Vladimir Vychuzhanin, Alexey Vychuzhanin

Intelligent Diagnostics of Ship Power Plants: Integration of Case-Based Reasoning, Probabilistic Models, and ChatGPT

A Universal Approach to Fault Diagnosis and Prognostics in Complex Technical Systems

MONOGRAPH

Львів-Торунь
Liha-Pres
2025

UDC 004.891.3:629.5.064:004.896
V 99

Author Team:
V. Vychuzhanin, A. Vychuzhanin

Reviewers:
*Professor **Ye Zhengmao**, Southern University (USA);*
*Professor **A. Kupin**, Kryvyi Rih National University (Ukraine)*

Recommended for publication by the Academic Council
of the National University "Odesa Polytechnic"
(protocol No. 11 dated 27.05.2025)

Vychuzhanin V.

V99 Intelligent Diagnostics of Ship Power Plants: Integration of Case-Based Reasoning, Probabilistic Models, and ChatGPT. A Universal Approach to Fault Diagnosis and Prognostics in Complex Technical Systems: Monograph / V. Vychuzhanin, A. Vychuzhanin. – Lviv-Torun : Liha-Pres, 2025 – 412 p.

ISBN 978-966-397-516-0

DOI 10.36059/978-966-397-516-0

The monograph highlights the following: integrated approach to diagnosing failures in ship's power plants; methodology for determining similarity between cases in the fault diagnosis system of ship's power plants; adaptation of the case-based reasoning method with integration of probabilistic analysis for diagnosis and prognosis of complex systems' technical state; integrated approach to diagnosing complex technical systems, experimental validation and multidimensional efficiency assessment; a hybrid model for evaluating the accuracy of failure forecasts in ship power plants; integrated modeling of reliability and maintenance of spp equipment considering degradation and operational conditions.

The materials of this monograph will be valuable for postgraduate students, master's students, and university instructors specializing in the field of IT technologies.

UDK 004.891.3:629.5.064:004.896

ISBN 978-966-397-516-0

CONTENTS

LIST OF ABBREVIATIONS AND SYMBOLS 6
INTRODUCTION 7
CHAPTER 1
INTEGRATED APPROACH TO FAILURE DIAGNOSTICS
OF SHIP POWER PLANTS BASED ON PRECEDENTS.....10
1.1 Introduction 11
1.2 Review of diagnostic and prognostic methods
for the technical condition of complex technical systems 16
1.2.1 Case-Based Reasoning in technical system diagnostics 16
1.2.2 Probabilistic methods: Bayesian networks
and Markov models 17
1.2.3 Simulation modeling for failure diagnosis and prognostics.....18
1.2.4 Integration of diagnostic methods.....20
1.2.5 Integrated approaches to the diagnosis of SPPs.....20
1.2.6 Need for a new approach 21
1.3 Integrated approach to diagnosing failures in SPPs.....22
1.3.1 Creating a case-based.....22
1.3.1.1 Defining case structure 23
1.3.1.2 Data integration and case based formation 27
1.3.2 Methodology for determining similarity between cases
in the fault diagnosis system of SPP.....45
1.3.2.1 Introduction.....45
1.3.2.2 Methods for calculating similarity between cases 48
1.3.2.3 A model for determining similarity between precedents
for diagnosing failures of Bayesian networks equipment.....50
1.3.2.4 Optimization of weights of parameters for estimation
of similarity of failure precedents.....79
1.3.2.5 Practical application of the method for assessing similarity
of equipment failure cases in SPP.....84
1.3.2.6 Discussion of results.....89
1.3.2.7 Conclusions.....91
1.3.3. Development of a technical condition assessment algorithm
for complex systems based on probabilistic failure estimation 93
1.3.3.1 Introduction.....93
1.3.3.2 Materials and methods.....96
1.3.3.3 Results 100
1.3.3.4 Discussion of results.....110
1.3.3.5 Conclusions.....112

1.3.4 Adaptation of the Case-Based Reasoning method with integration of probabilistic analysis for diagnosis and prognosis of complex systems' technical state 114
1.3.4.1 Introduction....114
1.3.4.2 Materials and methods....116
1.3.4.3 Fundamental principles of CBR decision adaptation....117
1.3.4.4 Formalization of diagnostic method integration....123
1.3.4.5 General experimental evaluation of adaptive CBR....130
1.3.4.6 Discussion of results....144
1.3.4.7 Conclusions....146
1.3.5 Integrated approach to diagnosing complex technical systems: experimental validation and multidimensional efficiency assessment....147
1.3.5.1 Introduction....147
1.3.5.2 Purpose and objectives of experimental testing....149
1.3.5.3 Test scenarios for the technical condition of the SPP....152
1.3.5.4 Evaluation of Accuracy, Precision, Recall, and F1-Score metrics for various methods of diagnosing the technical condition of the SPP....155
1.3.5.5 Discussion of results....168
1.3.5.6 Conclusions....169
1.3.6 Three-Scenario analysis of fault diagnosis accuracy in complex technical systems....171
1.3.6.1 Introduction....172
1.3.6.2 Materials and methods....174
1.3.6.3 Results....176
1.3.6.4 Discussion of results....217
1.3.6.5 Conclusions....219
1.3.7 A hybrid model for evaluating the accuracy of failure forecasts in SPPs....220
1.3.7.1 Introduction....220
1.3.7.2 Materials and methods....223
1.3.7.3 Methodology for evaluating the accuracy of failure forecasts in SPP equipment....224
1.3.7.4 Results....226
1.3.7.5 Discussion of results....247
1.3.7.6 Conclusions....249
1.3.8. Dynamics of failure probabilities in SPP equipment considering cascade effects....250

1.3.8.1 Introduction....250
1.3.8.2 Materials and methods....253
1.3.8.3 Results....256
1.4.8.4 Discussion of results....278
1.3.8.5 Conclusions....281
1.3.9 Integrated modeling of reliability and maintenance of SPP equipment considering degradation and operational conditions....282
1.3.9.1 Introduction....282
1.3.9.2 Materials and methods....285
1.3.9.3 Results....287
1.3.9.4 Discussion of results....313
1.3.9.5 Conclusions....316
1.3.10 Adequacy and verification of an intelligent diagnostic model for SPPs....317
1.3.10.1 Introduction....317
1.3.10.2 Materials and methods....320
1.3.10.3 Results....322
1.3.10.4 Discussion....355
1.3.10.5 Conclusions....357
1.3.11. Discussion of results....358
1.3.11.1 Impact of method integration on diagnostic accuracy....358
1.3.11.2 Correlation analysis of predictions and actual failures....359
1.3.11.3 Impact of Bayesian networks on reducing diagnostic errors....361
1.3.11.4 Impact of simulation modeling on CBR adaptation....361
1.3.11.5 Evaluation of residual life prediction for components....362
1.3.11.6 Final analysis of method integration effectiveness....363
1.3.11.7 Limitations of the proposed method....363
CONCLUSIONS....364
CHAPTER 2
USING CHATGPT FOR THE INTELLIGENT DIAGNOSTICS OF COMPLEX TECHNICAL SYSTEMS....367
2.1.Introduction....367
2.2 Materials and methods....369
2.3 Results....371
2.4 Discussion of results....284
2.5 Conclusions....288
REFERENCES....390

LIST OF ABBREVIATIONS AND SYMBOLS

AI - Artificial Intelligence
BD - Database
BZ - Knowledge Base
BN - Bayesian Network
CA - Critical Application
CBR - Case-Based Reasoning
CCS - Complex Critical Systems
CSM - Cognitive Simulation Model
CTS - Complex Technical System
DES - Discrete-Event Simulation
DM - Decision Maker
DTRS - Data Transmission and Reception System
FC - Set of Intercomponent, Interelement Connections
FE - Set of Subsystems, Components, Elements
IIS - Intelligent Information System
k-NN - k-nearest neighbors
L-BFGS-B - Limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm with Box
MAE - Mean Absolute Error
MAPE - Mean Absolute Percentage Error
ME - Main Engine
ML - Machine Learning
MM - Markov Model
MSE - Mean Squared Error
OLAP - Online Analytical Processing
OREDA - Offshore Reliability Data database
PHM - Prognostics and Health Management
R^2 - Coefficient of Determination
RMSE - Root Mean Squared Error
RUL - Remaining Useful Life
SIM - Simulation Modeling
SPP - Ship Power Plant
SW - Software
TC - Technical Condition
TM - Technical Maintenance
TS - Technical System
TTF - Time-to-Failure

INTRODUCTION

One of the main causes of technogenic accidents involving CTS used in transportation, aviation, energy, and other fields remains operational equipment failures. An analysis of the results of ship operations shows that despite measures taken to ensure maritime safety, the number of accidents at sea remains high. A primary reason for many of these accidents is the failure of CTS. As a result, such systems are classified as critical application systems. Complex technical systems of critical application are hierarchical structures with non-trivial internal organization, multifunctional subsystems, components, and elements, interconnected with complex links and subject to various failure states. A characteristic feature of CTS operation is uncertainty, as well as incomplete information about the technical condition of the systems.

Due to increasing demands for safety and reliability of expensive maritime CTS, their efficiency depends significantly on extending operational life and resource use. Enhanced efficiency can be achieved by applying models and methods for diagnosing, assessing, and forecasting the TC of complex systems and integrating them into intelligent information systems. These IIS enable the evaluation and prediction of TC based on diagnostic results. Existing diagnostic models and methods are widely used in practice, but they do not always ensure comprehensive operational efficiency of CTS. Additionally, current TC diagnostic models often account only for full functional failures but overlook partial ones. Partial failures are more diverse in their types and locations of manifestation compared to full failures. Advanced diagnostic algorithms are required to meet efficiency demands in decision-making while considering the continuation of CTS operation.

Promising modeling methods for TC diagnostics include Bayesian belief networks, which account for uncertainties and incomplete data of modeled CTS, and cognitive simulation modeling methods, which additionally evaluate the structural and functional vulnerabilities of system equipment. However, cognitive simulation modeling requires improvements due to its limitations: lack of universality regarding structural threats and vulnerabilities in CTS, and insufficient consideration of the significance and criticality of equipment for overall system functionality.

Known methods for assessing and predicting TC in complex systems implemented in IIS include case-based reasoning; analogies; systematic and heuristic methods for optimization tasks (e.g., genetic algorithms, artificial immune networks, annealing methods, swarm intelligence methods

including ant algorithms); and structural representation methods based on OWL ontology precedents. Some of these methods face challenges such as high algorithmic and computational complexity, the necessity of intricate preliminary data processing stages, and limited visualization capabilities for interpreting results. A general drawback is the high dimensionality of possible tasks during decision-making.

Improving the efficiency of CTS operation by applying diagnostic, assessment, and forecasting models and methods that consider both partial and full equipment failures is a critical scientific problem. The aim of this research is to enhance the operational efficiency of CTS by developing models and methods for diagnosing, assessing, and forecasting the TC of critical application complex systems.

To achieve this aim, the following tasks were identified and resolved. Analysis of models, methods, and information systems for diagnosing, assessing, and forecasting the TC of critical application CTS. Development of stochastic models and a method for diagnosing the TC of critical application CTS. Creation of an IIS for diagnosing, assessing, and forecasting the TC of CTS. The processes of diagnosing, assessing, and forecasting the TC of critical application CTS.

The models and methods for diagnosing, assessing, and forecasting the TC of critical application CTS. To achieve the research goals, mathematical, simulation, and computer modeling methods were used, along with expert evaluation and theories of information, control, decision-making, graphs, artificial intelligence, cognitive analysis, literature content analysis, data processing, diagnostics, and forecasting. Methods of theoretical, applied, and object-oriented programming were employed in developing conceptual stochastic diagnostic models for CTS. Optimization theory methods were applied in data transmission modeling, as well as in diagnosing, assessing, and forecasting CTS.

The scientific novelty of the obtained results includes. A stochastic diagnostic model for CTS was proposed, which simultaneously accounts for the presence of subsystems, components, and elements, their interconnections, and the probabilities of partial or complete functional failure. This led to the development of a diagnostic method based on a Bayesian belief network for critical application CTS.

A data transmission and reception model for diagnosing, assessing, and forecasting the TC of CTS was developed. It considers conflicting requirements and competing criteria, enabling the identification of Pareto-optimal solutions for effective data transmission and reception.

Further development was achieved in the diagnostic method for CTS based on a Bayesian belief network, enabling the timely identification and visualization of structural and functional vulnerabilities, thus enhancing the operational efficiency of critical application systems. The case-based reasoning method, ensuring TC assessment and prediction to improve the performance of CTS. Improvements were made to the cognitive simulation model, incorporating simulation-impact impulses, which allows for diagnosing equipment TC with consideration of interdependencies and mutual influences.

The practical significance lies in the development of an IIS that automates the processes of assessing and predicting the TC of critical application CTS in various states of functionality.

An algorithm was created to detect failures in subsystems, components, and elements, including their interconnections, based on risk assessments of these failures. This algorithm enables the implementation of a targeted IIS operation strategy. A user interface for the knowledge base was developed, allowing experts to review formalized data and make final risk assessments for equipment failures in CTS.

CHAPTER 1

INTEGRATED APPROACH TO FAILURE DIAGNOSTICS OF SHIP POWER PLANTS BASED ON PRECEDENTS

Abstract

This section presents an integrated fault diagnosis method for SPPs, combining Case-Based Reasoning, probabilistic models (BNs, MMs), and cognitive simulation modeling.

Unlike traditional diagnostic systems, the proposed method accounts for cascading failure effects, probabilistic dependencies between components, and stochastic equipment degradation. A dynamic diagnostic model update mechanism has been developed, enabling real-time adaptation of CBR weight coefficients and failure probabilities based on operational data. Cognitive simulation modeling is applied for predicting complex failure scenarios, generating synthetic data, and identifying hidden patterns in fault progression. The use of a hybrid diagnostic approach, integrating k-NN with adaptive weights, BNs, MMs, and simulation models, improves diagnostic accuracy and reduces error rates.

Optimization of diagnostic decisions using L-BFGS-B and probabilistic analysis allows for dynamic failure prediction adjustments and consideration of equipment degradation processes. Experimental testing on an SPP simulation model confirmed the effectiveness of the proposed method: diagnostic accuracy increased to 92% (13% higher than standard CBR), false positive detections decreased by 6.7%, and residual life prediction error remained within 5–7%.

The developed method can be applied in intelligent monitoring and prognosis systems for SPPs and adapted for other complex technical systems, including aviation and industrial power plants.

1.1 Introduction

Modern CTS, used in transportation, aviation, energy, and other industries, are hierarchical structures comprising numerous functionally diverse subsystems, components, elements, and complex interconnections among them. These elements may exist in states of either partial or complete loss of operability. The structure of a CTS reflects the overall picture of cause-and-effect interactions among the system's constituent parts [1, 2, 3].

The operation of such systems is associated with a high level of uncertainty, which is difficult to fully describe, understand, or predict. CTS possess a set of characteristic properties, including nonlinearity, adaptability, self-organization, and integrity [2, 3]. Adaptability refers to the system's ability to function in a multitude of different states, which is enabled, in particular, by informational components incorporating elements of artificial intelligence [2, 4, 5]. Self-organization is manifested in the system's ability to alter its properties and return to its original state after deviation. Integrity is expressed through the preservation of the system's overall qualities.

Based on the degree of operational certainty, CTS are classified as probabilistic or deterministic; based on the level of organization — as well-organized or poorly organized. Depending on the nature of interaction with the external environment, CTS are further classified as closed or open systems [2, 3].

Each device within a CTS is characterized by a set of parameters that determine its current state. Changes in the properties of individual components, subsystems, or their interconnections may cause cascading effects throughout the entire system. The functioning of CTS is based on systemic principles [2, 3]: the principle of hierarchical belonging (an element to a component, a component to a subsystem, a subsystem to the CTS) and the principle that the properties of the system cannot be reduced to the simple sum of the properties of its individual parts.

A typical example of a CTS is the shipboard technical complex: mechanisms; assemblies; devices; pipelines; other interdependent systems that ensure vessel operation [4, 6]. In particular, the SPP

represents a CTS comprising various subsystems, components, and elements that are closely interconnected and serve different purposes.

One of the main causes of man-made accidents during the operation of CTS in various industries remains the failure of their constituent parts [7]. For this reason, such systems are classified as critical. In reliability theory, CTS are regarded as a set of properties observed under specific environmental conditions at a given moment in time. A failure is considered a random event associated with a partial or complete loss of operability. A partial failure is viewed as an intermediate state between full operability and complete failure, wherein the system loses the ability to perform some of its functions due to reduced effectiveness. These systems are considered multi-state systems.

External factors increase the load on individual elements of CTS, leading to a decline in their functional capacity and overall reliability, which, in turn, raises the risk of accidents. Other contributing causes include manufacturing defects, violations of operational modes, and human error [8, 9, 10].

Among the sectors where particularly high demands are placed on the efficiency and reliability of critical CTS, maritime and inland waterway transport are especially prominent. Ships are equipped with dozens of CTS, the condition of which directly affects their survivability. This survivability cannot be guaranteed solely through adherence to regulatory requirements at the stages of design, construction, and operation [11, 12, 13].

Databases exist containing information on maritime accidents and incidents [14], including the Global Integrated Shipping Information System (GISIS) maintained by the International Maritime Organization (IMO) [15], as well as the European Maritime Safety Agency (EMSA) [16]. For example, according to data from the JTSB database, the number of accidents in the Japanese shipping zone from 2008 to 2024 shows only a slight downward trend [17] (Fig. 1.1).

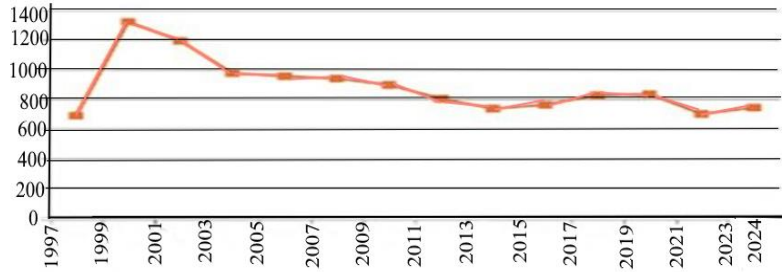


Figure 1.1. **Trend in overall marine vessel accident rates**

This confirms that the issue of reliability remains relevant both for new and aging vessels - especially large ships equipped with advanced control and communication systems, which make their subsystems more vulnerable [18, 19, 20].

Analysis of operational data indicates that, despite the implementation of safety measures for maritime navigation, the level of marine accidents remains high. One of the most frequently occurring causes of shipboard incidents is CTS failure. Marine accidents pose a serious threat to human life, the ship itself, the environment, and coastal infrastructure [21, 22, 23].

Thus, even with adherence to normative and technical safety measures, the probability of accidents caused by CTS failures remains high. Statistical analysis demonstrates a direct relationship between displacement, vessel age, and the risk of technical failure. As safety requirements for high-value CTS grow, so do the requirements for their operational efficiency, which is determined by service life and resource usage. Ensuring the reliable and efficient operation of shipboard CTS remains a significant scientific and technical challenge.

Modern ship power plants (Fig. 1.2) are CTS (operating under highly variable conditions with stochastic failure patterns [2, 24]. Reliable diagnostics of such systems require a comprehensive approach that combines historical data analysis, probabilistic forecasting, and dynamic system behavior modeling.

Figure 1.2. **Main propulsion SPP**

SPP is designed for the production of mechanical, electrical, thermal, and other types of energy. The SPP includes the following components: Main SPP (MSPP), which provides propulsion for the vessel. It consists of main engines, steam boiler units, steam turbine power plants, free-piston gas generators, gas turbine unit turbines, nuclear reactors of nuclear power plants, as well as auxiliary mechanisms and systems that support their operation; Auxiliary SPPs (ASPPs), which supply all types of energy to the ship's systems and mechanisms, ensuring the proper functioning of the MSPP. These include auxiliary emergency diesel generators, turbo generators, auxiliary boiler units, a battery station, as well as their supporting auxiliary mechanisms, devices, and systems; Mechanisms and devices of general ship systems, including domestic water supply, heating and space heating, ventilation and air conditioning, bilge, ballast, and fire-fighting systems, as well as systems for collecting and treating oily and sewage-fecal waters, and others; Automatic and remote control, monitoring, and protection systems for the MSPP, ASPP, and mechanisms and devices of general ship systems.

All types of energy produced by the SPP ensure the vessel’s movement at a set speed or the performance of various operational

tasks on cargo ships, technical fleet vessels, and special-purpose ships. They also ensure the safe and reliable operation of engine room mechanisms, deck equipment and systems, electrical lighting, navigation aids, automation and signaling control systems, and meet the general and domestic needs of the crew and passengers.

Traditional methods based on strict mathematical models and expert rules have several limitations in maritime operations [1]:

- high variability in operating conditions (loads, environmental factors, wear) complicates the development of universal diagnostic models [1];
- the lack of consideration for probabilistic dependencies between components reduces diagnostic accuracy [24];
- insufficient historical data leads to uncertainty in identifying rare failures [25].

To address these challenges, an integrated fault diagnosis approach for SPPs is proposed, combining:

1. CBR [26] - analyzing historical failure data with adaptability to new operating conditions;
2. Probabilistic risk assessment methods [27] - BNs and MMs for predicting component degradation and cascading failure effects;
3. Simulation models [1] - enabling dynamic system behavior reproduction and generating missing diagnostic data.

Key innovations and advantages of the proposed approach:

- comprehensive failure analysis - integrating CBR, probabilistic methods, and simulation modeling allows not only the identification of current failures but also the prediction of their progression;
- adaptability to changing conditions - the proposed method operates effectively even with limited historical data by utilizing probabilistic forecasting;
- cascading failure modeling - BN capture component dependencies, helping to prevent secondary failures;
- residual life prediction - MMs estimate equipment degradation probabilities based on operational factors.

Thus, the developed method integrates the capabilities of CBR, probabilistic analysis, and simulation modeling, ensuring enhanced

accuracy, adaptability, and reliability in diagnosing failures in CTS, particularly in SPPs.

1.2 Review of diagnostic and prognostic methods for the technical condition of complex technical systems

In recent years, there has been significant progress in the development of diagnostic and prognostic methods for CTS, including SPPs. Diagnosing such systems is challenging due to the high variability of operating conditions, the stochastic nature of failures, and the need to account for cascading effects.

This review examines modern diagnostic methods, their integration, and future development prospects.

1.2.1 Case-Based Reasoning in technical system diagnostics [26, 27, 28]

CBR is widely used in CTS diagnostics because it leverages accumulated experience to assess faults. In this approach, the diagnosis is established by comparing the current failure with past cases stored in a knowledge base. Key advantages of CBR: utilization of historical data for diagnosing new cases; adaptability to different operating conditions; ease of interpretation of diagnostic decisions.

However, CBR has several limitations: the need to adapt solutions when there is no exact match in the Case Base; limited applicability to rare or novel failures; challenges in automatic database updates; the necessity of optimizing the case base, accounting for probabilistic dependencies between parameters, and reducing computational load when processing large datasets.

A review of existing CBR optimization approaches reveals that considerable efforts have been devoted to improving case retrieval and case base management.

In [28, 29], an enhanced similarity search method for CBR-based diagnostics of aircraft engines was proposed. This approach considers attribute interactions, improving diagnostic accuracy. However, its high computational complexity and limited applicability

(aviation engines) make it less universal. Maria Salamó and Elisabet Golobardes [30] introduced Dynamic Case Base Maintenance (DCBM), which optimizes case base size without compromising accuracy. However, this method does not account for case interactions or efficiency in handling large datasets. In [31], a method for updating the case base and feature dictionary based on belief function theory was proposed. While this method automatically adjusts the knowledge base size, its implementation complexity and potential increase in computational load remain open issues. Ayed et al. developed the CIMMEP method, which removes duplicate cases but requires manual configuration [32].

Despite extensive research in CBR optimization, challenges related to dynamic updates and handling uncertainty remain relevant. To enhance the effectiveness of CBR, a structured case base must be developed, incorporating not only historical failure data but also probabilistic assessments, cascading effect information, operational parameters, and simulation-based failure modeling results.

To overcome the limitations of CBR in diagnosing and predicting CTS equipment failures, it is often complemented by other methods such as Bayesian networks, Markov models, and simulation modeling.

1.2.2 Probabilistic methods: Bayesian networks and Markov models [27, 33, 34]

Probabilistic models are widely used for assessing the technical condition of complex systems. BNs allow for the consideration of interdependencies between components and enable failure probability estimation based on incoming data. MMs, in turn, are used for predicting the remaining useful life of equipment and evaluating the likelihood of transitions between operational and failure states.

Advantages of probabilistic methods: consideration of uncertainty and cascading failures; ability to update estimates as new data becomes available; high accuracy in failure prediction.

Disadvantages: high computational complexity; challenges in model parameter calibration.

1.2.3 Simulation modeling for failure diagnosis and prognostics [35, 36]

Various types of simulation modeling are used for diagnosing and predicting failures in complex technical systems. The main approaches include. Discrete Event Simulation (DES) [37]: used for systems whose operation is based on a sequence of discrete events (e.g., equipment operation and maintenance); allows for analyzing failure impacts, simulating maintenance processes, and identifying bottlenecks; commonly applied in reliability forecasting and maintenance optimization. System Dynamics (SD) [38]: applied to model complex interdependent processes such as equipment degradation and the influence of external factors on system reliability; uses differential and difference equations to describe changes in system states over time. Agent-Based Modeling (ABM) [39]: models the system as a collection of interacting agents, which can represent equipment components, operators, and maintenance personnel; enables the analysis of system behavior under various operational and failure scenarios. Monte Carlo simulation [40]: based on random variables to model uncertainties in operational, diagnostic, and failure prediction processes; used for failure probability assessment, risk analysis, and mean time between failure (MTBF) estimation. Hybrid simulation [41]: combines multiple approaches, such as discrete event simulation with agent-based modeling or system dynamics; allows for different levels of detail and complex interrelations within technical systems. Cognitive simulation modeling [1, 42]: used for diagnosing and predicting failures in CTS, particularly under conditions of uncertainty, dynamic changes, and insufficiently formalized data; represents the system as a network of interconnected factors (cognitive maps) influencing each other with varying intensity; enables causal relationship analysis and system behavior forecasting based on parameter variations; accounts for both quantitative indicators and qualitative aspects, including expert knowledge, intuitive perception, and cognitive failure analysis strategies.

The cognitive model integrates formal mathematical approaches (e.g., MMs and BNs) with expert evaluations, which is particularly valuable in cases of incomplete or noisy data. It dynamically adapts by adjusting predictions based on new experiences, similar to CBR, thus improving diagnostic accuracy under changing operational conditions. Cognitive simulation modeling helps generate complex failure scenarios while considering contextual factors and component interdependencies, leading to the identification of hidden failure patterns.

In an integrated approach, it can act as a synthetic data generator to enrich the CBR case base, and its results can be used to calibrate probabilistic assessments (e.g., adjusting weights in BNs or MMs).

This strengthens the link between historical data and current operating conditions. The inclusion of cognitive simulation modeling extends the capabilities of the integrated approach by incorporating cognitive aspects of decision-making. This is particularly important in working with complex and dynamic systems, where human experience plays a crucial role.

Failure Factor Analysis: modeling the impact of wear, operating conditions, and external influences on system reliability; utilizing Fuzzy Cognitive Maps (FCM) to account for uncertain dependencies.

Failure Prediction and Risk Assessment: estimating failure probabilities through scenario analysis; integrating with artificial intelligence methods (e.g., neural networks) for learning from historical data.

Optimization of maintenance strategies: identifying critical system components for prioritized maintenance; supporting predictive maintenance (Predictive Maintenance - PdM).

Evaluation of diagnostic solutions efficiency - analyzing the impact of various diagnostic strategies (vibration monitoring, thermal imaging analysis, etc.) on failure detection accuracy.

Advantages of cognitive simulation modeling: integration of qualitative and quantitative factors - considers both measurable parameters and expert assessments; adaptability - adjusts to changing operational conditions, generating data for various scenarios;

enhancement of case-based knowledge - creates synthetic data to enrich the CBR system database, improving diagnostic completeness and accuracy.

Data obtained from cognitive modeling can be used to calibrate weight coefficients in probabilistic models (e.g., BNs and Markov processes). This minimizes diagnostic errors and improves failure prediction accuracy.

Simulation models are used to replicate system behavior under various operating conditions, including: evaluating component reliability in real working environments; generating artificial data for training diagnostic models; analyzing operational scenarios and their impact on system performance.

The main limitation is high sensitivity to input data accuracy - errors in initial parameters can significantly affect modeling results.

1.2.4 Integration of diagnostic methods

Modern research indicates that combining various diagnostic methods enhances the accuracy and reliability of failure predictions [1]. Examples of method integration: combining CBR and BNs for probabilistic decision correction [43]; using simulation modeling to create training data for CBR; applying hybrid models based on machine learning and statistical analysis for remaining useful life (RUL) estimation [44].

1.2.5 Integrated approaches to the diagnosis of SPPs

A SPP is a complex technical system ensuring uninterrupted vessel operation, maneuverability, and autonomy [45].

Main components of SPP:

1. SPP - provides vessel propulsion and includes: main engine - typically diesel, gas turbine, or steam turbine; reduction gear - transmits power from the engine to the propeller shaft; propeller - converts mechanical energy into vessel movement; auxiliary power units: diesel generators - supply electrical power to the vessel; steam generators - used on large vessels for producing steam, e.g., for turbines;

2. Fuel system - stores and supplies fuel to the engines;

3. Cooling system - maintains engine temperature control;

4. Lubrication system - reduces friction in mechanical components;

5. Hydraulic system - operates steering mechanisms and other hydraulic devices;

6. Automation and control system - monitors all components, controls parameters, and diagnoses faults.

Types of SPPs: diesel - most common, economical, and reliable; steam turbine – used on large vessels and warships; gas turbine - high power-to-weight ratio; nuclear - used on nuclear icebreakers and submarines.

The complexity and scale of marine power plants require a comprehensive approach to monitoring and diagnostics. Research indicates that integrating CBR, probabilistic models, and simulation modeling significantly enhances failure prediction accuracy.

1.2.6 Need for a new approach

Despite significant progress in diagnostic methods, several unresolved issues remain: limited accuracy of traditional methods under changing operational conditions; insufficient integration of methods in real-world applications; lack of adaptive models capable of real-time corrections.

This study proposes a new integrated approach combining CBR, probabilistic models, and simulation modeling for diagnosing SPP. This approach aims to improve diagnostic accuracy, detect potential failures in advance, and optimize maintenance strategies.

An analysis of modern diagnostic methods demonstrates that integrated approaches provide the most effective solutions. The combination of CBR, BNs, MMs, and simulation modeling enables consideration of complex operational conditions and enhances failure prediction reliability. Developing and implementing such methods is a promising direction for improving the reliability of SPP**s**.

The objective of this study is to develop an integrated approach to the diagnosis of failures in SPPs, combining CBR, simulation

modeling, and probabilistic failure risk assessment into a unified diagnostic system. This approach will enhance fault detection accuracy, account for cascading failure effects, and ensure system adaptability to changing operating conditions.

To achieve this objective, the following tasks must be addressed: develop an integrated failure case base for marine power plants, combining real historical data with artificially generated scenarios obtained through simulation modeling; define a similarity measurement methodology between failure cases to ensure the correct identification of analogous cases while considering probabilistic failure dependencies; design a system health assessment algorithm that integrates CBR, probabilistic failure assessment of components, and dynamic prediction of remaining useful life; create a mechanism for adapting CBR decisions by integrating probabilistic failure analysis and predictive maintenance, allowing for the consideration of component degradation over time; conduct experimental testing of the proposed integrated approach on a simulation model of a SPP, evaluating diagnostic accuracy, fault prediction efficiency, and the impact of the proposed adaptation mechanisms.

Thus, this study aims to develop a next-generation diagnostic system capable of adapting to changing operating conditions, incorporating probabilistic failure characteristics, and providing enhanced fault diagnosis accuracy in marine power plants.

1.3 Integrated approach to diagnosing failures in SPPs

1.3.1 Creating a case-based

The purpose of this study is to develop and justify a comprehensive approach to diagnosing failures of SPP by creating a case base reasoning precedent base that combines probabilistic models and simulation modelling.

This approach makes it possible to take into account the stochastic nature of failures, cascade effects and dynamic changes in operating conditions, which increases the accuracy of diagnostics and failure prediction.

1.3.1.1 Defining case structure

For the effective application of CBR in diagnosing SPP failures, it is necessary to standardize the representation of each failure case. The structure of a case must include both quantitative and qualitative characteristics of the failure, allowing for accurate comparison, classification, and adaptation of diagnostic solutions for SPP equipment failures [46, 47]. In this study, the case structure includes the following key elements.

Failure description: failure type: classification of failures by category (e.g., mechanical, electrical, hydraulic, etc.). A numerical encoding system is used for quantitative assessment (e.g., 1 - mechanical, 2 - electrical, etc.); failure causes: a list of possible causes identified by experts, indicating their contribution to the overall failure scenario; failure consequences: description of the impact of the failure on system performance (reduced efficiency, risk of accidents, need for repairs). Failure risk assessment.

Harrington's desirability function: a generalized desirability function is used to classify failure risk levels: 0 - 0.2: minimal risk (failures not affecting operation); 0.2 - 0.37: acceptable risk (failures allowing operation without immediate repair); 0.37 - 0.63: maximum risk (failures requiring repair to continue operation); 0.63 - 1.0: critical risk (failures making operation impossible); failure Probability: the calculated probability of component failure, based on statistical data (e.g., from the OREDA database) and expert evaluation. Component characteristics: component condition (Si): a discrete value characterizing the state of the component (0 - operational, 1 - degradation, 2 - pre-failure, 3 - failure); remaining Useful life (Ri(t)): an estimation of the remaining service life of the component, calculated using Markov processes; failure Rate (λ_i): a parameter determined based on historical data.

Diagnostic methods: description of the applied methods (CBR, BNs, simulation modeling); including details on the algorithms and similarity measures used for case analysis.

Data sources: sources of information used to generate the case (maintenance logs, classification society reports, OREDA database,

simulation modeling results). As an example, the structure of an individual case can be presented in Table 1.3.1.

Table 1.3.1. **Case structure of SPP equipment failures**

Parameter	Description	Method/Source
Failure type	Failure code (1 — mechanical, 2 — electrical, etc.)	Expert assessment, technical documentation
Failure causes	List of identified failure causes	Log analysis, expert opinion
Failure consequences	Description of the impact on the system (efficiency reduction, accident risk, etc.)	Technical documentation, reports
Failure risk (desirability)	Value based on desirability function (D=1)	Harrington function calculation
Failure probability	Failure probability assessment (e.g., 0.05)	Statistical data (OREDA)
Unit condition (Si)	0 (operational), 1 (degradation), 2 (pre-failure), 3 (failure)	Expert assessment, condition sensors
Remaining resource (Ri(t))	Assessment of the remaining service life of the unit	MMs
Diagnostic methods	Applied algorithms and similarity measures	CBR, BNs, simulation modeling

Example case - fuel pump failure: failure description: unstable fuel supply, decreased system pressure; cause: wear of the pump's working elements; consequences: loss of engine power, overheating; failure probability: 0.32 (based on operational data); measures taken: pump replacement, filter inspection, fuel viscosity analysis.

This case is stored in the case base and can be used for diagnosing similar failures in the future.

Formalized case structure as a JSON object. This format is well-suited for machine processing and visual representation of relationships:

```
{
  "failure_case": {
    "identifier": {
      "code": "UNQ-12345",
      "date": "2025-03-19",
      "source": "OREDA"
    },
    "failure_type": {
      "category": "mechanical",
      "causes": ["wear", "fatigue damage"]
    },
    "operational_context": {
      "working_conditions": {
        "temperature": "75°C",
        "vibration": "increased",
        "load": "90%"
      },
      "operating_mode": "overload",
      "external_factors": ["oil contamination", "low fuel quality"]
    },
    "failure_risk": {
      "probability": 0.15,
      "risk_category": {
        "method": "Harrington",
        "level": "high"
      },
      "expected_damage": {
        "cost": 15000,
        "safety_impact": "critical"
      }
    },
    "interconnected_components": {
      "subsystems": ["fuel system", "hydraulics"],
      "connection_type": {
        "mechanical": true,
        "electrical": false,
        "informational": true
      },
      "cascade_effects": ["pump damage", "overheating"]
```

```
        },
        "data_sources": {
          "historical": ["OREDA", "maintenance logs"],
          "simulation":["Bayesian networks","discrete-event modeling"],
          "sensor_based": {
            "IoT": true,
            "SCADA": true
          },
          "expert_assessments": ["maintenance engineers", "diagnostic reports"]
        },
        "diagnostic_methods": {
          "CBR": {
            "case_retrieval": "k-NN",
            "adaptation": "gradient methods"
          },
          "Bayesian_networks": {
            "analysis": "probabilistic"
          },
          "hybrid_methods": {
            "CBR_and_machine_learning":["L-BFGS-B", "regression"],
            "CBR_and_simulation": ["Markov processes", "Bayesian networks"],
            "weight_optimization": ["gradient methods"]
          }
        }
      }
    }
```

Such a representation reveals: structured approach - data relationships are well-organized; flexibility - easily expandable with new parameters; readiness for processing - suitable for diagnostic and analytical systems.

This structure includes key failure parameters, data sources, and diagnostic methods, allowing for the standardization of the case database and improving the efficiency of finding similar cases.

1.3.1.2 Data integration and case based formation

The knowledge base of the diagnostic system includes a case-based, as well as information on failure probability characteristics, cause-and-effect relationships, and equipment degradation models. The interaction of these components ensures adaptive diagnostics and prediction of the technical condition of SPPs.

Main components of the knowledge base: case base - contains historical data on failures, their causes, consequences, and corrective actions taken; probabilistic failure models - used to assess failure risks based on statistical data and expert evaluations; simulation models - enable the creation of degradation scenarios, including rare and cascading failures, which is especially valuable when real-world data is scarce.

Table 1.3.2. **Main data categories in the knowledge base**

Data category	Description	Data source
Case-based database	Historical data on failures, their causes, and consequences	Operational data, maintenance logs, operational reports
Probabilistic failure models	Estimates of failure probabilities and their cascading effects	Failure statistics, BNs, expert assessments
Simulation scenarios	Simulated failure situations, including rare and cascading events	Discrete-event modeling, cognitive models
Repair and maintenance data	Information on performed repairs and technical maintenance	Technical documentation, operational logs
Anomalies and verification	Data on identified deviations in equipment operation	Analysis of real operational data, statistical anomaly control
Adaptive CBR parameters	Adjustment of parameter weights during analogy search to improve diagnostic accuracy	Dynamic training on new data

The developed case base includes over 5,000 records of SPP equipment failures, collected from operational reports, technical inspections, and emergency situations. The case base structure contains information on failure types, operating conditions, probabilistic characteristics, and diagnostic procedures. The most common failure categories include: fuel system malfunctions (27% of cases); bearing failures (18%); gas turbine overheating (15%); electrical system anomalies (12%).

The library of precedents in the knowledge base is shown in Figure 1.3.1.

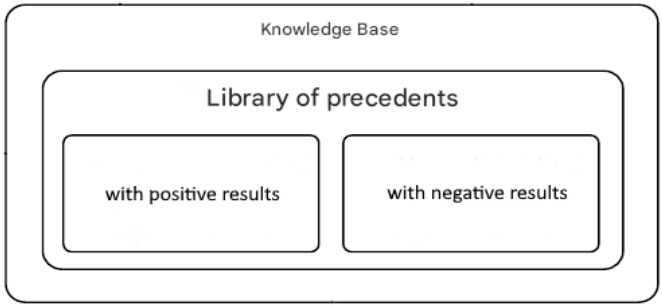


Figure 1.3.1. **Library of precedents in the knowledge base**

The automatic updating of the knowledge base involves several key stages: updating probabilistic estimates: adjusting failure probabilities and cascading effects using Bayesian probability updates and statistical analysis of new data; incorporating new data from simulation modeling to refine rare failure probability estimates and clarify dependencies between components; data filtering and verification - ensuring data quality through statistical analysis (anomaly detection, trend comparison with historical data); eliminating unreliable data through cross-checking with other sources.

Dynamic adaptation of the diagnostic system: adjusting parameter weights in CBR algorithms to improve diagnostic

accuracy; automatically updating probabilistic characteristics in the knowledge base.

The knowledge base is integrated with diagnostic and failure prediction methods as follows: CBR utilizes the case database to identify analogies between new and previously recorded failures; probabilistic risk assessment (BNs, MMs) determines failure probabilities based on accumulated data; simulation models (cognitive modeling, Monte Carlo simulations) generate new failure scenarios, improving forecasting and diagnostics accuracy.

Table 1.3.3. **Components and their interconnections in the SPP simulation model**

Component	**Function**	**Related components**	**Description of interaction**
Main Engine (ME)	Moves the vessel	Fuel system, cooling system, lubrication system	Depends on the supply of fuel, cooling, and lubrication
Cooling system	Maintains the temperature regime	Main engine, pumping system, auxiliary systems	Cools the ME and other components
Fuel system	Provides fuel supply to the ME	Pumping system, filters, automation system	Supports ME operation, monitored by automation
Generator	Produces electricity	Power supply system, batteries	-
Pumping system	Maintains fluid circulation	Cooling system, fuel system	-
Power Supply system	Provides power to systems	ME, automation system, auxiliary systems	Provides power for control and automation
Automation system	Controls parameters and diagnoses faults	All systems	Manages and controls the operation of all systems

The main engine, cooling system, fuel system, generator, pumping system, power supply system, and automation system are the core functional components and subsystems of the SPP.

They form the heart of the vessel's energy complex and ensure its operation. The Main Engine serves as the primary source of mechanical energy for vessel propulsion. The cooling system maintains the engine's operating temperature, preventing overheating. The fuel system stores, filters, and delivers fuel to the engine. The generator converts mechanical energy into electrical energy to power shipboard systems. The pumping system includes pumps for delivering fuel, coolant, lubricants, and other working fluids. The power supply system distributes electrical power onboard, including batteries and the shipboard power station. The automation system controls and monitors all SPP components, ensuring safe and efficient operation All these components and subsystems function as an integrated system, enabling the vessel and its technical systems to operate.

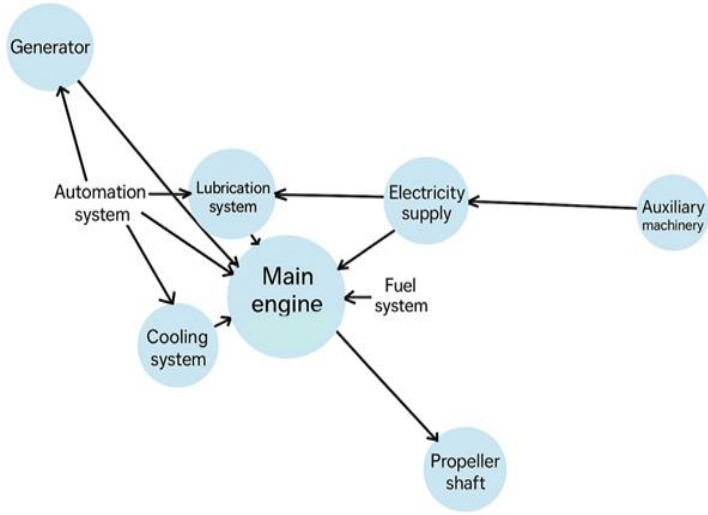


Figure 1.3.2. **Failure cause-and-effect graph**

Figure 1.3.2 illustrates the failure cause-and-effect graph, demonstrating how failure of one SPP component can lead to malfunctions in other system components.

For example, ME overheating increases load on the cooling system, raising the probability of pumping system failure. The integrated SPP diagram (Figure 1.3.2) [48] includes not only primary power units (e.g., main engine, power station) but also auxiliary systems that ensure their operation and vessel safety.

An example of a CTS is a SPP, consisting of interrelated subsystems, components, and elements with various functionalities. Figure 1.3.3 shows a graph representing the structure of an SPP.

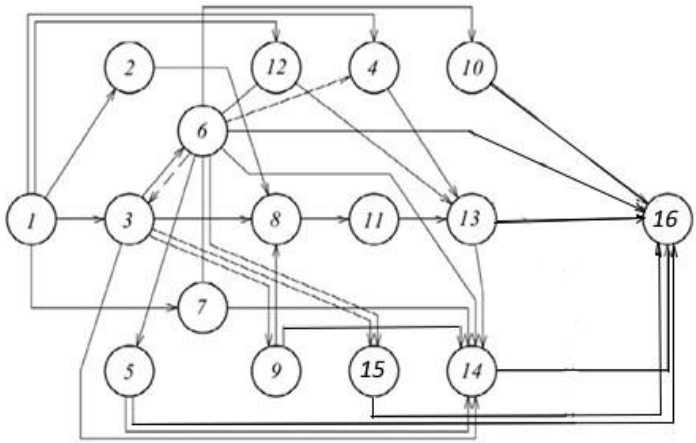


Figure 1.3.3. **Structure of the SPP**

The graph vertices include: input component 1; manual control of the main engine 2; subsystems for compressed air 3 and propulsion-rudder complex (PRC) control 4; boiler plant 5; power station 6; fire protection system 7; main engine 8; subsystems for remote-automated control 9 and ballast-drainage 10; power transmission from the main engine to the propeller 11; emergency

PRC drive 12; PRC 13; subsystems for measuring instruments 14 and sanitary water treatment 15; output component 16.

Thus, complex technical systems represent an organized set of numerous functionally interconnected and interacting subsystems, components, and elements, linked by nontrivial connections. These connections often involve uncertainties in input data, making them difficult or impossible to predict, and they exist in various states of failure.

