases for their broad multidisciplinary coverage of peer-reviewed publications and their strict indexing criteria, which prioritize high-impact journals and consolidated editorial processes.

Both databases are widely used in systematic reviews and bibliometric studies, since they provide structured metadata, detailed citation information, and tools that facilitate traceability and replicability of the literature search and selection processes (Mongeon & Paul-Hus, 2016; J. Li, Goerlandt, & Reniers, 2021). In particular, combining Scopus and WoS enables broader coverage of relevant literature and reduces biases arising from reliance on a single information source.

In the context of this research, these databases enabled the identification of studies on the application of artificial intelligence and machine learning within PHM-oriented condition monitoring, diagnostics, prognostics, and maintenance applications for maritime systems, including works focused on propulsion plants, condition monitoring, and predictive maintenance on ships.

Coverage of PHM-specific publication outlets was also considered when interpreting the scope of the search. The International Journal of Prognostics and Health Management (IJPHM), the flagship journal of the PHM Society, is indexed in both Scopus and the Web of Science Emerging Sources Citation Index (ESCI). Therefore, journal articles published in this outlet fall within the coverage of the selected databases. However, PHM Society conference proceedings constitute a separate publication channel and are not necessarily indexed comprehensively or uniformly by Scopus and Web of Science. Because no dedicated search of the PHM Society proceedings repository was conducted, conference-only PHM contributions

not indexed in the selected databases may be underrepresented in the present corpus. This is acknowledged as a coverage limitation rather than evidence of complete saturation of the PHM literature.

#### 3.2. Search Strategy

The search strategy was derived from the research questions and organized around four conceptual dimensions: (i) PHM functions and tasks, including fault detection, diagnosis, prognosis, condition monitoring, and health management; (ii) artificial intelligence and machine-learning approaches; (iii) maintenance strategies associated with the operational use of these models; and (iv) the naval and maritime application domain. Within each dimension, semantically related terms were combined using the Boolean operator OR, while the four dimensions were connected using AND to retrieve studies located at the intersection of PHM, AI, maintenance, and maritime applications.

The searches were executed independently in Scopus and the Web of Science Core Collection in October 2025. In Scopus, the query was applied to title, abstract, and keyword fields using `TITLE-ABS-KEY ( . . . )`, whereas in Web of Science the equivalent topic-field syntax `TS= ( . . . )` was used. No publication-date restriction was applied in either database. The searches retrieved 281 records from Scopus and 119 from Web of Science, yielding 400 records before duplicate removal. Table 2 reports the complete database-specific search strings, field syntax, and retrieval counts, thereby documenting the reproducible implementation of the search strategy in each database.

No publication-date filter was applied; therefore, the 1993–2025 publication span reported in the Results reflects the years of the eligible studies identified through screening rather than a predefined temporal window. Only English-language documents were eligible, as specified in Table 3.

The search was limited to records indexed in Scopus and Web of Science. No dedicated grey-literature search, backward or forward citation snowballing, or formal search-sensitivity or saturation analysis was performed. These decisions provided a reproducible search protocol based on two multidisciplinary bibliographic databases, while their potential effect on literature coverage is explicitly acknowledged in the limitations of the review.

#### 3.3. Screening and Sample Selection

Study selection followed sequential screening and eligibility stages. First, using the defined search query, the retrieved records were processed with the Bibliometrix R package (Aria & Cuccurullo, 2017) to generate a master matrix in Excel, supporting the extraction and standardization of metadata (title, abstract, authors, year, DOI, and document type), as well as

Table 2. Database-specific search strategy and retrieval results.

| Database       | Field syntax  | Exact search string                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Records |
|----------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| Scopus         | TITLE-ABS-KEY | ("Prognosis" OR "fault" OR "fault detection" OR "fault diagnosis" OR "condition monitoring" OR "anomaly detection" OR "failure detection" OR "malfunction detection" OR "predictive diagnostics" OR "PHM" OR "prognostic" OR "health management") AND ("artificial intelligence" OR "machine learning" OR "ML" OR "deep" OR "deep learning" OR "DL" OR "neural networks" OR "ANN" OR "physics-informed machine learning" OR "PINNs" OR "Physics-Informed Neural Networks" OR "hybrid physics-AI models" OR "physics-guided neural networks") AND (maintenance OR "prescriptive maintenance" OR "predictive maintenance" OR "preventive maintenance" OR "resilient maintenance" OR "smart maintenance" OR "maintenance strategies") AND ("naval" OR "marine" OR "shipboard" OR "maritime" OR "military vessel" OR "navy" OR "naval vessel" OR "ship" OR "warship" OR "marine vessel") | 281     |
| Web of Science | TS            | ("Prognosis" OR "fault" OR "fault detection" OR "fault diagnosis" OR "condition monitoring" OR "anomaly detection" OR "failure detection" OR "malfunction detection" OR "predictive diagnostics" OR "PHM" OR "prognostic" OR "health management") AND ("artificial intelligence" OR "machine learning" OR "ML" OR "deep" OR "deep learning" OR "DL" OR "neural networks" OR "ANN" OR "physics-informed machine learning" OR "PINNs" OR "Physics-Informed Neural Networks" OR "hybrid physics-AI models" OR "physics-guided neural networks") AND (maintenance OR "prescriptive maintenance" OR "predictive maintenance" OR "preventive maintenance" OR "resilient maintenance" OR "smart maintenance" OR "maintenance strategies") AND ("naval" OR "marine" OR "shipboard" OR "maritime" OR "military vessel" OR "navy" OR "naval vessel" OR "ship" OR "warship" OR "marine vessel") | 119     |

duplicate identification and removal.

With the cleaned database, the initial screening of studies based on titles, abstracts, and keywords was conducted independently by two reviewers using the predefined operational selection criteria. These criteria are summarized in Table 3, which presents the filtering framework used in terms of thematic relevance, document type, temporal coverage, and language (Hiebl, 2021).

Table 3. Selection criteria applied during screening and eligibility.

| Criterion         | Description                                                                                                                                                                                                                                                                                                                                                                         |
|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Relevance         | The selected articles must meet the following research points. <ul style="list-style-type: none"> <li>The topics must be established within the naval or maritime sector;</li> <li>Studies must address systems or functions associated with the selected SWBS groups: Hull Structure, Propulsion Plant, Electric Plant, Command and Surveillance, or Auxiliary Systems.</li> </ul> |
| Document type     | Articles, book chapters, and conference papers addressing the application of AI-based approaches to PHM-related problems.                                                                                                                                                                                                                                                           |
| Temporal coverage | No publication-date restriction was applied.                                                                                                                                                                                                                                                                                                                                        |
| Language          | Only documents written in English were reviewed.                                                                                                                                                                                                                                                                                                                                    |

The agreement between the authors at this screening stage was assessed using Cohen's kappa, resulting in a value of  $\kappa = 0.72$ . According to the interpretation criteria proposed by Landis and Koch (Landis & Koch, 1977), this value corresponds to substantial agreement. Disagreements between the two reviewers were discussed and resolved by consensus before proceeding to the eligibility assessment. The interpretation ranges used for Cohen's kappa are summarized in Table 4.

Subsequently, potentially eligible studies were independently assessed in full text by both reviewers to verify compliance with the inclusion and exclusion criteria. Discrepancies at this stage were likewise discussed and resolved by consensus before the final eligibility decision. Cohen's kappa was not recalculated at this stage; therefore, the reported  $\kappa = 0.72$  refers specifically to the initial title-, abstract-, and keyword-level screening.

Table 4. Interpretation of Cohen's kappa according to (Landis & Koch, 1977).

| Cohen's $\kappa$ | Strength of agreement |
|------------------|-----------------------|
| $\leq 0.00$      | Poor                  |
| 0.01–0.20        | Slight                |
| 0.21–0.40        | Fair                  |
| 0.41–0.60        | Moderate              |
| 0.61–0.80        | Substantial           |
| 0.81–1.00        | Almost perfect        |

Following full-text assessment, data from each included study were extracted and recorded in the master matrix. The extracted variables included the study objective, application context, shipboard system or equipment, reported methods, data type and source, PHM-related function, evaluation practices, and implementation characteristics.

For comparative purposes, PHM-related functions were classified according to the review-specific functional normalization adopted in this study: Detection, Diagnosis, Prognosis, and Prescription. Detection encompasses methods aimed at identifying deviations, anomalous conditions, or the occurrence of a fault without necessarily determining its specific cause or failure mode. Diagnosis refers to fault identification, isolation, or classification. Prognosis refers to the estimation of future degradation, health evolution, failure horizon, or Remaining Useful Life (RUL). Prescription refers to the explicit recommendation or prioritization of maintenance or opera-

tional actions based on the estimated asset condition. When a study explicitly addressed more than one PHM function, combined categories such as Detection + Diagnosis or Diagnosis + Prognosis were retained.

An additional boundary criterion was applied to distinguish asset-health PHM from broader operational monitoring. A study was considered core PHM evidence when the monitored condition, anomaly, or fault was explicitly related to the health, degradation, availability, reliability, or maintenance of a shipboard asset or technical system. Monitoring-oriented studies that did not fully satisfy this condition were retained as boundary cases when they fell within the selected SWBS domains and remained relevant to the methodological scope of the review. Studies in which anomalies were defined primarily at the level of vessel behavior, navigation, or situational awareness were explicitly identified as operational-monitoring boundary cases rather than conventional asset-health PHM applications. These cases were retained to characterize the interface between AI-based PHM and broader maritime monitoring but were interpreted separately during the synthesis.

For transparency, boundary cases remained represented in the SWBS mapping and in Detection counts when their analytical task satisfied the corresponding functional definition. However, they were not interpreted as evidence of prognostic capability, RUL estimation, or complete PHM implementation.

### 3.4. Verification, Validation, and Evidence Assessment

To assess the evidential strength of the included studies, verification and validation (V&V) were treated as distinct concepts. Following the general V&V distinction established in IEEE Std 1012, verification concerns evidence that an implemented model, software component, or analytical process conforms to its specified methodology or requirements, whereas validation concerns evidence that the resulting model or system is suitable for its intended use under representative conditions (IEEE, 2024).

For the purposes of this review, three related but distinct evaluation categories were used. First, *verification evidence* referred to explicitly reported checks of methodological or implementation correctness, including consistency or reproducibility checks and controlled comparison with known or simulated ground truth when these were used to confirm that the implemented analytical process behaved as specified.

Second, *internal model evaluation* included random holdout, train–test partitioning, cross-validation, k-fold procedures, and equivalent performance assessments conducted within the development dataset or a closely related data distribution. These procedures provide evidence of predictive performance but were not considered sufficient by themselves to demonstrate operational validation.

Third, *validation* was reserved for evidence addressing the intended-use context through meaningful independence from model development. This included chronological or campaign-based holdout, testing on previously unseen assets or vessels, external datasets, independently documented faults or maintenance events, prospective evaluation, or sustained onboard or operational testing.

In addition to this V&V distinction, evidential strength was characterized according to data provenance, coverage of operational regimes, traceability to confirmed faults or maintenance events, and deployment status. These dimensions were used for comparative synthesis and assessment of reporting completeness rather than as exclusion criteria or as a single numerical quality score.

No formal study-quality or risk-of-bias scoring instrument was applied to the included primary studies. This decision reflects the heterogeneous nature of the evidence base, which comprises simulation studies, controlled laboratory and test-bench experiments, retrospective analyses of shipboard data, and operational implementations addressing different PHM functions and performance outcomes. Applying a single retrospective scoring instrument across these heterogeneous designs could imply a degree of methodological comparability that is not supported by the corpus. Instead, study limitations were systematically assessed through the predefined dimensions of data provenance, V&V design, temporal or operational independence, fault- and maintenance-event traceability, reporting completeness, and deployment context. These dimensions were used to qualify the interpretation of the evidence rather than to exclude studies or generate an aggregate quality score.

### 3.5. Coding and Classification Rules

System and equipment classification was conducted using the high-level system groups defined by the Expanded Ship Work Breakdown Structure (ESWBS), a product-oriented decomposition scheme that organizes the ship into standardized system elements to support consistent identification and reporting throughout the life cycle (*Work Breakdown Structures for Defense Materiel Items*, 2022).

Each primary study was assigned to an SWBS group according to the shipboard asset, subsystem, or technical function that constituted the principal object of condition assessment, diagnosis, prognosis, or maintenance decision support. Classification was therefore based on the physical or functional location of the monitored asset rather than on the type of data, algorithm, or maintenance technique reported. Propulsion engines, shafts, propellers, and propulsion-support equipment were mapped to SWBS 200; electrical-generation, distribution, energy-storage, and associated power-system assets to SWBS 300; structural and hull-condition applications to SWBS 100; command, communication, monitoring, and surveillance functions to SWBS 400; and auxiliary service equipment to SWBS

500.

When a study evaluated primary assets belonging to more than one SWBS group and reported results for both domains, a combined SWBS classification was retained rather than forcing the study into a single group. For example, studies jointly evaluating propulsion and electric-generation assets were coded as SWBS 200 & SWBS 300. By contrast, when variables from other ship systems were used only as contextual inputs, the study remained assigned to the SWBS group of the primary monitored asset.

For SWBS 400, the PHM/monitoring boundary rule defined in Section 3.3 was applied. Operational-surveillance studies were retained in the system-level mapping when they fell within the selected command, communication, monitoring, navigation, or surveillance domain but were explicitly distinguished from conventional asset-health PHM when interpreting functional coverage.

PHM functions were coded according to the analytical output actually demonstrated in each primary study rather than according to terminology used in the title, abstract, or authors' general description. A study was coded as Detection when its output indicated whether a condition was normal, abnormal, or fault-related without identifying a specific failure mode. Diagnosis was assigned when the method identified, isolated, classified, or attributed the anomaly to a specific fault or health state. Prognosis required an explicitly forward-looking output, such as degradation evolution, future health state, failure probability over a specified horizon, time-to-failure, or RUL. Prescription was assigned only when analytical outputs were translated into an explicit recommendation, prioritization, maintenance action, intervention schedule, or decision-support rule.

When a study explicitly implemented more than one PHM function, the corresponding combined category was retained. A downstream function did not automatically imply that all upstream functions were present. For example, a study reporting RUL estimation was coded as Prognosis only unless it also explicitly implemented and evaluated a separate Detection or Diagnosis function. Similarly, discussion of potential maintenance use alone was not sufficient for classification as Prescription.

V&V practices were coded according to the evidence actually reported by each primary study rather than according to the terminology used by its authors. Thus, a procedure described by an author as "validation" was coded as internal model evaluation when it consisted only of random train-test partitioning, k-fold cross-validation, or equivalent resampling within the development dataset. Validation was coded when evaluation introduced meaningful temporal, campaign, asset-, vessel-, external-dataset, or prospective independence. Verification was coded separately when the study explicitly reported evi-

dence addressing the correctness, consistency, or conformance of the implemented model or analytical process. When insufficient information was available, the corresponding V&V practice was treated as not clearly reported rather than inferred.

Dataset characteristics were coded independently from the algorithm and V&V classifications using seven complementary dimensions. These dimensions and their corresponding operational categories are summarized in Table 5. The coding structure distinguishes data provenance, acquisition setting, signal modality, scale and temporal structure, run-to-failure evidence, accessibility, and PHM function, preventing dataset characteristics from being represented by a single aggregated descriptor. First, data provenance was classified as: (i) real measured or observational data; (ii) simulated or model-based data; (iii) mixed data combining real measurements with simulated or synthetic information; (iv) public or secondary benchmark data whose original acquisition context could not be represented by a single provenance category; or (v) no empirical dataset for conceptual or methodological studies.

Second, the acquisition setting was recorded separately as onboard/operational, controlled test-bench or laboratory, simulator/digital-twin based, public/secondary, mixed, or not clearly reported. This distinction was introduced because the use of real measurements does not necessarily imply onboard or operational data.

Third, the main signal or information modalities were recorded, including process and thermodynamic variables, vibration or acoustic measurements, electrical variables, oil or tribological information, maintenance or failure-event records, and navigation or environmental data. Fourth, dataset size, monitoring duration, sampling interval, and segmentation or window length were recorded when explicitly reported.

Fifth, the presence of an actual run-to-failure sequence or an explicit degradation trajectory was distinguished from synthetically generated degradation and from studies for which such information was not reported. Sixth, dataset accessibility was coded as public/open, available upon request, restricted/confidential, or not stated. Finally, each dataset was linked to the PHM function for which it was actually used.

Classification was based on the data effectively used for model development and evaluation rather than solely on the physical origin of the studied asset. Therefore, when real measurements were used only to calibrate or validate a simulator but the machine-learning experiment used simulator outputs, the dataset was coded as simulated/model-based. Conversely, studies combining real operational measurements with synthetic fault injection, simulated degradation, or simulation-domain transfer were classified as mixed. When a dataset characteristic was not explicitly reported, it was coded as not reported rather than inferred.

Algorithm coding followed a hierarchical taxonomy that

Table 5. Dataset coding dimensions used in the review.

| Dimension                    | Operational categories / coding rule                                                                                                                                        |
|------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data provenance              | Real measured/observational; simulated/model-based; mixed real + simulated/synthetic; public/secondary benchmark; no empirical dataset.                                     |
| Acquisition setting          | Onboard/operational; controlled test bench/laboratory; simulator/digital twin; public/secondary; mixed; not reported.                                                       |
| Signal/data modality         | Process/thermodynamic; vibration/acoustic; electrical/power; oil/tribological; maintenance/failure records; navigation/environmental; other explicitly reported modalities. |
| Scale and temporal structure | Number of samples or records, monitoring duration, sampling interval, and segmentation/window length, when explicitly reported.                                             |
| Run-to-failure evidence      | Actual failure/degradation trajectory; failure-event-aligned history; synthetic/constructed degradation; not reported.                                                      |
| Accessibility                | Public/open; available upon request; restricted/confidential; not stated.                                                                                                   |
| PHM function                 | Detection; Diagnosis; Prognosis; Prescription; or explicitly implemented hybrid category.                                                                                   |

grouped methods according to their methodological families and architectural characteristics. Specific algorithms were normalized to canonical names, while related methods were organized according to their broader model structure. In particular, recurrent neural network (RNN) methods were treated as a neural-network family, with long short-term memory (LSTM) and gated recurrent unit (GRU) identified as recurrent variants. One-class support vector machine (OCSVM) was treated as a specific variant of the support vector machine (SVM) family. Autoencoder-based representation-learning methods included the variational autoencoder (VAE), whereas convolutional neural network (CNN) methods were retained as a distinct neural architecture. This organization distinguishes general methodological families from specific architectures or variants and provides a consistent basis for study-level algorithm coding.

Algorithm frequencies represent study-level presence rather than mutually exclusive assignments. Each normalized algorithm was counted at most once per study regardless of the number of experiments or hyperparameter configurations in which it appeared. When a paper compared several algorithms, each explicitly evaluated method contributed one occurrence; therefore, algorithm frequencies and percentages are not expected to sum to the number of studies or to 100%.

For hybrid architectures, constituent algorithms were retained separately at the specific-method level when their individual roles were explicitly part of the implemented architecture. At the family level, hybrid architectures were interpreted according to their dominant analytical role or retained as hybrid approaches when no single dominant role could be identified.

Interrater agreement was calculated for the initial title-, abstract-, and keyword-level screening stage but not separately for the subsequent full-text eligibility assessment or thematic coding of SWBS groups, PHM functions, or algorithm families. Accordingly, the reported Cohen's kappa should not

be interpreted as a reliability estimate for these subsequent assessment and classification stages.

To reduce classification inconsistency, a rule-based consistency audit of the final extraction matrix was performed using the operational definitions described above. SWBS and PHM assignments were rechecked against the primary asset and the analytical output reported in each study, combined categories were retained for multi-system or multi-function applications, and category totals were reconciled against the 83-study analytical sample and the corresponding cross-tabulations. Cases for which the available information did not support a unique classification were not inferred beyond the evidence reported in the primary study.

### 3.5.1. Rationale for the Selection of SWBS as the Ship-System Classification Framework

A system-level classification framework was required to compare studies that describe shipboard assets using heterogeneous terminology and different levels of technical detail. The purpose of this classification is not to reproduce the complete equipment hierarchy of each vessel, but to assign the principal health object addressed by each study to a stable and reproducible ship-system domain. For this purpose, the Ship Work Breakdown Structure (SWBS), following the expanded ship-system decomposition used in the review, was selected over other technically valid maritime classification schemes such as the SFI Group System.

The SFI Group System constitutes a credible alternative. It is a function-oriented, three-digit decimal classification in which the ship or offshore unit is divided into main groups, groups, and sub-groups; additional detail codes may be used to reach component level. Its original purpose is broad life-cycle information management across shipping and offshore activities, including specifications, purchasing, maintenance, repair, technical records, cost control, and documentation. SFI was conceived to accommodate different ship and rig types and to provide a common technical account structure independent of individual company organization or operating method (SFI Group System, 2001). Its use has also been reflected and adapted in standardized offshore coding structures such as NORSOK Z-014 (NORSOK, 2002). Consequently, SFI offers strong repeatability and is particularly suitable when the analytical objective is equipment-level lifecycle management, maintenance databases, material administration, or cost coding.

SWBS/ESWBS, however, is more closely aligned with the analytical objective of this review. The U.S. Department of Defense defines a standard Work Breakdown Structure for sea systems in MIL-STD-881F, while NAVSEA documentation formalizes the Expanded Ship Work Breakdown Structure (ESWBS) as a standardized coding framework for ships, ship systems, and combat systems. NAVSEA further describes

ESWBS as a common language that integrates technical, configuration, logistics, and cost information through a hierarchical ship-oriented breakdown (*Work Breakdown Structures for Defense Materiel Items*, 2022; Naval Sea Systems Command, 2005). This institutional codification provides an externally defined hierarchy rather than a review-specific grouping and therefore supports transparent assignment rules and independent replication.

The decisive advantage for the present synthesis is the correspondence between SWBS high-level groups and the technical domains evaluated in maritime PHM. Hull structure, propulsion plant, electric plant, command and surveillance, and auxiliary systems can be represented directly by SWBS 100, 200, 300, 400, and 500, respectively. These categories are sufficiently broad to normalize heterogeneous terminology across studies, while remaining sufficiently specific to preserve the physical and functional location of the monitored asset. This balance is particularly useful for evidence mapping because it allows system-level concentration to be quantified without requiring the detailed configuration of every vessel.

By contrast, the top-level SFI architecture is optimized for a somewhat different information-management purpose. SFI separates Machinery Main Components (Main Group 6), Systems for Machinery Main Components (Main Group 7), and Ship Common Systems (Main Group 8). Thus, studies that belong to a single PHM domain from the perspective of the health object may be distributed across several SFI main groups. For example, a propulsion engine can be placed under Machinery Main Components, whereas its fuel, lubrication, cooling, starting-air, exhaust, or automation systems are organized under systems serving machinery; electrical distribution is included among ship common systems (SFI Group System, 2001). To reproduce the same macroscopic domains used in this review, SFI would therefore require an additional author-defined aggregation layer. Such an aggregation would be feasible, but it would introduce extra interpretive decisions between the original classification code and the final evidence categories.

This difference is relevant because the review seeks a cross-study map of where AI-enabled PHM is applied, rather than a detailed technical-account system for maintenance administration. The selected SWBS logic permits a direct mapping from the primary monitored asset to the analytical domain. A propulsion engine, shaft line, propeller, or propulsion-support asset is assigned to SWBS 200; electrical generation, distribution, and energy-storage assets are assigned to SWBS 300; structural applications to SWBS 100; command, communication, monitoring, and surveillance functions to SWBS 400; and auxiliary service equipment to SWBS 500. The algorithm employed, the type of sensor, or the presence of contextual variables from other systems does not alter this assignment. Only when the same methodological framework is explicitly

evaluated on primary assets belonging to more than one SWBS domain is a combined classification retained. This rule reduces artificial multi-classification and makes the coding decision dependent on the health object rather than on incidental features of the dataset.