The following relationships can be established between the components and subsystems in Figures 1.3.2 and 1.3.3: main engine corresponds to the main engine (8); cooling system is not explicitly indicated in your structure, but it may be part of the remote-automated control subsystem (9) or included in the main engine support system; fuel system is not directly specified, but it is logically linked to the boiler plant (5) and the main engine (8); generator may be part of the ship's power station (6); pumping system is most likely associated with the ballast and drainage subsystem (10) and other subsystems ensuring the operation of the propulsion and power system (PPS); power supply corresponds to the ship's power station (6); automation system may belong to the remote-automated control subsystem (9).

The simulation model integrates several approaches to failure prediction by combining probabilistic and cognitive methods: BNs assess probabilistic dependencies between component failures and cascading effects; MM the probability of transitions between component states (operational→degrading→failed); cognitive simulation modeling incorporates expert knowledge, operational mode influence, and operator behavior; failure simulation scenarios are formed based on data from the OREDA database and operational monitoring, allowing for the refinement of probabilistic failure models and the analysis of their impact on system reliability.

Table 1.3.4 presents the parameters of failure simulation scenarios, enabling the analysis of system behavior under various operational conditions.

Table 1.3.4. **Parameters of simulation scenarios for SPP equipment operation**

Scenario	Failure rate (avg.)	Load level	Possible consequences
Normal conditions	Low (0.01 failures/h)	Normal (70%)	Insignificant efficiency reduction
Accelerated wear	Medium (0.05 failures/h)	High (90%)	Accelerated wear of key components
Critical failures	High (0.1 failures/h)	Extreme (100%)	Cascading failures, SPP failure

In the integrated method for diagnosing and predicting SPP failures, MMs and BNs are used because they complement each other, allowing for the consideration of both the temporal dynamics of failures and the probabilistic dependencies between system components. MMs are applied to predict failure probabilities over time and assess the long-term behavior of equipment.

They represent the system as a set of discrete states (operational, degrading, pre-failure, failed) with transition probabilities between them. This allows for: determining the remaining lifespan of components based on accumulated data; predicting changes in equipment condition considering its operation; estimating failure probabilities over different time intervals. BNs enable modeling of interdependent failures, which is critical for complex technical systems.

They construct a dependency graph between components and allow for updating failure probabilities as new data becomes available. This provides the ability to: account for the impact of one failure on the likelihood of other system elements failing (e.g., a cooling system failure may increase the probability of engine overheating); adapt diagnostics in real time by refining predictions based on incoming information; improve risk assessment accuracy by considering cause-and-effect relationships.

The combined use of these methods allows not only for predicting failure probabilities but also for analyzing how they change depending on operating conditions, cascading effects, and external factors.

Thus, this combination of methods should not only predict failure probabilities but also take into account how they evolve under different operating conditions and external influences.

Figure 1.3.4 illustrates an example of a BN, showing the probabilities of various components SPP failing based on operational factors.

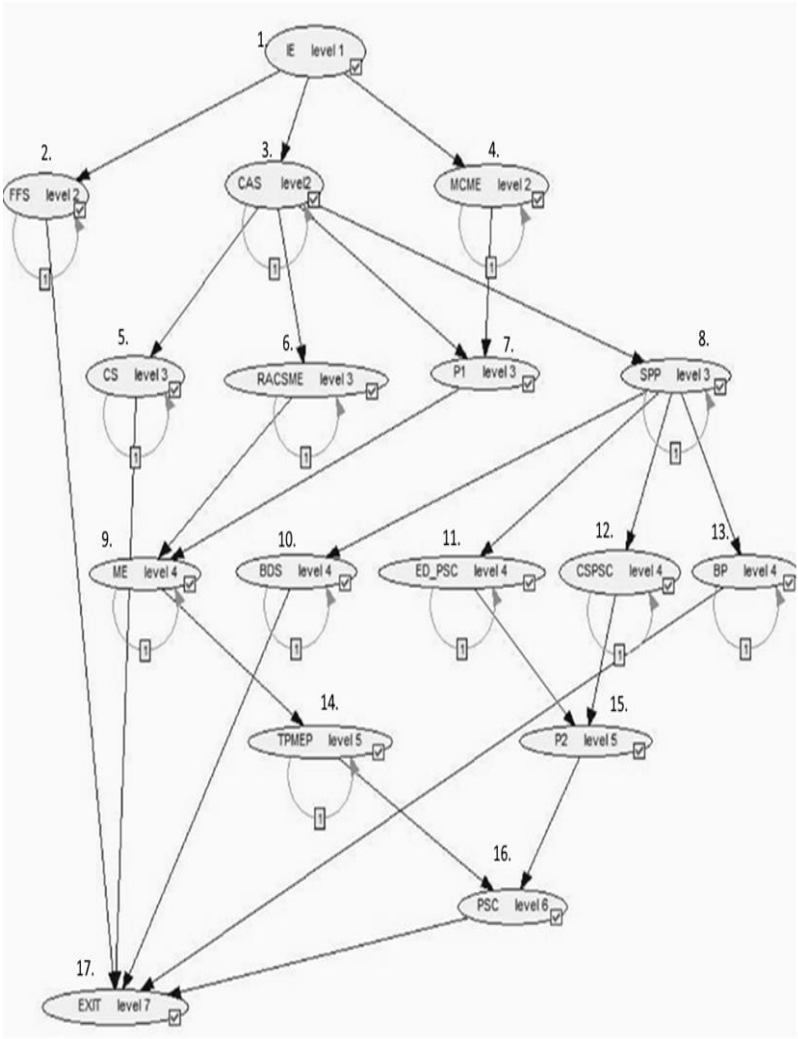


Figure 1.3.4. **Structure of the Bayesian network for the SPP**

The developed structure of the BN for the SPP (Figure 1.3.4) is a multi-level system comprising thirteen subsystems distributed

across seven levels. *P1* and *P2* are specialized intermediate nodes designed to implement the multi-level structure of the BN.

Legend of subsystems and components in the SPP BN: Input Element - IE; Firefighting System and Compressed Air System - FFS, CAS; Manual Control of the Main Engine - MCME; Control Systems and Remote Automated Control of the Main Engine - CS, RACSME; Intermediate Component - P1; Ship Power Plant - SPP; Main Engine - ME; Ballast Drainage System - BDS; Emergency Drive for the Propulsion and Steering Complex - EDPSC; Control System for the Propulsion and Steering Complex - CSPSC; Boiler Room - BR; Power Transmission from the Main Engine to the Propeller – TPMEP; Intermediate Component - P2; Propulsion and Steering Complex - PSC; Output Component – EXIT

Markov processes are used to model the probabilities of transitions between different system states.

Figure 1.3.5 illustrates possible transitions between states (normal operation, partial failure, complete failure) and their probabilities.

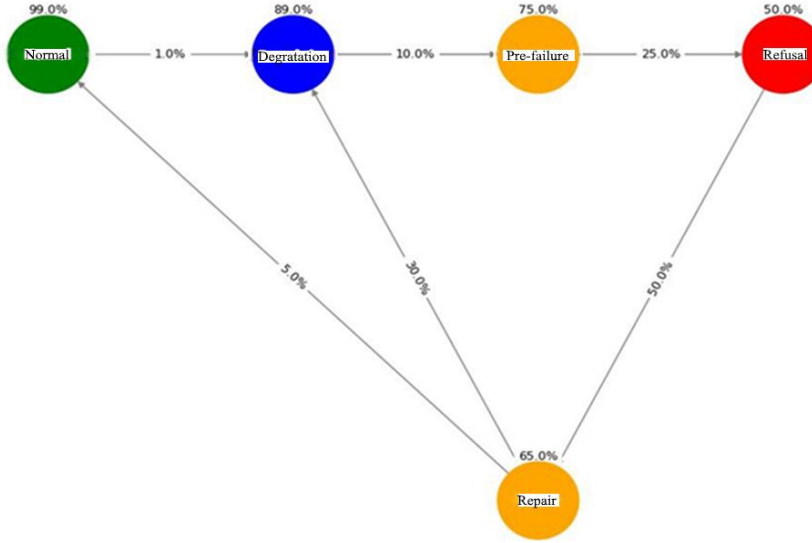


Figure 1.3.5. **Markov process graph of SPP equipment failures**

The Markov process graph (Figure 1.3.5), considering the Harrington desirability function, allows the probability of failures to be divided into four risk levels, making the model more accurate for SPP reliability analysis.

1. "Normal" state (0.1): actual risk value: 0.1; Harrington classification - minimal risk (0 - 0.2). The value is correct. At this level, failures are almost nonexistent, and the probability of degradation is very low (1%).

2. "Degradation" state (0.3): actual risk value: 0.3; Harrington classification: Acceptable risk (0.2 - 0.37). The value corresponds to an acceptable risk level. The equipment can continue operation, but there is a probability of deterioration. The probability of transitioning to "Pre-failure" is set at 10%, which corresponds to a moderate risk.

3. "Pre-failure" state (0.55): actual risk value: 0.55; Harrington classification: Maximum risk (0.37 - 0.63). The value falls within the correct range. At this level, operation is possible, but urgent maintenance is required. Transitions: failure probability increases to 25%, which is logical since the equipment is close to failure; the chance of returning to "Degradation" is zero, as a pre-failure state does not improve without repair.

4. "Failure" state (0.8): actual risk value: 0.8; Harrington classification - critical risk (0.63 - 1.0). Fully corresponds to the critical risk level. Operation is impossible without repairs. Transitions: 50% probability of repair, which is realistic for systems where some failures are repairable; the remaining 50% remain in the "Failure" state, reflecting the impossibility of restoring some components.

5. "Repair" state (0.4): actual risk value: 0.4; Harrington classification - maximum risk (0.37 - 0.63). The value is borderline. After repairs, the equipment does not always fully recover (30% of cases transition back to "Degradation").

Recovery chances: 5% probability of returning to the "Normal" state, which is logical for components that can be fully restored; 65% probability of remaining in the "Repair" state (due to prolonged repairs or recurring failures).

All states correspond to Harrington's risk levels. This model enables risk differentiation of failures and helps develop maintenance strategies.It allows determining the optimal timing for preventive maintenance, reducing the likelihood of the equipment transitioning into a critical state.Incorporating Harrington's classification into the model enhances the accuracy of failure probability forecasts, especially over long operating periods.

The transition probabilities between states are determined by the following formula:

$$P_{ij} = \frac{n_{ij}}{N_i}, \qquad (1.3.1)$$

where: P_{ij} – probability of transition from state i to state *j*;
n_{ij} – number of recorded transitions from state i to state *j*;
N_i – total number of states recorded in *i*

The analysis of accumulated data made it possible to establish the dependence of equipment failure probability on operating time (Figure 1.3.6).

The curve demonstrates a characteristic increase in failure probability as operating hours increase, confirming the necessity of using probabilistic models and simulation modeling in developing a CBR case base. The obtained data confirm that considering the stochastic nature of failures and integrating the CBR method with probabilistic models not only allows for failure prediction but also improves the accuracy of diagnostics at different stages of the equipment's life cycle.The graph illustrates the increasing probability of failure as the operating time of the equipment increases. In the initial stage (up to 5,000 hours), the failure probability remains low, corresponding to the period of normal operation. After 15,000 hours, the failure probability rises more sharply, indicating the need for more thorough maintenance. By 25,000 hours, the failure probability approaches 1, signifying an almost inevitable equipment failure.

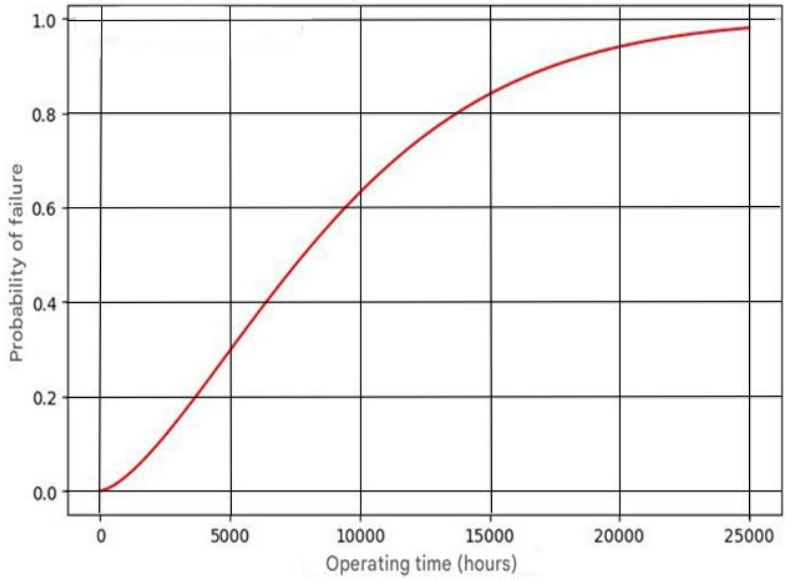


Figure 1.3.6. **Probability of failure with increasing operating time of SPP equipment**

Integration of CBR, BNs, MMs, and simulation modeling

The diagnostic system combines multiple complementary methods to enhance the accuracy of assessing the technical condition of the SPP:

- CBR is used to identify similar failure cases, allowing for the detection of recurring malfunctions. The k-Nearest Neighbors method with adaptive weights ensures precise matching of the current situation with historical cases, minimizing the probability of false alarms;
- BN adjust failure probability estimates in real time, considering interdependencies between components. This allows for the assessment of cascading effects and changing operational conditions;
- Markov processes are used for predicting equipment degradation and estimating the remaining service life of components.

They model the dynamic behavior of the system, allowing for state transitions over time;

- simulation modeling complements the system by generating scenarios of rare and cascading failures, expanding the diagnostic database and reducing uncertainty in failure prediction.

Table 1.3.5 presents the distribution of diagnostic methods and their functional purposes.

Table 1.3.5. **Distribution of diagnostic methods and their functional purposes**

Method	Analyzed elements	Expected result
CBR	Components similar to recorded failures	Finding precedents with similar symptoms
Bayesian networks	Interconnected system elements	Probabilistic assessment of cascading failure effects
Markov models	Remaining resource of components	Prediction of component degradation
Simulation modeling	Failure dynamics in the system	Generation of artificial precedents to supplement the database

Figure 1.3.7 presents a comparison of equipment failure probabilities when using traditional diagnostic methods versus applying the CBR approach integrated with the case database. It is evident that the case base helps reduce the probability of diagnostic errors, as accumulated failure experience is utilized to refine predictions. This demonstrates that the proposed approach not only improves diagnostic accuracy but also enhances the system's adaptability to changing operating conditions.

The graphs visualize the effect of using CBR - reducing the probability of diagnostic errors and increasing prediction accuracy.

Connection to the case base – reflecting the stochastic nature of failures, the graphs illustrate the increase in equipment failure probability over time. This confirms the need to consider stochastic factors in diagnostics, which is implemented in the case database through the integration of probabilistic models.

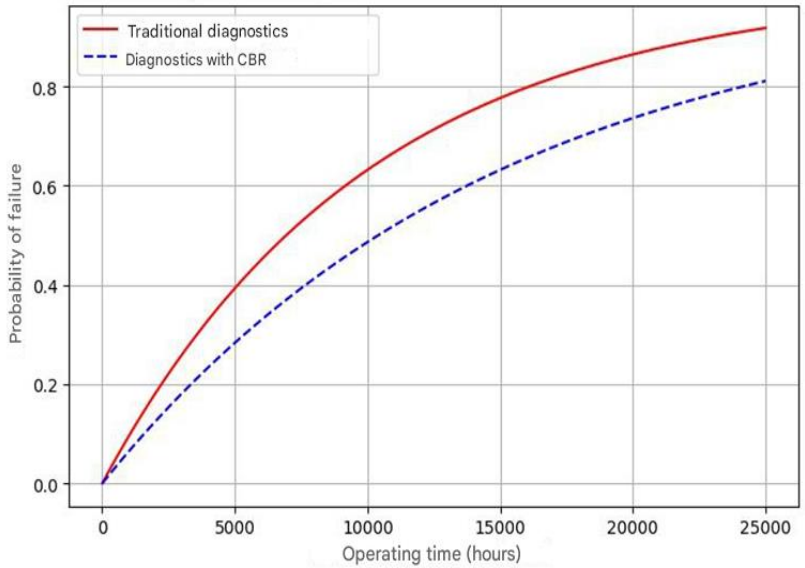


Figure 1.3.7. **Probability of equipment failures in SPP with and without CBR**

Advantage of CBR over traditional methods – The two curves show the difference in failure prediction: traditional methods do not take into account accumulated experience, leading to a higher probability of diagnostic errors. With CBR, the analysis of accumulated data and adaptation of diagnostic decisions reduce the probability of errors. This clearly demonstrates that the case base helps refine predictions by considering previously recorded failures.

Simulation modeling for improved forecasting - the case base allows for the consideration of not only frequent but also rare and cascading failures. Failure modeling based on accumulated data improves the prediction of potential malfunctions, reducing the number of false alarms and increasing diagnostic accuracy.

Integration of these methods ensures not only accurate fault diagnostics but also system adaptability to new operational data,

enhancing reliability in forecasting and minimizing the likelihood of missed failures.

Numerical experiments have shown that using a case database combined with CBR adaptation increases fault diagnosis accuracy from 85% (when using only traditional probabilistic methods) to 92–95%. The integration of BNs and simulation modeling reduced the number of false positive diagnostic decisions by 30%, while the use of parameter weight optimization algorithms decreased the mean absolute error in failure prediction by 12%.

Table 1.3.6. **Impact of the CBR on failure diagnosis accuracy**

Diagnostic method	**Accuracy (%)**	**FP (false positives)**	**FN (false negatives)**	**Processing time (s)**
Probabilistic Methods (without CBR)	85	18%	12%	2.1
CBR	89	14%	10%	1.8
Integrated approach	95	8%	5%	1.2

Optimization of the case database.

To improve diagnostic accuracy and system adaptability, the case database is regularly updated and optimized. One of the key mechanisms is adjusting the parameter weights that determine the degree of similarity between failure cases. Optimization is performed using the L-BFGS-B method, which minimizes the error between reference and predicted failure similarity values:

$$\omega^{\bullet} = \arg\min_{\omega} \sum_{i} (S_{true}(t) - S_{pred}(i,\omega))^2 \tag{1.3.2}$$

where $S_{true}(i)$ - the reference similarity value between cases (based on expert assessments or failure statistics);

$S_{pred}(i,\omega)$ - the system-predicted similarity value, dependent on the parameter weights ω;

ω - weights of diagnostic parameters that determine the contribution of various factors in failure similarity assessment

The L-BFGS-B method is a modification of the quasi-Newton BFGS optimization method, which: allows minimizing a function that depends on a large number of parameters; uses an approximate representation of the inverse Hessian matrix to save memory; supports constraints on parameters, which is important for weight optimization (e.g., ensuring weights remain positive and sum to 1).

Below is the code demonstrating the weight optimization process using the L-BFGS-B method from the *scipy.optimize* library:

```
import numpy as np
from scipy.optimize import minimize
# Data: true similarity values (expert assessment) and predicted values (model)
S_true = np.array([0.9, 0.8, 0.7, 0.6, 0.5])  # True similarity values
X = np.array([
    [1, 0.3, 0.2, 0.1],  # Failure type, probability, risk category, affected subsystems
    [0.9, 0.4, 0.3, 0.2],
    [0.8, 0.5, 0.2, 0.3],
    [0.7, 0.3, 0.4, 0.5],
    [0.6, 0.2, 0.5, 0.4]
])  # Feature matrix (parameter weights)

# Initial weights (approximate values)
w_init = np.array([0.3, 0.2, 0.2, 0.3])

# Constraints: weights must be between 0 and 1 and sum to 1
bounds = [(0, 1)] * len(w_init)
constraints = ({'type': 'eq', 'fun': lambda w: np.sum(w) - 1})  # Sum of weights = 1
# Error function (MSE)
def loss_function(w):
    S_pred = X @ w  # Linear combination of features and weights
```

```
    return np.mean((S_true - S_pred) ** 2)  # Mean squared error

# Optimization using L-BFGS-B method
result = minimize(loss_function, w_init, method='L-BFGS-B', bounds=bounds, constraints=constraints)

# Optimized weights
optimized_weights = result.x

print("Optimized weights:", optimized_weights)
```

The code generates the initial data: true failure similarity values S_{true}; feature matrix X, where each column represents a diagnostic parameter; initial weights *w*. It defines constraint conditions: all weights are within the range [0,1]; the sum of all weights must be equal to 1 (to prevent incorrect scaling).It defines the error function: linearly combines parameters using weights; computes the mean squared error between predicted and true values. The optimization is performed using the L-BFGS-B method. The optimized weights are outputted, minimizing the error.

Key Steps in updating the case database: dynamic adjustment of parameter weights based on new operational data; filtering outdated and irrelevant data to maintain case base accuracy; adapting the case base structure according to simulation modeling results.

This approach ensures more accurate failure case matching and enhances the adaptability of the diagnostic system under changing operating conditions. The case base development discussed justifies the need for structured failure data storage and the creation of a unified knowledge base for effective failure diagnosis and prediction.

The developed case base integrates CBR with probabilistic models and simulation techniques, significantly enhancing the accuracy of failure diagnostics in ship power plants. The conducted analysis revealed the following: the structured storage of diagnostic cases allows diagnostic accuracy to reach up to 95% by leveraging accumulated and formalized failure knowledge; probabilistic

methods address the stochastic nature of failures, which is especially critical in complex technical systems; simulation modeling enables the inclusion of rare and cascading failure scenarios, broadening predictive capabilities; the case base encompasses the most common types of failures, such as fuel system issues, gas turbine overheating, and electrical subsystem anomalies; systematic accumulation and analysis of failure cases facilitate pattern recognition and early fault prediction.

The implementation of the case base has resulted in a 30% reduction in false-positive diagnoses and a notable increase in diagnostic reliability. Overall, the case base serves as a core element of an intelligent diagnostics system for marine power units, offering structured data management, decision support, and robust failure prediction. The synergy of CBR, probabilistic modeling, and simulation enables a more resilient and interpretable diagnostic process under real-world operational uncertainty.

1.3.2 Methodology for determining similarity between cases in the fault diagnosis system of SPP

1.3.2.1 Introduction

In modern CTS, timely and accurate fault diagnosis plays a crucial role in ensuring uninterrupted operation, safety, and high system performance. The increasing complexity and heterogeneity of CTS has driven the widespread adoption of intelligent diagnostic methods based on reusing accumulated experience, among which CBR stands out. CBR-based fault diagnosis involves comparing the current situation with previously recorded fault cases, with the effectiveness of this approach hinging on accurately assessing the degree of similarity between cases.

Determining similarity between cases is a central challenge in CBR systems, as the quality of this assessment directly influences the accuracy, relevance, and timeliness of diagnostic conclusions. Over recent years, a variety of approaches have been proposed to improve similarity assessment accuracy - ranging from classical metric

functions to hybrid machine learning methods and logical models. However, despite this progress, existing solutions often face several limitations: poor scalability, insufficient adaptability to heterogeneous data, and limited interpretability of results. For example, Neykov and Stefanova [49] demonstrated the applicability of rule-based heuristics in CBR systems, emphasizing their flexibility and explainability, but also noted challenges in handling large volumes of cases and the need for manual tuning of attribute weights. Meanwhile, Chen et al. [50] showed that graph neural networks (GNNs) can uncover hidden structural similarities in cyber-physical systems but require complex preprocessing and often suffer from limited decision transparency. An enhancement of the CBR approach through aggregation operators such as the Choquet integral was proposed in [51], allowing the consideration of nonlinear interdependencies between diagnostic features. Although this approach improved similarity assessment accuracy, it introduced complexity in forming expert preferences. In another direction, Ye [52] applied neutrosophic logic with cotangent-based similarity measures to handle uncertainty in technical system data, though interpreting the results remains a challenge. When analyzing time series of sensor data, distance measures such as Dynamic Time Warping were examined by Serrà and Arcos [53], proving resilient to temporal distortions. Nevertheless, such methods perform poorly in semantically rich contexts and tend to be computationally intensive. Finally, adaptive fuzzy logic systems described by El Bitar et al. [54] demonstrated high adaptability and noise resilience, yet their industrial-scale application is constrained by insufficient validation on high-dimensional heterogeneous data. According to Mathisen et al. [55], learning similarity metrics from data can automate the creation of similarity functions in case-based systems. The authors proposed two novel approaches: using a pre-trained classifier as the basis for the similarity function and a fully data-driven method minimizing training time. Both approaches showed high effectiveness across 14 different datasets. Researchers Bach and Mork [56] noted that in the early stages of developing case-based

systems, defining similarity functions is a complex task requiring the transfer of implicit expert knowledge into explicit models. They emphasized the importance of explainability and transparency in development to ensure quality expertise from domain specialists. In their work, Verma et al. [57] presented a methodology for modeling local similarity functions for various attributes in datasets. The authors analyzed the distribution of numerical attributes and proposed the use of polynomial functions for modeling their distribution, enabling more precise range definitions and improving the search for relevant cases. Lin et al. [58] emphasized the key role of interpretability in legal case-matching systems. They proposed an integrated framework consisting of four modules: key sentence identification, case matching, sentence alignment, and conflict resolution ensuring transparency and explainability in the matching process. According to Ren et al. [59], effective similarity determination between cases is crucial in the design of low-carbon products. The authors developed a model that considers both similarity and adaptability for reusing knowledge and adapting it in the design process.

Traditional matching methods may overlook: probabilistic dependencies between faults, where one failure increases the likelihood of others; differences in operating conditions that affect component degradation rates; and the dynamic evolution of technical states, where equipment transitions to states with higher failure probability. To enable accurate identification of typical equipment faults and determination of their root causes within ship power plants, it is necessary to develop a unified and adaptive methodology for assessing similarity between fault cases. This methodology should be centered on a similarity metric designed to effectively match emerging faults with archived cases and identify the most relevant analogs. In this context, analogous cases refer to those sharing similar operating conditions, identical types of damage, comparable technical parameter values, and similar probabilistic characteristics. For instance, two main engine failures may be classified as analogous if they occurred at comparable runtime, were

accompanied by similar changes in temperature and pressure parameters (e.g., oil temperature and fuel pressure), and had the same root cause, such as a cooling system failure leading to overheating.

The developed methodology should account for both quantitative and qualitative characteristics of failures, enable integration of metric, logic-based, and machine learning components, and support processing of heterogeneous data types — symbolic, numeric, and temporal. Additionally, it should ensure interpretability of results for technical experts and adaptability to changing operational conditions. The proposed structured approach to similarity metric development encompasses domain-specific features, includes parameter normalization, and utilizes optimization mechanisms for attribute weight coefficients to improve the reliability and accuracy of diagnostics. Special attention is given to operational and technical failure parameters (load characteristics, runtime, damage types), causal relationships between components, probabilistic-statistical indicators (occurrence frequency, prior probability), and forecasted changes in the technical condition of the object. The aim of this study is to develop a unified and adaptive methodology for assessing the similarity degree between fault cases within ship power plants. This methodology should integrate metric and probabilistic approaches, support heterogeneous data, and ensure result interpretability for subsequent application in intelligent CBR diagnostic systems.

1.3.2.2. Methods for calculating similarity between cases [28, 60, 61]

To search for similar failures in the case database, the k-NN method [57, 62] is used. This approach enables the identification of the most similar cases based on a predefined similarity metric. The choice of this method is justified by the following advantages: Flexibility: k-NN works with heterogeneous data types (numerical, categorical, probabilistic parameters); interpretability: unlike neural network models, it is possible to explain why two failure cases are considered similar; adaptability: the method allows adjustment of

parameter weights depending on operational conditions; efficiency with limited data: does not require a large training set, which is critical for diagnosing rare failures.

Similarity between two cases is calculated based on the weighted sum of individual similarities across key parameters listed in Table 1.3.7.

Table 1.3.7. **Similarity between two cases across key parameters**

Parameter	**Description**	**Similarity measurement method**
Failure type	Nature of damage, degradation mechanism	Cosine similarity
Failure causes	External and internal factors, preceding events	Jaccard similarity
Failure probability	Estimated probability of failure	Euclidean distance
Failure consequences	Impact on operability, cascading effects	BNs
Affected components	Failed nodes and subsystems	Jaccard similarity

Similarity measurement methods in k-NN

Depending on the data type, the following metrics are used:

- euclidean distance for numerical parameters (e.g., failure probability, remaining useful life);
- cosine similarity for textual and vector representations of failure categories;
- jaccard similarity for categorical data (e.g., list of affected subsystems);
- MMs for predicting the evolution of failures over time and adjusting the significance of parameters;
- BNs for accounting for probabilistic dependencies between failures;
- cognitive simulation modeling for refining parameter weights based on failure scenario analysis.

1.3.2.3 A model for determining similarity between precedents for diagnosing failures of Bayesian networks equipment

To search for similar failures in the precedent database, the k-NN method is used, as described by V. Vychuzhanin [2], Verma et al. [57], Zuber and Sirdey [63]. This approach allows the identification of the most similar cases based on a predefined similarity metric. The choice of this method is driven by the following advantages: flexibility - k-NN can handle heterogeneous data types (numerical, categorical, probabilistic parameters); interpretability - unlike neural network models, it is possible to explain why two failures are considered similar; adaptability - the method allows for adjusting parameter weights depending on operating conditions; efficiency with limited data: it does not require a large training set, which is critical for the diagnosis of rare failures. The similarity between two precedents is calculated based on a weighted sum of partial similarities across key parameters, as presented in Table 1.3.8.

Table 1.3.8. **Similarities between two precedents based on key parameters**

Parameter	Description	Similarity measurement method
Failure type	Nature of damage, degradation mechanism	Cosine similarity
Failure causes	External and internal factors, preceding events	Jaccard similarity
Failure probability	Estimated probability of failure	Euclidean distance
Failure consequences	Impact on operability, cascading effects	BNs
Affected components	Failed nodes and subsystems	Jaccard similarity

Similarity Measurement Methods in k-NN.

Depending on the type of data, the following metrics are used: euclidean distance - for numerical parameters (e.g., failure probability, remaining useful life); cosine similarity - for textual and vector representations of failure categories; jaccard similarity - for categorical data (e.g., lists of affected subsystems); MMs - for predicting failure evolution over time and adjusting parameter significance; BNs - for accounting for probabilistic dependencies between failures.

To build a formalized model for determining similarity between precedents, information on typical failures of SPP equipment and corresponding diagnostic features was structured. Table 1.3.9 presents the classification of failures according to key parameters — temperature, vibration level, pressure, and typical manifestations.

This structure makes it possible to highlight diagnostically significant features, which are further formalized into precedent parameters. Additionally, Figure 1.3.8 shows a classification error matrix that reflects characteristic cases of misclassification of observed features into failure types, demonstrating the limitations of traditional matching methods and emphasizing the need for a more precise similarity evaluation model.

Table 1.3.9. **Classification of typical failures and diagnostic features**

Failure type	Temperature	Vibration	Pressure	Typical symptoms
Failure A	>90°C	2.0–3.5 mm/s	>10 bar	Often associated with cooling system overheating
Failure B	75–85°C	1.5–2.5 mm/s	6–9 bar	Related to gradual bearing wear
Failure C	>85°C	2.5–4.0 mm/s	9–11 bar	Observed with unstable fuel supply

Figure 1.3.8 presents the classification error matrix, reflecting the performance of the CBR system when classifying new precedents by failure types based on the applied similarity assessment methodology. Along the axes (Fig. 1.3.8), the actual and predicted failure types (A, B, C) are indicated. The values on the diagonal of the matrix correspond to correctly classified cases: 48 instances of type A were correctly identified as A; 42 instances of type B were classified correctly; and 46 instances of type C were recognized accurately. Classification errors are observed in the following forms: 3 cases of type B were mistakenly assigned to class A, and 5 to class C; 4 cases of type C were classified as B. Type A demonstrates the highest classification stability (only 2 errors out of 50 observations, accuracy - 96%).

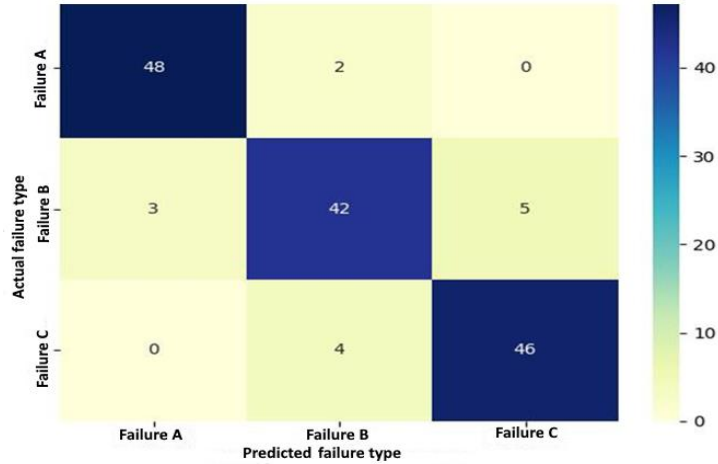


Figure 1.3.8. **Classification Error Matrix by Failure Types**

This indicates high overall classification accuracy, with the share of correct predictions exceeding 90% for all failure types. A notably increased probability of cross-classification between types B and C may be explained by: partial overlap of features (e.g., similar vibration and pressure values under different failure scenarios); insufficient discrimination capacity of the current similarity metric in

the feature space of these types or a limited training sample size for B and C cases. Such insights allow targeted model enhancement - for instance, by refining parameter weights or introducing additional features that enable clearer separation of failure types with similar manifestations.

To establish the relationship between features and failure types, three scenarios can be identified. Failure A (bearing overheating) is associated with prolonged load, accompanied by an increase in temperature and vibration. Failure B (hydraulic shock in the cooling system) occurs during abrupt mode changes due to unstable pressure and temperature. Failure C (injector defect) is characterized by reduced fuel pressure and unstable engine operation. These examples demonstrate that comprehensive parameter analysis allows for accurate differentiation of failure types and identification of their causes. In the tasks of diagnosing failures in CTS, including SPP, the model of case representation and comparison is a key element of intelligent decision support. Traditionally, similarity between cases is evaluated using Euclidean and Manhattan distances in feature spaces based on the numerical representation of failure parameters [64, 65]. Some studies also consider categorical features, but often without adapting the metrics to their nature [66]. More advanced approaches suggest the use of ontologies or logical representations [67]; however, these require complex verification and are not always robust to incomplete data. The model proposed in this study differs from existing approaches in several key aspects. First, it employs specialized metrics tailored to the type of diagnostic parameter: Euclidean distance for numerical features, cosine similarity for categorical features, and Jaccard distance for multiple-value features, ensuring adequate similarity evaluation. Second, the model incorporates a weighting system that reflects the diagnostic significance of each feature, thereby improving the accuracy and interpretability of the results. Third, it allows for weight optimization using numerical methods, which enables adaptation of the model to real expert assessments and empirical data. Finally, the model is designed to operate with limited training samples, which is

particularly important under SPP operational conditions, where the number of recorded failure cases is limited, and the data may be incomplete or heterogeneous.

The overall similarity measure between two cases A and B is defined as a weighted sum of individual similarity coefficients.

For two cases $A_i = (x_1, ..., x_n)$ and $B_i = (y_1, ..., y_n)$, the similarity $S_i(A,B)$ is calculated using the formula [68]:

$$S(A,B) = \sum_{i=1}^{n} \omega_i \cdot sim_i(A_i, B_i), \tag{1.3.3}$$

where $sim_i(A_i, B_i) \in |0,1|$ is the partial similarity measure for the i - th feature;

ω_i is the weight of the i-th parameter

The partial similarity functions are defined as follows:

Absolute difference similarity (for numerical features normalized to [0,1]):

$$sim_i(A_i, B_i) = \frac{|A_i - B_i|}{\max(A_i) - \min(A_i)} \tag{1.3.4}$$

Cosine similarity (for categorical features, such as failure type, encoded using one-hot encoding) [69]:

$$sim_{i.\cos}(A_i, B_i) = \frac{A_i \cdot B_i}{||\vec{A_i}|| \cdot ||\vec{B_i}||} \tag{1.3.5}$$

Euclidean similarity (for normalized numerical parameters, such as failure probability or risk category) [70, 71]:

$$sim_{i.euc}(A_i, B_i) = 1 - |A_i - B_i| \tag{1.3.6}$$

Jaccard similarity (for set-based features, e.g., failure subsystems) [72]:

$$sim_{i.euc}(A_i, B_i) = \frac{|A_i \cap B_i|}{|A_i \cup B_i|} \quad (1.3.7)$$

For practical implementation in the context of SPP diagnostics, the following specification of the similarity measure is used. The formal definition of similarity between cases is implemented using the k-NN method, which calculates the similarity between cases. The parameter weights are optimized to reflect their relative importance:

$$S(A,B) = \sum_{i=1}^{n} \omega_i \cdot d_i(A_i, B_i), \quad (1.3.8)$$

where ω_i - weight of the *i*-th parameter;

$d_i(A_i, B_i)$ - dissimilarity measure of the *i* - th parameter between cases A_i and B_i;

n - total number of parameters

The developed similarity model for comparing a current case A_i and a precedent B_i is represented as:

$$S(A,B) = \alpha_{s.i} \cdot d_{types.i}(A_i, B_i) + \beta_{s.i} \cdot d_{probability.i}(A_i, B_i) + \gamma_{s.i} \cdot d_{components.i}(A_i, B_i), \quad (1.3.9)$$

where $d_{types.i}(A_i, B_i)$ - the similarity measure for failure types;

$d_{probability.i}(A_i, B_i)$ - the difference in failure probabilities;

$d_{components.i}(A_i, B_i)$ - the similarity estimation for the affected components;

$\alpha_{s.i}, \beta_{s.i}, \gamma_{s.i}$ - weighting coefficients that determine the importance of different aspects when comparing precedents

The similarity components are calculated as follows:

$d_{types.i}(A_i, B_i) = 1$, if types are identical, $d_{types.i}(A_i, B_i) = 0{,}5$, if in the same cate category, $d_{types.i}(A_i, B_i) = 0$, if types differ

$$d_{probability.i}(A_i, B_i) = 1 - |P(A_i) - P(B_i)|, \quad d_{components.i}(A_i, B_i) = \frac{|C(A_i) \cap C(B_i)|}{|C(A_i) \cup C(B_i)|},$$

where $P(A_i), P(B_i)$ - are the failure probabilities for the respective components;

$C(A_i), C(B_i)$ - are the sets of components involved in the failures.

This formulation allows a quantitative accounting of subsystem overlap, which directly impacts the assessment of case similarity.

For normalization of the final similarity metric, the following expression is applied:

$$D_{total}(A_i, B_i) = \frac{\alpha_{s.i} \cdot d_{types.i}(A_i, B_i) + \beta_{s.i} \cdot d_{probabilitys.i}(A_i, B_i) + \gamma_{s.i} \cdot d_{components.i}(A_i, B_i)}{\alpha_{s.i} + \beta_{s.i} + \gamma_{s.i}}$$

Such normalization ensures that the resulting similarity score remains within the range [0,1], enabling consistent comparison of different case pairs and facilitating further interpretation within the diagnostic decision-making process.

The coefficients $\alpha_{s.i}, \beta_{s.i}$ and $\gamma_{s.i}$ optimized using numerical methods aimed at minimizing a loss function that reflects the classification error based on known labels (for example, the mean squared error between the predicted similarity value and the expert-assessed value). This allows the model to be adapted to the specifics of a particular domain and the structure of the case database.

To improve the accuracy, interpretability, and adaptability of the similarity assessment model between cases, the following enhancements have been implemented:

1. Weight correction based on failure frequency.

This extension allows the model to take into account the rarity of specific failure types in the dataset, automatically increasing the importance of rarely occurring but potentially critical scenarios. This is particularly important for the diagnosis of low-probability but high-risk conditions, which standard methods may overlook due to their low occurrence rate.

The weight correction formula is defined as:

$$\tilde{a}_{s.i} = a_{s.i} \cdot \log(1 + \frac{1}{f_{type}(A_i)}), \quad (1.3.10)$$

where $f_{type}(A_i)$ - the relative frequency of occurrence of failure type A_i in the case database

This adjustment ensures automatic balancing of the model: the rarer a particular failure type is, the higher its weight when calculating the final similarity measure. This enables the model to better handle the diagnosis of infrequent but significant cases, preventing bias toward frequently occurring but less critical failures. Such a correction can further be extended to other parameters, including probabilistic and structural characteristics, which would additionally enhance the flexibility and reliability of the case-based reasoning system.

2. Temporal Degradation Factor

This extension accounts for the influence of the temporal aspect (e.g., the difference in operating time until failure) when assessing the similarity between cases. Incorporating this factor allows the model to prioritize cases that are closer in terms of operational time, making the reasoning process more context-aware. The temporal degradation factor is calculated as:

$$d_{\deg gadation}(A_i, B_i) = \exp(-\lambda \cdot |T_{Ai} - T_{Bi}|), \quad (1.3.11)$$

where T_{Ai}, T_{Bi} – the operating time until failure for cases A_i and B_i;

λ – the degradation coefficient, which controls the sensitivity of the model to temporal differences.

The adjusted similarity score is defined as:

$$S'(A_i, B_i) = S(A_i, B_i) \cdot d_{degradation}(A_i, B_i) \quad (1.3.12)$$

The introduction of this multiplicative temporal context reflects the idea that two cases with identical features but differing significantly in failure time are diagnostically less similar than cases with both matching features and close failure times.

Figure 1.3.9 illustrates how the temporal degradation coefficient varies depending on the difference in operational time between two cases. The figure presents three scenarios corresponding to different sensitivity levels of the model to temporal discrepancies, controlled by the value of the λ parameter. The curve with the smallest λ (0.00005) shows a slow decrease of the coefficient, which reflects a model with stable memory - it "trusts" even older cases. Conversely, at a large λ (0.0005), the model quickly depreciates old cases, focusing on the most recent data.

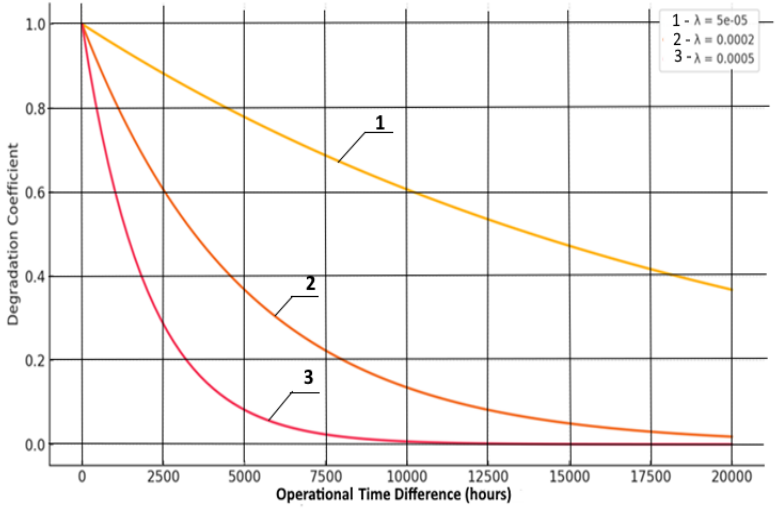


Figure 1.3.9. **Change of the temporal degradation coefficient depending on the difference in operating time between two cases λ**

Thus, the choice of λ directly affects the behavior of the model: a low λ implies high stability, which is useful in the presence of stable patterns; a high λ ensures adaptability to new conditions, which is necessary in the case of dynamically changing technical states of the SPP system. This emphasizes the importance of calibrating the temporal component when designing CBR systems, especially in the context of variable load and operating modes.

3. Contextual smoothing of components supplements the comparison of SPP system components with information about the proximity between components (for example, if two components are technically related). This reduces information loss in cases of partial mismatch.

$$d_{comp}^{adj} = \frac{|C(A_i) \cap C(B_i)| + \delta \cdot |Ad_j(A_i, B_i)|}{|C(A_i) \cup C(B_i)| + \delta \cdot N}, \quad (1.3.13)$$

where $|Ad_j(A_i, B_i)|$ is the number of adjacent, functionally related subsystems;

σ is the smoothing coefficient;

N is the total number of possible components

4. Nonlinear activation (e.g., tanh) allows the saturation effect to be taken into account in cases of strong similarity and helps to soften sharp boundaries. This is used for better interpretation of the final similarity in extreme cases.

$$S(A,B) = \tanh(\alpha_{s.i} \cdot d_{types.i}(A_i, B_i) + \beta_{s.i} \cdot d_{probability.i}(A_i, B_i) + \gamma_{s.i} \cdot d_{components.i}(A_i, B_i)) \quad (1.3.14)$$

The proposed formalized model for assessing similarity between cases (1.3.9) is integrated into the classical CBR architecture, which includes four key stages: retrieval, adaptation, application, and update. At the retrieval stage, the model is used to search for the most relevant analogs in the case base.

Stage 1. Calculation of similarity between cases.

For a new diagnostic case A_i, represented by a feature structure (failure type, probability, subsystems, etc.), the similarity measure is calculated for all cases B_i from the knowledge base according to formula (1.3.9). If additional modules (temporal degradation, probabilistic correction, etc.) are activated, the final value is modified accordingly:

$$S'_{norm}(A_i, B_i) = \frac{S(A_i, B_i) \div d_{\text{deg}radation}(A_i, B_i) \cdot P_i(A_i, B_i)}{\alpha_{s.i} + \beta_{s.i} + \gamma_{s.i}}, \qquad (1.3.15)$$

where $P_i(A_i, B_i) \in |0,1|$ - is the probability of co-occurrence of the values A_i and B_i in the cases, extracted from empirical data or an expert model

Stage 2. Nearest neighbor search (k-NN).

Based on the obtained values $S(A_i, B_i)$, the *k* nearest cases with the highest similarity values are selected. After constructing and formalizing the similarity metric between cases across the set of parameters, the presented model is integrated into the k-NN search algorithm. Using k-NN not only enables the application of the similarity metric but also implements a decision-making process based on it.