The selection is also consistent with the scope of the evidence base. The search strategy covers both general maritime terminology and explicitly naval terms such as navy, naval vessel, military vessel, and warship. SFI is not restricted to commercial shipping and can be adapted to a wide variety of ships and offshore units; nevertheless, its principal groups were developed primarily as a common technical-account structure for shipping and offshore operations. SWBS/ESWBS provides more explicit accommodation of naval mission-oriented domains, including command/surveillance and combat-system structures, while retaining generic groups that are equally applicable to propulsion, electrical, structural, and auxiliary systems found on non-combatant vessels. This characteristic reduces the need to change classification logic when the evidence base includes both commercial-maritime and naval applications.

Accordingly, the choice of SWBS should not be interpreted as indicating that SFI is intrinsically inferior. SFI may be preferable for detailed fleet asset management, maintenance planning, purchasing, documentation, or shipyard information systems. For the present systematic review, however, SWBS offers the more parsimonious and methodologically transparent solution because it provides: (i) an externally standardized hierarchical decomposition; (ii) direct correspondence with the ship-system domains required for the research questions; (iii) compatibility with both general ship systems and naval mission systems; and (iv) a simple coding rule based on the principal monitored asset. As summarized in Table 6, these properties reduce the number of review-specific transformations required between the original study and the final evidence map, strengthening traceability and reproducibility of the system-level synthesis.

The operational coding rules described above apply this rationale to the primary-study matrix and define how each included study was assigned to the corresponding ship-system domain.

### 3.6. Analytical Synthesis

The analytical synthesis combined frequency and proportional analysis with cross-dimensional evidence mapping. In addition to examining each extracted dimension independently, cross-tabulations were used to compare PHM functional coverage across publication periods and SWBS system groups. Dataset provenance, verification and validation practices, and deployment conditions were also analyzed comparatively to identify differences in evidential strength and implementation maturity across the reviewed corpus.

Table 6. Comparative suitability of SWBS/ESWBS and SFI for the system-level evidence mapping used in this review.

| Criterion                         | SWBS/ESWBS                                                                                                            | SFI Group System                                                                                                                        |
|-----------------------------------|-----------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|
| Hierarchical coding               | Standardized ship-system hierarchy with high-level groups and subordinate elements.                                   | Three-digit decimal hierarchy (main group, group, sub-group), with additional detail/material codes.                                    |
| Primary orientation               | Product/system decomposition for ships, ship systems, configuration, logistics, cost, and defense-materiel reporting. | Lifecycle technical-account classification for shipping/offshore information, maintenance, purchasing, documentation, and cost control. |
| Fit to this review                | Direct mapping of Hull, Propulsion, Electric, Command/Surveillance, and Auxiliary domains.                            | Comparable coverage is possible, but some PHM domains span several SFI main groups and require re-aggregation.                          |
| Naval mission systems             | Explicitly compatible with ship and combat-system structures.                                                         | Adaptable, but mission/combat domains are not represented as equally direct top-level categories.                                       |
| Coding rule for PHM evidence      | Primary monitored asset determines the SWBS group; contextual variables do not generate additional assignments.       | Replicable at code level, but reproducing the review's macroscopic domains would require an additional mapping rule.                    |
| Most suitable use in this context | Cross-study system-level evidence mapping and comparison.                                                             | Detailed asset, maintenance, documentation, purchasing, or lifecycle information management.                                            |

For comparative purposes, studies restricted to Detection, Diagnosis, or Detection + Diagnosis were grouped as detection/diagnosis-oriented applications, whereas studies including Prognosis and/or Prescription were grouped as extending toward downstream PHM functions. This aggregation was introduced only as a comparative analytical device and does not constitute a formal or validated PHM maturity scale.

The review was not designed as a statistical meta-analysis. The primary studies address heterogeneous systems, fault definitions, datasets, operating conditions, PHM tasks, V&V procedures, and performance metrics; therefore, directly comparable effect sizes could not be derived without imposing assumptions unsupported by the evidence. Similarly, this study does not constitute a formal bibliometric science-mapping analysis of citation, co-authorship, or keyword networks. Accordingly, the analytical contribution of the review lies in the systematic cross-comparison of system-level, functional, methodological, V&V, dataset, and implementation dimensions within the same evidence base, enabling relationships and evidence-concentration patterns to be examined across multiple dimensions of the 83-study corpus.

The database searches identified 400 records: 281 from Scopus and 119 from Web of Science. The two bibliographic exports were imported into Bibliometrix and merged using `mergeDbSources()` with duplicate removal enabled. This procedure removed 88 duplicate records, leaving 312 unique records for title-, abstract-, and keyword-level screening. At this stage, 170 records were excluded and 142 were retained for eligibility assessment.

Following eligibility assessment, 48 additional documents were excluded because they did not satisfy the predefined selection criteria, leaving 94 documents in the review corpus. Of these, 83 were primary studies included in the analytical synthesis, whereas 11 review articles were used exclusively to contextualize the state of the art. Thus, the final analytical and contextual corpus was reconciled as  $94 = 83 + 11$ . The complete selection flow is summarized in Figure 1.

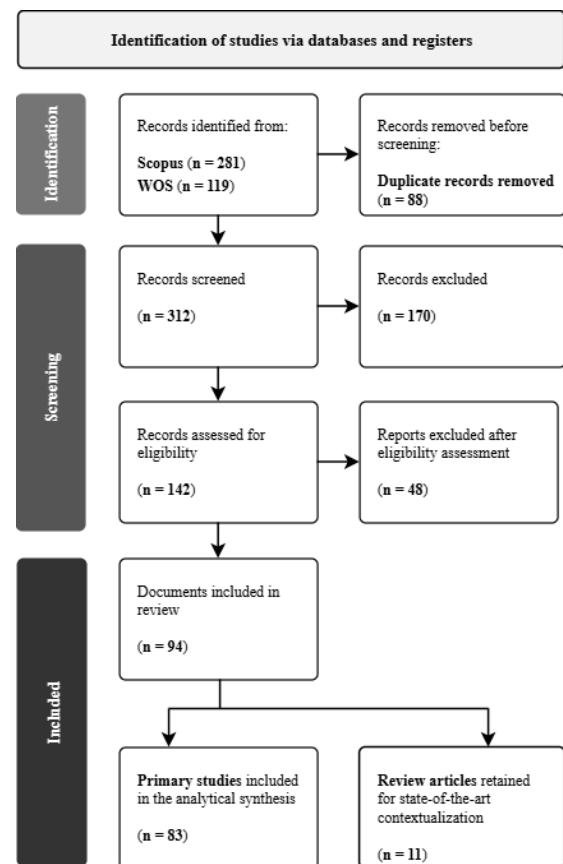


Figure 1. PRISMA flow diagram of the study identification, screening, eligibility, and inclusion process.

Before the quantitative synthesis, a final numerical consistency check was performed across the screening flow, extraction matrix, tables, and figures. The primary-study frequencies were reconciled to 83 records, the annual publication counts sum to 83, and the final retained corpus was reconciled as 94 documents comprising 83 primary studies and 11 contextual review articles.

#### 4. RESULTS

To answer the research questions, this section reports the evidence extracted directly from the included primary studies. The results are organized into three areas: (i) ship systems and equipment addressed by AI-based PHM applications; (ii) PHM functions, models, and algorithmic approaches; and (iii) reported data-management, V&V, and deployment practices.

Frequencies, distributions, study-level characteristics, and methodological practices explicitly reported in the primary studies are presented here as corpus-derived findings. Explanatory hypotheses, causal interpretations, implications, and recommendations derived by the authors from these patterns are reserved for Section 5.

##### 4.1. General Characterization of the Included Evidence

The final analytical sample comprises 83 primary studies published between 1993 and 2025. Of these, 67 studies (80.7%) were published from 2020 onward.

The highest annual publication output was recorded in 2023, followed by 2024 and 2021. During 2020–2025, the annual distribution was: 2020 ( $n = 5$ ; 6.0%), 2021 ( $n = 12$ ; 14.5%), 2022 ( $n = 7$ ; 8.4%), 2023 ( $n = 19$ ; 22.9%), 2024 ( $n = 16$ ; 19.3%), and 2025 ( $n = 8$ ; 9.6%). As shown in Figure 2, the publication record indicates a marked concentration of primary studies from 2020 onward.

##### 4.2. According to the literature, in which ship systems and equipment have PHM approaches and artificial intelligence models been applied?

Following the coding rules defined in Section 3.5, the included studies were mapped to SWBS system groups using the ESWBS logic (*Work Breakdown Structures for Defense Materiel Items*, 2022). Figure 3 provides a general representation of the principal SWBS functional groups used to organize the shipboard systems and equipment identified in the review.

The system-level distribution shows a marked concentration of evidence in propulsion and electric-power applications. SWBS 200 contains 56 primary studies and SWBS 300 contains 16. Four additional studies jointly address assets mapped to SWBS 200 and SWBS 300. SWBS 100, SWBS 400, and SWBS 500 contain two, three, and two studies, respectively. This distribution is summarized in Figure 4, which highlights the predominance of propulsion- and electric-power-related applications in the analytical sample.

Some contributions apply the same methodological framework to more than one ship-system domain. A representative example is Fernández-Barrero et al. (2021), who evaluated Neural-Network-Based Predictive Sustainment (SOPRENE), the Spanish Navy's predictive-maintenance tool, through applications involving both diesel propulsion equipment and diesel

power-generation equipment.

No primary studies in the final analytical sample were mapped to SWBS 600 (Outfit and Furnishings), SWBS 700 (Armament), SWBS 800 (Total Ship Integration/Engineering), or SWBS 900 (Ship Assembly and Support Services). Potential explanations for the observed concentration are examined separately in Section 5.

Alongside the SWBS classification, the functional scope of the reported applications was classified according to the review-specific normalization defined in Section 3.3: Detection, Diagnosis, Prognosis, and Prescription. Terms such as anomaly detection, fault detection, novelty detection, and deviation detection were retained when describing individual studies but were grouped under Detection for the quantitative synthesis. The functional normalization used to organize these categories is represented in Figure 5.

The sample also includes hybrid contributions in which two or more PHM functions were explicitly addressed within the same study.

Table 7. Distribution of studies by PHM functional approach.

| PHM approach type                    | Frequency | Proportion |
|--------------------------------------|-----------|------------|
| Detection                            | 34        | 41.0%      |
| Diagnosis                            | 18        | 21.7%      |
| Prognosis                            | 15        | 18.1%      |
| Prescription                         | 6         | 7.2%       |
| Detection + Diagnosis                | 4         | 4.8%       |
| Diagnosis + Prognosis                | 3         | 3.6%       |
| Detection + Prognosis                | 2         | 2.4%       |
| Diagnosis + Prognosis + Prescription | 1         | 1.2%       |
| Total                                | 83        | 100.0%     |

As shown in Table 7, Detection is the most frequently represented individual function, with 34 studies (41.0%), followed by Diagnosis with 18 (21.7%) and Prognosis with 15 (18.1%). Prescription appears as the primary classification in six studies (7.2%). Ten additional studies implement combined functional categories. Table 8 further disaggregates these functional categories by SWBS group.

##### 4.2.1. SWBS 100: Structure and Hull

Two primary studies were mapped to SWBS 100. One was classified as Detection and one as Prognosis. The identified applications address changes in hull- and propeller-related condition and their relationship with hydrodynamic performance. Coraddu et al. (2019) addressed biofouling and hull-fouling detection through anomaly-detection methods, whereas Spandonidis and Paraskevopoulos (2023) used deep-learning architectures to characterize condition changes associated with the hull and propeller.

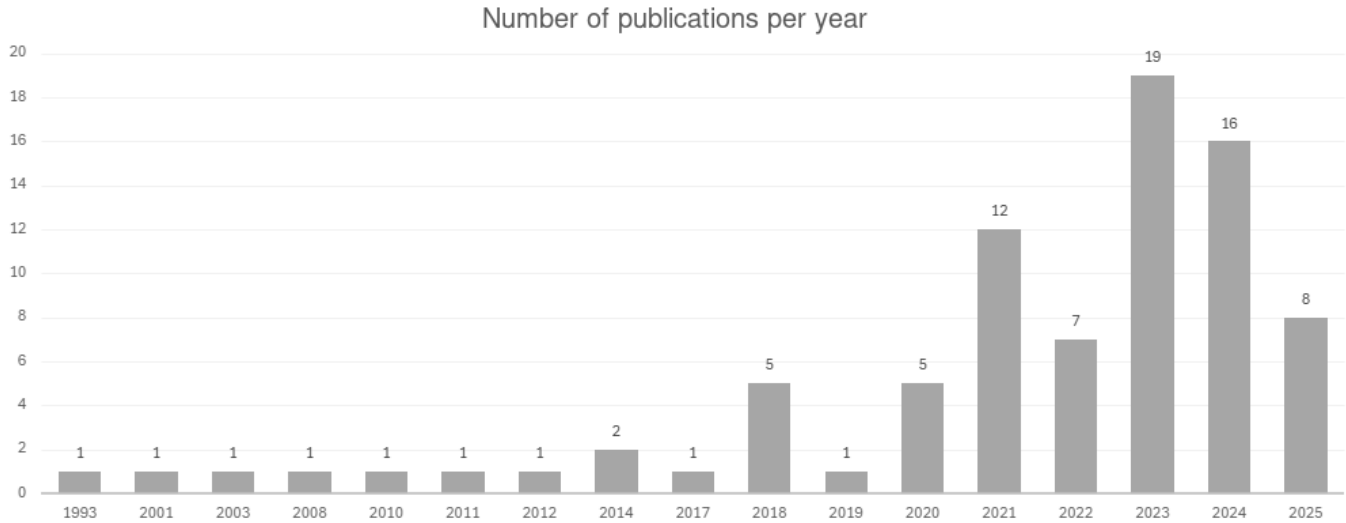


Figure 2. Number of publications per year in the included primary-study sample.

Table 8. Distribution of PHM functional approaches by SWBS system group in the included studies.

| PHM approach by SWBS system group | Frequency |
|-----------------------------------|-----------|
| SWBS 100                          | 2         |
| Detection                         | 1         |
| Prognosis                         | 1         |
| SWBS 200                          | 56        |
| Detection                         | 24        |
| Detection + Diagnosis             | 3         |
| Detection + Prognosis             | 2         |
| Diagnosis                         | 15        |
| Diagnosis + Prognosis             | 3         |
| Prescription                      | 3         |
| Prognosis                         | 6         |
| SWBS 200 & SWBS 300               | 4         |
| Detection                         | 1         |
| Detection + Diagnosis             | 1         |
| Diagnosis                         | 1         |
| Prognosis                         | 1         |
| SWBS 300                          | 16        |
| Detection                         | 6         |
| Diagnosis                         | 2         |
| Prescription                      | 2         |
| Prognosis                         | 6         |
| SWBS 400                          | 3         |
| Detection                         | 2         |
| Prescription                      | 1         |
| SWBS 500                          | 2         |
| Prescription                      | 1         |
| Prognosis                         | 1         |

#### 4.2.2. SWBS 200: Propulsion Plant

SWBS 200 contains 56 primary studies and represents the most frequently reported system group. Diesel engines constitute the dominant asset type, although gas turbines, controllable-pitch propellers, lubrication systems, fuel-injection systems, and air or supercharging systems are also represented.

Studies including Scaife et al. (1993), Z. Li et al. (2012), and Makridis et al. (2020) report fault-detection, diagnostic, or monitoring applications under controlled or instrumented operating conditions. Michałowska et al. (2021) addressed a controllable-pitch propeller system using an anomaly-detection and diagnostic framework.

Functionally, the SWBS 200 sample contains 24 Detection studies, 15 Diagnosis studies, three Detection + Diagnosis studies, two Detection + Prognosis studies, three Diagnosis + Prognosis studies, six Prognosis studies, and three Prescription studies.

Detection-oriented studies include applications formulated as anomaly detection, fault detection, and condition monitoring (Makridis et al., 2020; Cheliotis, Lazakis, & Theotokatos, 2020; Vorkapić, Radonja, Babić, & Mandić-Ilić, 2020). Diagnostic applications are reported in both early and more recent studies (Scaife et al., 1993; Ogaji & Singh, 2003; Z. Li et al., 2012; G. Wu, 2018). Prognostic and prescriptive applications are less frequently represented numerically in this group.

#### 4.2.3. SWBS 300: Electric Plant

SWBS 300 contains 16 primary studies. The reported applications are mainly associated with shipboard generation machinery, electrical components, and marine energy-storage systems.

Park and Oh (2023) addressed generator engines, whereas Michelena et al. (2023) reported fault-detection and maintenance-related analyses using sensor and operational records from diesel-generator components aboard the Spanish Navy's Álvaro de Bazán-class (F-100) frigates. Song et al. (2024) addressed capacity prognosis in marine lithium-ion batteries under dynamic operating conditions.

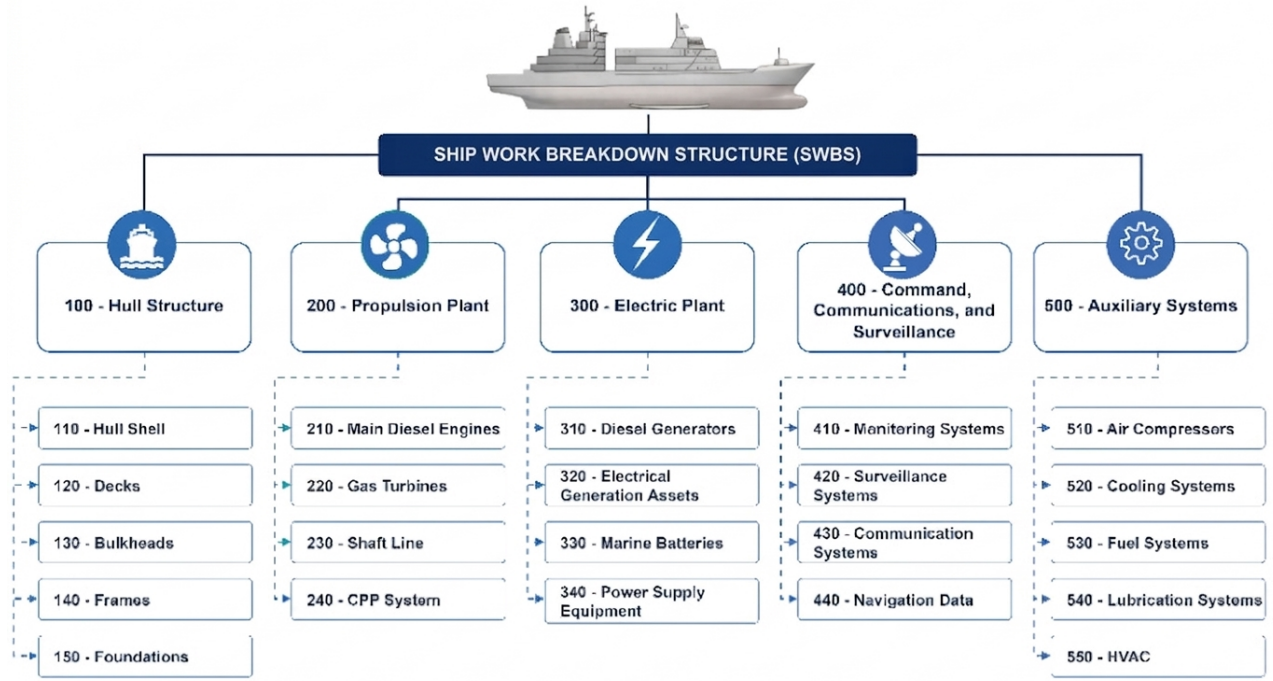


Figure 3. General mapping of the main SWBS functional groups.

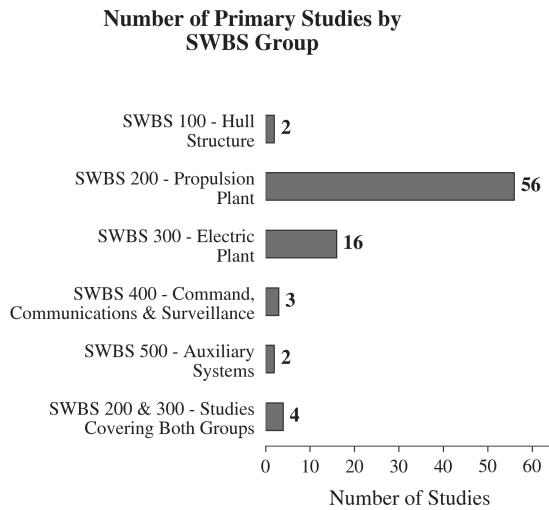


Figure 4. Number of studies by SWBS system group.

The functional distribution comprises six Detection studies, two Diagnosis studies, six Prognosis studies, and two Prescription studies. Reported prognostic applications include battery and degradation-tracking problems (Song et al., 2024; Díez-Oliván, Pagán, Sanz, & Sierra, 2017; Mandala et al., 2023; Velasco-Gallego & Lazakis, 2023), whereas decision-oriented applications include maintenance- or replacement-related support (Michelena et al., 2023; Karlsson, Arriagada, & Genrup, 2008; Daya & Lazakis, 2023; Yigin & Celik, 2024).

#### 4.2.4. SWBS 400: Command, Communications, and Surveillance

Three primary studies were mapped to SWBS 400. Two were classified as Detection and one as Prescription.

The sample includes both asset/system-oriented monitoring and operational-surveillance boundary cases. Benarfa et al. (2025) addressed fault detection associated with monitored communication and surveillance functions. B. Zhang et al. (2023) and W. Wu et al. (2025) include applications focused on anomalous vessel behavior or navigation patterns.

Following the boundary rule defined in Section 3.3, operational-surveillance applications remain represented in the SWBS and Detection mapping but were not coded as evidence of prognostic capability, RUL estimation, or complete asset-health PHM implementation.

#### 4.2.5. SWBS 500: Auxiliary Systems

Two primary studies were mapped to SWBS 500, one classified as Prognosis and one as Prescription. Reported applications include prognostic and maintenance-support problems involving auxiliary marine equipment. Gribbestad et al. (2021), for example, addressed PHM for marine air compressors through transfer learning, whereas other work considered intelligent operation and maintenance frameworks for auxiliary marine equipment (Wang, 2023).

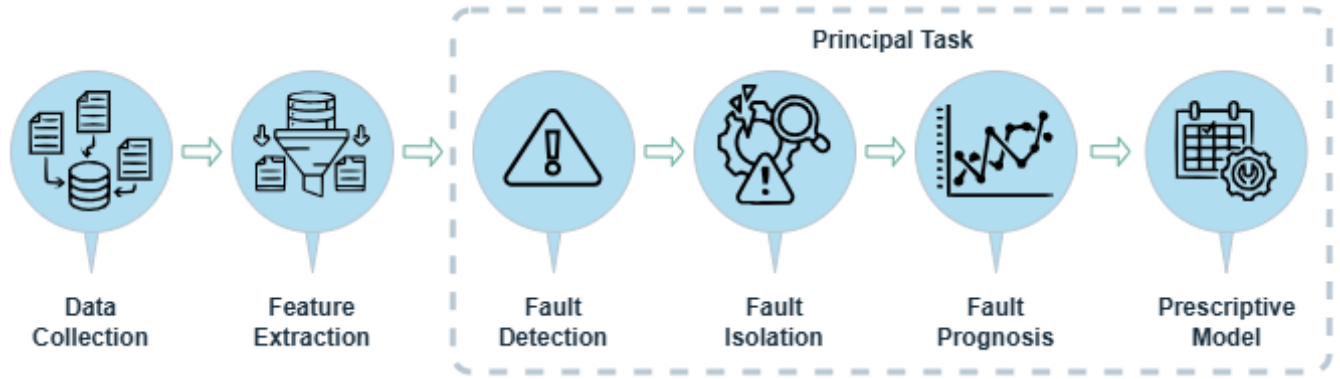


Figure 5. Functional normalization used to organize the PHM-related functions identified in the reviewed studies (adapted from Kumar and Flores-Cerrillo (2024)).