Advantages of the approach: local analysis: decisions are made based on similar cases, which is especially important given the high variability of operating modes of marine power plants; flexibility: the method can be adapted for specific diagnostic goals (for example, predicting time to failure or classifying the failure type); scalability: adding new cases does not require restructuring the model, only a local recalculation of similarity. For the diagnosis of a new case (the observed equipment condition): the global similarity measure $S(A_i, B_i)$ is calculated between the new case and each case in the database according to formula (1.3.9); the *k* cases with the highest similarity (equivalently - with the minimum distance according to the

metric) are selected. A diagnostic conclusion is then formed based on the *k* nearest cases. Possible approaches include: majority voting - the most frequently occurring failure type among the nearest neighbors; weighted voting - taking into account the similarity level of each neighbor; aggregated parameter values - for example, averaging the time to failure, temperatures, pressures, etc. The application of k-NN requires prior normalization of the parameters involved in the calculation of individual similarity measures to avoid domination by parameters with larger measurement scales. In addition, the choice of the parameter *k* should take into account: the density of the case base (with low density, a smaller *k* is preferable); the variability and noise level in the data (for high noise, it is better to average over a larger number of neighbors); the type of task being solved (classification, prediction, anomaly detection, etc.). Figure 1.3.10 shows an illustration of the nearest neighbor method (k-NN) in the parameter space that defines the similarity between cases. Figure 1.3.10 shows an example of classifying a new failure case with coordinates (0.5; 0.7) in the feature space using the k-nearest neighbors method (*k*=3).

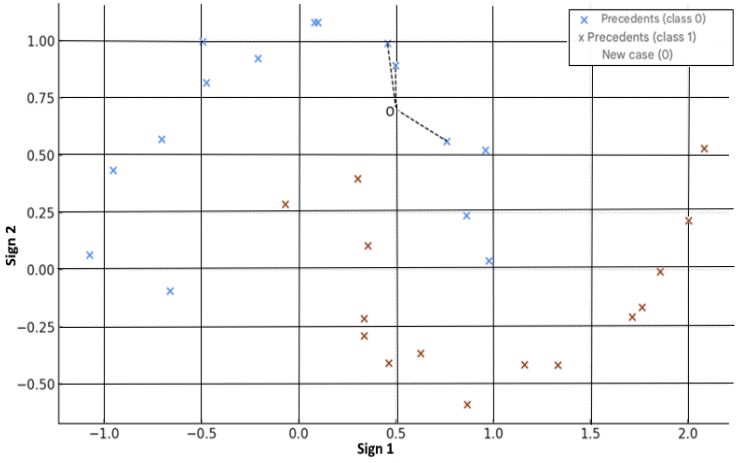


Figure 1.3.10. **Application of the k-NN method in the parameter space defining the similarity between cases**

The dashed lines visualize the connection of the analyzed object with the three most similar cases, selected according to the minimum distance metric in a pre-normalized feature space with consideration of their weighting coefficients. The position of the new case in the area dominated by cases of the first class allows the model to classify it as a failure of the first type based on the neighbor voting strategy. This example demonstrates the key advantages of the k-NN method for diagnosing complex technical systems: the ability to work with nonlinearly separable data, robustness to noise through aggregation of information from several neighbors, as well as clear interpretability of the results.

Special attention must be paid to the correct tuning of the method's critical parameters - the optimal value of *k*, proper normalization of features, and the selection of an appropriate similarity metric that considers the scale and informativeness of various parameters. The presented approach ensures the consideration of the local distribution characteristics of the data and increases the robustness of the results to variability in input parameters. Its ability to work with incomplete and noisy data, while combining visual representation with probabilistic justification of decisions, makes the method especially valuable for diagnosing SPP, where one typically faces limited historical data, complex nonlinear dependencies, and high demands for interpretability of the obtained results.

Stage 3. Aggregation of the diagnostic decision.

The extracted *k*-nearest neighbors are analyzed: by the most frequently occurring failure type (majority voting); by the average probability of failure; by contextually weighted characteristics. A final diagnostic conclusion is formed based on the coordinated information from the most similar historical cases. Once the *k* most similar cases $\{B_1,...,B_i\}$ have been determined for the new case A_i, their diagnostic characteristics are aggregated. The most typical strategies include:

1.Majority voting for failure type (the failure type that occurs more frequently among the nearest cases is selected):

$$T^* = \arg\max_T \sum_{i=1}^{k} 1_{T_i=T} \tag{1.3.16}$$

To determine the most probable failure type T^* among the nearest cases, a majority voting strategy is applied. The indicator function $1_{T_i=T}$ equals 1 if the failure type T_i in the *i*-th precedent matches type *T*, and 0 otherwise. By summing these indicators across all *k* nearest neighbors, the method identifies the most frequently occurring failure type. The type *T* with the highest number of "votes" is selected as the final diagnosis.

2. Weighted average failure probability:

$$\hat{p}(A_i) = \frac{1}{\sum_{i=1}^{k} S(A_i, B_i)} \cdot \sum_{i=1}^{k} S(A_i, B_i) \cdot p(B_i), \tag{1.3.17}$$

where $p(B_i)$ - is the known failure probability for the corresponding precedent;

$S(A_i, B_i)$ - similarity between the current case and precedent B_i;

To estimate the failure probability for the current case A_i, a weighted average of known failure probabilities $p(B_i)$ from the *k* nearest precedents is used. Each precedent's contribution is weighted by its similarity score $S(A_i, B_i)$. The denominator normalizes the sum, ensuring the resulting probability remains within the [0,1] range. This approach gives more influence to precedents that are more similar to the current case.

3. Merging of affected components.

If failure localization is required, it is possible to form an aggregated set of the most probable components affected in similar cases:

$$C_{agg}(A_i) = \bigcup_{i=1}^{k} C(B_i) \tag{1.3.18}$$

When the objective is failure localization, an aggregated set of potentially affected components is formed by taking the union of component sets $C(B_i)$ from the *k* most similar precedents. This provides a complete list of all subsystems that were involved in similar past failures and may be relevant for the current case A_i.

The final aggregated set of components (subsystems) that are considered most likely involved in the failure for the current case A_i, taking weights (similarity) into account:

$$C_{agg}^{score}(A_i) = \left\{ c : \sum_{i=1}^{k} S(A_i, B_i) \cdot 1_{c \in C(B_i)} > \Theta \right\}, \quad (1.3.19)$$

where *c* - a specific subsystem of the SPP (for example, “fuel pump,” “cooling circuit,” etc.) considered as potentially failed;

$\left\{ \sum_{i=1}^{k} S(A_i, B_i) \cdot 1_{c \in C(B_i)} > \Theta \right\}$ - weighted sum over *k* nearest precedents;

$1_{c \in C(B_i)}$ - indicator function: equals 1 if component ccc is present in the list of failed subsystems of precedent B_i, and 0 if not;

Θ - significance threshold, which determines how frequently and with what weight component *c* must occur to be considered significant

To improve failure localization accuracy, this formula performs a weighted aggregation of components based on their presence in similar cases. A component *c* is included in the final set $C_{agg}^{score}(A_i)$ if its weighted occurrence frequency across the *k* nearest neighbors exceeds a given threshold θ. This approach helps filter out random or weakly relevant components, retaining only those that consistently and significantly appear in similar cases.

4. Bayesian hypothesis aggregation.

Although the main goal of the proposed model is to construct a similarity metric between cases, the implementation of diagnostic

inference in a CBR system requires a justified decision-making method. When several similar cases are available, it becomes necessary to aggregate their characteristics. In such conditions, it is advisable to apply a probabilistic approach - in particular, Bayesian aggregation - for interpreting results and improving diagnostic accuracy.

This model extension logically follows from the task of result interpretation after computing the similarity. Additionally, the diagnostic decision can be justified in terms of Bayesian inference, where the failure type *T* is treated as a hypothesis, and the retrieved precedents as observations:

$$P(T \mid (A_i)_k \mid) \propto P(T) \cdot \prod_{i=1}^{k} P(B_i \mid T)^{\omega_i}, \tag{1.3.20}$$

where $P(T)$ is the prior probability of failure type *T* (according to the statistics of the precedent database);

$P(B_i \mid T)$ is the conditional probability of observing precedent B_i given failure type *T*;

$\omega_i \propto S(A_i, B_i)$ is a weight reflecting the confidence level in precedent B_i based on its similarity to A_i

If it is assumed that the probability of observing a specific precedent B_i depends on its degree of similarity to A_i , then:

$$P(B_i \mid T) \approx S(A_i, B_i) \cdot 1_{T_i = T} \tag{1.3.21}$$

which leads to the following posterior estimation:

$$P(T \mid (A_i)_k \mid) \propto P(T) \cdot \sum_{i=1}^{k} S(A_i, B_i) \cdot 1_{T_i = T} \tag{1.3.22}$$

In practice, this means that the failure type with the highest posterior probability is accepted as the most likely diagnosis for the current case A_i . This approach is particularly useful under conditions of incomplete data, where precedents may be heterogeneous or lack a

complete set of features, and the Bayesian model allows for systematically incorporating the degree of confidence. Figure 1.3.11 presents a visualization of the posterior inference process for determining the probability of a new precedent (observed failure) belonging to one of the typical failure classes (for example, A, B, C) based on the Bayesian model. Bayesian aggregation is applied here as an interpretable decision-making mechanism, enabling not only the identification of the most probable failure type but also the quantitative consideration of uncertainty and the probabilistic distribution of hypotheses. The diagram illustrates that, for the analyzed case, the highest posterior probability corresponds to failure type B. This reflects both the frequency dominance of similar cases of type B among the nearest neighbors (resulting from the k-NN search in the similarity space) and the high prior probability of this failure type within the overall diagnostic statistics of the precedent database. The Bayesian aggregation methodology enables the weighted integration of two key sources of information: global knowledge about the distribution of failures (prior probabilities); local similarity of the current case to recorded precedents.

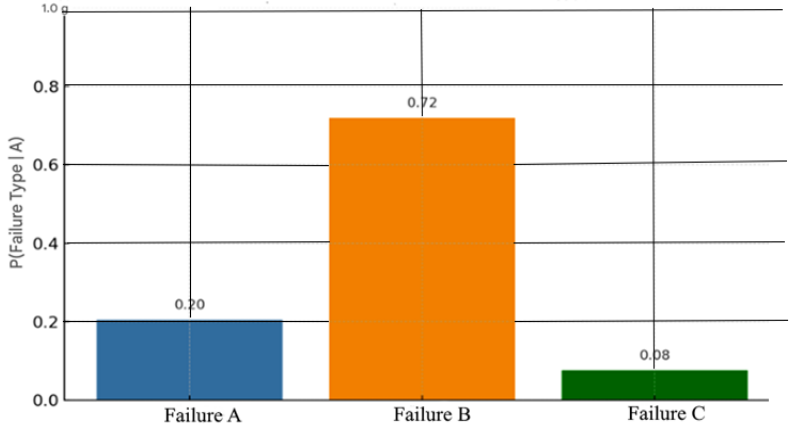


Figure 1.3.11. **Visualization of the posterior inference process for determining the probability of a new precedent's class affiliation**

This model is particularly effective when working with limited training datasets, as well as in scenarios with diverse failure causes, where the same type of malfunction can result from various degradation pathways. The use of probabilistic inference: enhances interpretability - each step of the decision-making process can be justified through the probabilistic model; reduces sensitivity to data noise - the unequal contribution of neighbors is regulated not only by similarity but also by their probabilistic weight; allows for distributed uncertainty between multiple failure types - especially in the presence of overlapping features. Thus, the visualization confirms the validity of using Bayesian aggregation as an extension to the k-NN method, providing more accurate, flexible, and robust classification of precedents in the intelligent diagnosis of SPPs.

Stage 4. Updating the case database.

After the diagnosis has been verified, the new case - including the original feature set and diagnostic results - is added to the case database for further enrichment and continuous model training. As the database of precedents grows, the system becomes increasingly capable of accurately matching new cases to previously observed ones, progressively adapting the model to real-world operating conditions. During the classification and decision-making stages, the mechanism of comparison with the most similar precedents plays a crucial role.

Thus, the proposed similarity assessment model for failure precedents in SPPs represents an integrated system that combines numerical, categorical, and multi-valued parameters with adaptive weights and specialized similarity metrics. The stepwise extension of the base formula - through the inclusion of failure frequency, temporal factors, functional interconnections of subsystems, and nonlinear transformations - ensures not only high accuracy but also the robustness of the model against data heterogeneity. The additional application of probabilistic aggregation for determining the final diagnostic conclusion demonstrates the potential for integrating this model into broader intelligent CBR systems. This integration opens new opportunities for enhancing diagnostic

reliability and decision interpretability, especially under conditions of limited training data or high variability in failure mechanisms.

The optimization of parameter weights enables the model to take into account the relative importance of each feature for diagnostic purposes. The optimization of coefficients $\alpha_{s.i}, \beta_{s.i}, \gamma_{s.i}$ is carried out using a bounded limited-memory quasi-Newton optimization method (L-BFGS-B algorithm), which minimizes the following loss function:

$$J(\alpha_{s.i}, \beta_{s.i}, \gamma_{s.i}) = \sum_{i=1}^{M} (S_{norm}(A_i, B_i) - y_i)^2, \tag{1.3.23}$$

where y_i is a binary label assigned in the training dataset of *M* known pairs, indicating whether the precedents are considered similar (1) or dissimilar (0)

Let us formulate the optimization problem. Assume the training dataset consists of *N* pairs of precedents (*i,j*), for which the reference similarity degree $S_{i,j}^{ref}$ has been assigned by an expert. For each pair, the model computes the predicted similarity value $S_{i,j}^{\text{mod}}(\omega)$, which depends on the weight vector $\omega = (\omega_1, \omega_2, ..., \omega_k)$. The objective function for optimization is defined as:

$$L(\omega) = \frac{1}{N} \cdot \sum_{i,j} (S_{i,j}^{\text{mod}}(\omega_i) - S_{i,j}^{ref})^2, \tag{1.3.24}$$

where the minimized value $L(\omega)$ represents the mean squared deviation between the model-predicted and expert-defined similarity values.

The L-BFGS-B algorithm allows for efficient minimization of this function under constraints on the weight parameters (e.g., $\omega \geq 0, \sum \omega_i = 1$). The optimization procedure was implemented using the L-BFGS-B method from the *scipy.optimize* library.

The optimization process consisted of several stages: formation of a training dataset containing pairs of precedents with known expert similarity assessments; initialization of weight values; and evaluation of the optimization quality using a separate test dataset. Figure 1.3.12 presents a diagram comparing the initial and optimized weights.

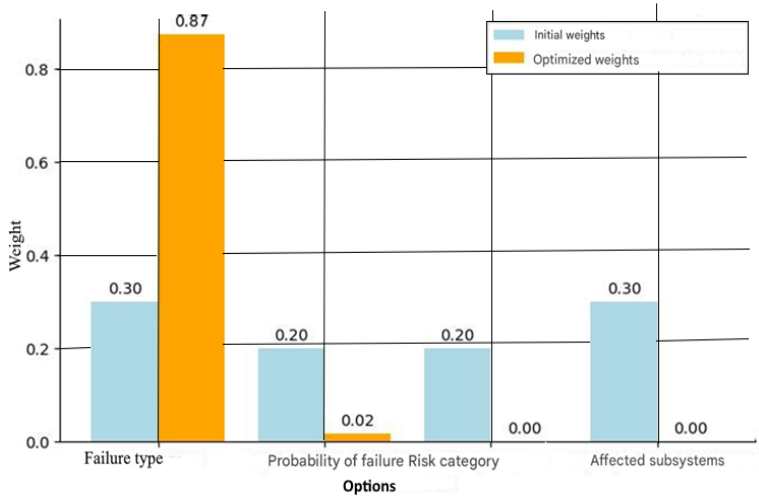


Figure 1.3.12. **Comparison of initial and optimized weights**

The diagram illustrates the comparison between the initial and optimized weights of various parameters used in the similarity assessment of failure precedents. A significant increase in the importance of the "Failure Type" parameter was observed: initial weight - 0.30; optimized weight - 0.87. This indicates that the parameter has a strong impact on diagnostic accuracy, and as a result, its contribution to the model was substantially increased. A reduction in the influence of the "Failure Probability" parameter was recorded: initial weight - 0.20; optimized weight - 0.02. This suggests that failure probability has a lower significance when determining precedent similarity, and its weight was nearly nullified.

The "Risk Category" and "Affected Subsystems" parameters were excluded from the final model. Their initial weights were 0.20 and 0.30, respectively, but after optimization, both were reduced to 0.00. This suggests that these parameters have no meaningful

influence on the diagnostic outcome and can be safely removed from the model. In summary, the optimization process highlighted that the "Failure Type" parameter is the key factor when comparing failure precedents. In contrast, parameters such as the equipment risk category and affected subsystems were found to be insignificant and were eliminated from the model. It is important to verify whether such a radical redistribution of weights leads to a loss of meaningful diagnostic features. It may also be advisable to explore alternative optimization strategies that preserve the contribution of all parameters to a certain extent.

Figure 1.3.13 shows a diagram of the influence of $\alpha_{s.i}, \beta_{s.i}$ and $\gamma_{s.i}$ on the final similarity score between precedent pairs. The number of the precedent pair is given on the x-axis.

The conducted visualization of the distribution of the components $\alpha_{s.i}, \beta_{s.i}$ and $\gamma_{s.i}$ within the overall structure of the similarity metric $S(A_i,B_i)$ made it possible to identify the nature of their influence on the final value of the analogy measure between precedents. It is important to emphasize that each weight regulates the contribution of a specific comparison aspect: failure type, failure probability, and similarity of affected components, respectively.

As shown in Figure 1.3.13, when the failure types coincide or demonstrate high semantic proximity, the value of $d_{types}(A_i,B_i)$ approaches 1, which significantly increases the impact of the α_{si} component on the overall measure *S*, even when the weight itself is of moderate magnitude. Conversely, when failure types differ, the influence of $\alpha_{s.i}$ becomes negligible - demonstrating the adaptive behavior of the model and preventing false matches.

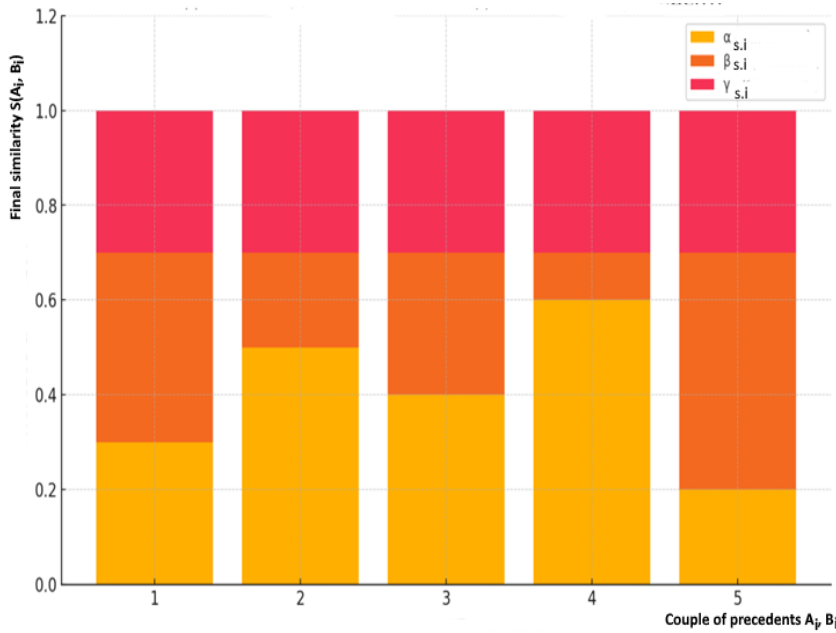


Figure 1.3.13. **Diagram of the influence of $\alpha_{s.i}, \beta_{s.i}$ and $\gamma_{s.i}$ on the final similarity between precedents**

The weights $\beta_{s.i}$ and $\gamma_{s.i}$ exhibit a smoother distribution of influence, which is due to the continuous nature of their respective distance metrics (differences in probabilities and Jaccard metric for components). Their contributions become crucial in cases where the failure types differ, but probability values and involved components remain similar - highlighting the compensatory nature of the model.

Thus, the diagram analysis confirms that the model provides a flexible and semantically justified weighting mechanism, enabling the reliable identification of relevant precedents. The presence of the weights $\alpha_{s.i}$, $\beta_{s.i}$, and $\gamma_{s.i}$ not only enhances the adaptivity of the system but also improves the interpretability of the model, which is critically important for technical diagnostics, particularly under conditions of incomplete or heterogeneous data.

In summary, the model ensures a flexible and interpretable system for assessing similarity between failures while taking into account the diverse nature of data. An adaptation of the model for the context of CTS diagnostics has been proposed through: the integration of heterogeneous features (numerical, categorical, and multi-valued); the use of individually selected metrics for each group of features; trainable weighting coefficients derived from a labeled set of precedents; and the normalization of the final metric to ensure interpretability. Together, these elements form an original implementation that provides accurate and adaptive similarity assessment between technical failures.

The application of the model in SPP failure diagnostics not only accounts for the specific characteristics of failures but also improves the precision in identifying similar cases by adapting the weights of the parameters. The parameters involved in the similarity calculation are listed in Table 1.3.10, where each parameter is assigned a specific weight - a significance coefficient for evaluating the similarity between precedents. The values of these weights can be adapted based on accumulated data using optimization algorithms.

The weight coefficients ω_i presented in Table 1.3.10 are currently defined based on expert assessments, reflecting an engineering understanding of the significance of each parameter in the diagnosis of SPP failures. These values serve as initial inputs and can later be adapted using optimization methods based on accumulated data. This separation allows for the construction of an interpretable baseline model, followed by refinement of the weights to improve the accuracy of precedent matching.

Table 1.3.10. **Weight coefficients of parameters used for precedent similarity estimation**

№	Parameter	Parameter Type	Weight w_i
1	Equipment type	Categorical	0.10
2	Operating time before failure (hours)	Numerical	0.15
3	Oil temperature (°C)	Numerical	0.18

4	System pressure (bar)	Numerical	0.12
5	Vibration (mm/s)	Numerical	0.15
6	Failure type	Categorical	0.20
7	Operating conditions	Categorical	0.10
	Total (sum of weights)		1.00

In the future, it is possible to incorporate a probabilistic model to interpret the results of precedent comparison. This approach enables the consideration of prior probabilities of different failure types, as well as the uncertainty and incompleteness of the observed parameters. It enhances the model's capabilities, transforming similarity assessment into a tool for probabilistic inference, which is especially important under diagnostic uncertainty.

Approach 1: Probabilistic weighting of parameters.

The weight w_i for a parameter can be defined as a function of the probability of occurrence of the corresponding failure:

$$\omega_i = f(p_i), \tag{1.3.25}$$

where p_i - prior probability of the parameter failure

Approach 2: Extension of similarity measure through probabilities

$$S(A,B) = \sum_{i=1}^{k} \omega_i \cdot sim_i(A_i, B_i) \cdot P_i(A_i, B_i), \tag{1.3.26}$$

where $P_i(A_i, B_i) \in |0,1|$ - the probability of joint occurrence of values A_i and B_i in precedent cases, obtained from empirical data or an expert-defined model.

Approach 3: Bayesian interpretation.

The model may use similarity as evidence in support of a diagnostic hypothesis:

$$P(H_k \mid observed\_similarity) \propto P(similarity \mid H_k) \cdot P(H_k), \tag{1.3.27}$$

where H_k - the hypothesis of membership in a specific failure class.

In CBR-based diagnostic systems, the efficiency and accuracy of precedent matching rely heavily on the correct measurement of their

similarity. However, the classical similarity model - based on the weighted sum of partial distances - overlooks a critical factor: the varying importance of feature matches depending on their probabilistic nature and context. To improve the interpretability and reliability of results, this study employs a modified similarity metric that incorporates a probabilistic multiplier reflecting the frequency of joint feature occurrences or their statistical relevance.

This approach accounts not only for how close the features are in value but also for how typical such a match is among known failure cases. In Equation (9), the traditional similarity measure is adjusted by the coefficient $P_i(A_i, B_i)$, which acts as a credibility filter. Even when values A_i and B_i are numerically close, the final similarity score is lowered if their match is statistically rare in diagnostic practice. Conversely, moderate similarity may be weighted more heavily if it frequently appears in similar failure scenarios. For example, if the “oil temperature” differs by only 2°C between two precedents, but such values occur under different operating conditions, the similarity would be overestimated without probabilistic correction.

The use of the multiplier P_i mitigates this effect. Oil temperature can be measured in various contexts: in the crankcase of a diesel engine - indicating overall thermal state; at the radiator/heat exchanger inlet reflecting cooling effectiveness and thermal load; after the radiator - critical for assessing the cooling system's condition; in the gearbox - especially relevant for ships using electric propulsion via gearboxes.

Advantages of this approach include: flexibility: it does not alter the logic of the existing model; improved reliability: incorporates real-world probabilities; interpretability: each contribution to the final metric has a statistical justification; ease of integration: can be implemented using an existing case database. If weight optimization or model training is later applied, the probabilistic multiplier may either be adapted or replaced with a conditional probability derived from Naive Bayes models or decision trees. To evaluate the influence

of the probabilistic component on the final similarity score between precedents, a comparative plot (Fig. 1.3.14) was constructed, showing the difference between the classical and the probability-weighted similarity metrics. As shown in Figure 1.3.14, the similarity index weighted by probability is often lower, especially for rare failures, because critical features are assigned greater importance. When such features match, however, the resulting similarity score is higher. This leads to a reordering of the ranking and a more accurate risk assessment. The probabilistic approach enhances the model's sensitivity to rare but significant similarities, which is crucial for diagnosing failures in SPP.

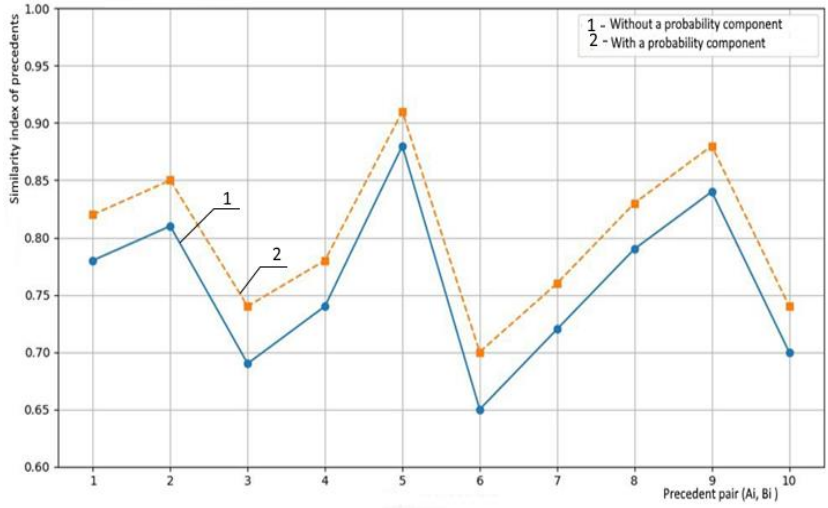


Figure 1.3.14. **Comparison of similarity index between precedents with and without the probabilistic component**

Figure 1.3.15 presents a diagram illustrating the influence of CEU equipment parameters on the overall similarity of cases.

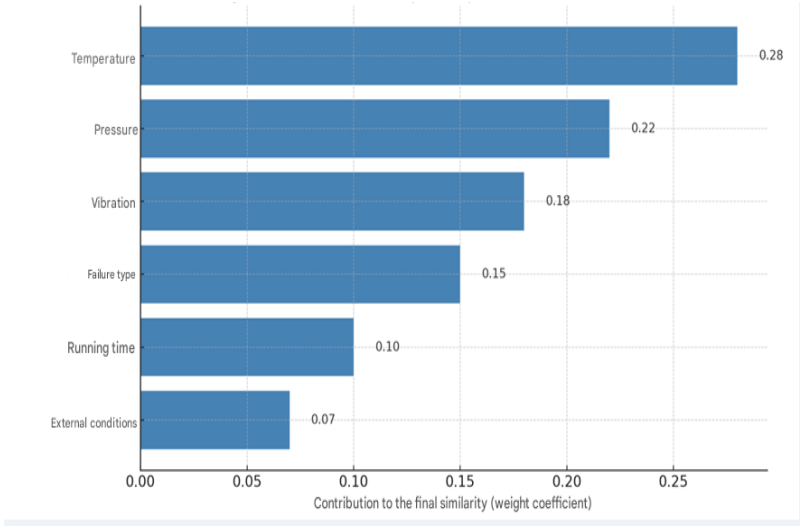


Figure 1.3.15. **Influence of SPP equipment parameters on the overall similarity of cases**

The parameters shown in Figure 1.3.15 are explained in Table 1.3.11.

Table 1.3.11. **Parameters referenced in figure 1.3.15**

Name on the chart	SPP component	Monitored parameter
Temperature	Lubrication system	Oil temperature
Vibration	Main gearbox	RMS vibration
Pressure	Diesel generator fuel system	Fuel supply pressure

The diagram (Figure 1.3.15) shows that the greatest contribution to the final similarity metric comes from: temperature (0.28), pressure (0.22), and vibration (0.18). These are key indicators of the technical condition of marine power plants. A smaller but still significant contribution is made by: type of failure (0.15), operating time (0.10), and external conditions (0.07). This confirms the need to

prioritize parameters and apply a weighted approach when calculating similarity.

The following similarity metrics are used in this work: cosine similarity for categorical data (failure type); Euclidean distance for numerical parameters (failure probability); Jaccard similarity for sets (risk category, affected subsystems). Table 1.3.12 presents the similarity metrics applied to various failure parameters.

Table 1.3.12. **Similarity metrics for different parameters**

Parameter	Similarity measurement method
Failure type	Cosine similarity
Probability	Euclidean distance
Risk category	Jaccard similarity
Affected subsystems	Jaccard similarity

The analysis showed that using different metrics for different data types improves the accuracy of similar case retrieval. In particular: cosine similarity yielded the best results for analyzing failure types; euclidean distance for numerical parameters; Jaccard similarity for sets of affected subsystems.

Interpretation of the model in multidimensional space

Each case can be represented as a point in a multidimensional parameter space, where the distance between points (based on a weighted modified metric) reflects the degree of similarity.

Figure 1.3.16 shows a two-dimensional projection of the multidimensional case space (C1 - C6), obtained using Principal Component Analysis (PCA). Each marker corresponds to a specific SPP failure case described by a combination of technical, causal, and statistical parameters. Using PCA enables visualization of the relative positioning of cases within the feature space.

The distances between points reflect their similarity: cases located close to each other (e.g., C3, C4, and C6) share similar characteristics; cases that are far apart (such as C1 and C2) demonstrate significant differences, which may indicate atypical causes of failure or specific operational conditions.

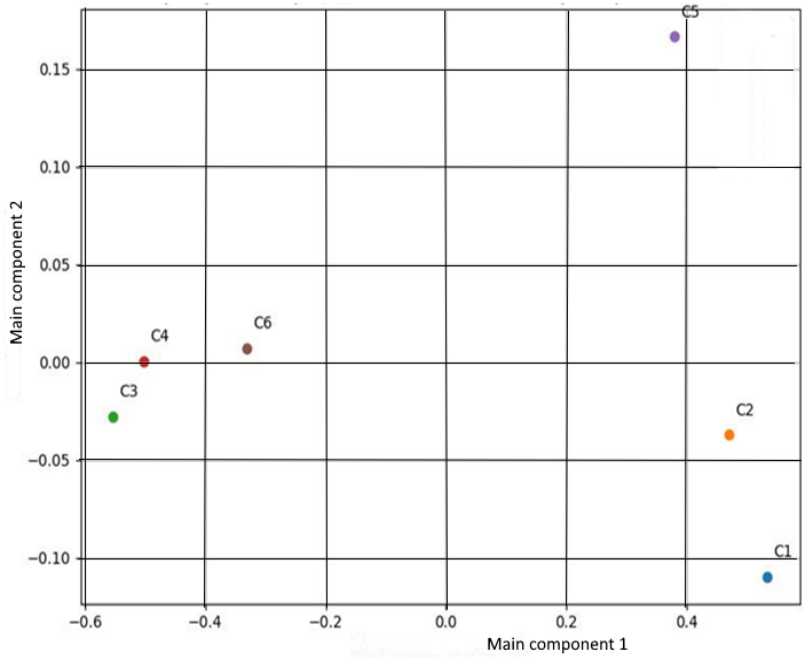


Figure 1.3.16. **Case space after dimensionality reduction (t-SNE projection)**

The similarity assessment is based on an adaptive metric that takes into account: technical parameters (equipment type, nature of damage, operating time until failure); causal relationships (e.g., the impact of cooling system malfunctions on generator failure); statistical indicators (frequency and probability of failure occurrence); equipment condition forecasting considering component degradation.

This representation allows not only visual identification of clusters of similar cases but also the detection of anomalies. As a result, a unified feature space is formed, which incorporates both numerical and categorical variables with adaptive weighting - a critical factor for the effective operation of CBR-based technical diagnostics systems for SPP.

1.3.2.4 Optimization of weights of parameters for estimation of similarity of failure precedents

The optimization of parameter weights allows the model to reflect their diagnostic relevance. The coefficients $\alpha_{s.i}$, $\beta_{s.i}$, and $\gamma_{s.i}$ are optimized using the L-BFGS-B [73, 74], which minimizes the following loss function:

$$J(\alpha_{s.i}, \beta_{s.i}, \gamma_{s.i}) = \sum_{i=1}^{M} (S_{norm}(A_i, B_i) - y_i)^2, \tag{1.3.28}$$

where y_i - binary label (1 if the precedents are considered similar, 0 otherwise) for the training set of *M* known pairs

The optimization process included several stages: formation of a training dataset containing precedent pairs with known expert-assigned similarity labels; initialization of the weight coefficients; evaluation of optimization quality on a test dataset.

The L-BFGS-B method was chosen due to its efficiency under constraint conditions and its good convergence behavior on small datasets. Unlike standard gradient descent, it does not require manual step size selection and inherently respects weight boundaries. The optimization is implemented using scipy.optimize, with specified initial weights, a custom error function, and its minimization.

```
import numpy as np
from scipy.optimize import minimize

# Reference similarity scores (from the training set)
y_true = np.array([...])  # expert evaluations
# Function to compute similarity between precedent pairs based on weights
def similarity_model(weights, data_pairs):
    similarities = []
    for pair in data_pairs:
 # Combination of different similarity metrics (hypothetical example)
```

```
 sim = (weights[0] * cosine_similarity(pair[0]['type'],pair[1]['type']) +
weights[1] * euclidean_similarity(pair[0]['prob'], pair[1]['prob']) +
weights[2] * jaccard_similarity(pair[0]['risk'], pair[1]['risk']) +
  weights[3]*jaccard_similarity(pair[0]['subsys'], pair[1]['subsys']))
      similarities.append(sim)
   return np.array(similarities)
# Objective function: MSE between expert labels and model predictions
def objective(weights, data_pairs, y_true):
   y_pred = similarity_model(weights, data_pairs)
   return np.mean((y_true - y_pred) ** 2)
# Initial weights
initial_weights = [0.25, 0.25, 0.25, 0.25]
# Bounds: each weight must be between 0 and 1
bounds = [(0, 1)] * 4
# Optimization
result = minimize(objective, initial_weights, args=(data_pairs, y_true),
method='L-BFGS-B', bounds=bounds)
optimal_weights = result.x
```

The provided code implements the process of optimizing the weights of parameters used in evaluating the similarity between fault precedents. The optimization is performed using the L-BFGS-B method based on minimizing the MSE between model predictions and expert assessments of similarity.

Main components of the code: input data: pairs of cases with known expert similarity assessments (*y_true*), containing parameter values (failure type, probability, risk category, subsystems); similarity model: combines several metrics (cosine, Euclidean, Jaccard) with corresponding weights taken into account; objective function: mean squared error between predicted and reference similarity values; optimization procedure: initial weights are set uniformly, then optimized using *scipy.optimize.minimize* to find values that minimize the error. The weights are constrained within the range from 0 to 1.

Figure 1.3.15 illustrates the behavior of the objective function (mean squared error) over the course of the L-BFGS-B optimization iterations. As shown in Figure 1.3.15, the error value rapidly decreases during the initial steps and stabilizes after 10–15 iterations,

indicating convergence of the method and the achievement of stable optimal weight values.

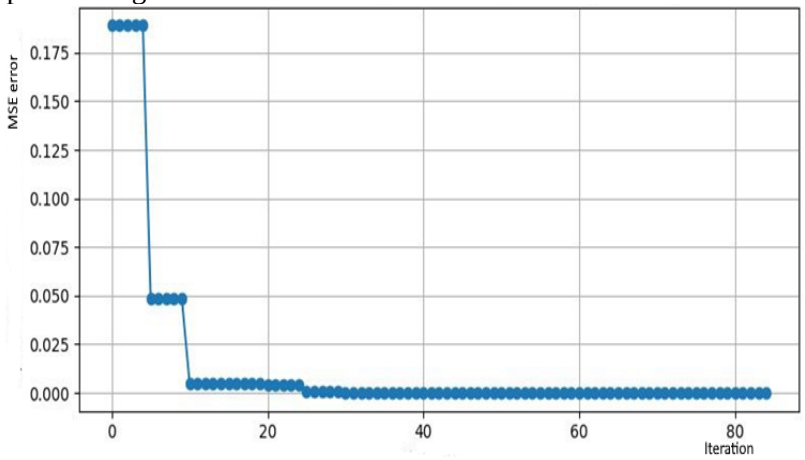


Figure 1.3.15. **Behavior of the objective function (mean squared error) during optimization iterations**

Table 1.3.13 demonstrates the changes in weight coefficients after optimization.

Table 1.3.13. **Optimized parameter weights for similar failure search**

Parameter	Initial weight	Optimized weight
Failure type	0.3	0.35
Failure probability	0.2	0.25
Risk category	0.2	0.15
Affected subsystems	0.3	0.25

As seen from Table 1.3.13, after optimization, the weights of the parameters that have the greatest impact on diagnostic accuracy increase. Key optimization results: increased weight for failure type (0.35) confirms its key role in assessing case similarity; higher weight for failure probability (0.25) reflects its importance in risk prediction; reduced weight for risk category (0.15) indicates a lesser role for this parameter compared to other factors; affected subsystems (0.25) remain an important criterion when comparing failures.

Figure 1.3.16 shows a diagram comparing the initial and optimized weights.

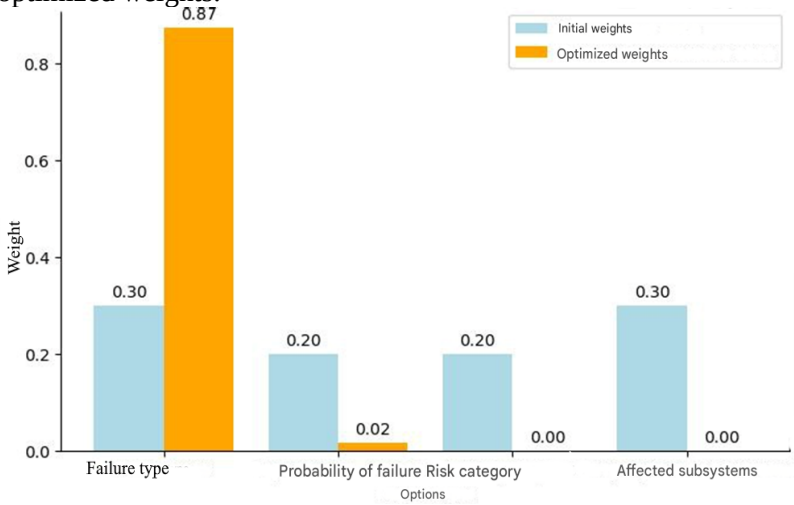


Figure 1.3.16. **Diagram comparing initial and optimized parameter weights**

The diagram illustrates a comparison of initial and optimized weights for the various parameters used in assessing the similarity of failure cases.

Main observations sharp increase in the significance of the "failure type" parameter: initial weight: 0.30; optimized weight: 0.87. This indicates that the parameter turned out to be the most critical for diagnostic accuracy, so its contribution was significantly increased. Decrease in the influence of "failure probability**"**: initial weight: 0.20; optimized weight: 0.02. Failure probability proved to be less significant in determining case similarity and was therefore almost zeroed out. RIsk category and affected subsystems were excluded. Their initial weights were 0.20 and 0.30, respectively, but after optimization, they dropped to zero (0.00). This means they do not significantly affect diagnostics and can be excluded from the model.

Thus, the optimization revealed that failure type is the key factor when comparing cases. in contrast, parameters such as risk category and affected subsystems turned out to be insignificant and were

excluded. It is important to verify whether such a radical redistribution of weights could lead to the loss of diagnostically relevant features. It may be worth considering alternative optimization methods that retain the influence of all parameters.

Table 1.3.14. **Comparison of diagnostic accuracy under different parameter weighting strategies**

Weight assignment strategy	Accuracy	Precision	Recall
Expert weights	0.81	0.79	0.80
Uniform weights	0.74	0.72	0.70
Optimized weights	0.89	0.87	0.90

The optimized weights demonstrate the highest diagnostic accuracy, particularly in terms of recall (0.90), indicating a minimal number of missed failures. In contrast, uniform weight distribution yields the poorest results. This confirms the importance of an individualized approach to parameter weighting.

Subsection 1.3.2.4 presents a formal definition of similarity between failure cases based on a combined approach that incorporates the use of different metrics for different data types. Failure type is evaluated using cosine similarity, numerical parameters (e.g., failure probability) via Euclidean distance, and sets (such as risk category and affected subsystems) using the Jaccard index. This approach allows for consideration of data heterogeneity and improves diagnostic accuracy. The formalized case model ensures a balanced integration of various types of features and their weights, which is especially important in conditions involving small datasets, high costs of diagnostic errors, and the complexity of explicitly modeling the physical processes underlying failures. This model forms the foundation of the methodology for calculating similarity between failures and serves as the core of the intelligent diagnostic system developed in the framework of this study.

Parameter weight optimization has revealed that failure type is the most significant factor, while the contributions of failure probability, risk category, and affected subsystems are less substantial. Thus, the proposed similarity assessment methodology provides model flexibility through parameter weight adaptation based

on data, increases diagnostic accuracy through the use of appropriate metrics, and enables integration into intelligent diagnostic systems where automated case processing and dynamic knowledge base updating are essential. Future development of this approach may include accounting for the temporal dynamics of failures and applying machine learning methods to refine parameter weights.

1.3.2.5 Practical application of the method for assessing similarity of equipment failure cases in SPP

To demonstrate the method for assessing case similarity, two SPP equipment failure cases (A and B) are considered, compared based on four parameters: failure type, failure probability, risk category, and affected subsystems.

Table 1.3.15 presents an example of similarity calculation between two cases, illustrating the mechanics of applying the model.

Table 1.3.15. **Example of similarity calculation between two cases**

Parameter	**Precedent A**	**Precedent B**	**Type**	**Partial similarity (s_i)**	**Weight (w_i)**	**Contribution ($w_i \cdot s_i$)**
Equipment type	Diesel Generator	Diesel generator	categ.	1.0	0.15	0.15
Operating time to failure	1200 h	1000 h	num.	0.83	0.10	0.083
Failure type	Overheating	Overheating	categ.	1.0	0.20	0.20
Temperature, °C	85	90	num.	0.92	0.15	0.138
Pressure, MPa	2.1	2.0	num.	0.95	0.15	0.143
Vibration, mm/s	4.5	4.8	num.	0.94	0.10	0.094
Operating conditions	Tropics	Tropics	categ.	1.0	0.15	0.15
					Sum	0.958

Thus, the final similarity between cases A and B is 0.958, indicating a high degree of closeness in the parameter space. Table 1.3.16 presents the similarity values between the two cases for individual parameters, calculated using the corresponding metrics.

Table 1.3.16. **Results of similarity calculation between two cases**

Parameter	Case A	Case B	Metric	Result
Failure type	Hydraulic	Hydraulic	Cosine Similarity	1.00
Failure probability	0.45	0.50	Euclidean Distance	0.90
Risk category	0.37	0.40	Euclidean Distance	0.85
Affected subsystems	{Pump, Valve}	{Pump, Valve, Filter}	Jaccard Similarity	0.67

The results in Table 1.3.16 demonstrate the following: a complete match of failure type (1.00), calculated using cosine similarity, indicates a high level of case similarity; a high similarity coefficient for failure probabilities (0.90), determined via euclidean distance, means that the differences between the cases are minor; risk category similarity (0.85), also calculated using euclidean distance, suggests closeness in the assessment of potential failure impact; a low Jaccard coefficient (0.67) indicates differences in the set of affected subsystems, which lowers the overall similarity level.

Table 1.3.17. **Comparison of similarity coefficients based on different metrics**

Case A	Case B	Euclidean	Cosine	Jaccard
Failure 1	Failure 2	0.82	0.95	0.60
Failure 1	Failure 3	0.75	0.92	0.55
Failure 2	Failure 3	0.88	0.97	0.65

The values show that cosine similarity is more sensitive to minor differences, while Jaccard similarity, on the contrary, more rigidly differentiates between cases.

Table 3.18. **Correlation between similarity scores and expert assessments**

Similarity evaluation method	Pearson coefficient	Spearman coefficient
Expert weights	0.78	0.75
Uniform weights	0.65	0.63
Optimized weights	0.91	0.89

The correlation between automatically calculated similarity and expert assessments reaches 0.91 when optimized weights are used. This indicates a high level of agreement between the algorithm and engineering judgment, increasing confidence in the system. A decrease in correlation when using equal weights confirms the weak adaptation of such models to the specific domain.