#### 4.3. What types of models and techniques are used in detection, diagnosis, prognosis, and prescription tasks?

This subsection presents an evidence map of the methods and algorithms reported in the primary studies. Since several studies compare or combine multiple models, frequencies indicate study-level presence rather than mutually exclusive assignments.

##### 4.3.1. Algorithms Mentioned in the Sample

Following the coding hierarchy defined in Section 3.5, algorithms were normalized to canonical method names and counted as study-level occurrences.

Table 9 organizes the reported algorithms according to their methodological families and architectural relationships. Indentation is used to identify specific variants within broader methodological structures. The hierarchy is conceptual rather than additive: frequencies represent study-level presence of each normalized method and should not be interpreted as parent-category totals. Because individual studies may report or compare several algorithms, the frequencies are not mutually exclusive.

**Strengths and Limitations of AI Model Families.** Traditional machine-learning methods, such as support vector machines and random forests, offer practical advantages for PHM applications involving structured operational data and engineered features. These approaches can be suitable when labeled data and computational resources are limited, although their performance often depends on feature selection and the representation of different operating conditions (Nguyen et al., 2023; Ucar et al., 2024).

In contrast, deep-learning models, including convolutional and recurrent neural networks, can learn complex representations and capture temporal dependencies from sensor data, making them particularly relevant for signal-based fault diagnosis

Table 9. Hierarchical organization and study-level frequencies of algorithms reported in the included studies.

| Method / architecture                   | Frequency | Proportion |
|-----------------------------------------|-----------|------------|
| <i>Neural-network-based methods</i>     |           |            |
| ANN                                     | 41        | 49.4%      |
| RNN                                     | 12        | 14.5%      |
| LSTM                                    | 25        | 30.1%      |
| GRU                                     | 5         | 6.0%       |
| CNN                                     | 24        | 28.9%      |
| MLP                                     | 10        | 12.0%      |
| RBF                                     | 5         | 6.0%       |
| Attention-based methods                 | 8         | 9.6%       |
| <i>Representation-learning methods</i>  |           |            |
| Autoencoder                             | 18        | 21.7%      |
| VAE                                     | 8         | 9.6%       |
| <i>Kernel-based methods</i>             |           |            |
| SVM                                     | 31        | 37.3%      |
| OCSVM                                   | 9         | 10.8%      |
| <i>Tree- and ensemble-based methods</i> |           |            |
| Random Forest                           | 14        | 16.9%      |
| <i>Instance-based methods</i>           |           |            |
| kNN                                     | 9         | 10.8%      |
| <i>Fuzzy methods</i>                    |           |            |
| Fuzzy logic                             | 9         | 10.8%      |
| <i>Probabilistic methods</i>            |           |            |
| Bayesian Network                        | 6         | 7.2%       |
| <i>Dimensionality-reduction methods</i> |           |            |
| PCA                                     | 16        | 19.3%      |
| <i>Statistical monitoring methods</i>   |           |            |
| EWMA                                    | 5         | 6.0%       |
| <i>Signal-processing methods</i>        |           |            |
| Wavelet                                 | 8         | 9.6%       |
| <i>Knowledge-based methods</i>          |           |            |
| Expert system or rules                  | 5         | 6.0%       |

and degradation prognostics (Ellefsen et al., 2019; Velasco-Gallego et al., 2023). However, their application may require larger and more representative datasets, greater computational resources, and careful assessment of model interpretability and generalization across operating conditions (Ucar et al., 2024; Liang et al., 2024).

Consequently, method selection should consider the intended PHM function, data availability and quality, computational constraints, and deployment requirements. The frequencies

reported in Table 9 therefore reflect research trends rather than comparative evidence of model superiority.

#### 4.3.2. Dataset Characteristics and Data Availability

The article-level dataset audit revealed substantial heterogeneity in both data provenance and acquisition setting. Table 10 summarizes the primary provenance classification used for the 83 studies. Of the included studies, 47 (56.6%) relied primarily on real measured or observational data, 17 (20.5%) used simulated or model-based datasets, and 15 (18.1%) combined real measurements with simulated or synthetically generated information. Two studies (2.4%) relied primarily on public/secondary benchmark datasets with heterogeneous or not fully traceable original acquisition contexts, while two (2.4%) were conceptual or methodological contributions for which no empirical dataset was used.

Importantly, provenance and acquisition setting were treated as different dimensions. Real measured data include both onboard or fleet-operational records and controlled physical experiments performed on test benches or laboratory equipment. Likewise, the mixed category includes different combinations, such as real healthy-operation baselines with synthetic fault injection, real ship data supplemented by engine simulations, and transfer-learning configurations combining experimental and simulated or secondary datasets.

Table 10. Primary dataset provenance in the 83 included studies.

| Dataset provenance                 | n  | %      |
|------------------------------------|----|--------|
| Real measured / observational      | 47 | 56.6%  |
| Simulated / model-based            | 17 | 20.5%  |
| Mixed real + simulated / synthetic | 15 | 18.1%  |
| Public / secondary benchmark       | 2  | 2.4%   |
| No empirical dataset / conceptual  | 2  | 2.4%   |
| Total                              | 83 | 100.0% |

Across the corpus, signal modalities include process and thermodynamic variables such as temperatures, pressures, flows, rotational speed, load, and fuel-related measurements; vibration and acoustic measurements; electrical and battery variables; oil and wear-related information; maintenance and failure-event records; and navigation or environmental data. Dataset scale also varies from small controlled experiments to multi-year shipboard histories. When explicitly reported, sampling intervals, monitoring duration, number of records, and segmentation windows were retained in the extraction matrix.

The distinction between real and simulated information is particularly important because several studies combine the two. For example, some models are trained on simulated fault conditions but subsequently evaluated using real measurements, while others use long-term onboard baseline data and synthetically introduce degradation or faults because real failure examples are unavailable. Therefore, the use of real

data should not be interpreted automatically as evidence of a fully real fault dataset or of onboard model validation.

#### 4.3.3. Verification, Validation, and Comparability Practices

The V&V evidence reported across the primary studies is heterogeneous. Formal verification of the implemented analytical process is seldom explicitly distinguished from predictive-performance evaluation, and the term “validation” is used inconsistently across the corpus. For this reason, the reported procedures were interpreted using the coding rules defined in Section 3.4.

Internal model evaluation is the most commonly reported form of performance assessment. Cross-validation or k-fold procedures were explicitly identified in 12 studies (14.5%), whereas fixed train–test partitions were explicitly reported in six studies (7.2%).

Evaluation introducing stronger independence from the model-development data was less frequently documented. Three studies (3.6%) explicitly reported temporal- or campaign-aware evaluation, while two studies (2.4%) included cross-vessel or external evaluation. These categories are not mutually exclusive and represent practices explicitly identified in the extraction matrix rather than exhaustive quality scores.

The evidence required for evaluation also differs according to the PHM function. Detection studies report normal/abnormal discrimination and threshold-related evaluation; diagnostic studies rely on labeled fault or health-state classes; and prognostic studies report degradation, future-health, failure-horizon, or RUL-related outputs. Maintenance-event, replacement, or end-of-life traceability is not reported consistently across the corpus.

Most model evaluations remain retrospective or offline, including several studies based on real shipboard measurements. A smaller subset reports real-time monitoring, onboard inference, ship–shore integration, or end-to-end operational architectures. These different implementation conditions are interpreted further in Section 5.

## 5. DISCUSSION

Section 4 reported the evidence directly extracted from the primary-study corpus, whereas the present section interprets the relationships, implications, and limitations arising from those findings. Numerical frequencies, proportions, and cross-tabulations reported below are derived directly from the 83 primary studies. By contrast, explanations concerning mechanisms that may underlie the observed distributions, implications for PHM development, and recommendations for future implementation represent interpretations of the authors and are identified accordingly.

### 5.1. Temporal and Functional Development of the Evidence

To move beyond marginal frequencies, the extracted evidence was examined across publication period, ship-system domain, and PHM functional coverage. For comparative purposes, Detection, Diagnosis, and Detection + Diagnosis were grouped as detection/diagnosis-oriented applications, whereas categories containing Prognosis and/or Prescription were considered as extending toward downstream PHM functions. This distinction is used only as a comparative analytical device and should not be interpreted as a validated PHM maturity scale. The resulting cross-dimensional comparison is reported in Table 11.

The temporal comparison shows that the strong increase in publication volume has not been accompanied by a proportional shift toward downstream PHM functions. Of the 67 studies published from 2020 to 2025, 47 (70.1%) remain focused on Detection and/or Diagnosis, whereas 20 (29.9%) include Prognosis and/or Prescription. In the smaller 1993–2019 subset, 7 of 16 studies (43.8%) include these downstream functions.

Given the limited size and temporal dispersion of the earlier sample, this difference should not be interpreted as evidence of regression in the field. Rather, the corpus indicates that recent growth in maritime AI-PHM research has been driven predominantly by Detection- and Diagnosis-oriented applications. Publication growth should therefore be distinguished from progression in PHM functional coverage.

A second pattern emerges when publication volume is compared with PHM functional coverage across the two dominant ship-system domains. SWBS 200 contains 56 studies and therefore represents the largest body of evidence; however, only 14 (25.0%) include Prognosis and/or Prescription. In SWBS 300, 8 of 16 studies (50.0%) extend toward these functions.

Thus, the larger volume of propulsion-related research does not necessarily imply greater functional depth. The electric-plant literature is smaller in absolute terms but exhibits a comparatively greater relative presence of prognostic and prescriptive functions. This distinction demonstrates why publication volume and functional PHM coverage should be interpreted separately.

### 5.2. Concentration of Evidence and Possible Explanatory Mechanisms

Studies mapped exclusively to SWBS 200 and SWBS 300 account for 72 of the 83 primary studies (86.7%). When the four studies jointly addressing SWBS 200 and SWBS 300 are also considered, 76 studies (91.6%) involve propulsion and/or electric-power systems. This concentration limits the extent to which conclusions derived from the corpus can be generalized to hull, auxiliary, surveillance, and other shipboard systems.

The mechanisms underlying this concentration were not coded as causal variables in the primary-study corpus and should therefore be interpreted as explanatory hypotheses rather than direct findings of the systematic review.

Operational criticality and instrumentation provide one plausible technical explanation because propulsion and electric-power systems have a direct impact on vessel availability, safety, and energy performance and are commonly connected to automation and monitoring infrastructure (Daya & Lazakis, 2023; Inal, Charpentier, & Deniz, 2022; Liang et al., 2024; Kalafatelis et al., 2025).

Organizational and economic conditions may reinforce this technical advantage. Propulsion engines, generators, and associated power systems are frequently supplied and supported by Original Equipment Manufacturers (OEMs) with established lifecycle-service and digital-service ecosystems. Such arrangements may lower the marginal effort required for remote monitoring and create incentives to use condition information for maintenance scheduling, availability management, spare-parts planning, and service delivery (Solem, Kohtamaki, Parida, & Brekke, 2022; M. Zhang, Wei, He, He, & Yan, 2023).

Data ownership and access may also influence which systems are represented in PHM research. Proprietary interfaces, vendor-specific formats, contractual access restrictions, and changes in data governance throughout the vessel lifecycle can affect whether operators and researchers obtain sufficiently granular and reusable historical data (Bronson, Fonseca, & Gaspar, 2024). Warranty arrangements, long-term service agreements, and performance-based maintenance contracts may additionally create stronger incentives for monitoring high-value machinery.

These mechanisms are therefore proposed as author interpretations consistent with the observed corpus and the broader literature rather than as causal relationships demonstrated directly by the 83 primary studies.

### 5.3. Dataset Availability and Implications for PHM Evidence

The dataset audit provides an additional perspective on the maturity of the evidence. Although 47 studies (56.6%) rely primarily on real measured or observational data and another 15 (18.1%) combine real information with simulated or synthetic data, the presence of real measurements alone does not imply equivalent evidential strength. Real data range from longitudinal onboard records and fleet maintenance histories to controlled physical experiments, whereas mixed datasets often use synthetic faults precisely because naturally occurring failure examples are scarce.

This distinction is particularly relevant for prognosis. Run-to-failure histories and complete degradation trajectories are much less consistently documented than healthy-operation or

Table 11. Cross-dimensional comparison of PHM functional coverage in the included primary studies.

| Dimension          | Category            | Studies | Detection/Diagnosis only | Includes Prognosis/Prescription |
|--------------------|---------------------|---------|--------------------------|---------------------------------|
| Publication period | 1993–2019           | 16      | 9 (56.3%)                | 7 (43.8%)                       |
|                    | 2020–2025           | 67      | 47 (70.1%)               | 20 (29.9%)                      |
| Ship-system domain | SWBS 200            | 56      | 42 (75.0%)               | 14 (25.0%)                      |
|                    | SWBS 300            | 16      | 8 (50.0%)                | 8 (50.0%)                       |
| Overall sample     | All primary studies | 83      | 56 (67.5%)               | 27 (32.5%)                      |

fault-classification datasets. Consequently, several prognostic studies construct degradation trajectories, use piecewise RUL labels, inject synthetic faults, or employ physics-based degradation models rather than relying on complete real-world failure histories. These approaches are valuable when failures are rare, but they should not be interpreted as equivalent to prospective run-to-failure evidence.

Dataset accessibility constitutes a separate limitation. Public benchmark datasets are used in some studies, whereas other datasets are proprietary, confidential, available only upon request, or have no explicit availability statement. This restricts independent replication and makes model comparisons dependent not only on algorithms but also on access to the underlying operational context. For this reason, future maritime PHM studies should report dataset provenance, acquisition setting, signal definitions, temporal coverage, sampling frequency, segmentation rules, maintenance-event traceability, and accessibility as minimum dataset descriptors.

#### 5.4. Verification, Validation, and Generalizability

The V&V synthesis reveals a substantial difference between internal model evaluation and evidence of generalization under independent or operational conditions. Cross-validation or k-fold procedures were explicitly identified in 12 studies (14.5%), while fixed train–test partitions were reported in six studies (7.2%). By comparison, only three studies (3.6%) explicitly reported temporal- or campaign-aware evaluation, and two (2.4%) included cross-vessel or external evaluation.

These categories are not mutually exclusive and should not be interpreted as exhaustive V&V quality scores. Nevertheless, the distribution shows that internal predictive evaluation is substantially more commonly reported in the literature than evidence of generalization across time, operating campaigns, vessels, or external datasets.

Accordingly, the findings of this review should not be interpreted as if all primary studies provided equivalent levels of evidence. Studies supported only by internal train–test evaluation provide more limited evidence of transferability than studies incorporating temporally independent campaigns, external assets or vessels, maintenance-event traceability, or operational validation. Similarly, results derived exclusively from simulated or controlled environments should be distinguished from evidence obtained under representative shipboard condi-

tions. These differences do not invalidate the reported studies, but they constrain the strength with which their results can be generalized to fleet-level PHM implementation.

This distinction is important from a V&V perspective. Internal predictive evaluation demonstrates model performance within or near the development distribution, whereas operational validation requires evidence that performance is retained when meaningful independence is introduced between development and intended-use conditions.

Formal verification of analytical-process or model implementation correctness is even less consistently distinguished in the maritime PHM literature. Consequently, terminology such as “validation” should not be interpreted uniformly across the primary studies. Random train–test partitions, cross-validation, temporal holdouts, cross-vessel testing, and sustained onboard evaluation represent materially different forms of evidence.

#### 5.5. Deployment and Operational Integration

Deployment reporting reveals a related distinction between computational capability and operational maturity. Several studies describe real-time or online inference, but these terms encompass heterogeneous situations, including laboratory experiments, simulated online phases, retrospective processing of onboard measurements, experimental prototypes, and a smaller number of operationally integrated systems.

Consequently, real-time computation should not be interpreted as equivalent to sustained onboard validation or deployment. The more demanding transition is from analytical model evaluation to persistent integration with shipboard data acquisition, ship–shore communication, maintenance records, model updating, cybersecurity requirements, and operational decision workflows.

Taken together, these findings indicate that the maturity of maritime PHM applications cannot be assessed solely through reported model accuracy. Table 12 organizes six complementary dimensions of verification, validation, and deployment evidence, highlighting what each type of assessment can establish and what additional evidence may be needed to support operational implementation. These dimensions provide a structured interpretation of the evidence rather than a formal or sequential maturity scale (IEEE, 2024).

Table 12. Evidence dimensions for assessing verification, validation, and operational maturity in maritime PHM applications.

| Evidence dimension                         | Evidence required                                                                                                                                                  | What the evidence establishes                                                                                                      | Further evidence needed                                                                                                         |
|--------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| Implementation verification                | Consistency of data processing, reproducibility of analytical results, and comparison with specified requirements or reference cases.                              | The analytical process behaves as intended under defined conditions and its implementation can be checked.                         | Evidence of predictive performance and generalization beyond the conditions used for verification.                              |
| Internal model evaluation                  | Train–test partitioning, cross-validation, and performance assessment using data drawn from the development distribution.                                          | The model can perform the intended detection, diagnosis, or prognostic task under the evaluated data conditions.                   | Independent testing across time, operating regimes, or assets to assess potential overfitting and limited generalization.       |
| Temporal and operational-regime validation | Evaluation using independent operating periods, monitoring campaigns, or different load, speed, and operating conditions.                                          | The model retains its reported performance across the temporal and operational variations actually evaluated.                      | Assessment under previously unseen operating conditions, failure events, or different equipment configurations.                 |
| Independent asset or vessel validation     | Evaluation on external datasets, different equipment units, or vessels not involved in model development, preferably supported by independently documented faults. | The model demonstrates transferability to the independent assets and conditions included in the evaluation.                        | Confirmation of robustness across broader equipment configurations, vessel classes, and operational contexts.                   |
| Onboard implementation and demonstration   | Integration with shipboard data acquisition, real-time processing, monitoring systems, and operational communication infrastructure.                               | The PHM solution demonstrates technical feasibility within the evaluated shipboard environment.                                    | Sustained performance monitoring, maintenance-event traceability, and evidence of integration into operational decision-making. |
| Sustained operational deployment           | Continued onboard operation, monitoring of model performance, traceability to maintenance actions, and documented procedures for system updates.                   | The PHM system demonstrates sustained functionality and supports maintenance decisions under the evaluated operational conditions. | Evidence of long-term effectiveness, independent replication, and transferability to broader fleet applications.                |

*Note:* These dimensions describe complementary forms of evidence rather than a formal numerical maturity scale. Their interpretation depends on data provenance, operating-regime coverage, failure-event traceability, and the intended PHM function.

### 5.6. Implications for PHM Prioritization

The limited representation of an SWBS group should not, by itself, be interpreted as sufficient justification for expanding PHM coverage. Underrepresentation identifies an evidence gap, but deployment decisions require an asset-level business case.

Candidate systems should be prioritized according to the expected operational and economic value of the additional information provided by PHM. Relevant criteria include: (i) operational criticality and the effect of failure on vessel availability or mission capability; (ii) safety, environmental, and regulatory consequences; (iii) expected failure cost, including repair, secondary damage, downtime, and logistical consequences; (iv) data readiness, including sensor coverage, historical depth, data quality, and maintenance-event traceability; (v) degradation observability and achievable warning horizon; (vi) maintainability leverage, including repair duration, spare-part lead time, accessibility, and dry-dock dependency; and (vii) implementation cost, including sensing, connectivity, data-access permissions, integration, V&V, cybersecurity, and lifecycle support.

This prioritization is also consistent with the lifecycle-support perspective adopted in Integrated Logistics Support (ILS). In the shipbuilding context, ILS integrates maintenance with supply support, technical data, support equipment, personnel, training, sustaining engineering, and lifecycle-cost considerations (Casadiego Miranda et al., 2021). Consequently, the value of a PHM application should not be assessed only through prediction accuracy or fault-detection performance, but also through its ability to provide actionable information

to these support processes during vessel operation.

From this perspective, and consistent with the lifecycle-support logic of ILS (Casadiego Miranda et al., 2021), the business case for PHM can be expressed as the expected value of avoided failures, avoided downtime, improved maintenance timing, enhanced support planning, and reduced operational and logistic risk relative to the lifecycle cost of sensing, integration, analytics, V&V, and model maintenance.

An underrepresented system with high criticality, observable degradation, sufficient data, and high failure consequences may therefore constitute a strong PHM candidate, whereas a low-cost, easily replaceable, poorly observable, or non-critical component may remain better suited to preventive or run-to-failure maintenance.

This prioritization also implies diminishing returns from expanding PHM coverage indiscriminately. Once high-criticality and data-ready assets have been addressed, the marginal benefit of extending PHM to additional components may decrease while sensing, integration, model maintenance, and false-alarm-management costs remain. Literature gaps should therefore be interpreted as opportunities for further assessment rather than as automatic recommendations for PHM deployment.

## 6. CONCLUSION

This study reviewed how artificial intelligence has been applied to shipboard predictive maintenance from a Prognostics and Health Management (PHM) perspective, focusing on the ship systems addressed, the analytical functions implemented,

the model families used, and the reported data, V&V, and deployment practices. The contribution of the review lies in integrating these dimensions within a common analytical structure rather than in introducing a new systematic-review methodology.

Three principal corpus-derived findings emerge. First, the evidence is strongly concentrated in propulsion and electric-power applications. SWBS 200 contains 56 studies and SWBS 300 contains 16, while four additional studies jointly address both groups. Altogether, 76 of the 83 primary studies (91.6%) involve propulsion and/or electric-power systems.

Second, the functional distribution remains dominated by Detection and Diagnosis. Detection accounts for 34 studies (41.0%), Diagnosis for 18 (21.7%), and Prognosis for 15 (18.1%), whereas Prescription is less frequently represented. The cross-dimensional analysis further shows that the rapid growth in publications since 2020 has not produced a proportional shift toward Prognosis and Prescription. Similarly, SWBS 200 dominates publication volume, but SWBS 300 exhibits a comparatively larger proportion of studies extending toward prognostic or prescriptive functions.

Third, dataset, V&V, and deployment practices remain heterogeneous. Although 47 studies (56.6%) rely primarily on real measured or observational data and 15 (18.1%) use mixed real and simulated/synthetic datasets, complete run-to-failure histories, public accessibility, and standardized dataset reporting remain uneven. Internal model evaluation through train-test partitioning or cross-validation is also reported more frequently than temporally independent, campaign-based, cross-vessel, external, or operational validation. The reviewed evidence therefore shows that real-time computation, use of real or on-board data, and sustained onboard deployment should not be treated as equivalent indicators of evidential or implementation maturity.

These findings indicate that maritime PHM is an unevenly developed field in which publication volume, functional coverage, evidential strength, and deployment maturity represent distinct dimensions. Explanatory mechanisms proposed in the Discussion—including OEM service structures, data access and ownership, contractual incentives, and integration architecture—should be interpreted as author-supported hypotheses rather than causal relationships demonstrated directly by the primary-study corpus.