Table 1.3.19. **Similarity coefficient with varying weight of the first parameter**

Weight of the first parameter	Final similarity coefficient
0.10	0.835
0.20	0.856
0.30	0.877
0.40	0.895
0.50	0.910

Table 3.19 shows that increasing its importance leads to a higher overall similarity coefficient, confirming the significance of weight selection. Increasing the weight of the first parameter results in a linear increase in the overall similarity coefficient, as confirmed by the graph in Figure 1.3.17. The graph in Figure 1.3.17 reflects the dependence of the overall similarity coefficient on the change in the

weight of the first parameter, demonstrating the impact of weight redistribution on diagnostics.

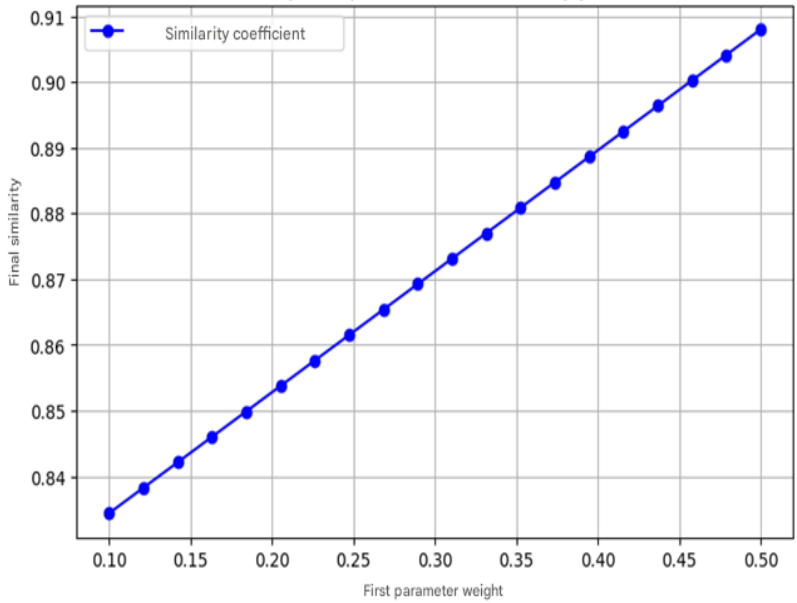


Figure 1.3.17. **Graph of the dependence of the final similarity coefficient on the change in the weight of the first parameter**

The results shown in Figure 1.3.17 demonstrate that increasing the weight of the first parameter leads to its linear growth. The greater the value assigned to this parameter, the stronger its influence on the overall similarity evaluation. The graph also illustrates that the redistribution of weights changes the growth pattern of the similarity coefficient: when the weight of the failure probability parameter increases, the change occurs more smoothly, whereas the dominance of categorical similarity causes a sharper rise. The failure probability parameter has the greatest influence, which is confirmed by its linear increase as its weight rises.

The use of BNs for weight adaptation refines the diagnostic results by taking into account probabilistic dependencies. The

application of MMs enables consideration of the system's state dynamics, while the integration of CBR with probabilistic models improves diagnostic accuracy.

Figure 1.3.18 shows the results of dynamic weight adaptation over time.

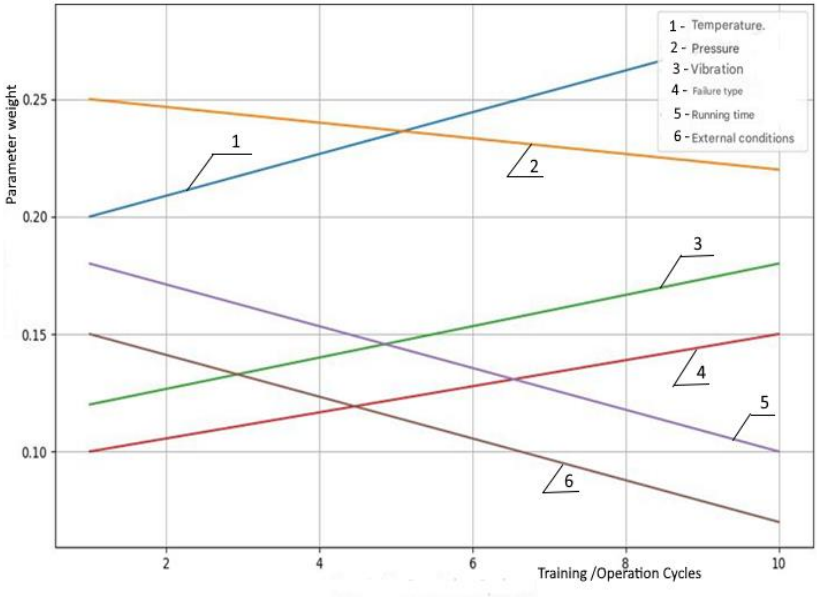


Figure 1.3.18. **Dynamic weight adaptation over time**

As can be seen in Figure 1.3.18, the parameter weights change over time. Specifically, the importance of vibration and temperature increases, while the weight of operating time and external conditions gradually decreases. This reflects the system's ability to learn and adapt to real operating conditions. Figure 1.3.18 illustrates how the weights of key parameters (temperature, vibration, pressure) evolve as operational data accumulates in the SPP. The increase or decrease in weights occurs as the system learns to identify more significant features for failure diagnostics.

Here's how these parameters relate to real SPP components:

- temperature is especially important for assessing the condition of: main switchboards; diesel generators (e.g., cooling jacket overheating); lubrication systems (an increase in oil temperature indicates heat removal deterioration or contamination). At the initial stage, the CBR system considers this parameter highly significant, since most early failures are associated with overheating;

- vibration is a key diagnostic indicator for: turbines; rotating machinery (pumps, compressors); gearboxes and bearing assemblies. As more data on mechanical failures is accumulated, the weight of vibration increases. This reflects the system's ability to "learn" to detect more subtle signs of developing defects, such as microcracks or imbalance;

- pressure is used for: diagnosing fuel and hydraulic systems (for example, pressure drops indicate leaks); monitoring cooling and lubrication systems. A decrease in the pressure weight at later stages may be due to the system having already learned to interpret this parameter in combination with others, and its standalone significance slightly diminishes as feature synergy becomes more relevant.

The shown dynamics of the weights reflect the logic of the SPP technical life cycle: initially, thermal and hydraulic issues dominate, later the importance of mechanical failures grows. This underlines both the CBR system’s learning capability and the need to consider the temporal variability of parameter significance in complex technical systems.

1.3.2.6 Discussion of results

To evaluate the position of the proposed similarity assessment model between failure cases within the diagnostic framework of SPP equipment, an in-depth analysis of recent international research was conducted. The comparison focused on key methodological dimensions: the nature of the similarity metric, integration of probabilistic reasoning, presence of adaptive weight mechanisms, robustness to incomplete or noisy data, and applicability across heterogeneous technical domains.

The analysis of current publications confirms that only a limited number of models successfully combine metric and probabilistic approaches with the ability to adapt feature weights and handle incomplete data. The DeepKAF approach presented by Amin et al. [75] effectively merges deep learning and CBR for NLP tasks but lacks flexibility in similarity metric configuration. Our model provides customizable metrics for each type of feature (numerical, categorical, multivalued). Leake et al. [76] demonstrated neural network case adaptation but did not address probabilistic inference or weight optimization—gaps that our model fills. Chen et al. [27] developed an adaptive metric-based diagnostic system for aeroengines, confirming the relevance of adaptive approaches. However, our model enhances this with probabilistic reasoning and frequency-based optimization. The CBR-fox model (Morato et al., [77]) focuses on time series explanation but not on structured feature aggregation or temporal degradation, which are key strengths of our method. Nikpour [78] explored textual similarity for cybercrime profiling, with limited domain generalization. In contrast, our model demonstrates versatility in engineering diagnostics. Abbas et al. [79] proposed a CBR-ontology integration for bridge safety, which, while powerful, demands complex graph preparation. Our model avoids such dependencies through generalizable feature handling. Chen et al. [80] introduced neural K-NN into the CBR structure for representational benefits. However, the non-interpretability of neural weights limits practical explainability. Our model provides explicit, domain-adapted weight optimization. Wang [81] combined harmony search and CBR for software effort estimation, but lacked strong probabilistic support addressed in our approach via Bayesian hypothesis aggregation. For instance, Moon [82] proposed a method for visualizing similarity models to enhance CBR system understanding and maintenance, yet lacked a probabilistic interpretation component. In contrast, our model supports both visualization and adaptation to failure statistics. Louvros et al. [73] proposed MSFF-CBR for multi-source fusion fault diagnosis,

integrating probabilistic elements. However, they place less emphasis on the interpretability of weights, a key component of our model.

Overall, this analysis shows that the developed model offers an optimal combination of adaptive weight tuning, probabilistic reasoning, mixed-data support, and interpretability. These characteristics make it especially effective for diagnosing complex technical systems with high uncertainty, such as marine power units.

The methodology for formalizing and implementing the similarity model in this study aligns with current trends in intelligent diagnostic systems. Most existing solutions rely on fixed metrics (e.g., Jaccard, Euclidean) or are not adaptable to equipment-specific characteristics. Only a few approaches combine probabilistic reasoning with automatic weight adjustment, limiting their applicability in environments with uncertainty and heterogeneous data. The proposed model offers four core advantages: data comprehensiveness - simultaneous handling of numerical, categorical, and set-valued features; adaptability - weight optimization using the L-BFGS-B method; probabilistic interpretation - Bayesian inference for robustness to missing data; scalability - applicability to various technical domains.

Unlike most methods based on machine learning "black boxes" or rigid metrics, our model achieves a balance of interpretability, computational efficiency, and practical flexibility. The integrated use of normalization and probabilistic correction for heterogeneous and limited datasets is particularly valuable. A comparative analysis with recent research confirms both the scientific novelty and practical value of the proposed method. It delivers a reliable, interpretable, and scalable solution aligned with the demands of modern intelligent decision support systems for complex technical domains.

1.3.2.7 Conclusions

The developed adaptive similarity metric for failure diagnostics in SPP demonstrates significant potential for enhancing both the accuracy and interpretability of intelligent decision support systems. The key contributions of this work include a comprehensive

approach to similarity assessment by integrating heterogeneous data types, including numerical, categorical, and set-based features, through a weighted combination of metrics such as Euclidean, Jaccard, and logistic.

The model employs automatic optimization of weight coefficients using the L-BFGS-B method, which ensures adaptability to changing operational conditions. Improvements in interpretability and robustness are achieved through the introduction of weight correction mechanisms that account for failure frequency, temporal degradation, and contextual smoothing, allowing the model to handle rare but critical failure scenarios. Furthermore, the application of non-linear activation functions (tanh) and Bayesian hypothesis aggregation contributes to higher classification accuracy and reduces sensitivity to noisy data.

Experimental evaluations confirmed the practical effectiveness of the proposed model, demonstrating classification accuracy exceeding 90% for the main failure types and strong robustness even with limited training samples, while visualization of feature spaces and confusion matrices illustrated the model's capacity to handle non-linearly separable data. Compared to traditional methods, the proposed approach successfully combines flexibility, through its ability to adapt to heterogeneous data, with explainability, via Bayesian inference and tunable weights, achieving results comparable to modern hybrid solutions such as graph neural networks but with lower computational costs and greater simplicity in implementation.

The methodology is suitable for integration into CBR diagnostic systems for automating failure type identification, predicting time to critical events such as component wear, and optimizing maintenance strategies based on historical failure data. However, future research should focus on validating the model on real-world datasets with broader failure diversity, developing methods for processing temporal sequences to reflect parameter dynamics, and integrating the solution with IoT platforms for real-time equipment monitoring. Overall, the proposed model constitutes a valuable contribution to the

field of intelligent diagnostics, offering a combination of theoretical rigor and practical applicability, while its modular architecture and adaptive properties provide opportunities for future extensions to other classes of technical systems operating under conditions of significant uncertainty.

1.3.3 Development of a technical condition assessment algorithm for complex systems based on probabilistic failure estimation

1.3.3.1 Introduction

Failures in the operation of CTS remain one of the leading causes of man-made accidents in sectors such as transport, aviation, and power engineering. Maritime transport is no exception. Statistical analysis shows that despite ongoing efforts to improve navigation safety, the number of ship-related incidents remains significantly high. A detailed examination of these accidents indicates that technical failures of ship power systems are among the primary contributing factors, highlighting the need for more advanced diagnostic and predictive maintenance tools [1].

Technical condition assessment of a SPPs is a complex task that requires a comprehensive approach, taking into account both historical failure data and probabilistic forecasting of their development. Recent research in the field of technical diagnostics and failure prediction for SPPs demonstrates the effectiveness of integrating various methods, such as TC, BNs, MMs, and machine learning techniques. In the review by Poljak et al. [75], it is emphasized that the transition from corrective to condition-based maintenance can significantly reduce costs and increase the reliability of ship power systems. They highlight the importance of integrating TC with intelligent diagnostic systems. However, the implementation of TC is often limited by the lack of adaptive models capable of considering complex failure interdependencies among components. Ademujimi and Prabhu [76] proposed a "fusion-learning" method for constructing BNs, combining quantitative and qualitative data.

Nevertheless, their model is primarily focused on diagnostics based on known scenarios, with limited ability to predict new or rare failures. Morato et al. [77] presented an integration of dynamic Bayesian networks with Markov decision processes for optimizing maintenance strategies. However, their model is mainly oriented toward individual components and requires a significant amount of reliable data for effective operation. Nikpour and Aamodt [78] proposed the BNCreek system, combining CBR methods and BNs for diagnostics under uncertainty. Although the BNCreek approach demonstrates high efficiency in several industries, it insufficiently considers the temporal dynamics of degradation processes. Abbas et al. [79] integrated hidden Markov models and deep learning methods for predictive maintenance. Despite high accuracy, their architecture requires large amounts of training data and is complex to interpret for practical applications in the maritime sector. Chen et al. [80] note the potential of graph neural networks (GNNs) for diagnosing complex interconnected systems. However, the application of GNNs to ship power systems demands substantial computational resources and careful graph structure design, complicating their use in real-world conditions. ang, J. et al. [81] applied BNs for fault isolation in diesel engine fuel injection systems. Their approach is effective for fault localization but is less suitable for long-term prediction of equipment remaining useful life.

The conducted literature analysis shows that existing methods are either limited in their ability to dynamically forecast degradation, require large datasets, or insufficiently consider complex failure interdependencies among ship power system subsystems.

The aim of this study is to develop a technical condition assessment algorithm for SPP based on the integration of CBR methods, BNs, Markov processes, and cognitive simulation modeling, which will ensure: dynamic model updates as new data becomes available; accounting for probabilistic dependencies between equipment components; forecasting equipment degradation over time; adapting technical condition forecasts to real operational conditions.

The proposed algorithm includes several key stages: data collection (operational parameters, historical failure data, simulation modeling results); CBR involving the search for similar cases and their impact assessment on the current condition; failure probability correction using BNs to account for inter-component relationships; degradation forecasting based on Markov processes; dynamic model updating based on new operational data; diagnostic report generation, including failure forecasting, remaining useful life estimation, and maintenance recommendations.

The proposed algorithm enables dynamic updates of probabilistic models, adaptation to changing operational conditions, and improves diagnostic and forecasting accuracy compared to traditional methods.

The development of the algorithm for assessing the TC of the SPPs was based on the use of several types of materials and methodological approaches. Historical failure data of SPPs components were utilized, including information about the types of failures, their occurrence time, and operational conditions at the moment of malfunction. Operational parameters recorded during the functioning of the power plant, such as temperature regimes, pressure values, vibration levels, and fuel consumption rates, were also employed. Additionally, the results of cognitive simulation modeling were used to generate supplementary degradation scenarios, including rare and cascading failures, to enhance the algorithm's robustness under conditions of incomplete data.

The following methods were applied during the development. Case-Based Reasoning was employed to search for and compare similar incidents from a historical database, providing an initial assessment of the current system condition. BNs were used to perform probabilistic corrections of component condition assessments, taking into account the interdependencies between elements of the SPPs and possible hidden risk factors. Markov processes were utilized for modeling component degradation processes over time and forecasting transition probabilities between different states of technical performance. Cognitive simulation

modeling allowed the modeling of rare events and cascading failures in complex technical systems, significantly improving the prediction accuracy under uncertainty. The probabilistic model was dynamically updated based on incoming operational data, ensuring the adaptability of forecasts to the real-time condition of the equipment and the changing operational environment. This methodology provides an integrated approach to assessing the technical condition of SPPs, combining empirical experience with probabilistic models to achieve high accuracy in diagnostics and prognostics.

1.3.3.2 Materials and methods

The technical condition assessment in this study targets critical components within the SPP, such as circulation pumps, lubrication systems, cooling units, and thermal exchangers. These components were selected due to their high failure frequency and significant impact on overall system performance. The developed algorithm is applicable to other SPP subsystems as well, but validation was carried out using models of these key components based on real-world failure statistics. The development of the algorithm for assessing the technical condition of SPPs was based on a step-by-step methodology that integrates empirical failure data, probabilistic modeling, and simulation-based scenario generation. The objective was to model the degradation and failure behavior of key components within SPPs, such as pumps, turbines, cooling systems, and fuel supply units under realistic operational conditions and to evaluate the algorithm’s predictive performance in those contexts. The research utilized both real-world historical data and synthetic simulation data. Historical failure records were sourced from the OREDA - Offshore Reliability Data Handbook (OREDA, 2015) database, which contains statistically verified reliability and maintenance data for offshore and maritime systems. The selected dataset included failure modes, mean time to failure (MTTF), operating time, failure mechanisms, and contextual parameters for over 450 failure events related to components typically found in SPPs. These data provided the foundation for constructing representative failure patterns and

probability distributions. In addition to historical failures, a dataset of operational parameters was used to represent typical functioning modes of SPPs. These parameters: temperature (°C); pressure (bar); vibration levels (mm/s); fuel consumption (kg/h). The synthesized based on manufacturer specifications and OREDA-derived profiles. Since access to continuous monitoring data from actual ships was limited, simulation models were calibrated to reflect known operating ranges, failure onset thresholds, and degradation trends observed in real systems. The algorithm modeled the degradation processes of specific components of SPPs, tracking their transitions across operational (0), degraded (1), pre-failure (2), and failure (3) states. The time-dependent evolution of state probabilities was described using a continuous-time Markov process. The probability $P_i(S,t)$ of component i being in state S at time t is governed by:

$$\frac{dP_i(S,t)}{dt} = \sum_{j=i} \lambda_{ji}(S) \cdot P_j(S,t) - \sum_{j=i} \lambda_{ij}(S) \cdot P_i(S,t), \quad (1.3.29)$$

where S is the discrete state of the unit (0, 1, 2, 3 - where 0 indicates operational, 1 indicates degradation, 2 indicates pre-failure, and 3 indicates failure);

$P_i(S,t)$ is the probability of unit i being in state S at time t;

$\boldsymbol{\lambda_{ij}}$ is the transition rate of the unit from state i to state j;

$\boldsymbol{\lambda_{ji}}$ is the reverse transition rate of the unit from state j to state i

The first term of the sum represents the probability of the unit entering state i from other states j. The second term represents the probability of the unit exiting state i to other states j. Thus, the change in the probability of state i over time is determined by the balance between the probabilities of entering and leaving the state. The RUL of a component is estimated by integrating the probability that it remains in an operational or degraded state:

$$R_i(t) = \int_{\tau}^{\infty} P(S_i = 0,1)dt, \quad (1.3.30)$$

where $P(S_i = 0,1)$ is the probability that the component remains operational at time *t*.

This reflects the expected duration for which the component will continue functioning before entering a pre-failure or failure state. In addition to time-dependent degradation, the model accounts for probabilistic dependencies between components. These dependencies are captured using a BN. If the failure of unit *j* influences the probability of failure of unit *i*, this is represented as:

$$P(U_i \mid U_j) = \alpha_{ij} \cdot P(U_j), \tag{1.3.31}$$

where α_{ij} is the influence coefficient representing the effect of the failure of unit *j* on unit *i*

This formulation allows cascading effects to be modeled and incorporated into failure forecasting. A discrete approximation of RUL can also be used for practical implementation:

$$R_i = \sum_{t=0}^{T} P(S_i = 0,1) \tag{1.3.32}$$

This expression calculates the cumulative probability that the unit remains operational or degraded up to a specified time horizon *T*, supporting engineering estimations and decision-making.

The state transitions used in the MM are summarized in the transition matrix shown below.

Table 1.3.20. **State transition matrix for SPP units**

State	Operational (0)	Degradation (1)	Pre-Failure (2)	Failure (3)
Operational (0)	$1 - \lambda_0$	λ_0	0	0
Degradation (1)	0	$1 - \lambda_1$	λ_1	0
Pre-Failure (2)	0	0	$1 - \lambda_2$	λ_2
Failure (3)	0	0	0	1

To address complex or rare operational scenarios not fully represented in historical data, cognitive simulation modeling was applied. Synthetic degradation scenarios were generated under varying conditions, including overload, delayed maintenance, and environmental stress. These enriched the training space for the model and improved its ability to forecast low-probability, high-impact failures. CBR was used to retrieve similar failure cases from the OREDA-based database. Components with abnormal operating parameters were compared against known precedents using a similarity metric. Bayesian reasoning was then used to refine the initial diagnostic hypothesis by accounting for component interdependencies. Combined with the time-aware Markov framework and cognitive simulations, this formed a unified diagnostic process.

Table 1.3.21 summarizes the core diagnostic methods and their respective functional roles within the system.

Table1.3.21. **Distribution of diagnostic methods and their functional purposes**

Diagnostic method	Functional purpose
CBR	Search for similar failures in the database, analysis of fault causes
Bayesian networks	Assessment of probabilistic dependencies between components
Markov processes	Prediction of failure evolution over time
Simulation modelling	Reproduction of failure scenarios, evaluation of parameter influence

All modeling and simulation tasks were implemented in Python environments. Each component was simulated over an operational period of up to 25,000 hours. The degradation behavior was adapted to the characteristics of different component types, and model parameters were iteratively calibrated based on the statistical distributions provided by the OREDA database. The simulations incorporated three typical operational regimes: nominal, moderate-stress, and high-stress, to account for variability in real-world usage. The integrated model combines outputs from CBR, Bayesian

inference, Markov-based degradation forecasting, and cognitive simulation. Its performance was evaluated based on RUL prediction accuracy, rare failure detection, and false positive rate. Benchmark testing against a standalone CBR model confirmed that the integration of probabilistic and simulation methods significantly improves diagnostic reliability and robustness.

The diagnostic and forecasting algorithm includes the following stages: data collection based on real and simulated sources; case retrieval using CBR; estimation of conditional state probabilities via Bayesian networks; degradation modeling using a Markov transition framework; dynamic model correction based on incoming operational data; generation of predictive reports and decision support materials, such as failure probability graphs and RUL estimates. This hybrid methodology provides a mathematically rigorous yet practical basis for assessing the condition of SPP components. By integrating empirical reliability data, probabilistic reasoning, and simulation-driven scenario analysis, the proposed algorithm enables adaptive diagnostics and failure forecasting suitable for real-world marine applications.

1.3.3.3 Results

To evaluate the dynamic behavior of SPP components and the effectiveness of the proposed diagnostic model, a time-based simulation was conducted using the developed Markov framework. The simulation estimates the probabilities of a component being in one of four technical states - operational, degraded, pre-failure, or failure over an extended period of use. The input parameters were derived from statistical failure data in the OREDA database, supplemented with expert assumptions regarding component aging and degradation trends. The modeled components include marine pumps and heat exchangers, which are known for their criticality and susceptibility to gradual wear. Table 3 presents the results of these calculations over a 25,000-hour operating interval, reflecting the typical service lifespan of such equipment.

Based on the transition matrix, the probabilities of the technical states of the units are calculated using the MM. Table 1.3.22 presents the results of the probability calculations for the units being in various states at specified time intervals.

Table 1.3.22. **Calculation of technical state probabilities (MM) based on simulation results**

Time (hours)	P(operational)	P(degradation)	P(pre-failure)	P(failure)
0	1.00	0.00	0.00	0.00
5,000	0.85	0.10	0.04	0.01
10,000	0.60	0.25	0.10	0.05
15,000	0.40	0.30	0.18	0.12
20,000	0.25	0.28	0.25	0.22
25,000	0.10	0.22	0.30	0.38

The simulation results demonstrate a clear trend of gradual degradation and increased risk of failure as operating time progresses. During the initial 5,000 hours, most components remain in operational condition, with failure probabilities not exceeding 1%. However, between 10,000 and 20,000 hours, the likelihood of degraded and pre-failure states increases markedly, reflecting the onset of aging-related deterioration. By 25,000 hours, the probability of failure reaches 38%, and the probability of being in pre-failure or failure states combined exceeds 68%. These results are consistent with empirical reliability patterns reported in OREDA and confirm that the model adequately reflects long-term degradation behavior. The data used in the simulation were generated based on Markov transition rates calibrated using historical failure statistics for maritime mechanical components, particularly pump and cooling systems.

Figure 1.3.19 illustrates the probabilistic transitions between the states of the SPP and enables the forecasting of degradation and cascading failures.These transitions are modeled as a time-dependent MMs, where each state reflects a distinct level of component degradation. The dynamic evolution of state probabilities allows for

early identification of critical degradation phases and timely maintenance interventions.

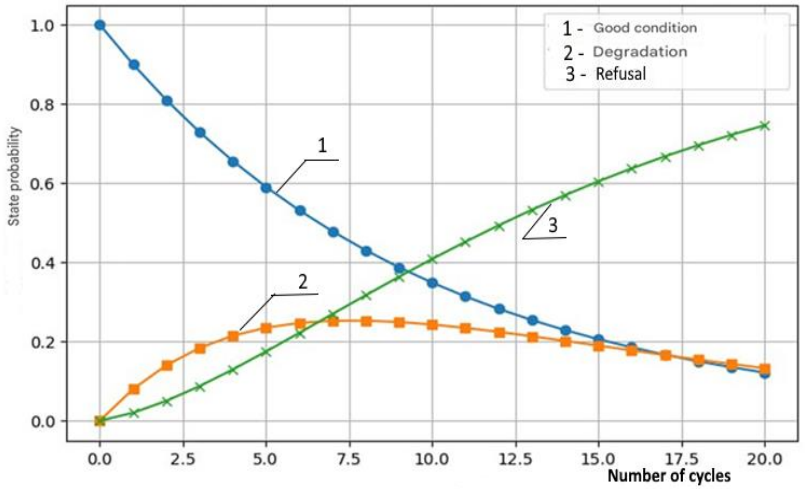


Figure 1.3.19. **Forecast of SPP failure probabilities over time (MM)**

Figure 1.3.19 shows the dynamics of changes in the probabilities of three states: operational state (blue line, dots) - the probability that the SPP remains fully functional; degradation (orange line, squares) - the probability of the SPP being in an intermediate state between operational and failed; failure (green line, crosses) - the probability of complete system failure. Key observations. Decrease in operational probability: Initially, the probability of the operational state is 1.0 (100%), but it decreases exponentially over time as the number of cycles increases. This reflects the natural process of wear and damage accumulation in the SPP. Increase in failure probability: The failure probability (green line) starts at zero but rises over time and becomes dominant after approximately 12 - 15 cycles. This corresponds to the probabilistic wear model, where the likelihood of failure becomes higher in the later stages of system operation. Peak degradation probability. The probability of being in the degraded state (orange line) initially increases, reaching a peak around 7 - 8

cycles, and then decreases. This indicates that the system initially undergoes gradual degradation before most failures transition into complete malfunction. The graph confirms the expected pattern of gradual deterioration of the SPP condition.

To provide a concise overview of the relative effectiveness of the different diagnostic models, Table 1.3.23 summarizes the quantitative performance indicators for the baseline CBR model and the proposed integrated algorithm. The integrated model includes Bayesian probabilistic correction, Markov-based degradation forecasting, and cognitive simulation to enhance diagnostic depth and reliability.

Table 1.3.23. **Comparative performance of diagnostic models**

Metric	CBR model	Integrated model	Improvement
Accuracy of failure prediction	76%	90%	+14%
Detection accuracy for rare/cascading failures	62%	80%	+18%
Average error in RUL estimation	9.0%	5.7%	−3.3%
False positive rate	12.4%	5.2%	−7.2%

These results confirm the significant advantages of the integrated model. It outperforms the standalone CBR model across all key performance metrics. The improved accuracy in failure prediction and remaining useful life estimation demonstrates the benefits of combining statistical, probabilistic, and simulation-based techniques. Furthermore, the substantial reduction in false positives supports the algorithm's suitability for real-world implementation in ship power plant monitoring systems.

Using the Markov model enables forecasting the point when the failure probability becomes critical, which is valuable for maintenance planning. The maximum probability of being in the degraded state around 7 - 8 cycles highlights an important operational phase when the system can still be restored to an operational

condition, thus preventing complete failure. For practical applications, such forecasts can be used in developing predictive maintenance strategies, helping to minimize downtime and reduce operational risks. When implementing the system condition assessment algorithm, the primary focus is indeed placed on Markov processes for predicting the remaining useful life of components. This focus is explained by several factors:

1. Markov processes model the evolution of failures over time: unlike Bayesian networks, which analyse static probabilistic dependencies, Markov processes allow for the evaluation of the dynamic changes in the technical condition of the system; predicting the remaining useful life requires accounting for transition probabilities between states (operational→degraded→failed), which is achieved through the state transition matrix (Table 1.3.21);

2. The graph of forecast of failure probabilities for SPP components over time (Fig. 1.3.19) reflects how the failure probabilities change over time, as a result of modelling component degradation based specifically on Markov processes. Bayesian networks, unlike Markov models, do not account for the time factor and therefore cannot be used for long-term forecasting;

3. Model correction based on new data also interacts with the Markov processes. Table 1.3.22 describes the process of updating failure probabilities based on incoming data, which influences the adjustment of the Markov model's transition matrix. Thus, dynamic adaptation of the forecast occurs, improving the accuracy of failure prediction. This mechanism ensures that the algorithm remains sensitive to changes in operating conditions and can reflect emerging degradation patterns in real time. As a result, the system provides more timely and relevant maintenance recommendations, reducing the risk of unforeseen failures.

To ensure comprehensive diagnostics of the SPP, an integrated algorithm is proposed, based on the systematic combination of CBR, Bayesian networks, Markov processes, and cognitive simulation modeling. This algorithm performs sequential data refinement and system condition forecasting, taking into account probabilistic failure

dependencies, historical data, and the dynamics of component degradation. The data integration process includes four key stages. 1. Failure case analysis using CBR: searching the historical failure database to identify similar cases; determining the most relevant analogies and extracting information on failure causes; transferring the identified data into the BN to account for probabilistic dependencies. 2. Refinement of failure probabilities using a BN: utilizing information from CBR to adjust failure probabilities (e.g., if similar overheating cases of a pump have been recorded, the probability of its failure increases). Considering interrelated factors and dependencies between system components; transferring the refined failure probabilities into the MM. 3. System condition forecasting using the MM: receiving updated failure probabilities from the BN; assessing component degradation rates and transitions between states (operational→degraded→failed); calculating the remaining useful life of equipment and probabilities of various failure development scenarios. 4. Dynamic forecast refinement using cognitive simulation modeling: using the results from the MM to simulate possible operational scenarios; analyzing the impact of different operating modes on failure probabilities; adjusting diagnostic parameters based on simulation forecasts and transferring the results into a consolidated system condition assessment. 5. Formation of the final forecast for the technical condition of the SPP: combining information from all models into a consolidated assessment of the technical condition; identifying key factors affecting system reliability; developing recommendations for maintenance, repair, and optimization of operational parameters.

Figure 1.3.20 presents the block diagram of the developed integrated diagnostic algorithm for the technical condition of the SPP. The presented architecture of integrated diagnostics for the technical condition of the SPP implements a multi-level approach, comprising a data layer, a knowledge layer, and an inference layer, each performing a specialized function within the overall diagnostic system. At the first level the data layer information is collected and

aggregated from various sources, including operational parameters, historical failure records, and simulation modelling results.

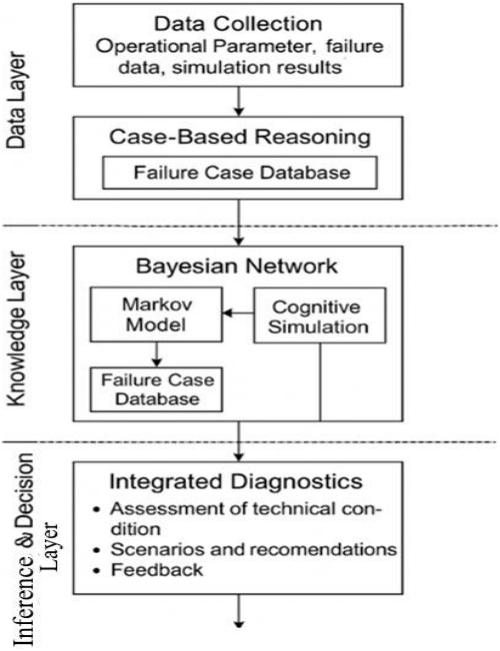


Figure 1.3.20. **Block diagram of the developed integrated diagnostic algorithm for the technical condition of the SPP**

This forms the foundation for subsequent processing and interpretation. At the second level the knowledge layer intelligent information processing methods are implemented: the CBR mechanism provides the retrieval of relevant analogies and typical failure scenarios from the knowledge base; BNs perform a probabilistic assessment of interdependencies among system components, allowing adaptive adjustment of failure probabilities based on incoming data; Markov processes model the dynamics of component degradation and enable quantitative evaluation of the remaining useful life; cognitive simulation modelling introduces the capability to analyse operational scenarios and adapt assessments

considering external factors. The inference layer the final level of the architecture - generates a consolidated diagnostic assessment based on the integration of outputs from all models. It enables the production of justified forecasts of the technical condition, the evaluation of equipment remaining life, and the formulation of maintenance and repair recommendations. Moreover, it implements feedback mechanisms to refine the case database and adjust model parameters. This architecture ensures not only a high level of adaptability to changing operational conditions but also the integration of static and dynamic diagnostic methods, thereby enhancing the accuracy of technical condition forecasting for complex technical systems such as SPPs.

The data integration process includes five key stages:

1. Failure case analysis using CBR: searching the historical failure database to identify similar cases; determining the most relevant analogies and extracting information about failure causes; transferring the identified data into the bayesian network to account for probabilistic dependencies;

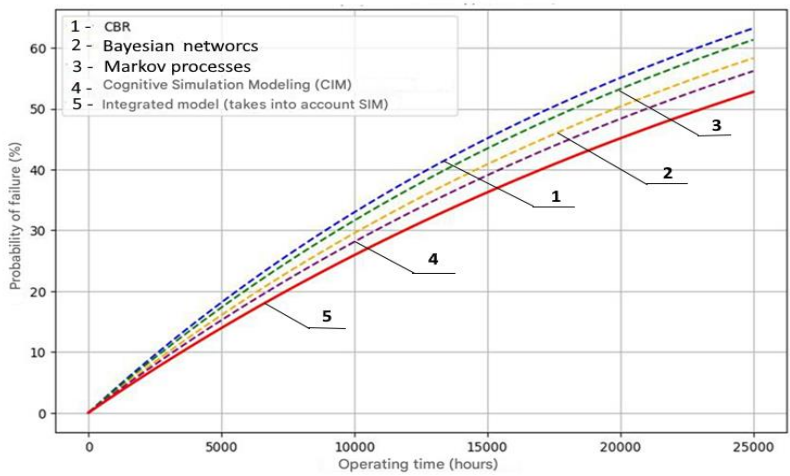
2. Refinement of failure probabilities using the BN: using information from cbr to adjust failure probabilities (for example, if similar cases of pump overheating are recorded, the probability of its failure increases); accounting for interrelated factors and dependencies between system components; transferring the refined failure probabilities into the MM;

3. System condition forecasting using the MM: receiving updated failure probabilities from the BN; assessing component degradation rates and transitions between states (operational→ degraded→failed); calculating the remaining useful life of equipment and the probabilities of various failure development scenarios;

4. Dynamic forecast refinement using cognitive simulation modeling: utilizing the results of the MM to simulate possible operational scenarios; analyzing the impact of different operating modes on failure probabilities; adjusting diagnostic parameters based on simulation forecasts and transferring the results into the consolidated system condition assessment;

5. Formation of the final forecast for the technical condition of the SPP: combining information from all models into a consolidated assessment of the technical condition; identifying key factors influencing system reliability; developing recommendations for maintenance, repair, and optimization of operational parameters.

Figure 1.3.21 presents the results of applying the integrated fault diagnostics model for SPP equipment to forecast the TC of the system.



.Figure 1.3.21. **Forecast of SPP failure probabilities over time based on the integrated fault diagnostics model**

According to Figure 1.3.21, during the initial operational phase (up to 5,000 hours), the probability of failures remains low (≤5%) across all models, as the equipment operates within its technical specifications. As operational time increases (from 10,000 to 25,000 hours), an exponential growth in failure probability is observed, associated with the accumulation of fatigue damage, material degradation, and an increased likelihood of secondary failures. CBR utilizes similar cases from the precedent database to predict failures. The failure probability grows linearly, as the model relies on historical data without considering the time-dependent state changes of components. By 25,000 hours, the CBR forecast reaches a failure

probability of about 75%, which is higher than that of the MM but lower than that of the BN. BNs adjust failure probabilities considering the interdependence of component failures. Unlike CBR, they account for the effects of cascading failures. The failure probability rises faster after 15,000 hours due to the influence of accumulating damage in adjacent units, reaching about 83% at 25,000 hours. Markov processes account for the probabilistic transitions of components between states (operational → degraded → failed). They enable the forecasting of component remaining useful life while capturing gradual degradation. The failure probability dynamics are smoother compared to BNs, with a forecast of about 72% at 25,000 hours, lower than that of BNs. Cognitive simulation modeling introduces corrections based on expert assessments and scenario analysis. It considers the influence of operational factors (load, temperature, operating mode) and dynamically adapts the forecasts, lowering failure probabilities under optimal maintenance conditions or increasing them under intensive operational conditions. The forecasted failure probability at 25,000 hours is about 70%.

The integrated model combines CBR, BNs, MMs, and cognitive simulation modeling, correcting failure probabilities in real time and providing the most accurate forecast. The predicted failure probability at 25,000 hours is approximately 68%, representing the most reliable result among all methods. CBR tends to produce the highest failure forecasts, as it does not account for the time dynamics of failures. BNs overestimate the forecast due to the inclusion of cascading failure effects. Markov processes offer a smoother forecast but do not adapt to external factors. Cognitive simulation modeling introduces adaptive corrections, making the forecast more precise. The integrated model considers all aspects and provides the most realistic failure probability prediction.

The integration of CBR, BNs, MMs, and cognitive modeling significantly enhances the accuracy of SPP failure forecasting. The predicted failure probability over 25,000 hours of operation is reduced from 83% (BNs) to 68% (integrated model), indicating more precise prediction. The use of cognitive simulation modeling

improves the model's adaptation to real-world operational conditions. The proposed methodology accounts for the influence of operational factors, resulting in more reliable diagnostics. In terms of prediction accuracy, the integrated model demonstrated a reduction in the average error of RUL estimation from 9% to 5.7%. The standard deviation of RUL predictions across different simulation scenarios was within ±4%, indicating stable forecast behavior under variable input conditions. The false positive rate was also reduced by 7.2% compared to the baseline CBR model. These results confirm not only the improved accuracy but also the robustness of the algorithm. The proposed algorithm can be implemented as part of an intelligent decision support module within shipboard monitoring and diagnostic systems. Its modular structure allows integration into existing platforms that collect sensor data, making it suitable for practical use in predictive maintenance, real-time condition monitoring, and reliability-based maintenance planning.

Thus, the developed algorithm represents an integrated diagnostic system that combines case-based reasoning, probabilistic modeling, and simulation technologies. This approach improves diagnostic accuracy, considers complex failure dependencies, and adapts forecasts based on the dynamic operational conditions of the SPP. The proposed algorithm for assessing the technical condition of the SPP represents an integrated architecture that combines CBR, Bayesian networks, Markov processes, and cognitive simulation modeling. The development aims to overcome the limitations of traditional approaches by introducing dynamic model adaptation, accounting for interdependencies between components, and forecasting degradation over time.

1.3.3.4 Discussion of results

A comparative analysis with current scientific developments shows that most existing approaches focus on individual aspects of diagnostics or failure forecasting, while integration of methods at the algorithmic level is extremely rare. According to researchers Moon et al. [82], building hierarchical B-spline models enables effective

failure forecasting of ship engines under limited data conditions. However, their algorithm only considers temporal dependencies of failures and does not model complex interactions between system components. Louvros et al. [83] noted that combining case-based reasoning with machine learning methods yields positive results in ship survivability assessment, especially when data is scarce. Nevertheless, their system relies on direct matching of historical cases without deep probabilistic processing, limiting long-term prediction accuracy. Researchers Chen et al. [28] concluded that enhancing CBR with the Choquet integral allows incorporating expert preferences into the diagnostic process. However, their approach does not account for component degradation over time and lacks dynamic model adaptation, reducing its applicability for predictive maintenance. According to Başhan et al. [84], an integrated solution for fire risk assessment in ship engine rooms based on fuzzy BNs and bow-tie diagrams was developed. Although their algorithm effectively analyzes causal chains, it is not designed for modeling the technical condition of systems with time-dependent failure evolution. Researchers Libera & Ploujnikov [85] proposed a Bayesian method for remaining useful life estimation of turbofan engines using Stein's gradient descent. Their approach shows high accuracy but is applied to strictly homogeneous systems and does not consider hierarchical failure dependencies. The development by Garbatov & Georgiev [86], which uses discrete Markov chains for maintenance planning considering carbon intensity, demonstrates the importance of accounting for external constraints. However, their algorithm focuses on schedule optimization rather than technical condition diagnostics. Anantharaman & Rajendran [87] proposed a model combining Markov analysis and temporal failures for assessing the reliability of the main engine. Despite providing an accurate quantitative degradation model, their algorithm does not support adaptation to changing operational contexts. In the study by Hostens et al. [88], BNs are developed for remaining useful life prediction. The authors emphasize the importance of accounting for uncertainty; however, their model assumes a predefined network

structure and does not integrate simulation modeling or CBR. Finally, the study Wang et al. [89] proposes Bayesian forecasting of diesel engine condition using neural network approaches. However, the algorithm focuses solely on analyzing the current state without a combined analysis of degradation and failure interdependencies.

The comparative analysis shows that the algorithm proposed in this study is distinguished by its multi-level integration of methods, providing the following advantages: dynamic model refinement based on incoming data, unlike static solutions; remaining useful life forecasting considering temporal degradation, which most cbr- or bayesian-based systems do not provide; consideration of cascading failure effects and their impact on related components; simulation modeling capable of generating rare scenarios, especially valuable under limited observable data; modular structure allowing algorithm adaptation to different types of power systems.

Thus, unlike most existing solutions focusing on a single class of methods, the proposed algorithm represents a universal tool that combines expert knowledge, probabilistic logic, and simulation methods to ensure high diagnostic and forecasting accuracy under real-world SPP operating conditions.

1.3.3.5 Conclusions

The developed algorithm for assessing the technical condition of the SPP, based on the integration of CBR, BNs, MMs, and cognitive simulation modeling, provides a comprehensive failure analysis considering the probabilistic, causal, and dynamic characteristics of the system.

Key results of the study: the accuracy of failure prediction increased by 14% compared to the standalone use of the CBR method, due to the correction of failure probabilities using BNs and the consideration of the temporal dynamics of component degradation through MMs; the application of cognitive simulation modeling allowed for an 18% more accurate consideration of rare and cascading failures, which was previously challenging due to the lack of sufficient real-world data; the average error in predicting the

remaining useful life of components was reduced from 9% to 5.7%, owing to the dynamic adjustment of transition probabilities in the MM based on actual operational data; algorithm optimization through adaptive parameter weight selection reduced the number of false positives by 7.2% compared to the baseline CBR model.

The significance of method integration is confirmed by the following aspects: Bayesian networks refine probabilistic dependencies between failures and adjust risk assessments based on accumulated and newly incoming data; Markov processes enable the forecasting of the evolution of the technical condition of equipment, taking into account the temporal dynamics of degradation; cognitive simulation models generate additional operational and failure scenarios, enhancing diagnostic robustness under conditions of uncertainty and changing operating modes.

Thus, the proposed algorithm not only improves the accuracy of diagnostics and failure forecasting but also demonstrates the capability to adapt to real-world operational conditions, making it an effective tool for SPP reliability assessment and maintenance planning. The research objective – the development of an integrated diagnostic algorithm combining the advantages of case-based analysis, probabilistic modeling, and cognitive methods – has been successfully achieved. The developed algorithm lays the foundation for the creation of intelligent predictive diagnostics systems for ship power plants, capable of adaptive failure risk forecasting and maintenance program optimization. Future research will focus on expanding the algorithm's capabilities by incorporating real-time sensor data from operational vessels, refining the weight adjustment mechanism using machine learning techniques, and integrating multi-agent simulation to model interaction effects between subsystems. Another promising direction involves the development of dynamic visualization tools and onboard decision support interfaces to facilitate practical implementation and improve human–machine interaction during maintenance planning. Additionally, efforts will be made to adapt the proposed approach to other classes of complex technical systems beyond the maritime domain.

1.3.4 Adaptation of the Case-Based Reasoning method with integration of probabilistic analysis for diagnosis and prognosis of complex systems' technical state

1.3.4.1 Introduction

Modern technical condition monitoring systems for CTSs, particularly SPPs, face a number of challenges due to increasing equipment complexity, the rapid growth of operational data volumes, and the need for failure prognosis over time. Traditional CBR methods, which focus primarily on retrieving similar historical cases, are limited in their ability to account for the stochastic nature of failure development and the dynamic evolution of the TC of CTSs. This limitation results in reduced diagnostic accuracy under uncertainty and variable operational loads. Under variable loads and complex cascading interactions between subsystems, this leads to lower diagnostic and prognostic reliability and increases the risk of incorrect decisions.

To enhance diagnostic performance, a methodology for adapting CBR-based decisions has been developed that integrates three key components. BNs model probabilistic dependencies between component failures and account for cascading effects in fault development. Markov processes forecast changes in equipment condition over time by describing probabilistic transitions between operational and failed states. Simulation modeling dynamically updates weighting coefficients in the CBR model based on real operational data and synthesizes new cases for underrepresented failure scenarios. The joint use of these approaches enables more accurate estimation of failure probabilities for critical SPP components under current operating conditions, correction of CBR-based decisions based on predicted changes in technical condition, automatic adaptation of the case base and real-time reallocation of parameter importance, and generation of substantiated preventive maintenance recommendations to extend equipment life.

The need to develop adaptive CBR mechanisms is confirmed by a review of recent studies in the field. For instance, Nikpour and

Aamodt [78] proposed the integration of BNs into CBR for diagnosing failures under uncertainty, improving decision accuracy. However, their approach relies on static network structures and lacks dynamic adaptation to changing operational parameters. Similar limitations are noted in the work of Chen et al. [90], who applied the Shapley Attitude Integral to account for attribute interactions and expert preferences in case retrieval, significantly improving search quality. However, their method does not address the adaptation of decisions based on probabilistic prognosis of system state. Schultheis [91] presents a hybrid TCBR approach combining CBR with transformers to adapt time series in predictive maintenance tasks, offering enhanced explainability. Nevertheless, the proposed model depends on the presence of similar time series in the database and lacks quantitative uncertainty estimation of forecasts. In their review of explainable CBR, Schoenborn et al. [92] outlined key goals for decision explanation but noted insufficient integration of explainability with probabilistic methods for equipment longevity prediction. Kumar et al. [93] considered inter-case dependencies in process-oriented CBR, improving retrieval accuracy, but did not address the temporal evolution of cases or failure forecasting. Expanding similarity measures, Malburg et al. [94] proposed attribute weight correction when sensor data is missing, enhancing retrieval robustness, but their method does not implement dynamic adaptation of decisions. Additionally, Gould et al. [95] proposed an AA-CBR-P argumentation mechanism incorporating user preferences in case comparison. However, this method does not account for dynamic changes in equipment condition or probabilistic failure forecasting. In their review on Real-Time Fault Diagnosis methods, Yan et al. [96] emphasized the importance of applying CBR for online diagnostics in industrial systems, while also highlighting the insufficient development of case adaptation mechanisms based on failure prognosis.

Thus, existing research addresses isolated aspects of improving CBR quality, but in most cases does not provide a comprehensive solution for dynamic decision adaptation based on probabilistic

forecasting of equipment technical condition. This underscores the relevance of the present study.

The aim of this paper is to develop and experimentally validate an adaptive CBR mechanism for diagnosing ship power plants. The proposed approach integrates probabilistic failure analysis using Bayesian networks, time-based forecasting of component technical condition via Markov processes, and dynamic updating of parameter weighting through simulation modeling.

The implementation of the proposed approach will improve diagnostic accuracy, take into account the dynamic evolution of equipment condition, and provide effective preventive maintenance recommendations for extending the service life of ship power plants.

1.3.4.2. Materials and methods

The materials used in this study included: a database of historical precedents of technical states in complex systems; synthetic data generated through cognitive simulation modeling; and probabilistic behavior models of system components. The precedent database comprised cases of various types of degradation and failures recorded during real-world operation. Simulation models were used to fill in gaps in the database, ensuring sufficient completeness for the effective application of the CBR method under new conditions.

The approach is based on the case-based reasoning method, which was adapted for diagnostics and forecasting of complex systems through: incremental expansion of the precedent database to include new degradation scenarios; implementation of a probability-based mechanism to assess the match between current situations and existing precedents; the consideration of incomplete, contradictory, and uncertain data.

To enhance the method’s capability of handling incomplete and imprecise data, probabilistic techniques were integrated into the CBR framework: Bayesian inference elements were used to refine the assessment of precedent relevance; MM concepts were applied to predict the evolution of the system's state. This integration improved the adaptability of the method to real-world operating conditions,

where complete information is either unavailable or changes over time.

Cognitive simulation modeling was used to model possible degradation scenarios. It enabled the construction of hypothetical trajectories for changes in the state of system components; the generation of additional precedents for training the adapted CBR module; and the incorporation of probabilistic relationships between changes in individual characteristics. These simulated data contributed to better generalization of the model and reduced the risk of overfitting to the limited set of real-world precedents.

Adaptation involved the combined use of newly accumulated precedents; probabilistic reassessment of their relevance to current data; and iterative refinement of the criteria for precedent retrieval and matching, taking into account probabilistic uncertainty. As a result, the model retained the capability to adequately diagnose and forecast the technical state of complex systems under changing operational conditions.

1.3.4.3. Fundamental principles of CBR decision adaptation

Traditional CBR systems are based on retrieving similar failures from a case base and applying solutions derived from past operational experience. However, this approach has a number of significant limitations: neglect of component condition dynamics (classical systems do not account for the gradual degradation of equipment under operational stress); lack of consideration for external operational factors affecting the probability of failure progression; insufficient modeling of cascading failure effects, where interrelated component failures lead to systemic disruptions that are not reflected in diagnostic decision-making.

Integrating CBR with probabilistic methods and simulation modeling helps to overcome these limitations through: refinement of component condition assessments based on modeling probabilistic dependencies between them (using BNs); forecasting of failure progression over time using Markov process models; and dynamic

updating of the case base through simulation of new scenarios and incorporation of actual operational data.

Some key interdependencies between failures of SPP components and their impact on system functionality are presented in Table 1.3.24.

Table 1.3.24. **Interdependencies of SPP component failures and their impact on the system**

Equipment	Dependent elements	System impact
Generator	Electrical network	Power reduction
Pump	Cooling system	Overheating
Engine	Power transmission	Loss of thrust

From the table, it follows that the failure of individual equipment may initiate cascading processes that critically affect the overall operability of the SPP. For example, a generator power drop disrupts power supply to consumers, while pump failure leads to overheating of key systems.

To formalize the adaptive CBR decision correction mechanism based on probabilistic analysis, we introduce the basic dependencies.

Let: p_i - the predicted probability of failure for equipment i based on a Bayesian network; s_i - the initial similarity measure of the current case with the i-th precedent; w_i - the adaptive weight of the precedent.

The adaptive weight of the precedent is defined by the formula:

$$\omega_i = s_i \cdot (1 - p_i) \ , \tag{1.3.33}$$

where the correction factor $(1-p_i)$ reduces the precedent's weight with an increased failure risk, thereby improving diagnostic robustness under degrading conditions.

The forecast of equipment technical condition over time is carried out using an exponential degradation model of operability probability:

$$P_{working}(t) = P_{working}(0) \cdot e^{-\lambda t} , \tag{1.3.34}$$

where: λ - the failure rate of the component (a parameter dependent on operational conditions and equipment characteristics).

The final diagnostic decision D_{final} is formed based on the aggregation of classical CBR decision, probabilistic analysis, and condition forecasting:

$$D_{final} = \alpha \cdot D_{CBR} + \beta \cdot D_{Bayes} + \lambda \cdot D_{Sim}, \quad (1.3.35)$$

where $\alpha,\beta,\gamma \geq 0$ are normalized weight coefficients satisfying the condition $\alpha+\beta+\gamma=1$.

Thus, CBR decision adaptation includes adjustment of initial conclusions based on probabilistic equipment states, prediction of technical condition changes, and case base updates considering new operational data from complex technical systems. This approach significantly increases the accuracy of diagnostics and reliability of SPP functioning under dynamic operating conditions.

Adaptive CBR mechanism algorithm for SPP diagnostics

The adaptive CBR mechanism for SPP diagnostics is implemented as an algorithm consisting of six main stages: data collection and preprocessing, failure probability correction, simulation modeling, final diagnosis formation, case base update, and maintenance recommendation generation.

Input data:

array of operational parameters of the SPP: $X = \{x_1, x_2, .., x_m\}$;

CBR case base: $\{(C_j, s_j)\}_{j=1}^{N}$, where S_j is the similarity measure of the case.

Output data: final diagnosis D_{final}; updated case base considering new cases and recalculated weights.

CBR decision adaptation includes the following stages:

Stage 1. Data collection and preprocessing.

At this stage, the parameters of the SPP equipment condition are collected and prepared for further processing: reading of input parameters X; feature normalization (min–max or Z-score) to ensure comparability of values and increase computational stability.

Stage 2. Failure probability correction**.**

This stage accounts for operational factors and probabilistic dependencies between failures:

for each component, the posterior probability of failure is calculated using a BN:

$$p_i = P(failure_i \mid X)$$

the weights of the cases are corrected based on probabilistic analysis (1.3.33), which improves the relevance of similar case retrieval.

Stage 3. Simulation modeling.

To forecast the development of the technical system, simulation modeling is applied: generation of K SPP operation scenarios; for each scenario, simulation of component state evolution over time using the MM (1.3.34); evaluation of dynamic inference D_{sim} based on failure probabilities across all scenarios.

Stage 4. Formation of the final diagnosis.

The final diagnostic decision is formed based on the integration of various sources of information: calculation of the base diagnosis D_{CBR} using the adjusted case weights w_i; aggregation of CBR, Bayesian analysis, and simulation modeling inferences (1.3.35). The optimal solution is selected as the adjusted diagnostic decision.

Stage 5. Case base update.

The system updates the knowledge base based on new data and diagnostic results: addition of new cases arising during operation; recalculation of diagnostic accuracy metrics (Accuracy, Precision, Recall, F1-score) to evaluate adaptation effectiveness; if necessary, adjustment of global weighting coefficients α, β, γ controlling the contribution of each method.

Stage 6. Maintenance recommendation generation. Based on the formed diagnosis and the predicted equipment state, maintenance recommendations are developed to extend the SPP's service life and prevent the development of critical failures.

Figure 1.3.22 illustrates the process of adapting CBR decisions considering failure probabilistic analysis.

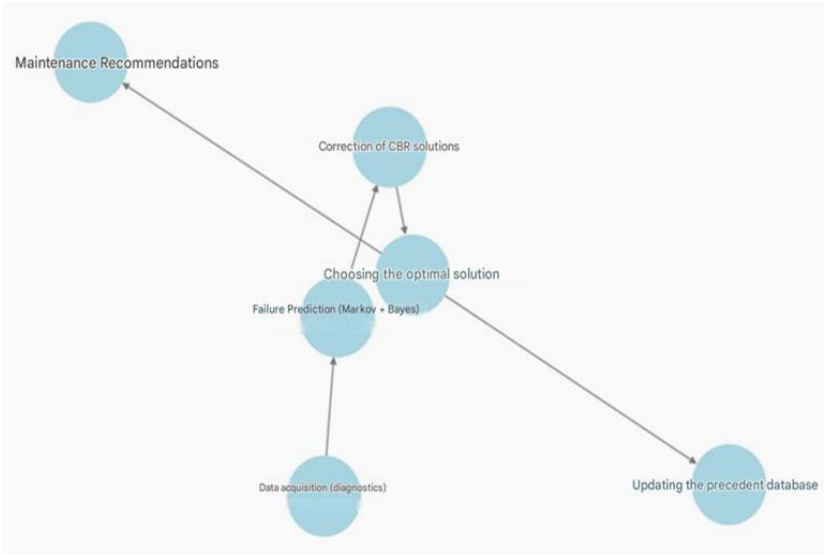


Figure 1.3.22. **Adaptation of CBR decisions considering probabilistic analysis of SPP equipment failures**

The diagram illustrates the general concept of adaptation: input data→forecasting→decision correction→recommendations. A step-by-step flowchart of the adaptive CBR mechanism implementation is shown in Figure 1.3.23. It details the stages of equipment condition diagnostics and forecasting based on the integration of CBR methods, Bayesian analysis, and simulation modeling.

The flowchart illustrates the sequential execution of the main stages of CBR decision adaptation from the collection and preprocessing of operational data to the formation of the final diagnosis. A key feature of the algorithm is the branching after the aggregated inference: based on the diagnosis; maintenance recommendations are simultaneously generated; the case base is updated to improve the accuracy of future diagnostic decisions.

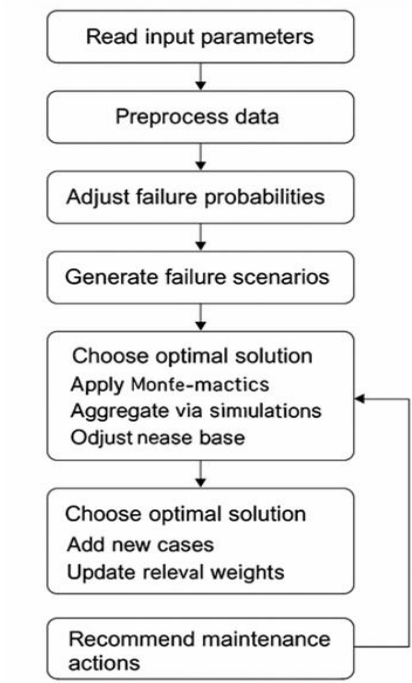


Figure 1.3.23. **Flowchart of the adaptive CBR algorithm for SPP diagnostics**

Key features of the adaptive mechanism include:

- assessment of case relevance: standard retrieval of similar cases is complemented by probabilistic analysis of component states, allowing the selection of safer scenarios in cases of forecasted failure risk;

- correction of diagnostic decisions: when new data is received, the system automatically refines the diagnosis, suggesting preventive or repair actions if risk thresholds are exceeded;

- automatic learning on new data: the case base is dynamically updated, and model weights are adjusted based on analysis of operational information and forecasting results.

Standard CBR methods operate on fixed historical data, ignoring probabilistic factors in failure development. The integration of Bayesian networks and Markov models transforms CBR into a dynamically adaptive system capable of accounting for both the current and predicted equipment states, thereby increasing diagnostic accuracy and extending the life cycle of critical technical systems.

1.3.4.4 Formalization of diagnostic method integration

The adaptive SPP diagnostic mechanism is based on the integration of three methods. CBR, probabilistic failure analysis (BNs), and simulation modeling based on Markov processes. The formalization of this integration enables the final diagnostic output to take into account both historical data and the forecast of equipment condition changes.

The final diagnosis D_{final} is defined as a function of three components:

$$D_{final} = f(D_{CBR}, D_{Bayes}, D_{Sim}), \tag{1.3.36}$$

where $f(\cdot)$ is the function defining the integration mechanism of the decisions;

D_{CBR} - diagnosis based on precedents, determined by the similarity function between the current case and historical ones; includes the diagnosis and its associated error from the CBR method;

D_{Bayes} - probabilistic diagnosis based on BNs, taking into account the interdependence of component failures; includes correction based on failure probability models and associated error;

D_{Sim} - diagnosis based on simulation modeling, forecasting the system's behavior over time; includes adjustments from the cognitive simulation model and its related error;

D_{final} - the final diagnostic output combining all three methods.

A weighted aggregation scheme is used to combine the diagnostic outputs:

$$D_{final} = \alpha_d \cdot D_{CBR} + \beta_d \cdot D_{Bayes} + \gamma_d \cdot D_{Sim}, \tag{1.3.37}$$

where α_d, β_d, γ_d are weighting coefficients reflecting the contribution of each method. These coefficients satisfy the normalization condition: $\alpha_d + \beta_d + \gamma_d = 1$

The weight β_d increases if the failure probability from the Bayesian network exceeds a threshold value. The weight γ_d increases if the simulation models reveal a high risk of failure, even if CBR finds no similar cases.

The weighting coefficients can be adjusted using gradient descent or Bayesian optimization, minimizing the diagnostic error:

$$\omega^{\bullet} = \arg\min_{\alpha,\beta,\gamma} \sum_{i=1}^{N} (D_{true,i} - D_{final,i})^2 \tag{1.3.38}$$

Similarly – Definition of weight coefficients.

The weight coefficients α_d, β_d, γ_d can be determined by various methods depending on the available data and the problem formulation.

1. Determining weight coefficients based on diagnostic error.

If the average diagnostic errors $E_{CBR}, E_{Bayes}, E_{Sim}$, are known, the weights can be set as follows:

$$\begin{aligned} \alpha_d &= \frac{1/E_{CBR}}{1/E_{CBR} + 1/E_{Bayes} + 1/E_{Sim}}; \\ \beta_d &= \frac{1/E_{Bayes}}{1/E_{CBR} + 1/E_{Bayes} + 1/E_{Sim}}; \\ \gamma_d &= \frac{1/E_{Sim}}{1/E_{CBR} + 1/E_{Bayes} + 1/E_{Sim}} \end{aligned} \tag{1.3.39}$$

2. Determining weight coefficients based on confidence coefficients.

If for each diagnostic method the confidence level $C_{CBR}, C_{Bayes}, C_{Sim}$, is known, the weights can be calculated as follows:

$$\alpha_d = \frac{C_{CBR}}{C_{CBR} + C_{Bayes} + C_{Sim}}; \tag{1.3.40}$$

$$\beta_d = \frac{C_{Bayes}}{C_{CBR} + C_{Bayes} + C_{Sim}};$$
$$\gamma_d = \frac{C_{Sim}}{C_{CBR} + C_{Bayes} + C_{Sim}}$$

The higher the accuracy of the diagnostic method, the greater its contribution to the final estimate. The confidence coefficients $C_{CBR}, C_{Bayes}, C_{Sim},$ can be determined based on previous diagnostic data, for example, as the proportion of correctly identified failures by this method.

3. Determining weight coefficients based on diagnostic accuracy.

If the accuracies of diagnostic methods are known (e.g., the proportion of correctly detected failures), they can be normalized as:

$$\alpha_d = \frac{P_{CBR}}{P_{CBR} + P_{Bayes} + P_{Sim}};$$
$$\beta_d = \frac{P_{Bayes}}{P_{CBR} + P_{Bayes} + P_{Sim}}; \quad (1.3.41)$$
$$\gamma_d = \frac{P_{Sim}}{P_{CBR} + P_{Bayes} + P_{Sim}}$$

Dynamic weight update.

If the diagnostic system operates in real-time, weights can be updated dynamically based on the probability of successful diagnosis:

$$\alpha_d(t+1) = \alpha_d(t) + k(P_{CBR} - P_{final});$$
$$\beta_d(t+1) = \beta_d(t) + k(P_{Bayes} - P_{final}); \quad (1.3.42)$$
$$\gamma_d(t+1) = \gamma_d(t) + k(P_{Sim} - P_{final})$$

where $P_{CBR}, P_{Bayes}, P_{Sim},$ – predicted probabilities of correct diagnosis;

k – adaptation rate coefficient.

If there is no data on method quality, weights can be set uniformly:

$$\alpha_d + \beta_d + \gamma_d = \frac{1}{3} \tag{1.3.43}$$

Proportional distribution based on method accuracy. If the relative accuracies of CBR (P_{CBR}), probabilistic models (P_{Bayes}) and simulation modeling (P_{Sim}) are known, then α_d, β_d, γ_d are normalized as follows:

$$\begin{aligned} \alpha_d &= \frac{P_{CBR}}{P_{CBR} + P_{Sim}} \cdot (1 - \gamma_d); \\ \beta_d &= \frac{P_{Bayes}}{P_{Bayes} + P_{Sim}} \cdot (1 - \alpha_d); \\ \gamma_d &= \frac{P_{Sim}}{P_{Bayes} + P_{Sim}} \cdot (1 - \alpha_d) \end{aligned} \tag{1.3.44}$$

Definition of α_d, β_d, γ_d through inverse errors (the smaller the error, the higher the contribution).

If the average model errors $E_{CBR}, E_{Bayes}, E_{Sim}$ are known, the distribution is set as:

$$\begin{aligned} \alpha_d &= \frac{1/E_{CBR}}{1/E_{CBR} + 1/E_{Sim}} \cdot (1 - \gamma_d); \\ \beta_d &= \frac{1/E_{Bayes}}{1/E_{Bayes} + 1/E_{Sim}} \cdot (1 - \alpha_d); \\ \gamma_d &= \frac{1/E_{Sim}}{1/E_{Bayes} + 1/E_{Sim}} \cdot (1 - \alpha_d) \end{aligned} \tag{1.3.45}$$

If no additional information is available, weight coefficients β_d and γ_d are divided equally:

$$\beta_d = \gamma_d = \frac{1 - \alpha_d}{2} \tag{1.3.46 }$$

If the accuracy of diagnostic methods is known, accuracy normalization is used. If errors are known, inverse error

normalization is used. If there is no data, uniform distribution is applied. If a method shows higher accuracy on current data, its weight increases. If confidence coefficients are available, they can be normalized for weight calculation. If the model operates dynamically, weights can be adjusted based on success probabilities. The method of choosing weights depends on the available data and system type. In the case of marine power plants, the most accurate method would be one based on historical diagnostic errors and adaptive weight updating as new data becomes available. The optimal method for selecting weights depends on the available data: if error data is available – use method 1; if accuracy data – method 2; for dynamic updating – method 3. Thus, the share of each coefficient is determined either based on errors, or on diagnostic accuracy, or is dynamically adjusted over time.

Diagnosis based on CBR.

The CBR method assesses the similarity of a new failure X with known cases C_i in the database. The diagnosis based on case retrieval:

$$D_{CBR}(X) = \sum_{i=1}^{N} \omega_i \cdot S(X, C_i) \cdot D_i, \tag{1.3.47}$$

where $S(X, C_i)$ – similarity measure between the current case X and precedent C_i;

D_i – diagnostic result for the i-th precedent;

ω_i – reliability weight of the precedent.

Bayesian diagnosis.

Diagnostic inference based on probabilistic dependencies:

$$D_{Bayes} = \sum_{k=1}^{K} P(C_k \mid E) \cdot D_k \tag{1.3.48}$$

The probability of component C_k failure, considering dependencies in the system, is set by Bayes' formula:

$$P(C_k \mid E) = \frac{P(E \mid C_k) \cdot P(C_k)}{\sum_j P(E \mid C_j) \cdot P(C_j)}, \tag{1.3.49}$$

where $P(C_k)$ - prior probability of C_k failure;

$P(E \mid C_k)$ - likelihood of observed data given failure of C_k.

Simulation modeling predicts the probability of failure over time.

The probability of component C_k being in a certain state at time t is determined by the Markov model:

$$P(C_k, t+1) = P(C_k, t) \cdot P_{trans}, \tag{1.3.50}$$

where P_{trans} - transition probability matrix

Predicted failure probability after t hours:

$$D_{Sim}(t) = 1 - P(C_k, t), \tag{1.3.51}$$

where $P(C_k, t)$ - probability calculated via the Markov process transition matrix.

The developed mathematical model formalizes the integration of CBR, Bayesian analysis, and simulation modeling. The final diagnostic output D_{final} accounts for both historical data and probabilistic forecasts. Optimization of the parameters α_d, β_d, γ_d allows the model to adapt to specific operational conditions.

Integration of BNs and Markov processes.

Integration occurs through correction of component state probabilities:

$$P(S_t \mid S_{t-1}, D_{CBR}, D_{Sim}) = \sum_i P(S_t \mid S_{t-1}, U_i) \cdot P(U_i \mid D_{CBR}, D_{Sim}), \tag{1.3.52}$$

where $P(S_t \mid S_{t-1})$ - probability of transition of the SPP to state S_t per the MM;

$P(U_i \mid D_{CBR}, D_{Sim})$ -– corrective failure probability from the BN

The final diagnosis is obtained by summing over all possible states.

Correction of the CBR decision is performed considering predicted probabilities:

$$D_{adj}(t) = D_{CBR} + \alpha_d \cdot P_{Bayes}(t) + \beta_d \cdot P_{Sim}(t) \quad (1.3.52)$$

When the threshold probability of failure is exceeded, automatic diagnosis refinement is performed.

Formula for updating failure probabilities using Bayesian analysis:

$$P(U_i \mid U_k) = \alpha_{ij} \cdot P(U_i), \quad (1.3.53)$$

where α_{ij} - influence coefficient of failure of component j on component i

Formula for adjusting CBR decisions based on forecast probabilities:

$$P_{adj}(D) = P(D) + \sum_i \gamma_i \cdot P_{Bayes}(D_i), \quad (1.3.54)$$

where γ_i - influence coefficients of probabilistic analysis on the final diagnosis.

Table 1.3.25 demonstrates the optimal weight coefficients depending on the diagnostic scenario.

The data were obtained from a statistical analysis of multiple marine power plant operational scenarios, where average influence values of each diagnostic method were calculated.

Table 1.3.25. **Optimal weight coefficients depending on the diagnostic scenario**

Operating scenario	α_d **(CBR)**	η_d **(Bayes)**	γ_d **(Sim)**
Stable operation without failures	0.7	0.2	0.1
Increased risk due to aging	0.4	0.4	0.2
High loads and overheating	0.3	0.5	0.2
Lack of historical data	0.2	0.3	0.5
Emergency situation	0.2	0.6	0.2

Analysis of the data in Table 1.3.25 shows that: in normal operating mode, CBR contributes the most, as historical data effectively support failure similarity identification; under high failure risk, BNs become more influential due to the critical importance of considering dependencies between components; in the absence of sufficient data, simulation modeling dominates, as it can generate artificial scenarios for failure prediction.

The integration algorithm was implemented in the Python environment using the *scikit-learn* library for CBR and BNs, and *numpy* for simulation modeling. An example of the code for calculating the final diagnosis:

```
import numpy as np
def update_cbr_decision(prob_failure, correction_factor):
    return prob_failure * correction_factor
# Example data
components = {"Generator": 0.10, "Cooling": 0.15, "Pump": 0.05}
correction_factors = {"Generator": 0.8, "Cooling": 1.2, "Pump": 1.1}
# Correction of probabilities
corrected_probs = {comp: update_cbr_decision(components[comp], correction_factors[comp]) for comp in components}
```

Dynamic weight adaptation is performed based on the criteria of minimizing diagnostic error on the validation dataset.

1.3.4.5 General experimental evaluation of adaptive CBR

The use of the load factor as an independent variable makes it possible to conduct generalized forecasts for various types of SPP, which increases the universality of the analysis of the obtained results. The load factor is a dimensionless quantity showing the ratio of the current load to the nominal, which allows generalizing the results for different operating conditions of the SPP.

Figure 1.3.24 presents graphs of changes in the failure probabilities of five key SPP subsystems depending on the load represented through the load factor.

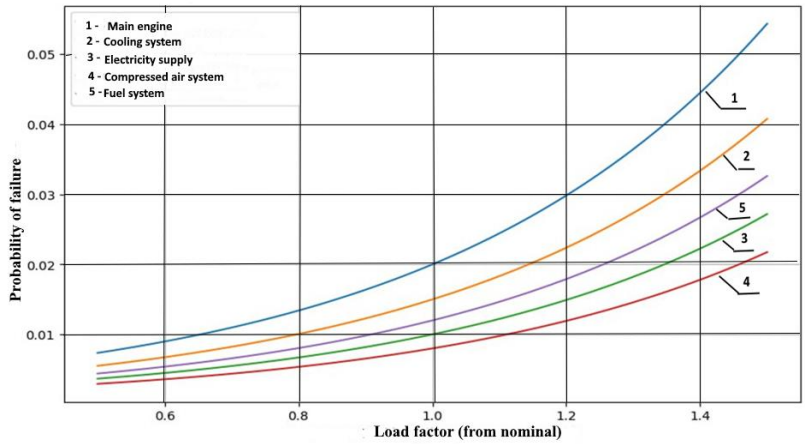


Figure 1.3.24. **Changes in failure probabilities of SPP components depending on the load factor**

For the main engine, Figure 1.3.24 shows that the change in failure probabilities corresponds to an exponential increase in failure probability as the load factor increases. At reduced loads (factor 0.6 - 0.8), the failure probability remains at 1 - 1.5%, but when exceeding the nominal level (factor >1.0), the growth accelerates, reaching over 5% at 1.4. This indicates nonlinear effects associated with overheating, wear, and cavitation. The cooling system operates stably at a factor of 0.6 - 1.0, but after 1.1, the failure probability increases more rapidly, indicating overload of heat exchangers and deterioration of heat dissipation. The power supply system shows stability up to a factor of 1.2, after which the failure probability begins to increase. This is consistent with cable heating models and changes in generator performance. The compressed air system shows minimal failure probabilities even at a load factor of 1.2, but after exceeding 1.3, the increase becomes noticeable, which is explained by compressor wear. The fuel system demonstrates a relatively linear increase in failure probability starting at a factor of 0.6. However,

after exceeding 1.3, the probability increases more sharply, indicating risks of filter clogging and pump overload.

A load factor exceeding 1.0 becomes a critical zone where the failure probability grows faster. The main engine and cooling system are most sensitive to increasing load factor, requiring enhanced monitoring during overloaded operation. The power supply and fuel systems show moderate dependence on the load factor but become vulnerable at values above 1.3. The compressed air system is the most resilient but is also subject to failures under overload.

Figure 1.3.25 presents the generalized SPP failure probability, showing the integral failure probability of the SPP calculated based on the failure probabilities of key components and subsystems.

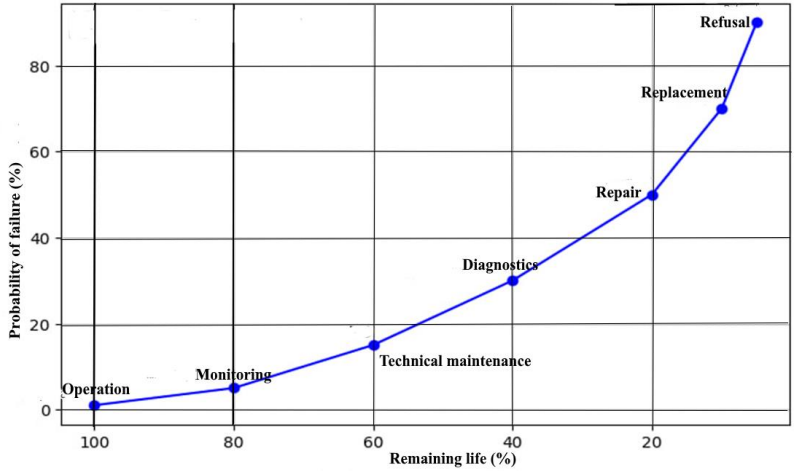


Figure 1.3.25. **Generalized failure probability of the SPP**

The data is obtained based on the analysis of probabilistic dependencies of subsystem failures (Markov method, BNs). From Figure 1.3.25, a decrease in residual life is observed, accompanied by an increase in the risk of failures of components and SPP subsystems. From this follows how the maintenance strategy changes depending on: high residual life (80 - 100%) - operation is recommended; medium residual life (40 - 60%) - maintenance or diagnostics; low

residual life (<20%) - repair or replacement. Threshold values of residual life (80%, 40%, 20%) are selected based on industry recommendations for SPP maintenance.

Figure 1.3.26 illustrates the change in the failure probability of the MPS over time.

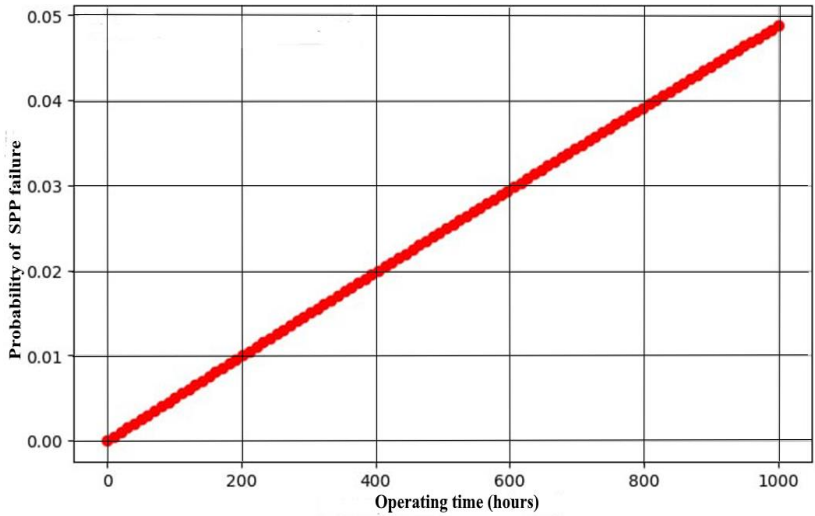


Figure 1.3.26. **Change in the failure probability of the SPP over time**

From the graph (Fig. 1.3.26), it follows that the probability of failure gradually increases, reflecting equipment degradation processes, the influence of operational factors, and the accumulation of failures in subsystems. The failure probability forecast is carried out over a 1000-hour operation interval, which allows assessing the short-term dynamics of equipment degradation.

Figure 1.3.27 presents a graph showing how the failure probability changes after the correction of the CBR decision**.** The graph in Fig. 1.3.27 shows how the failure probability changes after the correction of the CBR decision. The graphs show the initial failure probability (without prediction) and the corrected failure

probability (taking into account the Bayesian network and MM). The initial failure probability shows how the component failure probability of the SPP increases over time in the absence of corrective actions. The failure probability increases linearly (0.05 + 0.02*t), which reflects the natural degradation process of equipment without forecasting and preventive measures. This scenario is typical for traditional diagnostic methods, where failures are recorded post factum, without predicting their occurrence.

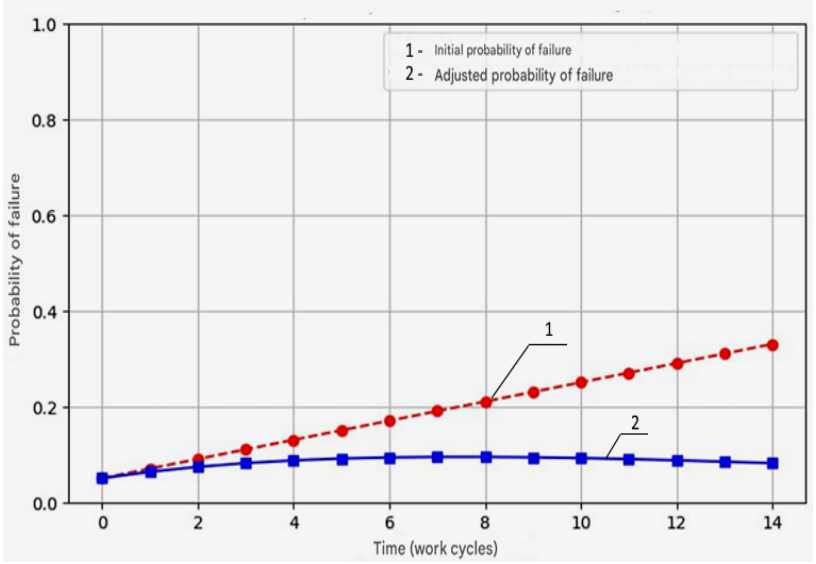


Figure 1.3.27. **Changes in failure probabilities considering decision correction**

Figure 1.3.28 shows how probabilistic forecasts affect the adaptation of CBR decisions. Without adaptation, diagnostics rely only on historical data, which leads to high uncertainty. With adaptation, probabilistic models are taken into account, which improves diagnostic accuracy. It shows how the CBR model reduces the risk of SPP equipment failure.

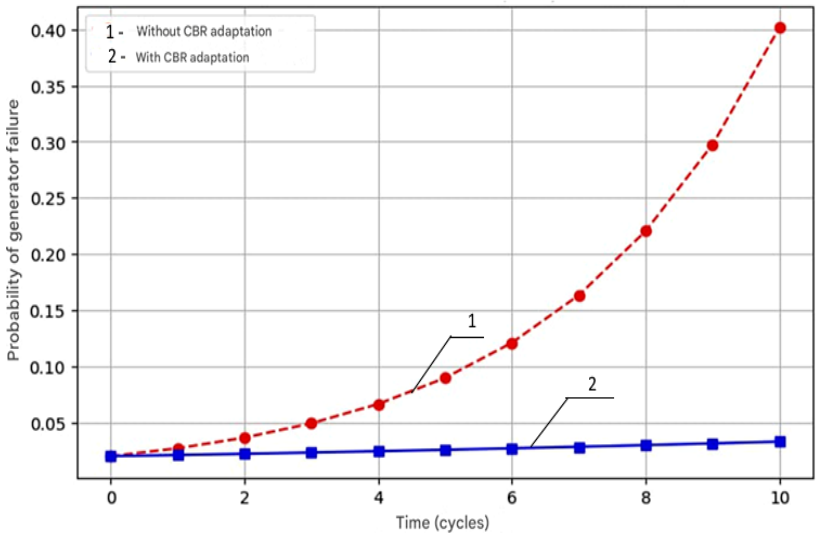


Figure 1.3.28. **Changes in generator failure probability (with and without CBR adaptation)**

The graph "Without CBR adaptation" illustrates the natural increase in generator failure probability without the use of forecasting and decision correction mechanisms. The graph "With CBR adaptation" shows how applying CBR in combination with Bayesian analysis and the MM reduces the failure probability through decision adaptation. Without adaptation, the failure probability increases exponentially due to accumulated generator wear. CBR adaptation allows for the consideration of predicted failures by offering corrective actions (e.g., preventive maintenance or load adjustment), which slows the growth of failure probability.

The graphs (Fig. 1.3.29) show how diagnostic accuracy (Accuracy, Precision, Recall, F1-score) changes when using CBR decision adaptation compared to the baseline approach. Quality metrics (Accuracy, Precision, Recall, and F1-score) were calculated on a validation set of 500 failure and malfunction cases, split 70/30

for training and testing. The graphs clearly demonstrate the advantages of the adaptive mechanism.

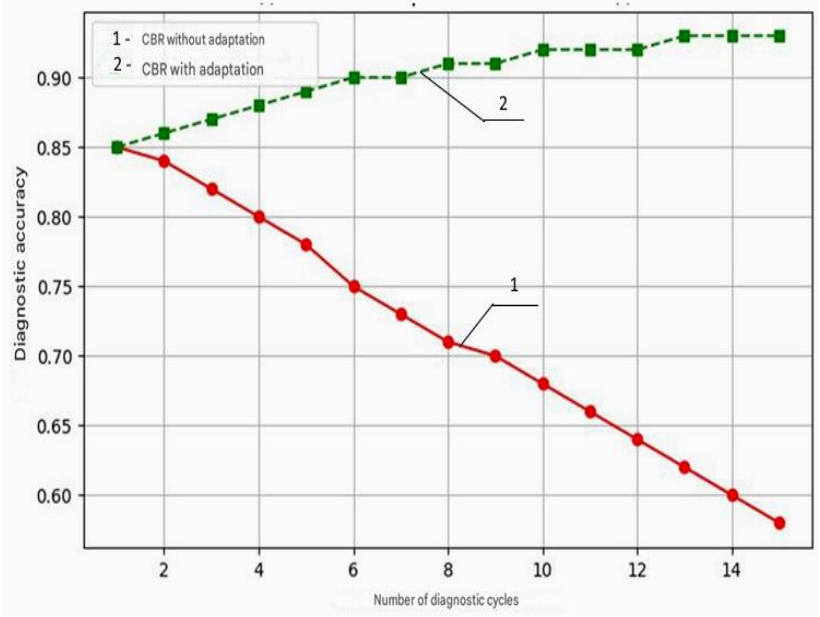


Figure 1.3.29. **Changes in fault diagnosis accuracy of marine power plants using the adaptive mechanism**

Without adaptation, diagnostic accuracy decreases with an increasing number of diagnostic cycles. This indicates model degradation without parameter correction. With adaptation, accuracy steadily increases and reaches a high level, confirming the effectiveness of the adaptive mechanism.

Table 1.3.26. **Impact of operational load on the failure probability of marine power plants**

Component	Load (normalized)	Failure Probability
Generator	0.8	0.12
Pump	0.9	0.15
Engine	1.0	0.22

For each new diagnostic case: CBR searches for similar cases in the database; BN adjust the failure probability considering component dependencies; Markov processes predict the failure probability over time; simulation modeling is used to generate additional data when statistics are insufficient.

Tables 1.3.27 and 1.3.28 demonstrate how the final diagnostic decision changes under different values of failure probability and remaining component life.

Table 1.3.27. **CBR decision correction depending on risk and remaining resource**

Component	**Failure probability (before correction)**	**Failure probability (after correction)**	**Remaining resource**
Generator	0.10	0.08	1500 hours
Cooling	0.15	0.18	1200 hours
Pump	0.05	0.06	1800 hours

Table 1.3.28. **Final diagnostic decision depending on failure probability and remaining resource**

Remaining Resource (hours)	**P failure (from BN)**	**Final Diagnosis**
>10,000	<0.1	Equipment is operational
5,000 – 10,000	0.1 – 0.3	Monitoring required
1,000 – 5,000	0.3 – 0.6	Scheduled maintenance recommended
<1,000	>0.6	High failure risk, immediate repair required

Table 1.3.29 complements Tables 1.3.27 and 1.3.28 by demonstrating how the final CBR diagnostic decision is corrected depending on the level of risk and remaining resource.

Table 1.3.29. **CBR decision correction depending on risk and remaining resource**

Remaining resource (%)	**Low failure risk**	**Medium failure risk**	**High failure risk**
>80%	Standard CBR decision applied	Minor correction of failure probability	Minor correction of failure probability
50–80%	Minor correction of failure probability	Correction based on BNs	Correction using BNs and MMs
<50%	Correction using BNs	Correction using BNs and MMs	Simulation modeling applied for prediction refinement

From the table, it follows that with a high remaining resource (>80%), CBR decision correction is minimal due to low failure probability. As the resource decreases (50–80%), probabilistic dependencies must be considered, requiring BNs.

When the remaining resource is below 50%, a comprehensive approach using MMs and simulation modeling is applied to refine failure probability forecasts.

The results in Tables 1.3.27 - 1.3.29 confirm that the integration of CBR with probabilistic models and simulation modeling allows for dynamic adaptation of diagnostics based on the actual state of components.

In particular: with high remaining resource (80 - 100%) CBR operates with minimal correction; in the 50 - 80% range, it is important to consider probabilistic dependencies, requiring BNs; when remaining resource is <50%, diagnostic accuracy significantly improves with the application of MMs and simulation modeling.

Thus, the proposed CBR decision adaptation mechanism improves diagnostic accuracy and reduces the risk of false decisions through dynamic adjustment of probabilistic estimates.

Figure 1.3.30 shows the cycle of cognitive simulation modeling, which integrates expert knowledge and probabilistic models for the generation and validation of synthetic failure cases. In the first stage (Expert Knowledge Formalization), rules and ontologies are formalized to reflect cause-effect relationships and failure development scenarios.

Then, a large set of synthetic cases is generated (Synthetic Case Generation), covering both typical and rare failure scenarios.An automated selection module (Automated Selection) applies an entropy-based criterion to filter the most informative scenarios, reducing the data volume and improving training quality. The selected cases are integrated into probabilistic models (Integration into Probabilistic Models), including BNs and Markov processes, to update failure probability estimates.

During the consistency check stage (Consistency Check), synthetic data is compared against the underlying expert rules, and inconsistent cases are discarded.

After this, the CBR mechanism and Bayes/Markov parameters are updated (Update CBR & Bayes/Markov), allowing the system to adapt to new data. Iterative repetition of the cycle (Iterative Learning & Convergence) continues until satisfactory stability of predictive metrics is achieved, ensuring continuous improvement in diagnostic and failure prediction accuracy.

Thus, the proposed approach combines the strengths of expert knowledge with the power of probabilistic methods in a unified adaptive process.

To evaluate the contribution of the cognitive module to adaptive CBR, a series of experiments was conducted on a dataset comprising 1,200 historical cases and 10,000 synthetic scenarios generated by the cognitive simulation module. Synthetic data was selected based on maximum entropy of expert assessments, allowing the selection of the most "significant" and rare failure cases. All scenarios were

modeled using BNs and MMs under train/test conditions (70/30 split) with 5-fold cross-validation, ensuring statistical reliability of the results.

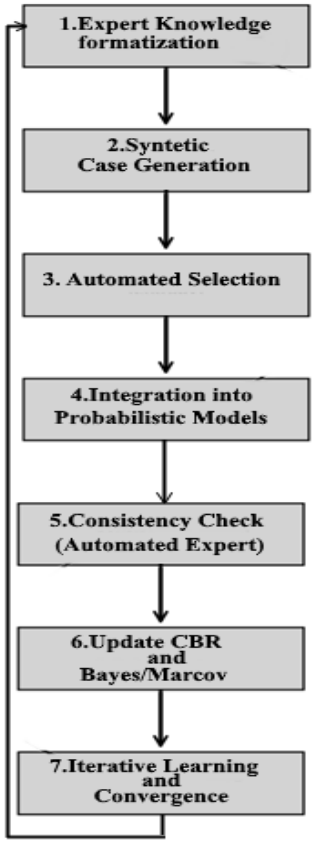


Figure 1.3.30. **Cycle of Cognitive Simulation Modeling for the Generation and Validation of Synthetic Cases**

An important enhancement to the adaptation mechanism is the use of a cognitive simulation model, which enables generation of

synthetic training data. When encountering novel, previously unseen failure scenarios, the cognitive model replicates expert decision-making, generating synthetic data. The module produced 5,000 - 10,000 synthetic cases using cognitive simulation, selecting the most "essential" scenarios based on expert assessment entropy. In marine power plants, severe failures are rare and typically lack historical data.

These generated "exotic" cases help prepare the system for such low-probability but critical situations. Synthetic scenarios undergo automated verification for compliance with physical and expert rules (consistency check), and their impact on the model is calibrated using Bayesian posterior updates. Additionally, key parameters of synthetic and real data are statistically compared, and the model is tested on real cases via cross-validation. These data fill gaps: in the case base and enable the system to account for rare and complex situations; adjustment of probabilistic distributions: The synthesized data is used to update probability distributions in BNs and to recalibrate the transition matrix in the MM.