This review has several limitations. First, eligibility was restricted to English-language publications, which may have underrepresented research disseminated in other languages and regions and may therefore have introduced a language-related selection bias into the evidence map. Second, the search was limited to Scopus and the Web of Science Core Collection. Although these databases provide broad multidisciplinary coverage, relevant studies indexed exclusively in other biblio-

graphic sources, as well as conference proceedings, technical reports, institutional documents, and other grey literature, may not have been captured. No dedicated grey-literature search, backward or forward citation snowballing, or formal search-sensitivity or saturation analysis was performed. Third, terminology in maritime PHM is not fully standardized. Related studies may describe comparable analytical functions using terms such as condition monitoring, fault detection, predictive maintenance, health management, or equipment-specific terminology without explicitly using “PHM.” Similarly, shipboard systems and equipment may be identified through different technical, commercial, or classification-based designations. Although the search strategy incorporated multiple synonyms and the subsequent coding applied operational definitions independent of the terminology used by the authors, relevant studies employing unanticipated terms may still have been omitted. Consequently, the final corpus should be interpreted as a structured representation of the literature retrieved under the defined language, database, and terminology boundaries rather than as an exhaustive inventory of all maritime PHM research.

In addition, no formal study-quality or risk-of-bias scoring instrument was applied. Methodological limitations were instead systematically assessed through the predefined dimensions of data provenance, V&V design, temporal and operational independence, maintenance-event traceability, reporting completeness, and deployment context. The review therefore does not provide quality-adjusted estimates of the evidence.

The analytical strategy combines frequency distributions, proportional comparisons, and cross-dimensional analysis across publication periods, ship-system domains, PHM functions, dataset characteristics, V&V practices, and deployment conditions. This approach enables evidence concentration, functional coverage, evidential strength, and operational implementation patterns to be examined within the same study corpus.

Future research should therefore prioritize not simply underrepresented SWBS domains but assets for which PHM presents a defensible technical and economic case. Candidate systems should be assessed according to operational criticality, safety and regulatory consequences, expected failure and downtime costs, data availability, degradation observability, maintainability leverage, integration effort, and expected lifecycle value.

For selected systems, future work should place greater emphasis on independent operational validation, including chronological and campaign-based evaluation, cross-vessel testing where feasible, traceability to confirmed maintenance events, and prospective onboard assessment. Deployment studies should also report explicitly how inference, model updating, data synchronization, cybersecurity, and decision support are integrated into sustained shipboard and shore-side workflows.

## ACKNOWLEDGMENT

We thank COTECMAR for providing the space and resources needed to carry out this research. We also thank the Technological University of Bolívar (UTB) for access to its facilities and computing services used to run the models. We further acknowledge the support of Minciencias through Convocatoria 950 (2024), which funded the scholarship support that made this work possible.

## NOMENCLATURE

|                |                                                |
|----------------|------------------------------------------------|
| <i>AI</i>      | artificial intelligence                        |
| <i>ANN</i>     | artificial neural network                      |
| <i>CBM</i>     | condition-based maintenance                    |
| <i>CNN</i>     | convolutional neural network                   |
| <i>DL</i>      | deep learning                                  |
| <i>EWMA</i>    | exponentially weighted moving average          |
| <i>GRU</i>     | gated recurrent unit                           |
| <i>ILS</i>     | integrated logistics support                   |
| <i>kNN</i>     | k-nearest neighbors                            |
| <i>LSTM</i>    | long short-term memory                         |
| <i>ML</i>      | machine learning                               |
| <i>MLP</i>     | multilayer perceptron                          |
| <i>OCSVM</i>   | one-class support vector machine               |
| <i>PCA</i>     | principal component analysis                   |
| <i>PdM</i>     | predictive maintenance                         |
| <i>PHM</i>     | prognostics and health management              |
| <i>RAM</i>     | reliability, availability, and maintainability |
| <i>RBF</i>     | radial basis function                          |
| <i>RNN</i>     | recurrent neural network                       |
| <i>RUL</i>     | remaining useful life                          |
| <i>SOPRENE</i> | Neural-Network-Based Predictive Sustainment    |
| <i>SVM</i>     | support vector machine                         |
| <i>SWBS</i>    | ship work breakdown structure                  |
| <i>VAE</i>     | variational autoencoder                        |
| <i>V&amp;V</i> | verification and validation                    |
| <i>WoS</i>     | Web of Science                                 |

## REFERENCES

- Ahmad, R., & Kamaruddin, S. (2012). A review of condition-based maintenance decision-making. *European Journal of Industrial Engineering*, 6(5), 519–541. Retrieved from <https://www.inderscienceonline.com/doi/abs/10.1504/EJIE.2012.048854> doi: 10.1504/EJIE.2012.048854
- Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. doi: 10.1016/j.joi.2017.08.007
- Benarfa, A., Dahmane, S., Brik, B., & Yagoubi, M. B. (2025). Predictive Maintenance for UWSNs: A Practical Implementation of Fault Detection Using NDN and Deep Learning. *Internet Technology Letters*, 8(2), e630. doi: 10.1002/itl2.630
- Bronson, J. A., Fonseca, I., & Gaspar, H. M. (2024). Challenges towards an integrated digital twin platform for maritime systems: Tackling shifts in data ownership. In *Proceedings of the asme 2024 43rd international conference on ocean, offshore and arctic engineering*. doi: 10.1115/OMAE2024-136483
- Casadiago Miranda, L. P., Bastos Blandón, E. A., Paipa Sanabria, E., Ruiz Pianeta, M., Pinzón Ramírez, J. S., & Zúñiga Portillo, N. F. (2021). Analysis of Integrated Logistics Support (ILS) in the Shipbuilding Industry: Advantages and Improvement Opportunities for the Naval Industry. COTECMAR Case Study. *Ship Science & Technology*, 14(28), 9–17. doi: 10.25043/19098642.211
- Cheliotis, M., Lazakis, I., & Theotokatos, G. (2020). Machine learning and data-driven fault detection for ship systems operations. *Ocean Engineering*, 216, 107968. doi: 10.1016/j.oceaneng.2020.107968
- Condition monitoring and diagnostics of machines: Data processing, communication and presentation - part 1: General guidelines* (No. ISO 13374-1:2003). (2003). ISO.
- Coraddu, A., Oneto, L., Ghio, A., Savio, S., Anguita, D., & Figari, M. (2019). Machine learning approaches for improving condition-based maintenance of naval propulsion plants. *Ocean Engineering*, 180, 18–27. doi: 10.1016/j.oceaneng.2019.01.054
- Daya, A. A., & Lazakis, I. (2023). Developing an advanced reliability analysis framework for marine systems operations and maintenance. *Ocean Engineering*, 272. doi: 10.1016/j.oceaneng.2023.113766
- Díez-Oliván, A., Pagán, J. A., Sanz, R., & Sierra, B. (2017). Data-driven prognostics using a combination of constrained k-means clustering, fuzzy modeling and lof-based score. *Neurocomputing*, 241, 97–107. doi: 10.1016/j.neucom.2017.02.024
- Ellefsen, A. L., Æsøy, V., Ushakov, S., & Zhang, H. (2019). A comprehensive survey of prognostics and health management based on deep learning for autonomous ships. *IEEE Transactions on Reliability*, 68(2), 720–740. doi: 10.1109/TR.2019.2907402
- Elmdoost-Gashti, M., Shafiee, M., & Bozorgi-Amiri, A. (2023). Enhancing resilience in marine propulsion systems by adopting machine learning technology for predicting failures and prioritising maintenance activities. *Journal of Marine Engineering & Technology*. Retrieved from <https://www.tandfonline.com/doi/pdf/10.1080/20464177.2023.2243748> doi: 10.1080/20464177.2023.2243748
- En 13306:2017 maintenance - maintenance terminology*. (2017). (European Standard)
- Fernández-Barrero, D., Fontenla-Romero, O., Lamas-López,

- F., Novoa-Paradela, D., R-Moreno, M. D., & Sanz, D. (2021). SOPRENE: Assessment of the Spanish Armada's Predictive Maintenance Tool for Naval Assets. *Applied Sciences*, 11(16), 7322. doi: 10.3390/app11167322
- Gharib, H., & Kovács, G. (2023). A Review of Prognostic and Health Management (PHM) Methods and Limitations for Marine Diesel Engines: New Research Directions. *Machines*, 11(7), 695. Retrieved from <https://doi.org/10.3390/machines11070695> doi: 10.3390/machines11070695
- Gribbestad, M., Hassan, M. U., & Hameed, I. A. (2021). Transfer Learning for Prognostics and Health Management (PHM) of Marine Air Compressors. *Journal of Marine Science and Engineering*, 9(1), 47. doi: 10.3390/jmse9010047
- Hiebl, M. R. W. (2021). Sample selection in systematic literature reviews of management research. *Organizational Research Methods*. (Advance online publication) doi: 10.1177/1094428120986851
- Huang, C., Bu, S., Lee, H. H., Chan, C. H., Kong, S. W., & Yung, W. K. C. (2024). Prognostics and health management for predictive maintenance: A review. *Journal of Manufacturing Systems*, 75, 78–101. doi: 10.1016/j.jmsy.2024.05.021
- IEEE. (2024). *IEEE Standard for System, Software, and Hardware Verification and Validation*. IEEE Std 1012-2024. IEEE.
- IEEE Standard Framework for Prognostics and Health Management of Electronic Systems (No. IEEE Std 1856-2017). (2017). IEEE.
- Inal, O. B., Charpentier, J.-F., & Deniz, C. (2022). Hybrid power and propulsion systems for ships: Current status and future challenges. *Renewable and Sustainable Energy Reviews*, 156, 111965. doi: 10.1016/j.rser.2021.111965
- International Electrotechnical Commission. (2014). IEC 60300, Dependability Management [Computer software manual].
- International Electrotechnical Commission. (2018). IEC 60812:2018, Failure Modes and Effects Analysis (FMEA and FMECA) [Computer software manual].
- International Organization for Standardization. (2015). ISO 13381-1:2015, Condition Monitoring and Diagnostics of Machines – Prognostics – Part 1: General Guidelines [Computer software manual].
- Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483–1510. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0888327005001512> doi: 10.1016/J.YMSSP.2005.09.012
- Kalafatelis, A. S., Nomikos, N., Giannopoulos, A., Alexandridis, G., Karditsa, A., & Trakadas, P. (2025). Towards predictive maintenance in the maritime industry: A component-based overview. *Journal of Marine Science and Engineering*, 13(3), 425. Retrieved from <https://www.mdpi.com/2077-1312/13/3/425> doi: 10.3390/jmse13030425
- Karlsson, C., Arriagada, J., & Genrup, M. (2008). Detection and interactive isolation of faults in steam turbines to support maintenance decisions. *Simulation Modelling Practice and Theory*, 16(10), 1419–1434. doi: 10.1016/j.simpat.2008.08.013
- Kumar, A., & Flores-Cerrillo, J. (2024). *Machine learning in python for process and equipment condition monitoring, and predictive maintenance: From data to process insights*. MLforPSE. Retrieved from <https://play.google.com/store/books/details?id=7JXtEAAAQBAJ> (Machine Learning for Process Industry, Book 3)
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. doi: 10.2307/2529310
- Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014). Prognostics and health management design for rotary machinery systems: Reviews, methodology and applications. *Mechanical Systems and Signal Processing*, 42(1–2), 314–334. doi: 10.1016/j.ymssp.2013.06.004
- Li, J., Goerlandt, F., & Reniers, G. (2021). A bibliometric analysis of maritime safety research using web of science and scopus. *Safety Science*, 138, 105234.
- Li, Z., Yan, X., Yuan, C., & Peng, Z. (2012, August). Intelligent fault diagnosis method for marine diesel engines using instantaneous angular speed. *J. Mech. Sci. Technol.*, 26(8), 2413–2423.
- Liang, Q., Knutsen, K. E., Vanem, E., Aesøy, V., & Zhang, H. (2024). A review of maritime equipment prognostics health management from a classification society perspective. *Ocean Engineering*, 301, 117619. doi: 10.1016/j.oceaneng.2024.117619
- Lim, S., Oh, J., & Park, J. (2025). Maintenance time prediction for predictive maintenance of ship engines. *Applied Sciences*, 15(9), 4764. Retrieved from <https://www.mdpi.com/2076-3417/15/9/4764> doi: 10.3390/app15094764
- Makridis, G., Kyriazis, D., & Plitsos, S. (2020). Predictive maintenance leveraging machine learning for time-series forecasting in the maritime industry. In *2020 IEEE 23rd international conference on intelligent transportation systems (itsc)* (pp. 1–8). doi: 10.1109/ITSC45102.2020.9294450
- Mandala, V., Senthilnathan, T., Suganyadevi, S., Gobhinath, S., Selvaraj, D., & Dhanapa, R. (2023). An optimized back propagation neural network for auto-