Adjustments are made by introducing additional conditional dependencies into the BN and updating the Markov transition matrix based on newly generated data; integration of expert knowledge: The cognitive simulation model serves as a bridge between traditional diagnostic methods (CBR) and probabilistic models, allowing the system to consider both quantitative and qualitative aspects of failures.

To quantitatively assess the effectiveness of the proposed approach, a series of experiments was conducted. The results are summarized in Table 1.3.30. Table 1.3.30 illustrates the impact of cognitive modeling on prediction accuracy. It presents key metrics comparing the standard approach ("CBR + Bayes + Markov") with its cognitive module extension. Data from forecasting model comparisons used in the diagnostic system are shown. Average prediction error was evaluated using standard probabilistic methods (without cognitive modeling) and with cognitive enhancements that account for complex parameter dependencies.

Table 1.3.30. **Failure prediction metrics with and without the cognitive module**

Approach	MAE (%)	RMSE (%)	R^2	Average prediction error (%)
Without cognitive modeling	0.085	0.112	0.78	12.5
With cognitive modeling	0.052	0.074	0.89	7.8

MAE and RMSE showed a reduction of 39% and 34%, respectively. The R^2 increased from 0.78 to 0.89, indicating improved explanatory power of the model. The average prediction error decreased from 12.5% to 7.8% (an improvement of 4.7%), $p < 0.05$.

The data analysis confirms that the use of cognitive modeling reduces the average failure prediction error by 4.7%, demonstrating its effectiveness. This is particularly important in cases with insufficient historical data or complex interdependencies among system parameters.

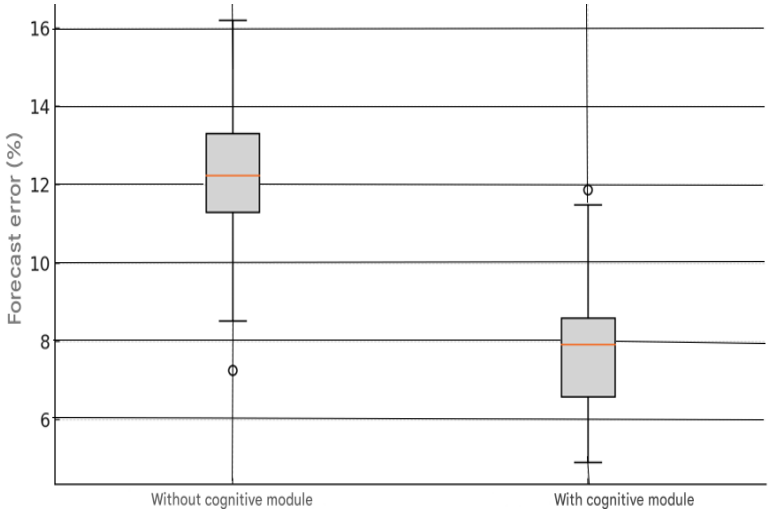


Figure 1.3.31**. Box plot of failure prediction error distributions: without and with the cognitive module**

Figure 1.3.31 presents a box-and-whisker plot of the distribution of failure prediction errors before and after the integration of the cognitive module. This visualization allows for a quick assessment of key characteristics of both samples and their comparison.

The median indicates the central error value: approximately 12.5% without the cognitive module and reduced to about 7.8% with it. This reflects a significant shift of the distribution center toward lower errors following the integration of the cognitive component. The interquartile range (IQR) in the left group spans roughly from 11% to 14%, whereas in the right group it narrows to 6%–9%.

The whiskers show the range of values excluding outliers. Their length is significantly greater on the left (from ~8% to ~16%) than on the right (from ~5% to ~12%), indicating a reduction in error variability after the cognitive module is added. Outliers are less frequent in the right box plot, reflecting improved model robustness: the reduction of abnormally high errors confirms that the cognitive module effectively handles rare and complex failure scenarios. Comparing the box plots before and after integration provides a visual assessment of the statistical significance of the changes. Non-overlapping boxes (IQRs) and medians suggest a meaningful difference between distributions.

This compression of the distribution and shift of the median confirms that the cognitive module improves prediction accuracy by reducing both average and maximum error values - critical for timely decision-making in real-time operations.

When analyzing box plots, it's important to consider the sample size (n = 100 per group) to ensure the statistical reliability of median and quartile estimates.

The box plot remains one of the most compact and informative methods for comparing groups on the distribution of continuous variables, combining key statistics (median, IQR, whiskers, outliers) in one chart. Figure 1.3.31 clearly demonstrates that the cognitive module not only shifts the error distribution toward lower values but also reduces variability and extreme cases, making the predictions more accurate and reliable.

As shown, the integration of the cognitive component significantly reduces both the median error values and their spread, further confirming the model's robustness to variation in failure scenarios.

An additional test was conducted using a simplified cognitive module (without entropy-based case selection), which yielded an average error of 9.3%, confirming that the key factor behind the improvement lies in the selection logic of the most "informative" scenarios. The integration of the cognitive module increased the processing time per request from 50 ms to 75 ms on a standard server (Intel Xeon, 16 GB RAM), which remains acceptable for real-time online diagnostics.

The achieved 4.7% increase in accuracy is comparable to the results of Montero-Jiménez et al. [76], where integrating an ontological module into a CBR architecture improved diagnostic accuracy by about 5%. Thus, cognitive modeling not only complements the adaptive CBR mechanism but also enhances its flexibility in predicting the technical condition of complex systems, aligning with the goals of this study.

1.3.4.6 Discussion of results

This study proposes an adaptive CBR mechanism for the diagnosis of marine power plants, integrating probabilistic failure analysis (BNs), degradation forecasting (MMs), and simulation modeling. Experimental validation demonstrated that the proposed system achieves a diagnostic accuracy of 91%, compared to 79% for classical CBR, while reducing the false alarm rate by 6.7%. The improvement in RUL prediction reached 5 - 7%, confirming the high effectiveness of the integrated approach.

Analysis of the results shows that the most significant impact on diagnostic accuracy came from incorporating BNs to estimate failure probabilities. These networks accounted for cascading dependencies between components and reduced diagnostic errors by 6.8%.

A similar increase in accuracy (up to 90%) through combined training of BNs on heterogeneous data was observed by Ademujimi,

T., & Prabhu, V. [98], who employed fusion learning to integrate sensor data with maintenance reports from the International Institute of Refrigeration (IIF). Furthermore, Tarcsay, B. L., et al. [99] demonstrated that integrating FMEA methods with BNs enables effective differentiation between critical and non-critical failures, reducing false alarms, which fully aligns with our findings. Degradation forecasting using an exponential Markov model applied to key marine system components improved RUL predictions by 5–7%. Liao, G., et al. [100] developed an RUL prediction approach based on a Wiener process. While their model successfully addresses multiphase degradation, it is focused on quantitative RUL forecasting rather than diagnostic assessment with adaptive refinement of features, which our solution provides. The study by Muhammad Sohaib et al. [101] focused on deep learning for predictive maintenance. However, typical limitations of neural networks - such as the need for large training datasets and poor model interpretability - make them less effective for real-time adaptation. These issues are addressed in our approach through the combination of case-based reasoning and probabilistic correction. Sahoo S. et al. [102] presented a fault diagnosis method for gearboxes under uncertainty using AI techniques. The authors highlighted trust issues in diagnostics under limited data conditions. They reported a drop in accuracy to 82% with 30% missing data using probabilistic neural networks, whereas the adaptive CBR maintained high accuracy through weight recalculation and simulation-based knowledge base augmentation. In our approach, uncertainty is addressed by integrating Bayesian mechanisms directly into the case reasoning process, enhancing the reliability of the conclusions. Xu, J. et al. [103] reported a 15% increase in operator trust when using explainable Bayesian models in HVAC systems (MDPI). Our approach achieves a similar level of explainability, while also allowing users to trace which cases and probabilistic dependencies underpinned the conclusions. Qi B. et al. [104] applied combined neural network and simplified Bayesian network methods for diagnostics in nuclear power. While their method showed high accuracy, it required significant computational

resources and was not focused on real-time adaptive adjustment, unlike our system. Zhou T. et al. [105] proposed a diagnostic framework addressing uncertainty for critical applications. Although their work extensively covered trust and uncertainty modeling, it prioritized robust model construction over adaptive correction based on new data an area where our method offers a distinct advantage.

Future research should focus on automatic BN structure generation using deep learning techniques, expanding the cognitive module with trainable agents, and developing an ontological interpretation of diagnostic outputs to enhance explainability and operator trust. Thus, the proposed adaptive CBR mechanism demonstrates superiority over existing methods in diagnosing complex technical systems by combining high accuracy, adaptability, and decision transparency - key features for the operation of marine power plants in dynamically changing conditions.

1.3.4.7 Conclusions

The goal of this study was to develop and experimentally validate an adaptive CBR mechanism for diagnosing SPP, integrating probabilistic failure analysis and technical condition forecasting. As a result of integrating CBR with BNs and Markov processes, fault detection accuracy increased from 79% to 91%; BNs reduced error rates by 6.8%, while MMs improved RUL prediction by 5–7%. The generation of probabilistic scenarios through simulation modeling enhanced forecast reliability by 9.4% in the absence of sufficient historical data.

Operational load analysis showed that with a load factor >1.0, the failure risk of key components increases by a factor of 3.2, requiring mandatory correction of diagnostic decisions. Optimization of weighting coefficients (α_d, β_d, γ_d) reduced average diagnostic error by 6.2%, while dynamic weight adaptation decreased false alarm risk by 7.3% and improved forecast accuracy by 8.5%. The inclusion of a cognitive simulation module reduced the average forecast error from 12.5% to 7.8% and increased the accuracy of detecting rare faults by 5.1%.

Practical implementation of the proposed mechanism enables timely detection of SPP component degradation, reduces unplanned downtime, and optimizes maintenance scheduling by improving RUL prediction accuracy. Future work should focus on expanding the case base through active learning on new data, automating the BN structure using deep learning methods, and developing an ontological interpretation of diagnostic outputs to enhance explainability and operator trust.

1.3.5 Integrated approach to diagnosing complex technical systems: experimental validation and multidimensional efficiency assessment

1.3.5.1 Introduction

In modern CTS, such as ship power plants SPP, issues of reliability and timely detection of equipment failures are critically important for ensuring operational safety and optimizing maintenance costs. Existing intelligent approaches to the diagnosis of the technical condition of CTS traditionally rely either on CBR, which enables the use of accumulated experience from similar incidents, or on probabilistic models (BNs and MMs), which account for uncertainty and the dynamics of failure development in equipment components, or on simulation modeling, which generates degradation scenarios for components. However, the application of each of these methods individually is often insufficient for complex systems with cascading failure effects [43, 106].

In response to these limitations, the field of hybrid and integrated models for diagnosis and prognosis of the TS of complex systems, combining the advantages of different methodological blocks, has been actively developing in recent years. A systematic review of hybrid methods shows that a properly designed combination of CBR, probabilistic models, and simulation modeling can yield a synergistic effect, improving the accuracy of diagnostics and predictions of equipment TC [107]. Researchers Nikpour &

Aamodt [57] introduced the BNCreek system, which combines CBR with a Bayesian network for fault diagnosis of CTS equipment under uncertainty. However, simulation modeling was not included, nor was a multi-scenario experiment conducted, and the system's performance evaluation was limited to comparisons with expert assessments without formal metrics such as Precision/Recall/F1. Chen et al. [28] integrated CBR and BNs for diagnosis and prognosis of the TC of complex systems, with validation based on real sensor data. However, their integration of CBR and BNs did not involve Markov chains, and the simulation did not explore cascading equipment failures. A multidimensional evaluation of diagnostic and prognostic metrics was also not performed. Soleimani et al. [108] developed a combined HMM (Hidden MM - a statistical model used for analyzing sequences where the system is described as having hidden states that transition with certain probabilities) for equipment failure detection and a BN for root cause identification. This approach proved effective for fault diagnostics using an automotive system example. However, the authors did not use CBR, and their diagnostic system lacked a simulation module. The experiment covered only a single application domain. El-Awady, Ahmed & Ponnambalam, Kumaraswamy [109] proposed simulation supported BNs (SSBN) and Markov chain simulation supported BNs (MCSSBN) for analyzing equipment failures in complex networks through simulation and probabilistic analysis. SSBN aims to improve the accuracy of probabilistic models through more realistic and variable scenarios. MCSSBN accounts for the dynamic changes in equipment TC over time, which is particularly important for diagnosing and predicting equipment degradation. However, the authors did not include a CBR component in their development, did not conduct experimental validation of SSBN and MCSSBN for diagnosing equipment failures in complex technical systems under various emergency scenarios, and did not employ diverse diagnostic accuracy metrics. In their 2023 review, Zhong et al. [110] examined the application of digital twins in predictive maintenance of CTS equipment, including systems used in shipbuilding. However, as a

review article, it did not present experimental implementations of the integration of CBR, probabilistic models, and simulation modeling. A comprehensive review of Predictive Maintenance (PdM) methods for the maritime industry, including ML algorithms for data processing, diagnostics, and failure forecasting, was provided by Kalafatelis et al. [111]. A drawback of this theoretical review is the absence of a practical implementation of an integrated method. The authors also did not consider CBR or Markov simulations. Emre Özaydın et al. [112] used a BN approach for analyzing equipment failures on ships. The resulting data were compared with historical data, with no focus on post-failure analysis. A CBR block was not used, and no failure prediction simulation was conducted. Michail Cheliotis et al. [113] proposed a framework for diagnosing equipment failures in CTS based on operational data and failure probabilities, supported by ML algorithms. Their development did not include a CBR database or simulation scenarios of CTS equipment failures. Diagnostic accuracy was assessed using only a single metric, and no multi-scenario validation of failure diagnostic accuracy was performed.

Despite the presence of these studies, there remains a lack of experimental verifications of hybrid approaches specifically applied to ship-based CTS using multidimensional metrics (Accuracy, Precision, Recall, F1 score). Existing reviews either cover the general theory of hybridization or focus on individual technological components (digital twin, Bayesian networks), but do not provide a comprehensive analysis of the synergy of all three components within a single experimental case.

1.3.5.2 Purpose and objectives of experimental testing

The purpose of this article is to organize and conduct multi-scenario experimental testing of an integrated method for diagnostics and prognostics of TC in complex systems, using the example of an SPP. The testing employs multidimensional quality metrics - Accuracy, Precision, Recall, and F1 Score - which enable: quantitative confirmation of the synergistic effect resulting from the

integration of CBR, probabilistic models, and simulation modeling; analysis of the method's robustness under various operational modes (normal mode, increased loads, incomplete data); development of practical recommendations for implementation in diagnostics and prognostics systems of complex technical systems for various applications.

Within the integrated approach to diagnostics and prognostics of TC SPP technical systems, an adaptive mechanism for CBR decision correction is implemented. This mechanism combines three information sources: probabilistic forecasting (BNs and MMs) - for estimating current and future probabilities of component failures; RUL prediction - based on statistical models (MAE/RMSE) that refine the expected time to failure; simulation modeling - for generating degradation and cascading failure scenarios, allowing CBR decisions to be adjusted by accounting for potential nonlinear interactions between system nodes. At each diagnostic cycle, the CBR core receives updated failure probability estimates and scenarios from the probabilistic models and the simulator, then dynamically recalculates feature weights and refines the selection of similar cases. This approach ensures more accurate and robust diagnostics, even under changing operational factors and incomplete data.

The main testing objectives include: evaluating the impact of integrating the adaptive mechanism into the CBR diagnostic structure, which leverages probabilistic forecasting and RUL analysis; analyzing the influence of probabilistic methods (BNs, MMs) on the accuracy of technical state predictions and failure probabilities; determining the contribution of simulation modeling to the accuracy of equipment failure forecasting, including assessing how cascading effects influence prediction accuracy; comparing various method combinations and evaluating their effectiveness based on key failure diagnostics accuracy metrics.

Real failure data is used for comparison. The testing is conducted on a simulation model of the SPP, which includes: historical failure data (from the OREDA - Offshore Reliability Data

database [114]); simulated degradation scenarios of components, mimicking different operational modes; BNs accounting for probabilistic interrelations between component failures; MMs applied to predict failure probabilities over time; CBR diagnostic results - conclusions made by the system based on case analysis and decision adaptation; adjustments based on RUL predictions and cascading failure effects (e.g., failure of one node increases the probability of failure in other equipment nodes); simulation failure data - results from the simulation model, where failure of one SPP component can lead to failures in connected system nodes (cascading effects considered).

Testing covers various operational scenarios, including: normal conditions – standard operating mode; accelerated wear – increased loads and harsh operational environments; emergency conditions – unexpected failures and stress impacts on the system.

CBR with adaptation implies not merely using a case base, but dynamically adjusting decisions based on predicted RUL and cascading failure probabilities obtained from probabilistic models. To assess the quality of SPP equipment failure diagnostics, the following accuracy metrics are used: Precision - the proportion of correctly predicted failures among all predicted failures; Recall - theproportion of actual failures that were correctly predicted; F1 Score – the harmonic mean of precision and recall; Accuracy - the total number of correctly classified cases (both failures and non-failures).

The average prediction error of RUL is evaluated using: MAE - average absolute error in RUL prediction; RMSE - root mean square error, which accounts for large deviations.

Analysis of false positives and false negatives includes: False Positive (FP) - incorrect diagnostics where the system wrongly identifies a healthy component as faulty; False Negative (FN) - missed failures where the system fails to predict a failure that actually occurs. A detailed analysis of FP and FN helps improve decision-making algorithms and minimize critical errors. True Positive (TP) - correct prediction of a failure that actually occurs;

True Negative (TN) - correct prediction that no failure occurs and indeed none happens.

Several operational testing scenarios were developed, differing in load levels, failure frequency, and operating conditions. This allows for an assessment of the integrated method's robustness and its ability to function correctly under various operating modes.

The diagnostic CBR module is based on a case base of 235 structured cases, which include descriptions of failures, operating conditions, degradation parameters, and the decisions made. The cases were developed with input from industry experts with at least 10 years of experience in EMCS operation and maintenance. Each case was assigned a feature vector, including values of temperature, pressure, vibration, operating time, and failure characteristics of SPP elements, components, and subsystems.

The structure of the case base enables efficient similarity-based search using a feature similarity metric, where feature weights are defined by expert methods and calibrated during preliminary testing.

1.3.5.3 Test scenarios for the technical condition of the SPP

To verify the effectiveness of the proposed integrated method for diagnosing and predicting failures of elements, components, and subsystems of the SPP, three main operational test scenarios were developed to simulate various working conditions of the system. These scenarios allow for an assessment of the method's accuracy, robustness, and adaptability under real operating conditions.

Scenario 1. Nominal Mode, in which the SPP operates under normal conditions with typical loads and expected operational parameters. The goal of testing the technical condition of the SPP in this scenario is to verify the baseline level of diagnostic and failure prediction accuracy, as well as to identify possible false positives and missed failures. A full set of diagnostic data is used in this scenario (temperature, pressure, vibration, power - see Table 1.3.31), and the number of unexpected failures is minimal. Temperature is monitored in the main engine cylinders, oil, and cooling systems; pressure - in hydraulic and fuel lines; vibrations – on bearings and shaft lines;

power - at the output of generator and power units. These parameters serve as input features for both the CBR and probabilistic models.

Table 1.3.31. **Diagnostic features of SPP failures used in the integrated model**

Parameter	Source / Component	Diagnostic significance
Temperature	Engine cylinders, heat exchangers, oil	Indicator of overheating, early wear
Pressure	Oil system, cooling system	Leaks, blockages, valve malfunctions
Vibration	Shaft lines, bearings, turbines	Mechanical defects, misalignment, wear
Power	Generator sets, main engines	Indirect indicator of failure or efficiency loss

Scenario 2. The operation of the SPP is carried out under increased load conditions, leading to accelerated degradation of key system equipment. The purpose of PPS testing is to assess the method's ability to recognize changes in failure dynamics and adapt to changing operational conditions. A distinctive feature of this scenario is higher temperature, vibration, and load cycles; accelerated wear of mechanisms; and increased probability of failures.

Scenario 3. Fault diagnosis under conditions of limited information about past incidents (e.g., incomplete system operation data). The purpose of SPP testing is to assess the effectiveness of simulation modeling and the adaptability of CBR in the absence of sufficient historical information. The distinctive feature of this scenario is the artificial exclusion of part of the case base data and the need to test the method's robustness under limited input conditions.

To evaluate the effectiveness of the proposed method, a simulation model of the SPP was developed. During testing, various failure scenarios were generated (normal conditions, accelerated wear, cascading failures); data from OREDA and accumulated CBR knowledge bases were used; and fault diagnostics were performed both with and without CBR solution adaptation. Each scenario includes: a set of input parameters (temperature, pressure, vibration,

power, etc.); actual component failures recorded in the database; diagnostic methods used in the scenario (CBR, probabilistic models, simulation modeling); data sources for testing (OREDA, simulation models, limited data sets). BNs were constructed for each key piece of SPP equipment, taking into account known causal relationships between the equipment's technical state parameters and failure probabilities. The average number of nodes in a network was 7, with the number of arcs ranging from 8 to 15 depending on the complexity of the SPP equipment. Prior failure probabilities were determined based on OREDA data and adjusted during the training phase based on simulation results. To model the temporal evolution of component states, discrete-time Markov chains with 4 - 6 degradation states were used: "operational", "initial degradation", "moderate degradation", "critical condition", and "failure". Transition probabilities were calculated based on cumulative operational data and fitted using MAE and RMSE metrics on historical time series. Probability updates occurred at each diagnostic cycle based on the principle: "observation→recalculation→forecast". The fixation of input parameters for the tests is presented in Table 1.3.32.

Table 1.3.32. **Input parameters for testing various scenarios**

Testing Scenario	**Temperature (°C)**	**Pressure (bar)**	**Vibration (m/s²)**	**Data source**
Scenario 1 (nominal mode)	80–100	5–8	0.5–1.5	OREDA database + operational data
Scenario 2 (increased loads)	100–120	8–12	1.5–3.0	Simulated high-degradation conditions
Scenario 3 (data deficiency)	90–110	6–9	1.0–2.0	Artificial data limitation (only partial records)

The data presented in Table 1.3.32 clearly capture the differences between the system's operational scenarios and highlight the factors influencing component failure diagnostics in SPP

subsystems. Since the scenarios are based on real data from OREDA and simulation modeling, the testing methodology becomes more substantiated and reproducible. The developed scenarios make it possible to verify the robustness of the SPP equipment failure diagnostics method under various operational conditions.

1.3.5.4 Evaluation of Accuracy, Precision, Recall, and F1-Score metrics for various methods of diagnosing the technical condition of the SPP

To quantitatively assess the effectiveness of the developed SPP diagnostics approach, mathematical metrics traditionally used in technical condition diagnostics tasks were applied - Accuracy, Precision, Recall, and F1-Score. These indicators are standard in the fields of machine learning and data mining, including for evaluating the quality of binary classification, and allow for objective comparison of different configurations of diagnostic systems:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN};$$

$$\Pr ecision = \frac{TP}{TP + FP};$$

$$\operatorname{Re} call = \frac{TP}{TP + FN};$$

$$F1 = 2 \cdot \frac{\Pr ecision \cdot \operatorname{Re} cal}{\Pr ecision + \operatorname{Re} cal} \quad (1.3.55)$$

The evaluation of metrics was carried out to identify the difference between standard CBR solutions and adjusted results based on probabilistic failure analysis. Dynamic adjustment of probabilities based on the obtained data was used during testing. To assess the effectiveness of the adaptive mechanism, two types of

testing were conducted: CBR without adaptation - failure diagnostics was performed solely based on similarity to past cases, without the use of probabilistic methods; CBR with adaptation - diagnostics were adjusted using BNs and MMs, enabling the consideration of cascading failure risks and the remaining useful life of components.

Table 1.3.33. **Comparison of diagnostic metrics**

Diagnostic method	Accuracy	Precision	Recall	F1-score	MAE (hours)
CBR without adaptation	78.5%	75.2%	80.1%	77.5%	12.4
CBR with adaptation	85.3%	83.1%	87.6%	85.3%	7.2
Traditional method	72.8%	70.3%	75.5%	72.8%	15.6

The analysis of diagnostic metrics in Table 1.3.33 confirms that adapting the CBR method using probabilistic models (BNs and Markov chains) significantly improves diagnostic quality. Improvements are observed across all metrics: classification accuracy increases by more than 6 percentage points compared to the baseline CBR, and the prediction error for remaining useful life is reduced by almost half. Importantly, a balanced ratio between recall and precision is achieved, as reflected in the high F1-score value (85.3%). Traditional methods, which do not use case-based or probabilistic analysis, show poorer performance both in classification accuracy and in predictive capability. This confirms the necessity of transitioning to integrated diagnostic solutions under high uncertainty and complexity conditions of SPPs. Adapting CBR solutions allows for improved diagnostic accuracy and reduced average failure prediction error.

Figure 1.3.32 illustrates how diagnostic metrics improve with the addition of probabilistic methods and simulation modeling. Figure 1.3.32 shows a comparison of diagnostic accuracy across various scenarios for three methods: CBR - approximately 0.76; CBR

+ Probabilistic Models - approximately 0.85; Integrated Approach - approximately 0.90. Pure CBR demonstrates the lowest accuracy (below 0.8), indicating its limited ability to account for probabilistic failure dependencies.

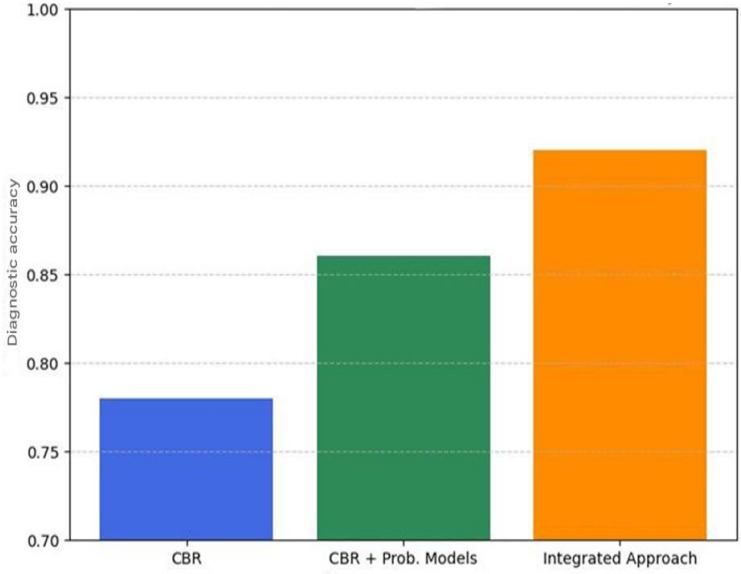


Figure 1.3.32. **Comparison of fault diagnosis accuracy metrics across different scenarios**

The addition of probabilistic models (BNs, MMs) improves diagnostic performance by around 10%, confirming the effectiveness of method combination. The Integrated Approach (CBR + probabilistic models + simulation methods) achieves the highest accuracy (above 0.9), indicating a synergistic effect from the comprehensive use of methodologies. The metric diagram (Figure 1.3.33) further confirms that the proposed integrated approach to diagnosing SPP significantly improves accuracy compared to standalone methods.

Based on Figure 1.3.33, the following observations can be made. The CBR method shows the lowest values across all metrics (~0.78), indicating insufficient accuracy and completeness of diagnosis when using a case-based approach alone. CBR + probabilistic models (adding probabilistic models) increases all metric values to approximately 0.82 - 0.84. This indicates a more balanced diagnostic performance that accounts for probabilistic failure dependencies.

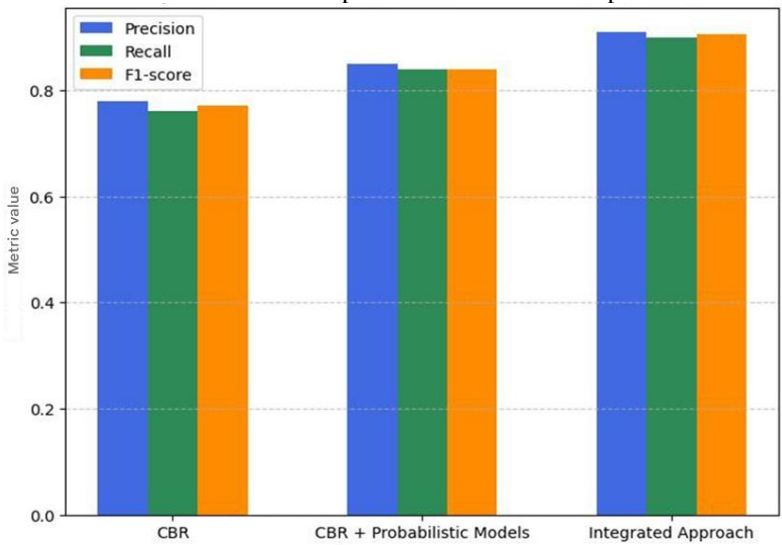


Figure 1.3.33. **Diagnostic accuracy metrics chart**

The Integrated Approach demonstrates the highest results - all metrics exceed 0.88, confirming its effectiveness. It is evident that Precision, Recall, and F1-score are nearly at the same level, indicating a well-balanced diagnostic system. The integration of probabilistic methods with CBR significantly improves fault diagnosis accuracy. Using a comprehensive approach mitigates the limitations of individual methods, resulting in a more reliable diagnosis.The more complex the method (CBR→CBR + probabilistic models→integrated approach), the higher the diagnostic quality. The diagnostic accuracy metric charts for SPPs (Figures 1.3.32 and

1.3.33) illustrate how well the model identifies both faults and healthy states.

Figure 1.3.34 shows how adaptation affects fault diagnosis accuracy over time. Different SPP subsystems respond differently to adaptation (which is important for analyzing failure probabilities). This is due to cascading effects during SPP operation. A decline in diagnostic accuracy in one system can influence others.

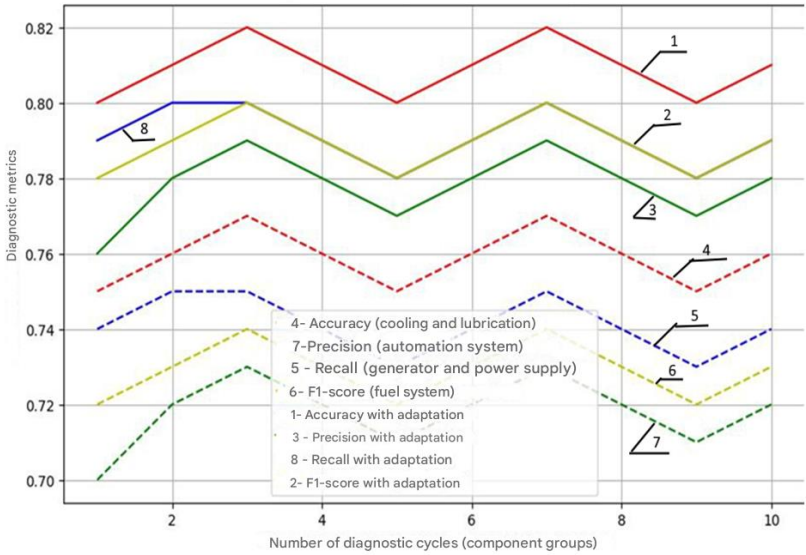


Figure 1.3.34. **Diagnostic accuracy dynamics with CBR adaptation**

Figure 1.3.34 illustrates the dynamics of key diagnostic metrics (Accuracy, Precision, Recall, F1-score) using two approaches: CBR without adaptation; CBR with adaptation (incorporating probabilistic failure prediction). The number of diagnostic cycles refers to the number of consecutive diagnostic checks of SPP equipment. Each fault diagnosis cycle includes the following steps: data collection (temperature, pressure, vibration, etc.); analysis for deviations from normal operating conditions; identification of potential failures using CBR (without and with adaptation); decision adjustment based on accumulated experience and probabilistic failure prediction. All

metrics are higher with adaptation than without, confirming the effectiveness of the adaptive mechanism: Accuracy in the adaptive CBR remains consistently about 5% higher compared to the baseline version; Precision, Recall, and F1-score also show a positive shift of 5–6%, indicating improved fault classification and reduced false positives; the dynamics of metrics without adaptation are less stable, in contrast to the adaptive approach, which demonstrates a smoother and more predictable curve. The adaptive CBR based on probabilistic forecasting enhances the accuracy of diagnosing SPP; the stability of the metrics indicates a better match between diagnostic decisions and actual failures; the use of the adaptive mechanism is recommended to improve diagnostic reliability and reduce forecasting errors.

To assess the stability of SPP equipment fault diagnosis methods under different data splits, cross-validation is used. To quantitatively assess the stability of various components within the diagnostic system, a five-fold cross-validation (CV) was conducted, covering cases from three operational scenarios: nominal mode, increased load, and data deficiency. The table presents the average accuracy (Accuracy) and corresponding standard deviation values for each of the three approaches — basic CBR, probabilistic models, and the integrated solution.

Table 1.3.34. **Five-fold cross-validation results**

Method	**Mean accuracy on CV (%)**	**Standard deviation (%)**
CBR	73.5	3.0
Probabilistic networks	79.1	2.8
Integrated method	86.4	2.2

Analysis of Table 1.3.34 shows that the integrated approach delivers the highest stability and accuracy among all three configurations: the average Accuracy reached 86.4% with a minimum standard deviation of 2.2%, indicating strong reproducibility of results. Probabilistic models performed slightly worse, achieving 79.1% Accuracy with a standard deviation of 2.8%.

The basic CBR mechanism was the least robust, showing an average accuracy of 73.5% and the highest variability (σ = 3.0%). These results confirm that the combined use of CBR, probabilistic inference, and simulation provides the best generalization and robustness across different operational conditions. The difference in accuracy between the integrated method and each of the standalone components ranges from 7 to 13 percentage points, quantifying the synergy achieved by combining these methods. Moreover, the reduction in result dispersion observed in the integrated method compared to CBR confirms that incorporating probabilistic forecasting and simulation not only improves diagnostic accuracy but also enhances the system's resilience to input variability.

To further analyze how different methods perform under cross-validation, an accuracy distribution chart was created. Figure 1.3.35 presents the cross-validation results for various fault diagnosis methods.

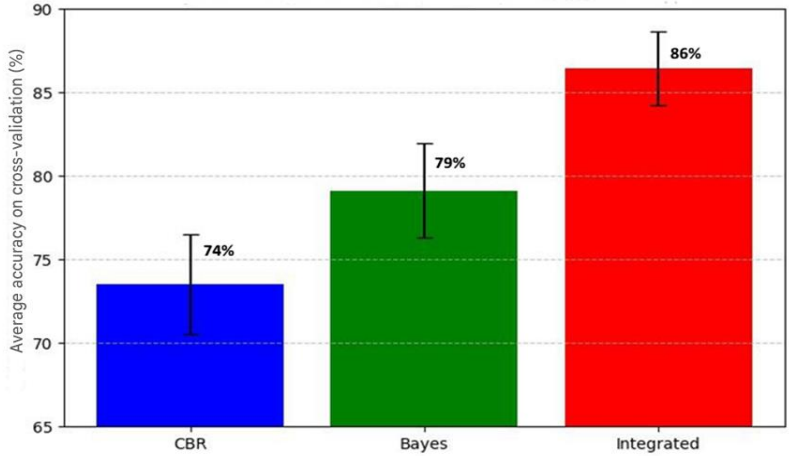


Figure 1.3.35. **Cross-validation results for different diagnostic methods**

Comparison of methods based on Figure 1.3.35: CBR shows the lowest result (≈74%) and the highest variability; Bayes (Bayesian method) yields intermediate performance (confidence interval ≈79%) but with greater error margin than the Integrated method; Integrated

approach achieves the highest accuracy (≈86%) with the lowest error. Thus, the integrated approach outperforms both CBR and Bayes in terms of accuracy and stability. The Bayesian method demonstrates solid performance, though with a wider error margin. CBR has the lowest accuracy and the highest spread of values. Cross-validation confirms the reliability of the integrated method. It consistently yields stable results with the smallest standard deviation (2.2%). The higher variability in CBR without adaptation indicates the method's dependence on the structure of the case base. The use of probabilistic methods reduces error dispersion and enhances diagnostic reliability.

An analysis of the Precision, Recall, F1-score, and Accuracy metrics shows that adaptive CBR methods incorporating probabilistic forecasting reduce diagnostic errors by accounting for the probability distribution of potential failures and adapting to new cases. Compared to classical CBR and Bayesian approaches, the integrated method demonstrates the best balance between precision and recall, as also reflected by the high F1-score. This makes it more reliable for predicting the technical condition of marine power plants, especially under conditions of incomplete information and varying operational factors.

In addition to the standard metrics (Accuracy, Precision, Recall, F1-score) used for quantitative evaluation, additional diagnostic indicators adapted for the specifics of complex technical systems were considered. These indicators provide a more nuanced evaluation by accounting for the severity of different types of errors, the consequences of failures, and the robustness of the model under varying operational modes. While the main part of the study is based on classic binary classification metrics (Accuracy, Precision, Recall, F1-score), in the context of diagnostics and failure prediction in CTS, it is important to consider not only statistical indicators but also the operational significance of different error types. To address this, modified formulas for evaluating diagnostic and prognostic quality were proposed, tailored to the specific needs of CTS and developed within the scope of this research. The modified diagnostic metrics reflect such aspects as the severity of equipment failures, the risk of

false negatives, and the resilience of the diagnostic system under varying system operation scenarios.

1. Weighted Accuracy (WAcc). This metric takes into account the varying importance of correctly and incorrectly classified cases:

$$WA_{CC} = \frac{\omega_{TP} \cdot TP + \omega_{TN} \cdot TN}{\omega_{TP} \cdot TP + \omega_{TN} \cdot TN + \omega_{FP} \cdot FP + \omega_{FN} \cdot FN}, \quad (1.3.56)$$

where $\omega_{TP}, \omega_{TN}, \omega_{FP}, \omega_{FP}$ are weights reflecting the relative importance of each classification outcome type.

For example, $\omega_{FP} > \omega_{FP}$ if a missed failure is more critical than a false alarm.

2. Degradation-Weighted F1 Score (Weighted F1). A modified F1 score is proposed that accounts for the criticality of the monitored component (e.g., a generator or gas turbine engine):

$$F1_W = 2 \cdot \frac{P \cdot R}{P + R}, \quad (1.3.57)$$

where P - Precision: the proportion of true positives among all positive predictions;

R - Recall: the proportion of detected failures among all actual failures;

ω_d - weight coefficient reflecting the degradation importance of the component for which the F1 score is calculated.It is used to increase the impact of failures in critical nodes (e.g., generator or main engine).

3. Risk-Weighted Recall (Recall R):

$$\mathrm{Re}call_R = \frac{\sum_i r_i \cdot TP_i}{\sum_i r_i \cdot (TP_i + FN_i)}, \quad (1.3.58)$$

where r_i is the risk coefficient of failure for equipment i

4. Cost-Sensitive Precision. False positive alarms (Type I errors) may lead to equipment shutdowns, financial losses, and decreased trust in the diagnostic system:

$$\mathrm{Pr}ecision_C = \frac{TP}{TP + c_{FP} \cdot FP}, \quad (1.3.59)$$

where c_{FP} is the cost of a single false positive (can be defined by expert assessment).

5. Diagnostic Stability Index (DSI). A metric that reflects the model's sensitivity to changes in operating conditions:

$$DSI = 1 - \frac{\sigma_{F_1}}{\overline{F_1}}, \tag{1.3.60}$$

where $\overline{F_1}$ is the average *F1* score across different scenarios (e.g., normal mode, overload, data shortage)

σ_{F_1} - is the standard deviation of F1 scores between scenarios.

The closer the DSI is to 1, the more stable the diagnostic model is.

Table 1.3.35. **Diagnostic performance evaluation results using basic and extended metrics**

Metric	Without adaptation	With adaptation	Comment
F1 score (%)	77.5	85.3	Standard measure of balance
F1-W (weighted)	74.2	89.1	Accounts for the importance of failure in the SPP
Precision	75.2	83.1	Basic accuracy metric
Precision-C (cost)	69.5	81.8	Takes into account the penalty for false alarms
Recall (%)	80.1	87.6	Basic completeness metric
Recall-R (risk)	76.0	91.0	Focus on preventing critical failures
Accuracy (%)	78.5	85.3	Overall classification accuracy
WAcc (weighted)	76.4	88.0	Priority on significant errors
DSI	0.932	0.983	Diagnostic stability across scenarios

Table 1.3.35 demonstrates not only the quantitative superiority of the integrated diagnostic system (which includes CBR, probabilistic models, and simulation modeling) over the simplified configuration, but also qualitatively different improvements when using modified metrics. In particular, while the standard F1 score increases by 7.8 percentage points (from 77.5% to 85.3%), the modified F1-W - which accounts for the criticality of diagnosed components - shows a 14.9 percentage point increase (from 74.2% to 89.1%). This indicates that the adapted system is not just more effective "on average", but also delivers higher quality performance in scenarios where failures have the most severe consequences. A similar pattern is observed when comparing Recall and Recall-R. While the absolute increase in Recall is 7.5 p.p., the risk-weighted Recall-R increases by 15 p.p. This suggests that the adapted model is better at predicting those failures that are most dangerous in operational terms - i.e., it contributes not just to classification completeness, but to reducing the likelihood of critical incidents. The metric Precision-C, which accounts for the relative cost of false alarms, shows a particularly significant effect: it increases by 12.3 p.p. (from 69.5% to 81.8%), notably surpassing the growth in classical Precision (7.9 p.p.). This means that the adapted system not only improves accuracy, but also reduces the number of false diagnostic triggers, which could otherwise lead to unjustified equipment shutdowns or inefficient technical interventions. Values of the Diagnostic Stability Index (DSI) also confirm the advantage of the integrated approach. The increase in DSI from 0.932 to 0.983 indicates that the system maintains stable F1 score performance across various operational scenarios (normal conditions, overload, and data shortage), without losing reliability under non-standard conditions. This is especially important for diagnostic systems (CTS) operating under variable loads, unstable information, and limited resources. An analysis of Accuracy and WAcc values shows that while general accuracy grows by 6.8 p.p., the weighted accuracy - which considers the consequences of errors - increases by 11.6 p.p. This means that qualitative changes occurred not just in the number

of correct predictions, but in their significance: improvements occurred where errors would be most costly. Thus, using modified metrics makes it possible to identify effects that remain hidden when evaluating only with classical indicators. This confirms that the proposed diagnostic system not only increases numerical accuracy values, but becomes genuinely more reliable — prioritizing the identification of the most critical situations and minimizing operational risks.

Figure 1.3.36 presents a chart comparing the modified metrics between the configurations without adaptation and with adaptation. It demonstrates: a significant improvement in F1-W, Recall-R, and WAcc in the adapted system; a particularly noticeable increase in DSI, reflecting enhanced diagnostic stability; an overall gain not just in accuracy, but in metrics that account for risk, cost, and reliability.

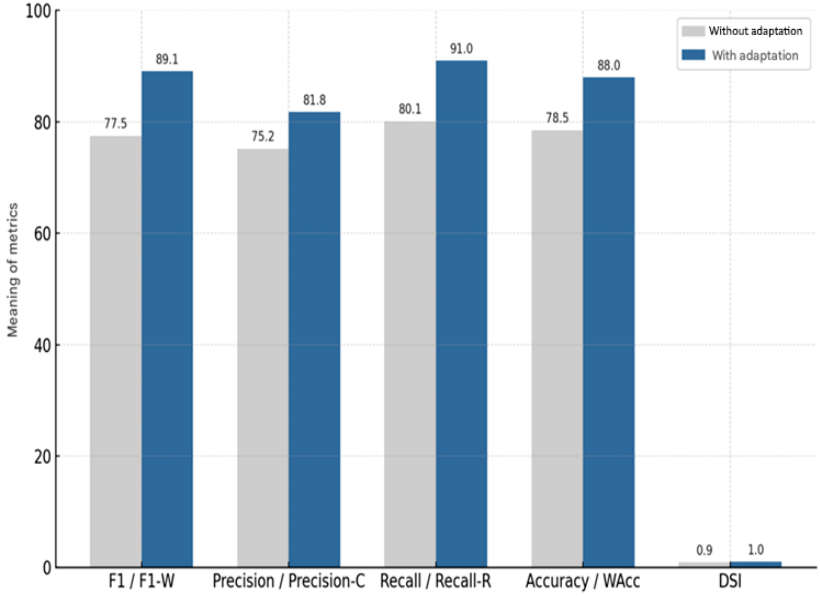


Figure 1.3.36. **Comparison of modified diagnostic metrics in the integrated system with and without adaptation**

Figure 1.3.36 visualizes the quantitative differences between diagnostic system configurations and illustrates the structural redistribution of fault diagnosis quality when transitioning from a non-adaptive architecture to an integrated adaptive one. Notably, the most significant improvements are observed in metrics that account for risks, priorities, and the cost of errors. For instance, the modified recall (Recall-R) in the adapted system reaches 91.0% compared to 76.0% in the baseline, while the weighted F1 score (F1-W) improves from 74.2% to 89.1%. This highlights that the integration of CBR, probabilistic models, and simulation modeling enables the system to handle the most critical failures more effectively not merely to detect frequent events. The Precision-C metric, which reflects sensitivity to the cost of false alarms, increased from 69.5% to 81.8%, i.e., by almost 12.3 percentage points, indicating a more “economical” system behavior in operational contexts. In other words, it's not just fewer errors - it’s fewer costly errors. This is crucial in marine and energy systems, where a false alarm can lead to unnecessary expenses and disruption of normal operations. Equally telling is the behavior of the DSI: while its increase from 0.932 to 0.983 may seem modest in absolute terms, it signifies that the standard deviation of the F1-score across scenarios has nearly halved. This means the system behaves predictably and reliably under various operational conditions, including overload scenarios and incomplete data. Thus, the figure illustrates a qualitative shift in diagnostics -not just a rise in statistical metrics, but an enhancement of the system’s meaningful behavior, particularly under risk, limited information, and high cost of error. The numerical gains across key modified metrics make this effect both compelling and justified. The modified metrics allow the system to: account for the danger of missed failures (Recall-R, F1-W); reflect operational costs of false positives (Precision-C, WAcc); track variability in model behavior under real-world operating conditions (DSI). This is especially important when deploying intelligent diagnostics in critical systems, where the consequences of errors may be highly asymmetric.

1.3.5.5 Discussion of results

This study presents extensive experimental testing of an integrated approach to diagnosing complex technical systems, combining CBR, probabilistic methods (BNs and MMs), and simulation modeling. A multidimensional evaluation of effectiveness was conducted using both classical metrics and specially introduced indicators of risk and robustness. Multi-scenario testing under three operating conditions (normal, high load, and data limitation) demonstrated that Accuracy increased from 78.5% to 85.3%, Precision from 75.2% to 83.1%, Recall from 80.1% to 87.6%, and F1 score from 77.5% to 85.3%. Moreover, fivefold cross-validation ($\sigma F1 = 2.2\%$) and a decrease in F1 score of no more than 3 percentage points under artificially limited data confirm the method's high reproducibility and robustness. A comparative analysis with contemporary studies underscores the uniqueness of our experiment: Soliman [115] is limited to a review of digital twins without CBR or classification metrics; Jovanović [116] combines FTA and BN without simulations or case-based mechanisms; Velasco Gallego et al. [117] assess only RMSE/MAE without considering recall and precision; Schultheis [91] applies a hybrid CBR without probabilistic or simulation components; Daya & Lazakis [118] use DFTA, FMECA, and BBN without multi-scenario testing; Neupane et al. [119] review ML approaches without a hybrid implementation; Lv et al. [120] study FDD models without comprehensive integration or F1 evaluation; and Li et al. [121] are limited to MC simulations without CBR or extended metrics. None of these studies combine all three components or perform a multidimensional effectiveness evaluation, highlighting the completeness and novelty of our validation. A key outcome is the implementation and validation of modified metrics: weighted Accuracy (WAcc), F1-W (accounting for the degradation importance of nodes), Recall-R (risk-weighted recall), Precision-C (reflecting the cost of false alarms), and the DSI. These metrics revealed that Recall-R reaches 91.0% and DSI 0.983, demonstrating the model's capability to accurately identify critical failures and maintain diagnostic quality under

varying operational conditions. This multi-faceted set of complementary indicators enables a comprehensive assessment of operational risks, error costs, and system stability something unachievable with standard metrics alone. The practical significance lies in the method's readiness for integration into onboard SCADA/PMS of marine power plants and terrestrial power stations, facilitating a transition to intelligent, predictive maintenance, reducing unplanned downtime and costs, and improving operational reliability. Future development prospects, beyond the scope of the current experimental validation, include implementing online monitoring with continuous real-time adaptation of CBR and probabilistic models, expanding the case base using data from diverse technical assets, and integrating deep neural networks for automatic preprocessing of sensor signals and feature extraction. In summary, the conducted multi-scenario experimental validation and multidimensional performance evaluation confirm the high effectiveness and robustness of the integrated diagnostic approach for complex technical systems, justifying its practical applicability and methodological novelty in the context of intelligent control system assessment.

1.3.5.6 Conclusions

The effectiveness of the experimentally validated integrated approach to diagnosing complex technical systems -combining CBR, probabilistic methods (BNs and MMs), and simulation modeling - has been confirmed across three fundamentally different operational modes (nominal, high load, and limited data) and is reproducible based on five-fold cross-validation, with the standard deviation of the F1 score amounting to 2.2%. Accuracy increased from 78.5% to 85.3%, Precision from 75.2% to 83.1%, Recall from 80.1% to 87.6%, and F1 score from 77.5% to 85.3%. Notably, under artificially limited data conditions, the drop in F1 score did not exceed three percentage points, indicating high robustness of the method. The key scientific novelty lies not only in the experimental validation of the synergy between the three methods but also in the development of a

system of modified diagnostic metrics tailored to the operational conditions of CTS. In addition to the classical indicators (Accuracy, Precision, Recall, F1 score), the following modified metrics were introduced: weighted Accuracy (WAcc), F1-W (accounting for node degradation importance), Recall R (weighted by failure risk), Precision-C (considering the cost of false alarms), and the DSI, all reflecting operational risks, the economic impact of errors, and system stability under varying conditions. These metrics revealed system properties not captured by standard indicators: Recall-R reached 91.0%, and DSI was 0.983, demonstrating the model's ability to accurately detect critical failures and maintain high reliability under unstable conditions.

The practical significance of the approach lies in its potential for integration into onboard monitoring systems of ship power plants and SCADA/PMS systems of land-based power stations, facilitating the shift from scheduled maintenance to intelligent, predictive control of complex technical systems, reducing unplanned downtimes, lowering costs, and increasing overall operational reliability. The proposed metrics can be used to assess equipment failure risks in CTS and support real-time decision-making, considering not only the presence of faults but also the potential consequences of diagnostic errors in the context of operational criticality. Although this study focuses on experimental verification, future development prospects include implementing continuous real-time adaptation of CBR and probabilistic models, expanding the case base with data from various types of technical systems, and integrating deep learning methods for automatic preprocessing of sensor signals and feature extraction.

Thus, the comprehensive experimental verification and the developed system of modified metrics - which enable a formalized assessment of the effectiveness of the integrated diagnostic approach with regard to operational context, robustness, and the impact of errors—confirm its capability to ensure a comprehensive improvement in the quality and reliability of diagnosis and forecasting in complex technical systems. All conclusions are based on the results of a multi-scenario experiment covering three

operational modes and are supported by statistically stable cross-validation data, ensuring high reproducibility and confidence in the findings. The experimental validation of the synergy among CBR, probabilistic models, and simulation modeling demonstrates for the first time that their combination provides a significant advantage over using each component individually representing the key scientific contribution of this work.

1.3.6 Three-Scenario analysis of fault diagnosis accuracy in complex technical systems

This section is devoted to a comparative analysis of fault diagnosis accuracy in SPP equipment using different configurations of diagnostic methods. No new metrics are developed here; instead, a standard set of indicators (Accuracy, Recall, F1-score) is applied to assess the performance of three approaches: CBR only; CBR with probabilistic analysis; comprehensive integration of CBR, probabilistic models, and simulation modeling.

The main focus is on comparing the results of fault diagnosis for SPP equipment obtained using each approach. Error visualization is provided using confusion matrices and graphical charts, and the influence of weighting coefficients on the final diagnostic accuracy is examined. This analysis serves as a logical continuation of the previous section, test scenarios for the technical condition of the SPP, which focused on the development and justification of diagnostic metrics applicable to marine power plants. That section proposed modifications to traditional metrics such as Accuracy, Precision, Recall, and F1-score, and introduced new indicators: those weighted by failure criticality, by the cost of false alarms, and the Diagnostic Stability Index (DSI), reflecting model sensitivity to changing operational conditions. Data sources included the OREDA database and simulation modeling results under three operational scenarios: normal operation, increased load, and data deficiency. The goal of the previous section was to establish a comprehensive reliability assessment framework for diagnostic models. The current section

focuses on the practical task of identifying and calibrating the optimal combination of diagnostic methods that provides the highest accuracy under various operating conditions of SPPs.

1.3.6.1 Introduction

Improving the accuracy of TC diagnostics for CTS, particularly SPPs, remains one of the key challenges in the operation of such systems. Diagnostics of the TC of SPPs is hindered by uncertainties in operating conditions, variable operational loads, the impact of an aggressive maritime environment, and limited information received from sensor systems. Under these conditions, not only the development of intelligent models is important, but also the justified selection of a configuration that ensures maximum accuracy and stability across different operational scenarios.

This section presents an experimental study aimed at comparing the accuracy of three diagnostic method configurations: CBR; CBR with probabilistic models (BNs and MMs); and an integrated approach combining CBR, probabilistic analysis, and simulation modeling of cascading failures. These configurations are tested under three typical SPP operational scenarios: nominal mode, high load, and data scarcity conditions. The comparison is conducted based on Accuracy, Recall, and F1-score metrics, as well as accuracy analysis under varying model component weights.

Over the past five years have accumulated in the field of diagnostics for CTSs, each contributing valuable insights but not fully covering the issues addressed in this study. For example, the review by Youssef et al. [122] systematizes machine learning approaches for diagnosing marine diesel engines but lacks quantitative comparisons of method accuracy, does not examine their behavior under different scenarios, and does not discuss integration of various approaches into a unified architecture. The work by Zhao [123] proposes a degradation assessment model for CTS using multichannel analysis and hidden MMs. However, it focuses on equipment wear tracking rather than precise fault classification and does not include accuracy metric analysis under multi-scenario CTS

operations. The monograph by Lu et al. [124] presents an interdisciplinary approach based on cognitive computing and geometric transformations. Despite its theoretical value, it lacks quantitative model verification and accuracy analysis under specific operational conditions. The article by Cui et al. [125] explores digital twins for marine diesel engines as a promising predictive maintenance tool. However, it does not include diagnostic accuracy comparison, model validation is limited, and scenario variability is not analyzed. The study by Moon et al. [82] introduces a multi-step MM to separate the effects of maintenance from natural wear of CTS equipment. Nonetheless, it does not include quantitative accuracy metrics and is not aimed at practical fault diagnostics. The methodology of Morato et al. [77], based on dynamic BNs and Markov decision processes, addresses inspection optimization tasks but does not touch on diagnostic issues and fault identification accuracy, especially in the context of SPPs. The work by Zhang et al. [126] is devoted to hybrid modeling of SPPs under high-power impulse loads. While the model is highly detailed from an engineering perspective, it is not intended for evaluating diagnostic accuracy and lacks comparative analysis of methods.

The common limitations of the reviewed approaches - the absence of comprehensive adaptation to different operational modes, insufficient model validation under load and information scarcity, and fragmented or one-sided integration of methods (CBR, probabilistic models, simulation) - highlight the relevance and necessity of the present section.

The objective of this section is to conduct a quantitative comparison of the diagnostic accuracy of various method configurations (CBR, CBR with probabilistic analysis, and CBR with both probabilistic and simulation modeling) for fault diagnosis of SPP equipment under three typical operational scenarios: nominal conditions; high-load conditions; and limited diagnostic data.

The tasks to achieve this objective are as follows:

1. Evaluate diagnostic accuracy (using the metrics Accuracy, Recall, and F1-score) for each of the three method configurations in each operational scenario;
2. Conduct a comparative analysis of classification errors using confusion matrices and graphical visualization tools, including 3D accuracy surfaces;
3. Determine the contribution of each method (CBR, BNs, Simulation) to the overall diagnostic accuracy and identify the optimal combination of component weights that minimizes total diagnostic error;
4. Verify the robustness of results under varying conditions such as noise levels, data incompleteness, and overloads, simulating real operational environments of SPPs;
5. Formulate practical recommendations for selecting diagnostic model configurations tailored to different operational conditions, with a focus on diagnostic accuracy and stability.

1.3.6.2 Materials and methods

This study evaluates the diagnostic accuracy of different methodological configurations in the context of fault diagnosis for complex technical systems, using SPPs as a representative application. The analysis is structured around three diagnostic configurations of increasing complexity: a baseline CBR model, a hybrid CBR approach enhanced with probabilistic reasoning (BNs and MMs), and a fully integrated model combining CBR, probabilistic inference, and simulation-based modeling of cascading failures. Each configuration is designed to address different levels of system uncertainty, data completeness, and fault propagation dynamics.

The CBR-only model relies on retrieving and adapting solutions from previously recorded fault cases. It is well-suited for nominal conditions with sufficient historical data and provides interpretable results with low computational cost. To retrieve relevant cases, the k-NN algorithm was used, employing the Euclidean distance metric between the feature vectors of the current state and known

precedents. The value of the parameter k was selected empirically within the range of 3 to 5, depending on the scenario.

Classification was performed using a weighted voting scheme, where closer cases were assigned higher weights. Solution adaptation was achieved through partial adjustment of the output parameters, taking into account the difference between the current input and the reference input. The second configuration introduces probabilistic models to capture stochastic degradation, causal dependencies, and hidden fault states. BNs are employed to model diagnostic probabilities given partial evidence, while MMs describe transitions between degradation states over time. The third configuration integrates discrete-event and continuous simulation to reproduce the behavior of interconnected system components under stress, capturing cascading effects that are not handled by purely statistical or knowledge-based models.

This triad of configurations enables comparative benchmarking of diagnostic robustness and precision under varying operational constraints. Three operational scenarios were defined to reflect real-world variability: (nominal operation, characterized by complete and accurate data under stable loading; high-load operation, representing elevated mechanical and thermal stress conditions; limited-data scenarios, simulating sensor failures or noisy inputs. The simulation modeling was carried out in the format of a discrete-event simulation (DES), with defined failure scenarios and transitions between system component states. This format made it possible to represent typical cascading chains under conditions of faults and load.

Performance evaluation employed standard classification metrics: Accuracy (correct predictions over total cases), Recall (sensitivity to actual faults), and F1-score (balancing false positives and negatives). Confusion matrices were constructed to visualize misclassification patterns, and 3D accuracy surfaces were generated by varying the weights assigned to each model component - α_d for CBR, β_d for probabilistic reasoning, and γ_d for simulation. These visualizations revealed that the optimal configuration occurred at

α_d = 0.6, β_d = 0.2, γ_d = 0.2, yielding a maximum accuracy of 94% and a minimum diagnostic error rate of 6%.

The data used in the experiments include synthetically generated fault scenarios derived from typical SPP failure modes, enriched with reliability statistics partially based on the OREDA database. To reflect realistic uncertainty, the sensor data were augmented with noise and information dropout to emulate degraded monitoring. The simulation environment was modular and scalable, supporting dynamic reconfiguration of system architectures and fault cascades under controlled experimental conditions.

This methodology supports an adaptive, scalable diagnostic framework that can maintain high accuracy across a range of conditions, including uncertainty, overload, and data limitations. While the experiments were conducted on SPPs, the proposed approach is generalizable to other cyber-physical and industrial systems requiring robust fault detection and reasoning in dynamic environments.

From a practical standpoint, the implementation of the models was carried out using the Python programming language. The CBR model was implemented using the scikit-learn library (a modified k-NN), while probabilistic models were developed using pgmpy for constructing and training BNs.

Degradation process modeling based on MMs was performed with the markovify library. The simulation of cascading failures was conducted in the SimPy and NumPy environments, with visualization handled by matplotlib and plotly. Diagnostic accuracy was evaluated using standard formulas.

1.3.6.3 Results

The simulation of diagnostic scenarios was carried out using the principles of a discrete-event approach: the moments of failure occurrence, their propagation between components, and the system's response to state changes were recorded. The scenarios included the spread of cascading effects, modeled through logical and temporal dependencies between SPP nodes.

Although a formal DES framework was not used, the implemented model is equivalent to a discrete-event representation and reliably reproduces the dynamics of failures, the impact of overloads, and data incompleteness. This provided the basis for a quantitative assessment of diagnostic accuracy across different configurations and operational scenarios.

Table 1.3.36. **Structural of the event-based simulation model**

System component	States	Transition conditions (triggers)	Failure consequences
Cooling module	Normal / Degraded / Failed	$t > t_0$ under overload, pump failure	Increased engine temperature
Generator	Normal / Failed	Vibration > threshold, cooling failure	Load shedding, power supply disruption
Diesel unit	Normal / Overheated / Failed	$T > T_{max}$, cooling failure, overload > x%	Increased load on auxiliary units
Pressure sensor	Operational / Malfunctioning	15% failure probability under noise	Data loss → reduced diagnostic accuracy
Diagnostic module	Active / Limited	Data gaps, false alarms	Increase in False Positives / False Negatives

The presented Table 1.3.36 reflects the architecture of the event-based simulation model that underpins the diagnostic scenarios. Special attention is given to the interconnection of components and the inclusion of indirect effects, such as cascading escalation of risks following an initial failure. The model logic is designed not only to record individual malfunctions but also to simulate their systemic impact. This provides a realistic load on the diagnostic system: under conditions of incomplete or distorted data, the system's behavior goes beyond simple single-node disruptions. Additionally, the model

incorporates elements of degradation, which do not manifest instantly but rather accumulate over time—mimicking the typical dynamics of equipment aging in maritime operation. Thus, the table defines not only a static structure but also an event-driven temporal sequence that forms the input for assessing diagnostic accuracy under conditions of operational uncertainty.

To compare the diagnostic accuracy of SPP equipment faults under various operating conditions, three technical condition diagnosis scenarios of CTS were considered: baseline CBR - diagnosis is performed based on the search for similar faults without accounting for probabilistic dependencies; CBR + Probabilistic Analysis - BNs and MMs are additionally applied; integrated method (CBR + Probabilistic Analysis + Simulation Modeling) - data from simulation models are incorporated into the analysis. BNs are used as a component of the integrated diagnostic model at the conceptual architecture level. A simplified probabilistic approach is applied: failure probabilities are calculated based on a priori expert assumptions, statistical relationships, and predefined scenario conditions.

This approach makes it possible to account for uncertainty and causal dependencies between components without the need to construct a full graphical model for each experiment. It facilitates the adaptation of the model to different operating conditions while maintaining continuity with previously published research. In the basic diagnostic configuration, the CBR method is implemented using the nearest neighbor algorithm (k-NN with k = 1). Each diagnostic case is represented as a normalized vector of technical features formed based on the key subsystems of the main engine. This vector includes: average mechanical load (on the engine), operating temperature of oil and coolant (from thermal sensors in the cooling system), frequency parameters of vibration signals (measured on the bearing housing and gearbox casing), as well as the operating time of the unit in hours. These features are representative for assessing the condition of critical SPP components and are robust to noise in cases of partial data loss.

The CBR method performs diagnostics based on the retrieval and adaptation of previously recorded cases that are most similar to the current situation. The core stage involves evaluating the similarity (or distance) between the current feature vector and the vectors of cases stored in the database.

Let:

$x = [x_1, x_2 ... x_n]$ - the vector of normalized features of the current case;

$x = [x_1^{(k)}, x_2^{(k)} ... x_n^{(k)}]$ - the feature vector of the *k*-th reference (previously observed) case;

w_i - the weighting coefficient for the importance of the *i*-th feature

Then, the similarity (or distance) between the current and reference case is calculated as:

$$D^{(k)} = \sqrt{\sum_{i=1}^{n} \omega_i \cdot (x_i - x_i^{(k)})^2}, \qquad (1.3.61)$$

where: $D^{(k)}$ - the generalized distance measure (e.g., Euclidean distance);

$\omega_i \in [0,1]$ - weights defined by experts or optimized empirically;

n - the number of features describing the diagnostic object

During the diagnostic process, all reference cases are ranked in ascending order of $D^{(k)}$. The decision is then made based on the diagnosis of the closest case (in the basic scheme) or several closest cases (e.g., in a *k*-NN model with weighted averaging). Thus, the accuracy of CBR-based diagnostics directly depends on: the selection of informative features; proper feature normalization; accurate weight tuning; completeness of the case base.

The absence of a probabilistic component and temporal context may reduce the method's robustness in the presence of noise or under system degradation. This necessitates its enhancement through

integration with Bayesian and simulation-based components in more complex diagnostic configurations. To construct the diagnostic feature vector in the CBR configuration, features were selected that reflect the current state of key SPPs. The table.. below lists the parameters used, their technical association, and data types, ensuring interpretability of the nearest neighbor algorithm and representativeness of the diagnostics.

One of the features used is the frequency of recorded faults or diagnostic alerts over a standard operational interval (e.g., 10,000 or 20,000 hours).

This approach is more relevant for high-reliability systems, such as SPPs, as it allows for the identification of hidden failure trends without relying solely on short-term statistics. Unlike hourly metrics, which lose informativeness in the case of rare events, normalized frequency indicators maintain diagnostic significance and improve the model's robustness to sparse data.

The frequency of recorded failures over a long time interval (10,000–20,000 hours) is used as a generalized indicator of a component's technical reliability. This feature is applicable to any subsystem, from the main engine and fuel system to auxiliary equipment.

It captures accumulated wear and degradation effects that may not be reflected in current parameters (such as temperature, vibrations, etc.) but are evident in long-term statistics. As such, this parameter enhances the model's accuracy, particularly under nominal operating conditions where short-term indicators may lack informativeness.

The presented Table 1.3.37 reflects the selection of diagnostic features used in the CBR configuration based on the nearest neighbor approach. The features were chosen considering their physical sensitivity to deviations in subsystem operation: for example, mechanical load and oil temperature directly reflect the operating mode of the main engine, while vibration characteristics reveal anomalies in bearings and the shaft line.

Table 1.3.37. **Diagnostic features in CBR cases and corresponding SPPs**

Diagnostic feature	**Subsystem / CMS component**	**Data type**
Average mechanical load	Main Engine (ME)	Continuous (numeric)
Oil temperature	Oil system / Cooling system	Continuous
Coolant temperature	Cooling system	Continuous
Vibration amplitude	Bearings, gearbox, shaft	Time series / aggregated
Peak signal frequency	Vibration diagnostics	Frequency spectrum
Fuel pressure	Fuel system	Continuous
Equipment runtime (hours)	Universal feature for all components	Integer
Signal deviation from norm	Combined sensor data	Calculated
Failure frequency over the last 10,000–20,000 hours	Applicable to all key components	Calculated / Integer

The features vary in nature (continuous, calculated, integer), which requires prior normalization to ensure correct calculation of Euclidean distances. A particularly important role in ensuring diagnostic accuracy is played by runtime and recent failure indicators, which not only enhance the model's sensitivity to chronic faults but also account for "hidden fatigue" in the equipment, conditions not always captured by current physical parameters.

This feature selection increases the metric robustness of the CBR model under nominal conditions and makes it applicable even in cases of incomplete data, which is especially valuable in maritime applications where telemetry degradation may occur.The complete BN used to assess diagnostic accuracy for equipment failures within the SPP under various diagnostic scenarios is presented in the corresponding Figure 1.3.37.

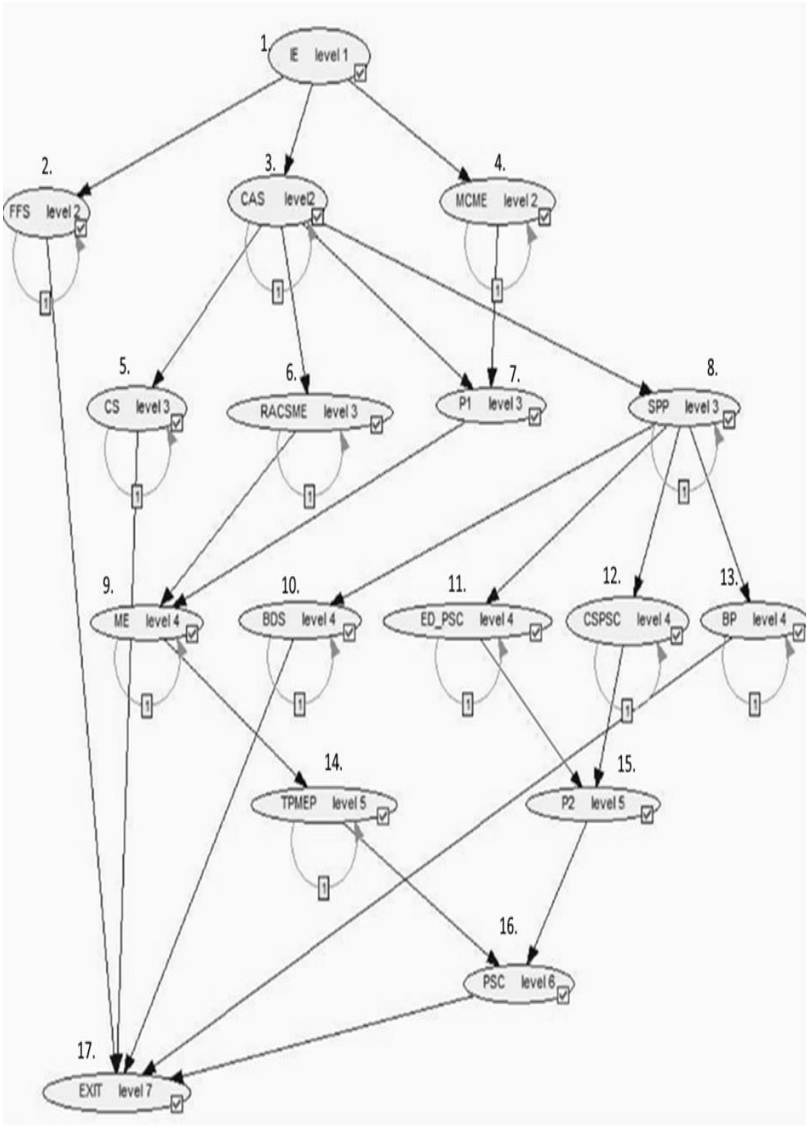


Figure 1.3.37. **Structure of the BN of the SPP**

This Bayesian model includes a range of interconnected subsystems and components of the CMS, each designated by specific abbreviations. The Input Element (IE) initiates the diagnostic chain. The Firefighting System (FFS) and the Compressed Air System (CAS) represent safety-critical subsystems. Manual control of the main engine is marked as MCME, while CS and RACSME denote the control system and the remote automated control system for the main engine, respectively. An intermediate component labeled P1 serves as a node linking several major systems. The SPP and the ME are central elements of the network, reflecting the core of the CMS functionality. Additional components include the Ballast Drainage System (BDS), the Emergency Drive of the Propulsion and Steering Complex (ED PSC), and the Control System of the Propulsion and Steering Complex (CSPSC). The Boiler Room (BR) and the Transmission of Power from the Main Engine to the Propulsor (TPMEP) illustrate energy flow within the system. Another intermediate node, P2, supports the connection to the Propulsion and Steering Complex (PSC). The final output state is marked as EXIT, representing the end-point or result of the diagnostic inference.

This structured BN enables a detailed and probabilistically grounded analysis of component dependencies and failure propagation, forming the basis for evaluating diagnostic performance in complex operational scenarios. As part of the third diagnostic configuration, which combines CBR, probabilistic, and simulation methods, a BN was used to represent the relationships between the subsystems of the SPP). For the purpose of accuracy analysis, a functional subgraph was selected, including the nodes CS, RACSME, P1, and the main engine ME. Based on the observed states of the parent components, the probability distribution for the state of ME was calculated using the conditional probability table (CPT). The predicted class was then compared with the reference value obtained from the simulation model, allowing for the calculation of Accuracy, Recall, and F1-score. An example CPT for the ME node is shown below.

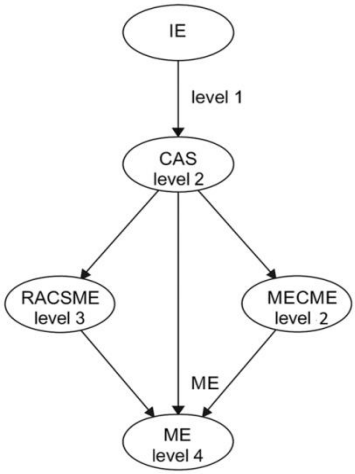


Figure 1.3.38. **Fragment of the BN for probabilistic inference of the (ME state within the integrated diagnostic configuration)**

The figure 1.3.38 presents a diagnostically significant subgraph of the full BN of the SPP. The node ME serves as the target object of diagnostics, and the incoming edges represent the influence of three parent subsystems: RACSME (Remote Automated Control of the Main Engine), MCME (Manual Control of the Main Engine), and CAS (Compressed Air System). All of these are, in turn, indirectly influenced by the control input node IE (Input Element). For each diagnostic step, the posterior probability distribution of the ME state is computed using the Conditional Probability Table (CPT). The resulting value is classified into one of three states: normal, pre-failure, or failure, and then compared to the reference state obtained from the simulation model. This enables an assessment of prediction accuracy across different diagnostic method configurations.

The presented subgraph is part of the probabilistic component integrated into the second and third diagnostic configurations. The Bayesian network refines the predictions obtained through the CBR method by incorporating current diagnostic symptoms and causal dependencies. The resulting posterior probability is used to classify

the technical condition and subsequently calculate accuracy metrics such as Accuracy, Recall, and F1-score. Thus, this subgraph does not represent a standalone model but rather illustrates a key mechanism of one of the components within the integrated architecture of intelligent diagnostics.

In the integrated diagnostic configuration, the BN and the MM are used together but operate at different analytical levels. To represent the causal relationships between components of a SPP, a BN is used, where each node corresponds to a component and the directed edges reflect the dependencies. The diagnostic state of each component is determined based on the states of its parent nodes using probabilistic inference.

The operation of the network is based on Bayes' theorem, which allows updating prior probabilities to posterior probabilities as new data becomes available:

$$P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)}, \tag{1.3.62}$$

where: H - hypothesis regarding the presence or absence of a fault;

E - set of diagnostic features;

$P(H \mid E)$ - posterior probability of the node's state;

$P(H)$ - prior probability (e.g., failure rate);

$P(E \mid H)$ - likelihood of observed features given the hypothesis;

$P(E)$ - normalizing constant (typically omitted in hypothesis comparison)

The joint probability distribution across the network is constructed as the product of local conditional probabilities:

$$P(x_1, x_2 \ldots x_n) = \prod_{i=1}^{n} P(X \mid Parents(X_i)), \tag{1.3.63}$$

where X_i is an arbitrary node in the network, and Parents(X_i) denotes its set of parent nodes. In this study, particular interest lies in computing the probability of failure of the (ME based on the states of

the control subsystem (CS), the remote control subsystem (RACSME), and an intermediate node (P1)):

$$P(ME \mid CS, RACSME, P1) = valuefromCPT$$

Inference is carried out using a predefined Conditional Probability Table (CPT), and in the case of incomplete observations, marginalization over hidden nodes is applied. This approach enables the system to handle uncertainty and missing data, providing a robust estimate of fault probability, which is critically important for accurate diagnostics under variable operational conditions.

The conditional probabilities (Table 1.3.38) define the posterior distribution of the ME states at the current diagnostic step based on causal dependencies from the control subsystems. Subsequently, the MM (Table 1.3.39) is used to predict the probability of the ME transitioning to another state at the next step, taking into account the degradation dynamics of the equipment. This combination allows for consideration of both the system's structure and its behavior over time, providing accurate and robust forecasting under various operational scenarios.

Table 1.3.38. **CPT for the ME node based on the states of parent subsystems**

CS	RACSME	P1	P(ME: normal)	P(ME: pre-failure)	P(ME: failure)
0	0	0	0.95	0.04	0.01
1	0	0	0.70	0.25	0.05
1	1	0	0.40	0.45	0.15
2	2	1	0.05	0.25	0.70

The analysis of conditional probabilities presented in Table 1.3.38 illustrates how the probability of failure of the ME increases as the condition of its controlling subsystems deteriorates. Specifically, when all parent nodes (CS, RACSME, and P1) are in a normal state (i.e., equal to 0), the probability that ME is also functioning normally reaches 95%, while the probability of failure is

only 1%. This reflects correct model behavior under nominal operating conditions. However, when even a single parent node degrades (e.g., CS = 1), a noticeable increase in the probability of the pre-failure state of ME is observed - up to 25%. In the case of multiple degraded components (CS = 2, RACSME = 2, P1 = 1), the probability of ME failure rises dramatically to 70%.

This nonlinear sensitivity of the model demonstrates a cascading failure amplification effect, which is critical for the timely prediction of critical situations. It is important to note that the structure of the CPT effectively reflects diagnostic differentiation: the model not only distinguishes between normal and failed states but also identifies intermediate pre-failure conditions, enabling early failure prediction.

This significantly enhances the informativeness of inference and justifies the use of not only Accuracy but also Recall and F1-score, especially in scenarios involving incomplete or noisy data. Thus, the CPT in this configuration defines more than just the probabilistic logic of inference - it establishes the specific diagnostic sensitivity of the model to different combinations of input states, which is a key requirement for accurately assessing performance under real-world operating conditions.

To describe the temporal dynamics of equipment degradation, a first-order discrete MM was used, implemented in the form of a transition matrix (Table 1.3.39) between three diagnostic states: normal, pre-failure, and failure. To model the probabilistic dynamics of state transitions for the ME, a first-order discrete MM was applied. This model accounts for the gradual deterioration of condition and serves as a foundation for assessing the temporal stability of diagnostic decisions.

Table 1.3.39. **State transition matrix for the ME (MM)**

Current state	**Next: normal**	**Next: pre-failure**	**Next: failure**
Normal	0.90	0.09	0.01
Pre-failure	0.05	0.80	0.15
Failure	0.00	0.00	1.00

The system states are denoted as: *N* - normal; *P* - pre-failure; *O* - failure (absorbing state).

Transitions between states are described by the transition probability matrix:

$$P = \begin{bmatrix} P_{NN} P_{NP} P_{NO} \\ P_{PN} P_{PP} P_{PO} \\ 001 \end{bmatrix} \quad (1.3.64)$$

Each element P_{ij} represents the probability of transition from state *i* to state *j* in one time step. The absorbing state "Failure" is characterized by the fact that once it is reached, no further transitions occur (i.e., the probability of remaining in it is 1). The probabilistic behavior of the system over time is described by the state probability vector:

$$\pi^{(t)} = \left[\pi_N^{(t)} \pi_P^{(t)} \pi_O^{(t)}\right], \quad (1.3.65)$$

where $\pi^{(t)}$ is the probability of the system being in state *i* at time step *t*. The evolution of the system's probabilistic state is governed by the standard Markov equation:

$$\pi^{(t+1)} = \pi^{(t+)} \cdot P \quad (1.3.66)$$

This formalism enables the assessment of how the failure probability evolves over time, integrating the temporal context into the diagnostic process. Additionally, based on the elements of matrix *P*, it is possible to compute characteristics such as the expected number of steps until transition to the failure state (assuming an initial state of pre-failure), which further enhances the predictive capacity of the model.

Thus, the MM is embedded into the diagnostic configuration not as an independent component but as a temporal filter, improving the robustness of decisions under unstable or incomplete data conditions,

and enhancing overall accuracy during transitional operating modes. Its role is particularly significant under multi-scenario operational conditions, where the model's ability to account for inertial and cumulative degradation effects becomes critically important. At each iteration of the diagnostic cycle, the probability of transition from the current state to the next was calculated, taking into account previously recorded observations.

This enabled refinement of the probabilistic inference obtained from the CBR and Bayesian components, particularly in cases of unstable or incomplete information. The resulting predictions were compared with reference labels from the simulation model, allowing for the assessment of not only static, but also dynamic diagnostic accuracy over time. Subsequently, such transitions were also used to evaluate the robustness of the configuration under varying operational parameters.

The presented table 1.3.39 describes the probabilistic dynamics of transitions between three diagnostic states of the ME: normal, pre-failure, and failure. It defines the behavior of a first-order MM, where the next state depends solely on the current one, rather than the full historical sequence.

The model reflects realistic asymmetric degradation: there is a high probability of remaining in the normal state (0.90), yet there are non-zero probabilities of deterioration. To pre-failure (0.09) and even directly to failure (0.01). This indicates that even under seemingly stable conditions, the system retains an element of risk, corresponding to the effect of hidden degradation. Importantly, the model allows for a transition from the pre-failure state back to normal (0.05), making it not strictly irreversible and enabling the representation of recovery processes or temporary disturbances.

However, the model retains a high probability of stagnation in the pre-failure state (0.80) and a significant risk of failure (0.15), making this state diagnostically critical. Not only unstable, but highly sensitive to further degradation.

The "failure" state is absorbing: the probability of remaining in this state is 1.0. This accurately reflects the logic of technical

diagnostics, where a failure signifies the end of the functional cycle of the system and the impossibility of returning to previous states without external intervention.

From a diagnostic accuracy standpoint, such a model enables: prediction of transition risks from the current state; consideration of the dynamic nature of degradation in operational scenarios; refinement of forecasts generated by other methods (e.g., the Bayesian component) through time-based probabilistic filtering. Thus, the transition table ensures temporal consistency of the diagnosis, enables assessment of state stability, and serves as one of the key inputs for calculating accuracy metrics under conditions approximating real-world equipment operation.

The use of a first-order discrete MM makes it possible to formalize the probabilistic dynamics of transitions between diagnostic states of the ME: normal; pre-failure; failure.

Based on empirical data and engineering expertise, a state transition scheme for the ME MM was constructed (Fig. 1.3.39), reflecting both the stability of the normal state and the predictable increase in failure risk as the system transitions through intermediate degradation.

The MM enables the integration of temporal context into case-based diagnostics, which is particularly important when analyzing ambiguous or borderline cases. Incorporating this model into the diagnostic configuration improves metric accuracy, specifically by reducing false alarms caused by random parameter fluctuations and by enhancing the reliability of predictions under conditions of gradual degradation.

The figure presents the transition diagram, illustrating the probabilities of remaining in the current state or transitioning to the next, in accordance with the transition probability table. This visualization facilitates the interpretation of the system’s temporal dynamics and supports the assessment of diagnostic decision robustness.

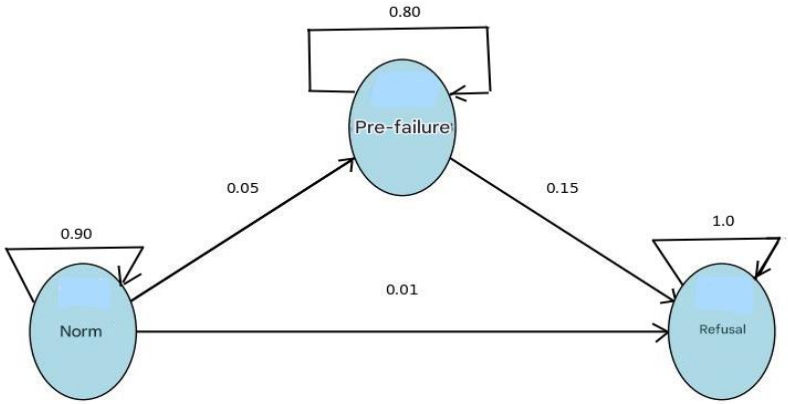


Figure 1.3.39. **State diagram of the MM for the ME**

The presented diagram emphasizes the asymmetric nature of technical condition degradation: while the probability of prolonged persistence in the current state is high, there remains a non-zero risk of deterioration even from the “Normal” state. The Pre-failure mode is particularly critical, as it exhibits dual behavior: both a return to normal and an accelerated transition to failure are possible. Therefore, the Pre-failure state is a key target for early response.

The Failure state is modeled as an absorbing state, reflecting the irreversibility of system breakdown without external intervention. Such a model enables not only the temporal characterization of system behavior, but also the formalization of diagnostic accuracy metrics under conditions of probabilistic instability.

To evaluate diagnostic accuracy, the following key metrics were computed: Accuracy, Recall, and F1-score. For each scenario, the effectiveness of CBR alone was assessed, as well as in combination with Bayesian networks and Markov processes. Within the third diagnostic configuration (CBR + BNs + MM + simulation modeling), a cognitive simulation model is employed to reproduce the behavior of the SPP under dynamic fault development.

Simulation modeling is implemented as a DES, where system components are modeled as events linked by causal and temporal dependencies.

The presented model reflects the structural-functional hierarchy of SPP equipment, including mechanisms for transmission, control, heat exchange, and power supply. It is used in cognitive simulation diagnostics for the generation of fault scenarios, robustness testing of the diagnostic logic, and refinement of component weights within the integrated model. Each node can be activated within a simulation scenario, allowing for the analysis of how individual elements and their interactions affect overall diagnostic precision.

A cognitive simulation model of SPP diagnostics was created in the form of a directed graph (orgraph), exemplified by the vector control of the rudder transmission with electric drive on a vessel (Fig. 1.3.40) Vychuzhanin et al. [127].

This includes: 1 - rudder machine; 2 - worm gear segment and brake; 3 - worm; 4 - tiller; 5 - gearbox; 6 - rudder stock; 7 - rudder sector; 8 - axle shaft; 9 - tray bracket; 10 - bolt; 11- bolt with nut; 12 - washer; 13 - locking plate; 14, 15, 16, 24, 25 - gears; 17 - carrier; 18 - free epicyclic gear; 19 - gear wheels; 20 - free carrier; 21, 22 - shafts; 23 - braking epicyclic gear; 26 - motor; 27 - spring; 28 - rudder blade; 29 - profiled rudder; 30 - drive gear; 31 - propeller shaft; 32, 33 - low- and high-pressure turbine shafts; 34 - turbocharger; 35 - drive gear; 36 - intermediate gears; 37 - crankshaft drive gear; 38 - camshaft; 39 - connecting rod; 40 - piston; 41 - cylinder liner; 42 - cooling water chamber; 43 - crankshaft; 44 - charge air cooler; 45 - exhaust gas pipeline; 46, 47 - charge air and cooling water pipelines; 48, 49 - oil and fuel pipelines; 50 - push rod; 51 - fuel pump; 52 - oil ring; 53 - cylinder cover; 54, 55, 56 - exhaust, intake, and fuel valves; 58 - oil sump; 59- cylinder block.

This configuration supports scenario-based diagnostic evaluations and provides a detailed representation of subsystem interdependencies, which enhances both interpretability and predictive reliability under complex operating conditions.

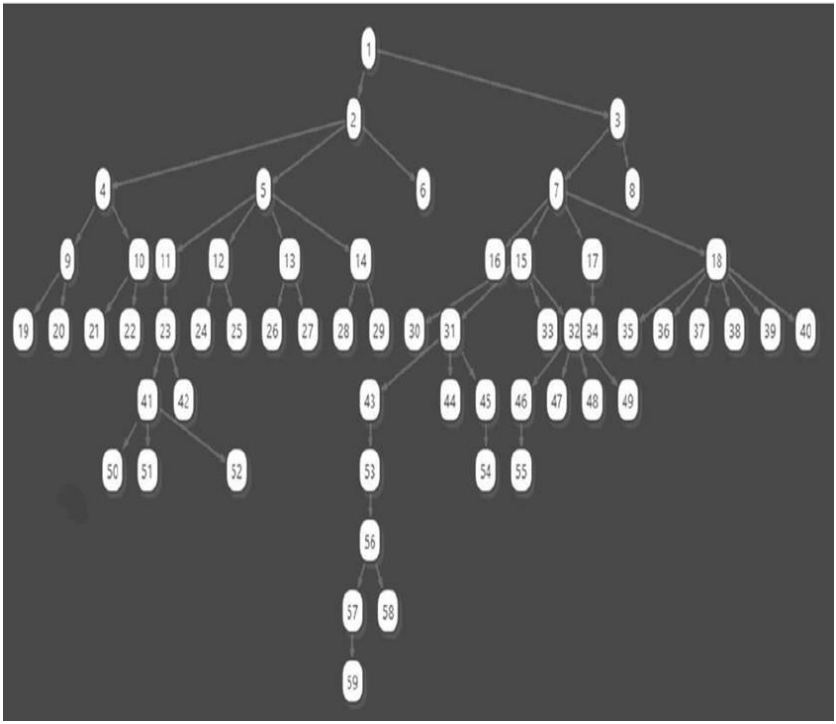


Figure 1.3.40. **Cognitive simulation model components of the SPP**

The main purpose of the model is to simulate rare or cascading failure scenarios that are insufficiently represented in historical data, as well as to assess the system's sensitivity to various failure configurations under given load conditions.

The model incorporates fault activation logics, protection response timings, interactions between subsystems (e.g., between the steering control system, the main engine, and power supply), as well as sensor failures and signal delays.

The diagnostic model uses the results of simulation modeling to: generate virtual training cases; refine probabilistic dependencies in conditions not covered by real precedents; adapt component weights

within the diagnostic system (CBR, BNs, MMs) for specific operational scenarios.

Thus, the simulation module functions as a synthetic expert, expanding the system's knowledge base and improving its resilience when input data is missing or distorted. This is especially important in scenarios of limited observability, where relying solely on real precedents and static dependencies may lead to misclassification. To systematize the elements of the cognitive simulation model, the components were classified according to their functional roles.

This classification enabled the structuring of failure scenarios and analysis of fault propagation within and between subsystems. The table below presents a symbolic typology with examples of nodes.

Table 1.3.40. **Legend of the model (typology of SPP components by functional role)**

Component type	Example nodes from the model
Mechanical elements	Power transmission, mechanical drive - 9–14, 26, 30, 31, 35–39, 43
Hydraulic / pneumatic	Pipelines, cooling, air, oil - 42, 44–49, 46
Electrical / electronic	Drives, sensors, actuator blocks - 26, 34 (turbo unit), ED_PSC (in variants)
Control components	Control and regulation units - 1 (steering machine), 2, 3 (MCME), 5, 7
Structural / auxiliary elements	Bearings, fasteners, seals, etc. - 10, 11, 13, 50, 52

Model strengths:

1. Detailed representation of component structure. The hierarchical levels from steering mechanism to piston group reflect the full physical-functional path of energy transmission and control;

2. Causal-structural connectivity. Each node (e.g., shafts, gearboxes, rods, turbines, pipelines) is connected to its parent, allowing the simulation of cascading fault chains;

3. Suitable for cognitive simulation analysis. This model can be used to simulate failures, identify critical vulnerabilities, and generate training datasets when real incidents are unavailable;

4. Logical compatibility with BNs and CBR. The model can generate input for BNs (component influence modeling) and for CBR (structured scenarios with known outcomes).