- mated evaluation of health condition using sensor data. *Measurement: Sensors*, 29, 100846. doi: 10.1016/j.measen.2023.100846
- Michałowska, K., Riemer-Sørensen, S., Camilla, S., & Hjellset, O. M. (2021). Anomaly detection with unknown anomalies: Application to maritime machinery. *IFAC-PapersOnLine*, 54, 105–111. doi: 10.1016/j.ifacol.2021.10.080
- Michelena, Á., López, V., López, F. L., Arce, E., Mendoza, J., Suárez-García, A., ... Quintián, H. (2023). A fault-detection system approach for the optimization of warship equipment replacement parts based on operation parameters. *Sensors*, 23(7), 3389. doi: 10.3390/s23073389
- MIMOSA. (2010). *Open system architecture for condition-based maintenance (osa-cbm)*. MIMOSA Open Standards for Physical Asset Management. (OSA-CBM Version 3.3.1)
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of web of science and scopus: a comparative analysis. *Scientometrics*, 106(1), 213–228.
- Naval Sea Systems Command. (2005). NAVSEA Cost Estimating Handbook [Computer software manual]. Washington, DC. (Cost Engineering and Industrial Analysis Division (SEA 017); includes Expanded Ship Work Breakdown Structure (ESWBS) cost-estimating guidance)
- Nguyen, K. T. P., Medjaher, K., & Tran, D. T. (2023). A review of artificial intelligence methods for engineering prognostics and health management with implementation guidelines. *Artificial Intelligence Review*, 56, 3659–3709. doi: 10.1007/s10462-022-10260-y
- Noot, D., Birmele, R., & Rey, J. (2024, 10). Maritime data quality for predictive maintenance: Navigating efficiency and reliability. In *Metc/iomass 2024*. Retrieved from <https://oar.a-star.edu.sg/communities-collections/articles/21528> (Conference abstract or preprint record)
- NORSOK. (2002, October). NORSOK Standard Z-014: Standard Cost Coding System (SCCS) [Computer software manual]. Oslo, Norway. Retrieved 2026-08-30, from <https://standards.globalspec.com/std/9966210/z-014> (Rev. 1)
- Ogaji, S. O. T., & Singh, R. (2003). Advanced engine diagnostics using artificial neural networks. *Applied Soft Computing*, 3(3), 259–271. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1568494603000383> doi: 10.1016/S1568-4946(03)00038-3
- Oldfield, M., McMonies, M., & Haig, E. (2014). Next generation condition based predictive maintenance. In *Sps14 conference proceedings*. Retrieved from <https://conferences.chalmers.se/index.php/SPS/SPS14/paper/viewFile/1742/414> (Online proceedings file)
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, n71. doi: 10.1136/bmj.n71
- Park, J., & Oh, J. (2023). A machine learning based predictive maintenance algorithm for ship generator engines using engine simulations and collected ship data. *Energy*, 285, 129269. doi: 10.1016/j.energy.2023.129269
- Scaife, M., Charlton, S., & Mobley, C. (1993). A neural network for fault recognition. In *Sae technical paper*. doi: 10.4271/930861
- SFI Group System. (2001, May). SFI Group System: Product Description [Computer software manual]. (Version 05-2001)
- Solem, B. A. A., Kohtamaki, M., Parida, V., & Brekke, T. (2022). Untangling service design routines for digital servitization: Empirical insights of smart pss in maritime industry. *Journal of Manufacturing Technology Management*, 33(4), 717–740. doi: 10.1108/JMTM-10-2020-0429
- Song, Q., Yang, X., Tang, T., Liu, Y., Chen, Y., & Liu, L. (2024). Capacity prognostics of marine lithium-ion batteries based on icpo-bi-lstm under dynamic operating conditions. *Journal of Marine Science and Engineering*, 12(12), 2355. doi: 10.3390/jmse12122355
- Soualhi, A., Lamraoui, M., Elyousfi, B., & Razik, H. (2022). PHM SURVEY: Implementation of Prognostic Methods for Monitoring Industrial Systems. *Energies*, 15(19), 6909. Retrieved from <https://www.mdpi.com/1996-1073/15/19/6909> doi: 10.3390/en15196909
- Spandonidis, C., & Paraskevopoulos, D. (2023). Evaluation of a deep learning-based index for prognosis of a vessel's propeller-hull degradation. *Sensors*, 23(21), 8956. doi: 10.3390/s23218956
- Ucar, A., Karakose, M., & Kırımca, N. (2024). Artificial intelligence for predictive maintenance applications: Key components, trustworthiness, and future trends. *Applied Sciences*, 14(2), 898. doi: 10.3390/app14020898
- Velasco-Gallego, C., & Lazakis, I. (2023). Mar-rul: A remaining useful life prediction approach for fault prognostics of marine machinery. *Applied Ocean Research*, 140, 103735. doi: 10.1016/j.apor.2023.103735
- Velasco-Gallego, C., Navas De Maya, B., Matutano Molina, C., Lazakis, I., & Cubo Mateo, N. (2023). Recent advancements in data-driven methodologies for the fault diagnosis and prognosis of marine systems: A systematic review. *Ocean Engineering*, 284, 115277. Retrieved from <https://doi.org/10.1016/j.oceaneng.2023.115277> doi: 10.1016/j.oceaneng.2023.115277
- Vorkapić, A., Radonja, R., Babić, K., & Mandić-Ilić, S. (2020).

Machine learning methods in monitoring operating behaviour of marine two-stroke diesel engine. *Transport*, 35, 462–473. doi: 10.3846/transport.2020.14038

Wang, R. (2023). Intelligent Operation and Maintenance System of Marine Equipment Based on PHM. In *Proceedings of the 2023 4th international conference on artificial intelligence in electronics engineering* (pp. 36–45). doi: 10.1145/3586185.3586203

Work breakdown structures for defense materiel items (No. MIL-STD-881F). (2022, 5). Retrieved from <https://www.humphreys-assoc.com/evms/evms-documents/dod/DoD\MILSTD-881F\May2022.pdf> (Revision F, issued 13 May 2022)

Wu, G. (2018). Fault detection method for ship equipment based on bp neural network. In *2018 international conference on robots & intelligent system (icris)* (pp. 556–559). Changsha, China: IEEE. doi: 10.1109/ICRIS.2018.00143

Wu, W., Yan, D., Yan, J., & Wang, X. (2025). Ship behavior prediction and anomaly detection using lstm-dcross model based on ais and remote sensing data. *International Journal of Digital Earth*, 18(1), 2515251. doi: 10.1080/17538947.2025.2515251

Yigin, B., & Celik, M. (2024). A prescriptive model for failure analysis in ship machinery monitoring using generative adversarial networks. *Journal of Marine Science and Engineering*, 12(3), 493. doi: 10.3390/jmse12030493

Zhang, B., Hirayama, K., Ren, H., Wang, D., & Li, H. (2023). Ship anomalous behavior detection using clustering and deep recurrent neural network. *Journal of Marine Science and Engineering*, 11(4), 763. doi: 10.3390/jmse11040763

Zhang, M., Wei, K., He, S., He, Z., & Yan, W. (2023). Performance-based maintenance contract for mission-critical equipment considering spare parts inventory cost-sharing and suppliers' risk attitudes. *International Journal of Production Research*, 61(21), 7524–7539. doi: 10.1080/00207543.2022.2151661

Zhang, P., Gao, Z., Cao, L., Dong, F., Zou, Y., Wang, K., ... Sun, P. (2022). Marine systems and equipment prognostics and health management: A systematic review from health condition monitoring to maintenance strategy. *Machines*, 10(2), 72. doi: 10.3390/machines10020072

## BIOGRAPHIES



**Luis F. Mendoza Cardona** is a Mechatronics Engineer. He is currently pursuing the M.S. degree in Engineering with an emphasis on Computing at the Universidad Tecnológica de Bolívar in Cartagena, Bolívar, Colombia (since early 2025). He has worked as a reliability engineer at COTECMAR under the Integrated Logistic Support (ILS) frame-

work, focusing on reliability-centered practices and maintenance decision support. His research interests include reliability engineering, maintenance, and data analytics, with a particular focus on Prognostics and Health Management (PHM) approaches.



**Joan Suarez** is a mechanical engineer who graduated from the Technological University of Bolívar in 2023. He has developed software prototypes for COTECMAR and web-based interfaces for engineering decision support. His work includes optimization libraries for freight-transport routing (e.g., genetic algorithms) and discrete-event simulation for decision-making under uncertainty using pseudo-random number generation and RAM-based analysis. He currently works at COTECMAR as a Reliability Engineer. He is pursuing the M.S. degree in Engineering with an emphasis on Computing at the Universidad Tecnológica de Bolívar, Cartagena, Bolívar, Colombia (since early 2024).



**Edwin Puertas** is an Artificial Intelligence software architect and Natural Language Processing (NLP) researcher with 20 years of experience spanning academia and industry. He is currently an Associate Professor at the Technological University of Bolívar, where he also serves as Head of the Master's and Doctoral Programs in Engineering. He is an active member of the Artificial Intelligence Standards Committee. His work focuses on bridging academic research and real-world adoption by leading innovative AI projects and designing scalable, production-grade systems that apply advanced machine learning to complex challenges across multiple sectors. His research interests include NLP, human-computer interaction, natural language understanding, and AI-enabled software architectures for data-intensive applications. He has taught courses in AI, NLP, Big Data, and Software Engineering and has led initiatives that connect academic research with industry needs to foster technology transfer and practical impact. Multiple scholarships and awards have supported his contributions. He is a Senior Member of the IEEE and regularly participates in international conferences and workshops, contributing to global standards and best practices in Artificial Intelligence.



**Edwin Paipa** is a Naval Engineer and a Navy officer with professional experience in integrated logistics support (ILS) frameworks and naval sustainment. He has worked on ILS-related initiatives at COTECMAR, contributing to the structuring of lifecycle support strategies for naval assets. He currently serves as a Division Head, leading innovation projects to strengthen operational readiness and main-

tenance capabilities. His most recent work is focused on Maintenance 5.0 initiatives, aligning digital transformation and human-centered technologies with next-generation maintenance practices in maritime and defense contexts.



**Juan C. Martínez-Santos** received the B.Sc. degree in Electronic Engineering (2001) and the M.Sc. degree in Electrical Power (2004) from the Universidad Industrial de Santander, Bucaramanga, Colombia, and the Ph.D. degree in Computer Engineering from Northeastern University, Boston, MA, USA (2013).

He was a Fulbright Scholar — DNP — Colciencias (2007). He has been with the Technological University of Bolívar since 2004 and has been an Associate Professor

since 2007. His research focuses on computer architecture and organization, with applications in hardware support for computer security in multicore and multiprocessor architectures, as well as advanced digital design techniques, including hardware description languages, intellectual property (IP) modules, programmable systems, embedded systems, and hardware/software co-design. Since 2012, he has led an interdisciplinary group with an ICT backbone. He currently teaches in the Faculty of Engineering. He is responsible for courses in Computer Architecture and Assembly (Computer Engineering), Microprocessors (Electrical and Electronics Engineering), Microcontrollers (Mechatronics Engineering), and Advanced Digital Design Techniques (graduate-level, electronics and computer track).