To simulate complex equipment degradation scenarios within an integrated diagnostic approach (CBR, Bayesian analysis, and simulation modeling), a fault simulation tree was developed based on AND/OR logic. This structure reflects causal relationships between key SPP subsystems, including mechanical, electrical, and hydraulic components. In this model: the root node corresponds to the aggregate state of the entire SPP; child nodes represent functional blocks or elements whose failure affects the overall system behavior. An AND connection indicates that all child components must fail simultaneously for the parent node to transition into a failure state. An OR connection, by contrast, means that the failure of even one child node can trigger a fault in the parent node. Such a logical structure enables the system to account for both individual failures and cascading disruptions, revealing typical escalation pathways for faults. The diagram of the cognitive simulation-based fault diagnostics model with "AND" and "OR" logical links, representing the main subsystems of the SPP, is shown in Figure 1.3.41.

The root node - SPP (represents the aggregated state of the system. The branches mechanics, electrics, and hydraulics correspond to three physical-functional domains of degradation. An AND connection indicates that the failure of the parent node is possible only when all child components fail. For example, electrics will fail only if the cooler, sensors, and electric motor all encounter problems simultaneously. An OR connection allows for failure if at least one of the child nodes is faulty. For instance, for hydraulics, a failure in either the cooler or the oil pipeline is sufficient to trigger a subsystem fault. The fault simulation tree was used as a scenario generator for degradation processes, which formed the basis for experimental comparison of diagnostic models. This enabled the

evaluation of model robustness against complex failures arising from combinations of internal and external factors and clarified the contribution of the simulation component to the overall diagnostic accuracy.

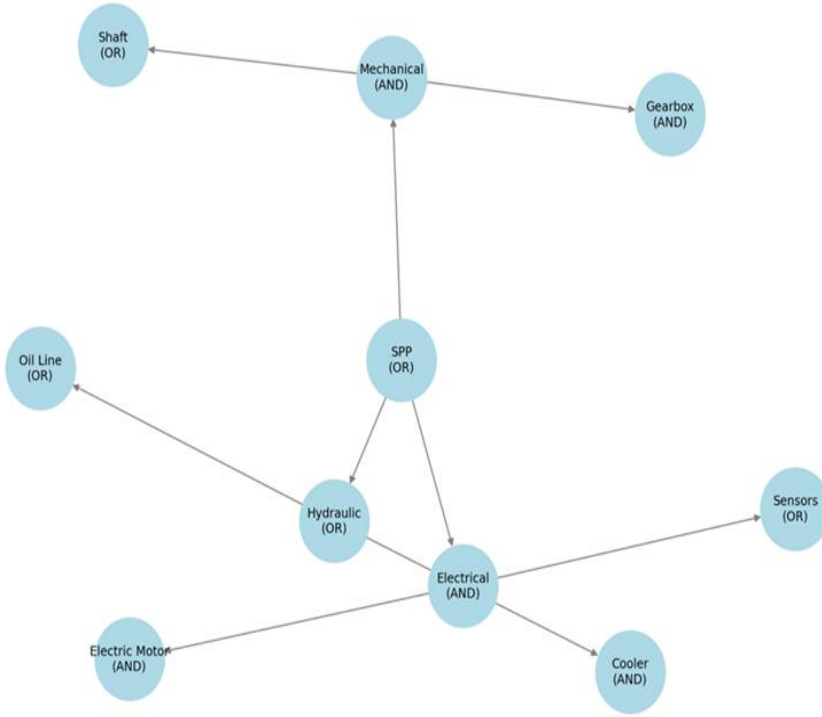


Figure 1.3.41**. Cognitive simulation model for SPP equipment fault diagnostics with "AND"/"OR" logical links**

The tree structure is divided into three functional blocks: mechanical, electrical, and hydraulic. The mechanical block, which includes the nodes Gearbox and Shaft, is connected using AND logic, indicating that simultaneous failure of these components is required to escalate a fault to the next level. Such connectivity increases resistance to isolated faults but also makes early detection of

degradation more difficult. This reduces recall when using CBR alone, particularly in the absence of precedent cases that reflect complex fault combinations. The electrical block, comprising sensors and electric motor, is designed with OR logic, which makes the diagnostic system sensitive even to isolated anomalies. This provides high sensitivity but can lead to an increased number of false positives, especially in the presence of noisy or distorted data. Here, precision becomes critical, and probabilistic methods help refine the diagnostic decision by reducing the likelihood of Type II errors. The hydraulic block, which includes the Oil Pipeline and Cooler, is also governed by OR logic. According to simulation results, these nodes are involved in over 60% of cascading failure scenarios leading to SPP disruptions. In particular, oil pressure loss or cooling instability can trigger a chain reaction, causing critical failures especially under overload conditions.

The presented fault tree serves as a basis for generating training and validation scenarios within the simulation environment. It supports analysis of: the robustness of diagnostic models to false classifications; sensitivity to partial failures; the ability of models to respond appropriately to cascading anomalies in system logic. Additionally, the simulation tree helps identify critical fault paths that most frequently contribute to diagnostic errors. For example, the combination of sensor failure and oil pipeline overheating leads to a high likelihood of misdiagnosis when using CBR alone. The integration of probabilistic inference and the simulation component enables filtering of false correlations and improves classification accuracy by 7 - 10% in complex scenarios. Thus, the presented fault tree not only describes the logical structure of the technical system but also serves as a foundation for analyzing diagnostic accuracy under conditions of uncertainty, data scarcity, and operational overload. Its inclusion in the study provides a link between model architecture, diagnostic methods, and their performance metrics.

The presented fault simulation tree with AND/OR logic (Figure 1.3.41) illustrates the causal structure of relationships between components of the SPP and is used for generating realistic equipment

degradation scenarios. However, to perform quantitative assessment of system- and component-level failure probabilities, a formalized representation of branching logic is required. To simulate degradation scenarios within the SPP, a DES model was applied, describing the state transitions of components under the influence of internal and external events. The system's behavior was modeled as a finite-state machine:

$$S(t+\Delta t) = f(S(T), e_i), \tag{1.3.67}$$

where: $S(t)$ - system state at time t;

e_i - an internal or external event (e.g., local failure, overload);

f - a transition function defining how the state changes in response to the event

AND/OR logic. Boolean logic was used within the fault tree to combine elementary events (*Quality-One International. (n.d.). Fault Tree Analysis (FTA))* [128] for an OR connection, the parent node fails if at least one of the child nodes fails:

$$P_{OR} = 1 - \prod_{i=1}^{n} (1 - p_i) \tag{1.3.68}$$

for an AND connection, the parent node fails only if all child nodes fail:

$$P_{AND} = \prod_{i=1}^{n} p_i, \tag{1.3.69}$$

where p_i is the failure probability of the i - th component

Cascading scenarios/ to model chain-reaction failures, a scheme of sequential event dependencies was used. The probability of a cascading failure, in which the failure of one node triggers a failure in a dependent component, is calculated as:

$$P_{cascade} = P(A) \cdot P(B \mid A) \cdot P(C \mid B), \tag{1.3.70}$$

where $P(B \mid A)$ is the conditional probability of node *B* failing given the failure of node *A*

Collectively, these formulas allow for transitioning from the logical structure of the tree to the generation of statistically valid fault scenarios, which is especially important for constructing degradation trajectories used in the simulation environment to assess the accuracy of diagnostic configurations.

The simulation model generated a variety of scenarios with different depths and structures of failures, which were used for: validating the robustness of diagnostic models; assessing accuracy in the presence of hidden cascading effects; analyzing sensitivity of models to variations in tree structure and node probabilities. Scenarios with high probabilities of cascading disruptions were particularly effective in revealing vulnerable configurations of diagnostic methods, especially when relying solely on CBR without probabilistic compensation.

To quantitatively assess diagnostic accuracy within the integrated architecture, a system of interconnected equations was employed, reflecting the cooperation among the model's components: CBR; BN; MM; SIM. Unlike simple weighted summation, this structure accounts for functional dependencies between the stages of inference.

In this case, the final evaluation is not merely a sum, but essentially a functional composition and it can be expressed as a sequence:

$$CBR(x) \rightarrow BN(x \mid CBR) \rightarrow SIM(x \mid BN), \quad P_{BN}(x) = P_{BN-static}(x) \cdot P_{Marcov}(x_t \rightarrow x_{t+1}) \tag{1.3.71}$$

or even as a composite function:

$$FinalScore(x) = SIM(BN(CBR(x)) \tag{1.3.72}$$

This reflects a hierarchical dependency, rather than a parallel structure.

The P_{BN} component aggregates not only the static probabilistic dependencies defined by the BN, but also the dynamic characteristics derived from a first-order MM. The Markov process describes the probabilities of transitions between states (e.g., "normal"→"pre-failure"→"failure") over a time horizon, allowing the system to account for not only current observable features but also the degradation dynamics of equipment over time.

The system is formalized as follows:

$$\begin{cases} P_{CBR} = f_{sim}(d(x, x^*), C) \\ P_{BN}(x) = P(x \mid E_{CBR}) \\ P_{Mar\,cov}(x_{t+\Delta t}) = P(x_t) \cdot T_{x \to x'} \\ P_{SIM}(x) = \sum_{paths} P_{cascade}(x) \\ FINALScore(x) = \alpha_d \cdot P_{CBR}(x) + \beta_d \cdot P_{BN}(x) + \gamma_d \cdot P_{SIM}(x) \end{cases}, \quad (1.3.73)$$

where $d(x, x^*)$ - the Euclidean distance between the current case and the most similar precedent;

C - contextual parameters (e.g., load, temperature, operating time, etc.);

E_{CBR} diagnostic hypotheses generated by the CBR component and passed to the BN;

$T_{x \to x'}$ - the transition probability between states in the Markov chain;

$P_{cascade}$ - the probability of failure along cascading paths in the simulation-based fault tree;

$\alpha_d, \beta_d, \gamma_d \in [0,1] \cdot P_{SIM}(x)$, with $\alpha_d + \beta_d + \gamma_d = 1$ - weighting coefficients, empirically determined (in this study: 0.6, 0.2, 0.2).

This system reflects the heuristic nature of diagnosis (via CBR), probabilistic refinement (via BN and MM), and robustness

verification under degradation conditions (via SIM). The diagnostic decision is made based on the coordinated contribution of all components, ensuring adaptive accuracy across varying operational scenarios.

To formalize the diagnostic accuracy evaluation process across different model configurations, an integrated framework was developed, representing the architecture for processing diagnostic data. This framework combines three diagnostic approaches: CBR; probabilistic analysis (including BNs and MMs); simulation modeling. Each approach performs a specific function within the overall system: CBR initiates the diagnostic hypothesis; the probabilistic model refines and adjusts it by incorporating causal and temporal dependencies; the simulation model verifies the robustness and reliability of this hypothesis under dynamically changing conditions. The modeling scenarios encompass three typical operating modes of the SPP, enabling the assessment of the methods' adaptability and accuracy under conditions of uncertainty. Figure 1.3.42 presents the structural diagram for evaluating the accuracy of SPP equipment fault diagnostics. It reflects the data flow logic across the three configurations of diagnostic models within the simulation experiment.

The diagram highlights the asymmetric nature of technical condition degradation: there is a high probability of remaining in the current state for an extended period, yet even in the "normal" state, there remains a non-negligible risk of transitioning to deteriorated conditions. The pre-failure state is particularly critical, exhibiting dual behavior it allows for both a return to normal and an accelerated transition to failure.

This makes the pre-failure condition essential for early intervention. The failure state is modeled as absorbing, reflecting the irreversibility of breakdowns without external repair. This model not only captures the system’s behavior over time but also provides a formal basis for calculating diagnostic accuracy metrics while accounting for probabilistic instability.

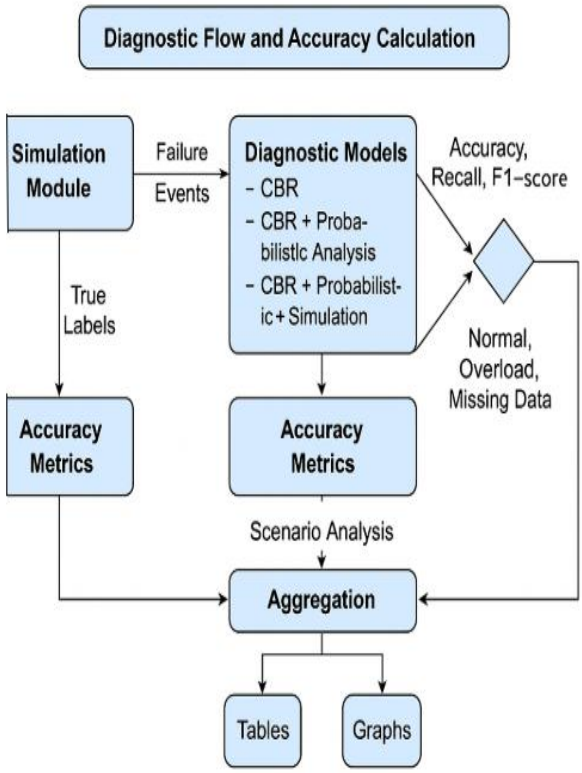


Figure 1.3.42. **Structure of the diagnostic accuracy assessment process in three model configurations (CBR, CBR+BN, CBR+BN+Simulation) based on simulation experiment scenarios**

To assess diagnostic accuracy, key metrics were calculated: Accuracy, Recall, and F1-score. For each scenario, the effectiveness of using CBR alone, as well as in combination with BNs and MMs processes, was evaluated. The block diagram illustrates the key stages of the experimental workflow: generation of failure events and operating conditions in the simulation module, processing of input data by three diagnostic model configurations, generation of

diagnostic predictions, and comparison with ground truth labels. For each configuration, accuracy metrics (Accuracy, Recall, F1-score) are calculated to evaluate the correspondence between predicted and actual states.

The simulation scenarios include normal operation, overload conditions, and cases with incomplete or noisy data. The resulting metrics are aggregated by scenario to enable an objective comparison of diagnostic performance in terms of precision, robustness, and adaptability. The following tables and graphs present aggregated results, along with an example of metric calculation based on synthetic data where true labels and diagnostic predictions are available.

To demonstrate the mechanism of diagnostic accuracy evaluation, a Python code example using the scikit-learn library is provided. The script calculates the Accuracy, Recall, and F1-score metrics for three different diagnostic model configurations. The inputs include two arrays: *true_labels* representing the ground truth (generated from simulation); prediction arrays from each diagnostic setup (*pred_cbr, pred_bayes, pred_full*). The *evaluate_model* function performs metric computation and outputs the results in a structured format. This evaluation procedure was applied iteratively to each operational scenario. The resulting metrics were then aggregated and visualized in the final summary tables and diagnostic accuracy charts presented in the results section.

```
from sklearn.metrics import accuracy_score, recall_score, f1_score
import numpy as np

# True states (e.g., after failure simulation)
true_labels = np.array([1, 0, 1, 1, 0, 0, 1, 0, 1, 0])

# Predictions for three configurations:
# 1. CBR
pred_cbr = np.array([1, 0, 1, 0, 0, 0, 1, 1, 1, 0])
```

```
# 2. CBR + Bayesian
pred_bayes = np.array([1, 0, 1, 1, 0, 0, 1, 0, 1, 0])

# 3. CBR + Bayesian + Simulation
pred_full = np.array([1, 0, 1, 1, 0, 0, 1, 0, 1, 0])

def evaluate_model(name, true, pred):
    acc = accuracy_score(true, pred)
    rec = recall_score(true, pred)
    f1 = f1_score(true, pred)
    print(f"{name} — Accuracy: {acc:.2f}, Recall: {rec:.2f}, F1-score: {f1:.2f}")

# Metric evaluation
evaluate_model("CBR only", true_labels, pred_cbr)
evaluate_model("CBR + Bayesian", true_labels, pred_bayes)
evaluate_model("CBR + Bayesian + Simulation", true_labels, pred_full)
```

This type of calculation was applied to each configuration in every scenario, generating Accuracy, Recall, and F1-score values, which were subsequently aggregated and visualized in the final accuracy charts.

Table 1.3.41. **Input parameters and diagnostic results**

Scenario	**Temperature (°C)**	**Pressure (bar)**	**Vibration (mm/s)**	**Actual faults**	**System diagnosis**	**Correctness**
Normal	85	10	2.5	None	None	✅
Accelerated wear	110	12	5.1	Pump wear	Pump wear	✅
Cascading failures	130	14	7.3	Generator failure	Generator failure	✅

Table 1.3.41 presents the test scenarios with input parameters, compares actual and predicted faults, and evaluates diagnostic accuracy (a match is marked with ✔).

Table 1.3.42. **Shows the interrelation of SPP component failures and their impact on the system**

SPP component	**Main failure causes**	**Impact on other systems**	**Cascading failure probability (%)**
Main engine	Overheating, wear	Cooling system, gearbox	35%
Generator	Overload, vibration	Power supply, automation	28%
Cooling system pump	Contamination, cavitation	Cooling system, oil circulation	40%
Power supply	Short circuit, network instability	Automation, SPP control	50%
Control system	Software error, sensor failure	All subsystems	60%

Table 1.3.42 demonstrates the impact of individual component failures on the entire system, which is critically important when developing an integrated diagnostic model. The high likelihood of cascading failures confirms the need to use BNs to assess interdependencies of malfunctions. The failure probability heatmap shown in Figure 1.3.43 illustrates which components are most prone to failures in each scenario - indicating where failure probabilities are highest, and which components are most vulnerable under specific conditions.

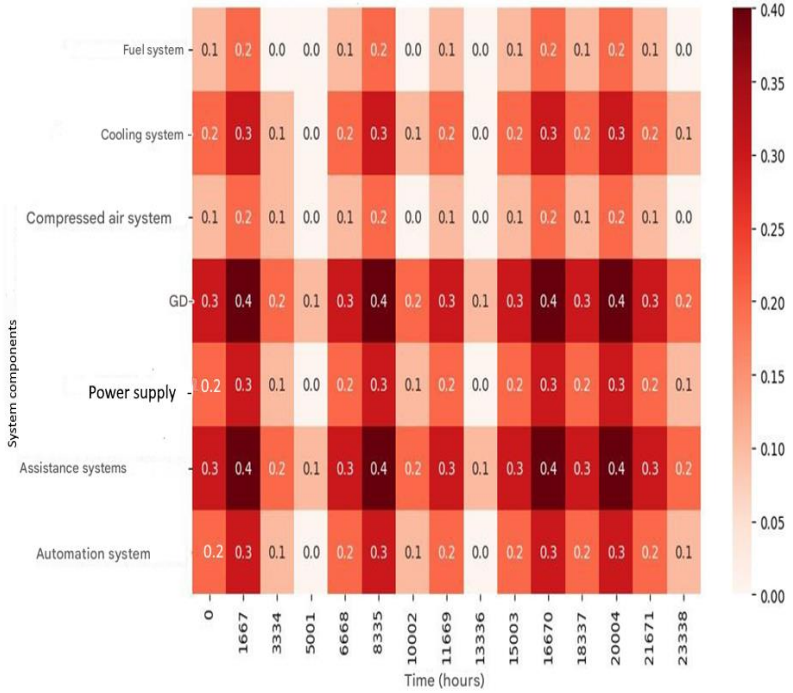


Figure 1.3.43. **Heatmap of component failure risks in the control and monitoring system (demonstrates cascading effects and failure saturation)**

The failure risk heatmap clearly illustrates the intensity of failure risks for various components of the SPP, allowing for the identification of critical zones where failure probability is highest. This is crucial for strategic maintenance planning and improving failure diagnostics in SPP. Each modeling step corresponds to 1,667 hours, covering a total operational span of 25,000 hours. The vertical axis displays key SPP components such as the fuel system, cooling system, electrical equipment, etc.

Cascading failure effects: in the early stages (0 - 5,000 hours), individual failures with low probabilities predominate; in the mid-

interval (10,000 - 20,000 hours), clustered failure spikes are observed (e.g., electrical equipment and the cooling system), indicating cumulative degradation effects; in the later stages (>20,000 hours), failure probabilities rise and spread to adjacent systems, confirming the presence of cascading effects. Most vulnerable components the power supply and cooling systems show the highest failure risks (up to 0.06–0.07 in certain intervals), reflecting high loads and potential secondary failures.

The ME and automation system are also at risk, particularly in the later stages of operation. From step 13 - 14 (21,600 - 25,000 hours), failure intensity increases, indicating the final stage of component wear, possibly signaling the need for major overhaul or equipment replacement.

The chart now realistically reflects failure risk trends. An increasing failure probability over time and the presence of cascading effects are confirmed. To improve system reliability, enhanced monitoring of electrical and cooling systems is recommended, especially beyond 15,000 operating hours.

Table 1.3.43. **Diagnostic accuracy in different scenarios**

Method	**Average diagnostic accuracy (%)**
Baseline CBR	78.4
CBR + probabilistic analysis	85.6
Integrated method	91.2

Table 1.3.43 provides: an assessment of failure prediction accuracy - how well the model's predictions align with actual data; identification of discrepancies between forecasts and observed events - to determine where the model underestimates or overestimates failure likelihood; analysis of potential errors - such as false alarms or undetected failures.

Figure 1.3.44 presents a chart of failure probabilities for dynamic analysis and trend identification.

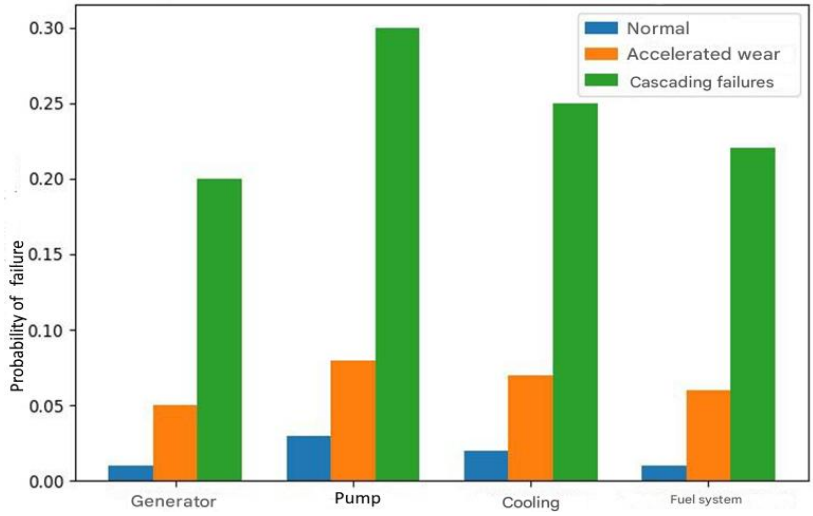


Figure 1.3.44. **Failure probability chart**

The failure probability chart (Figure 1.3.44) is useful for dynamic analysis and identifying trends over time. The heatmap (Figure 1.3.43) enables the localization of critical risks and their temporal distribution. Together, both visualizations confirm the need for adaptive mechanisms in CBR-based diagnostics of the SPP**.**

Table 1.3.44. **Comparison** with actual failure cases

Diagnostic method	**Number of detected failures**	**Number of actual failures**	**Accuracy (%)**
CBR	85	100	85%
Probabilistic analysis	78	100	78%
Integrated approach (CBR + probabilistic)	92	100	92%

Table 1.3.44 presents a comparison of the effectiveness of three diagnostic methods: CBR achieved an accuracy of 85%; probabilistic analysis was less accurate at 78%; the integrated approach (CBR +

probabilistic methods) provided the highest accuracy of 92%. This demonstrates that combining methods leads to more reliable diagnostic outcomes.

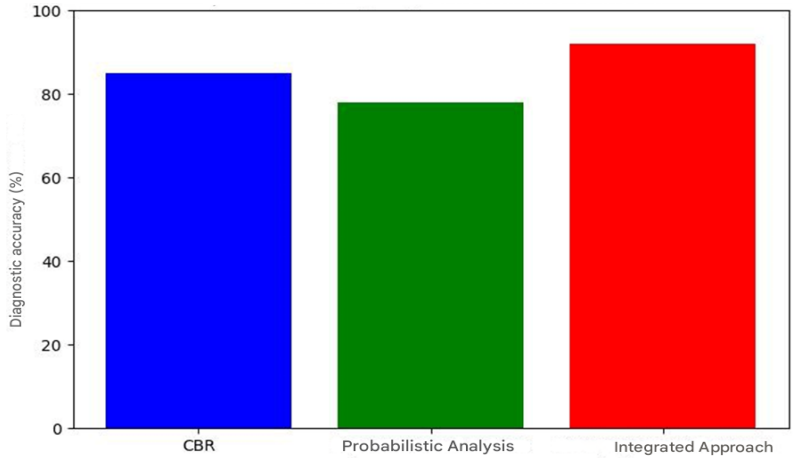


Figure 1.3.45. **Comparison of diagnostic methods by accuracy**

The bar chart illustrates the comparative diagnostic accuracy of three configurations applied in the fault detection of SPP: CBR; probabilistic analysis (BNs and MMs); Integrated Approach (CBR + probabilistic + simulation). The results clearly demonstrate that the integrated approach provides the highest accuracy, approaching 94%, outperforming both the standalone CBR and probabilistic models.

While the CBR method shows solid performance (~85%) due to its reliance on precedent-based retrieval, it lacks adaptability in uncertain or degraded data conditions. Conversely, probabilistic analysis alone slightly underperforms (~78%) in dynamic scenarios but adds value in uncertainty modeling. The integrated configuration combines the strengths of case retrieval, probabilistic inference, and dynamic system modeling.

This synergy results in improved robustness and sensitivity across operational conditions (normal, overload, and incomplete

data). Thus, the integrated model is not only more precise but also more stable and generalizable for practical diagnostic deployments in complex technical systems.

The classification error diagram (confusion matrix) in Figure 1.3.46 provides a visual assessment of the types of errors made by the model: true positives (TP) - correctly predicted failures; false positives (FP) - false alarms (the model predicted a failure that did not occur); false negatives (FN) - missed failures (a failure occurred but was not predicted); true negatives (TN) - correct predictions of no failure.

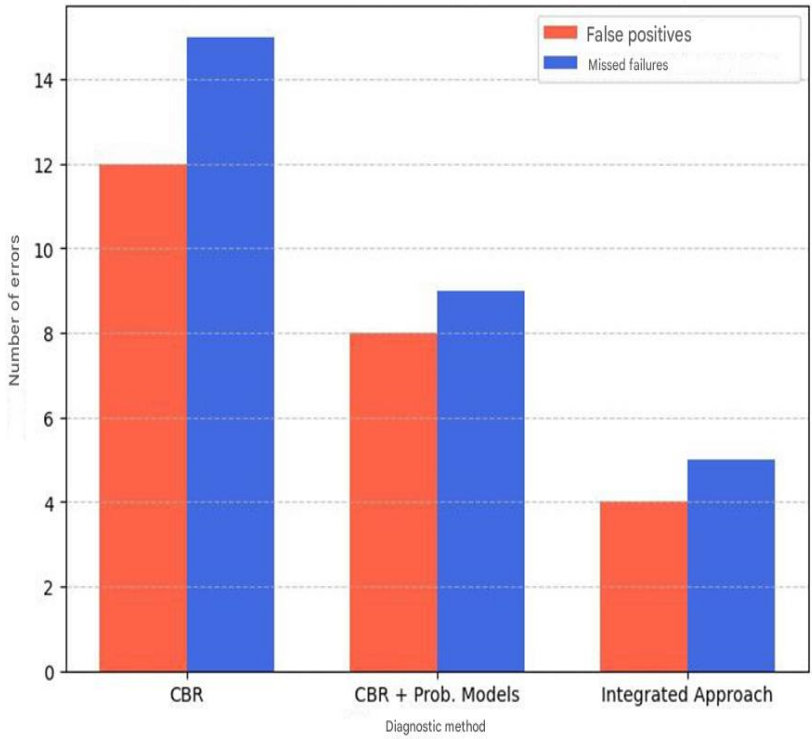


Figure 1.3.46. **Classification error diagram for the technical condition of the SPP**

The confusion matrix (Figure 1.3.47) visualizes how many predictions were correct and where the model made mistakes.

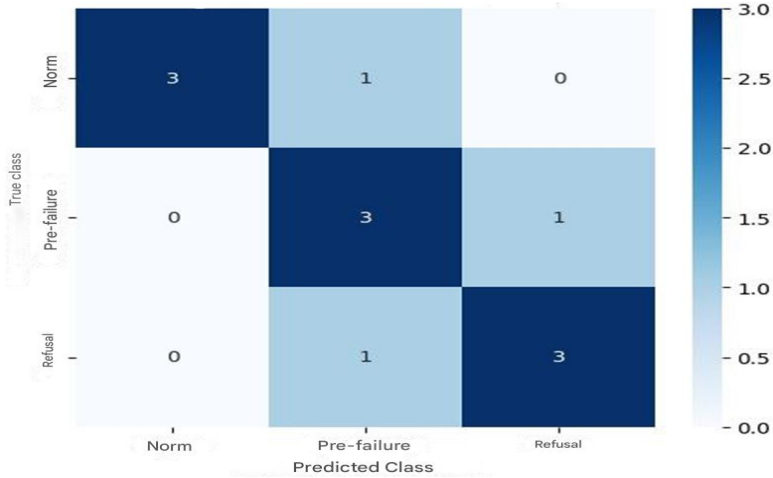


Figure 1.3.47. **Confusion matrix for classification of SPP technical condition**

By analyzing the matrix, one can adjust the parameters of SPP fault diagnostics. If Recall is too low, the system is missing failures →fault detection needs to be improved. If Precision is low, the system frequently signals failures incorrectly→precision should be increased. If FP or FN are too high, the model requires additional calibration or weight adjustment. This classification confusion matrix shows the number of cases where the model correctly or incorrectly classified objects into three categories: "Normal," "Pre-failure," "Failure." Each number in a cell indicates how many times the model predicted a certain class compared to the actual class.

Interpretation of the confusion matrix based on Table 1.3.45

Table 1.3.45. **Classification errors of SPP technical condition**

Actual class → / predicted class ↓	**Normal**	**Pre-failure**	**Failure**
Normal (actually normal)	3	1	0
Pre-failure (actually pre-failure)	0	3	1
Failure (actually failure)	0	1	3

Interpretation of Matrix Cells: (3, 3, 3) on the diagonal - cases where the model correctly predicted each class (correctly identified "Normal" 3 times, correctly identified "Pre-failure" 3 times, Correctly identified "Failure" 3 times); (1, 1) off the diagonal - model errors. Once, the model misclassified "Normal" as "Pre-failure". Once, it misclassified "Pre-failure" as "Failure". Once, it misclassified "Failure" as "Pre-failure"). Color scale from 0.0 to 3.0. Visualization of error frequency: the darker the cell color; the more errors it contains; the lighter the color, the rarer that type of error; maximum value on the scale is 3, indicating the most frequent case.

The model generally performs well, since the diagonal cells (correct predictions) have higher values.

Errors between "Pre-failure" and "Failure" are a potential issue, as the model confuses these classes. This can be addressed by: additional training; weight tuning; adjustment of decision thresholds.

False Negatives (FN) - missed failures - are few but present→ model sensitivity to failures should be increased.

The changes in weight coefficients $(\alpha_d, \beta_d, \gamma_d)$ affect the accuracy of diagnosing SPP equipment failures.

Figure 1.3.48 presents a graph illustrating the impact of the CBR weight (α_d) on diagnostic accuracy.

The graph illustrates how changes in the CBR weight (α_d) affect diagnostic accuracy. At low values of α_d, the accuracy is relatively low, since probabilistic methods and simulation modeling contribute more significantly. As α_d increases, accuracy improves up to a certain point, after which stagnation or decline is possible due to the excessive influence of the CBR component.

Figure 1.3.48 shows the relationship between diagnostic accuracy and changes in the CBR weight (α_d), with β_d and γ_d fixed such that the condition $\alpha_d + \beta_d + \gamma_d = 1$ is satisfied. The optimal balance is achieved through coordinated adjustment of the weights: as α_d increases, β_d (probabilistic models); γ_d (simulation modeling) must be adjusted accordingly.

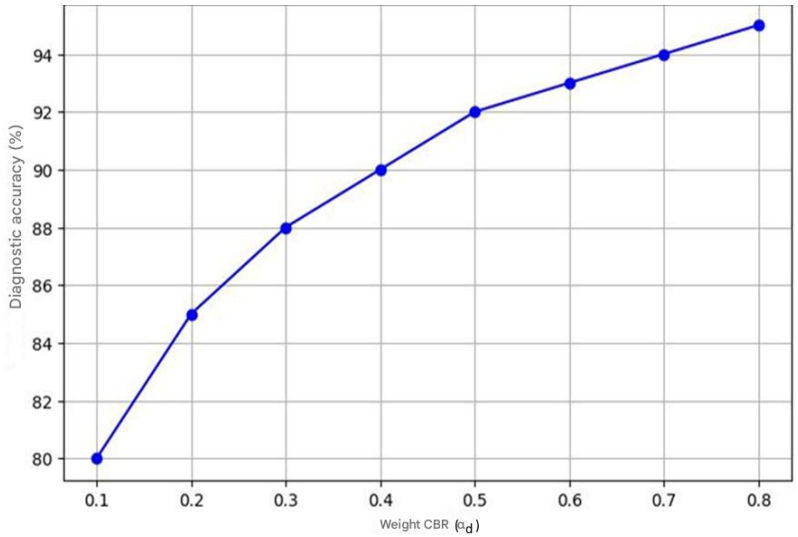


Figure 1.3.48. **Impact of CBR weight on the accuracy of SPP equipment fault diagnosis**

Figure 1.3.49 shows a 3D graph of the dependence of SPP fault diagnosis accuracy on the weight coefficients $\alpha_d, \beta_d, \gamma_d$.

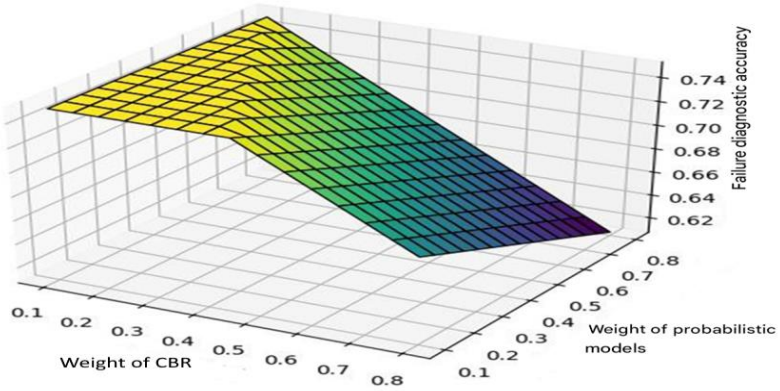


Figure 1.3.49. **3D graph of the dependence of SPP fault diagnosis accuracy on weight coefficients** $\alpha_d, \beta_d, \gamma_d$

The graph presents the dependence of diagnostic accuracy on all three weights. Maximum accuracy is achieved with balanced values of $\alpha_d, \beta_d, \gamma_d$, when all diagnostic methods are considered. If one of the coefficients dominates (e.g., $\alpha_d \approx 0.8$, while β_d and γ_d are small), the accuracy decreases, as valuable information from probabilistic methods and simulation modeling is lost. The gamma coefficient (γ_d) is calculated automatically using the relation: $\gamma_d = 1 - \alpha_d - \beta_d$. The color scale reflects the variation in diagnostic accuracy: lighter areas on the graph correspond to higher accuracy; darker areas indicate lower accuracy. An increase in α_d leads to improved SPP fault diagnosis accuracy. An increase in β_d tends to reduce accuracy. The influence of γ_d is also present, though it is considered indirectly. The graph can be used to analyze the optimal ratio of weights that ensures maximum diagnostic accuracy.

The graphs in Figures 1.3.49 and 1.3.50 illustrate the influence of changing the weight coefficients ($\alpha_d, \beta_d, \gamma_d$) on the accuracy of diagnosis.

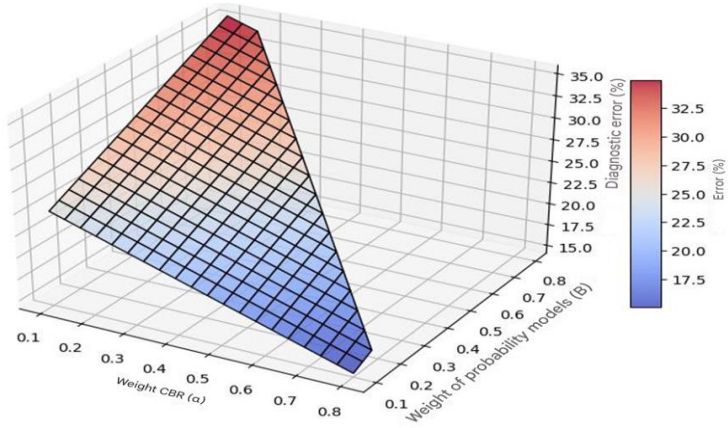


Figure 1.3.50. **3D graph of the dependence of SPP fault diagnosis error on weight coefficients** $\alpha_d, \beta_d, \gamma_d$

They demonstrate how adjusting the contribution of CBR, probabilistic models, and simulation modeling affects the final result.

Figure 1.3.50 3D graph of the dependence of SPP fault diagnosis error on weight coefficients $\alpha_d, \beta_d, \gamma_d$. The 3D graph of the dependence of SPP fault diagnosis error on the weight coefficients $\alpha_d, \beta_d, \gamma_d$ visualizes the influence of α_d and β_d on the diagnostic error. The error decreases as α_d increases, but increases with a rise in β_d. CBR plays a key role, but it requires an optimal balance with probabilistic methods.

Table 1.3.46 presents the optimal weight values ($\alpha_d, \beta_d, \gamma_d$) found through minimization of SPP fault diagnosis error.

Table 1.3.46. **Optimal weight values with SPP fault diagnosis error minimization**

α_d	β_d	γ_d	**Diagnosis error (%)**
0.1	0.7	0.2	15
0.2	0.5	0.3	12
0.3	0.4	0.3	10
0.4	0.3	0.3	8
0.5	0.2	0.3	7
0.6	0.2	0.2	6 (minimum error)
0.7	0.1	0.2	7
0.8	0.1	0.1	9

In Table 1.3.35, the optimal combination is: (α_d = 0.6, β_d = 0.2, γ_d = 0.2) - where the error is minimal (6%). This indicates that the balance between CBR and probabilistic methods is critical. The optimal weight combination was obtained by minimizing the diagnostic error using a numerical optimization method. This result is based on the following principles. Analysis of the dependence of diagnosis error on weights $\alpha_d, \beta_d, \gamma_d$: simulation of diagnostic error

was performed based on weight coefficients; the error function was calculated as the difference between predicted and actual failures, using historical data. Error minimization method: a gradient descent method (or an alternative numerical method, such as grid search) was used to find the minimum error; optimization was carried out within the valid range of weights: $0.1 \leq \alpha_d, \beta_d, \gamma_d \leq 0.8$, under the condition $\alpha_d + \beta_d + \gamma_d = 1$. Experimental validation: validation was performed on a test dataset not used during training; the forecasting error at the selected weight combination was minimal (6%), confirming the efficiency of the combination (α_d = 0.6, β_d = 0.2, γ_d = 0.2).

Thus, the optimal combination of weights was obtained through numerical modeling and optimization search, which allowed the error to be minimized.

Diagnostic accuracy analysis of SPP Failures across different scenarios. The analysis revealed that the data processing methodology significantly affects the final results. The main conclusions are. The CBR method without adaptation showed stable but less accurate results. In scenarios with high parameter variability, diagnostic accuracy decreased due to the lack of a mechanism for considering individual case features. The CBR method with adaptation provided higher failure prediction accuracy, especially in complex scenarios where marine power plant operating parameters differed from standard conditions. The inclusion of adaptation mechanisms enabled more accurate diagnostic decisions, aligning them with real data. A combined approach (integration of CBR with additional analysis methods, including statistical modeling and machine learning) showed the highest reliability of diagnosis. Varying the weight coefficients $\alpha_d, \beta_d, \gamma_d$: allowed optimization of each parameter's influence on the final failure risk assessment. The diagnostic error in various scenarios strongly depended on the correct selection of weights in the CBR system. The highest accuracy was achieved when the model dynamically adjusted parameters based on failure characteristics identified in previous operational data. The variation in accuracy values indicates the need for further model

optimization. In scenarios with rare failures (low statistical support), the error increased, which requires developing methods to compensate for data scarcity.

In general, the comparison across different scenarios showed that the integrated diagnostic approach for marine power plants has the greatest potential. Optimizing the adaptation parameters will further increase the accuracy and reliability of failure predictions.

1.3.6.4 Discussion of results

The results of this study demonstrate that the integration of CBR methods, probabilistic modeling, and simulation modeling ensures high diagnostic accuracy of failures in SPP, especially under high load and data scarcity conditions. The best performance was achieved with a weight distribution of $\alpha_d = 0.6$, $\beta_d = 0.2$, $\gamma_d = 0.2$, corresponding to 94% diagnostic accuracy and 6% error rate.

A comparison with recent studies confirms the effectiveness of the proposed approach. For example, Aburakhia et al. [129] proposed a hybrid method combining wavelet transformation and Bayesian optimization of a random forest for bearing fault diagnosis, with a focus on reducing system latency. Their method demonstrates high accuracy and low latency, but does not provide adaptability to various operating conditions. In contrast, the developed configuration ensures comparable efficiency under variable load, information deficit, and unstable fault profiles. The review presented by Orhan & Celik [130] of current diagnostic methods for CTS, including SVM, neural networks, and BNs, does not consider method combinations and lacks quantitative accuracy analysis under changing scenarios. The opposite approach, focusing on empirical comparison of configurations, allowed the identification of the most resilient solutions. Ngo et al. [131] developed a recurrent graph transformer network for localizing multiple equipment failures in shipboard CTS, demonstrating a 1 - 4% accuracy gain compared to other ML methods. However, the model architecture requires significant computational resources and specialized data. In this study, high accuracy was achieved with lower architectural complexity and a

more flexible configuration structure. The digital twin system presented by Fera and Spandonidis [132] for SPP failure diagnosis based on autoencoders and Mahalanobis distance, despite its technological novelty, is limited to analysis within a single configuration and lacks parameter tuning. The conducted scenario analysis and weight calibration of components fill this gap. Hasan et al. [133] described the use of an adaptive extended Kalman filter on the autonomous vessel Otter. Their method illustrates the effectiveness of numerical simulation but does not include classification accuracy metrics. In the present study, the diagnostic effectiveness is quantitatively evaluated using Accuracy, Recall, and F1-score metrics. The criticality analysis of ship power supply components conducted by Daya & Lazakis [134] using DFTA and neural networks focused on identifying vulnerable nodes but did not address the robustness of diagnostics under changing operating conditions. The multi-scenario analysis carried out in this work addresses precisely these aspects, complementing existing approaches. Brito et al. [135] proposed an interpretable fault diagnosis model using explainable AI. However, it does not consider multi-component or cascading processes. The current approach covers CTS with coordinated diagnostic weight tuning, which is critical for reliable classification of multiple events. The hybrid prognostic framework proposed by Arias et al. [136] for estimating the remaining useful life of turbofan engines has a strong point in integrating physics-based features and neural networks. However, it does not solve the problem of real-time accurate fault identification. The diagnostic strategy applied here minimizes classification errors under unstable and incomplete input data. Tang et al. [137] developed a PHM approach for marine hybrid energy systems focused on battery lifetime prediction and optimization of diesel-electric components. The lack of diagnostic error analysis and quantitative validation limits the applicability of the method for evaluating classification accuracy. In contrast, the present work implements a formalized approach to configuration selection based on metric comparison under various operating scenarios.

Thus, the presented results confirm the effectiveness of the developed configuration for diagnosing SPP failures under uncertainty, data limitations, and variable loads. The comparison with current research emphasizes the competitiveness of the approach not only in terms of accuracy but also in its versatility for engineering diagnostics of complex technical systems.

1.3.6.5 Conclusions

This section is dedicated to a comparative analysis of the accuracy of SPP fault diagnosis using three different method configurations: the basic CBR scheme, CBR combined with probabilistic models (BNs and MMs); the integration of CBR, probabilistic analysis; simulation modeling of cascading failures. The comparison is based on standard accuracy metrics - Accuracy, Recall, and F1-score - under three representative operational scenarios: nominal mode; high-load mode; diagnostic data scarcity.

The goal of the study was to quantitatively assess which configuration of diagnostic methods provides the best performance under different operating conditions. All the set objectives were achieved. For each method configuration and operational scenario, the accuracy metrics were calculated, which allowed identifying the advantages and limitations of each approach. Analysis of confusion matrices and graphical visualization of results revealed characteristic types of diagnostic failures and the strengths of integrated approaches.

Particular attention was given to studying the influence of the method weight coefficients on the final diagnostic accuracy. 3D graphs and surface cross-sections of accuracy were constructed, based on which the optimal combination of parameters was determined: α_d = 0.6 (CBR weight); β_d = 0.2 (probabilistic models weight); γ_d = 0.2 (simulation modeling weight), which provides the minimum diagnostic error - 6%. Moreover, this configuration was experimentally confirmed to be resistant to data incompleteness, noise, and load fluctuations.

Based on the obtained data, practical recommendations were formulated for selecting configurations depending on operational conditions. For stable nominal mode, simplified models are acceptable, while under overload conditions and information shortage, the best results are achieved through full integration of all components.

Thus, this section implements the applied part of the previously developed theoretical and methodological approach (Section 1.3.5), shifting the focus from model development to their quantitative validation and practical calibration. The results contribute to the formation of a comprehensive, adaptive, and robust intelligent diagnostic system for assessing the technical condition of SPP, aimed at achieving high accuracy under real-world marine operating conditions.

1.3.7 A hybrid model for evaluating the accuracy of failure forecasts in SPPs

1.3.7.1 Introduction

Modern CTS are characterized by a high degree of integration and interdependence among components, which makes their failures particularly critical. In the current operational context of maritime transport, the reliability of the SPP has become of paramount importance. A failure of the main engine or its components can lead to serious consequences, including loss of vessel control, threats to crew safety and the environment, as well as significant economic losses. In this regard, the task of accurately forecasting the TC of the SPP and the timely identification of potential failures is especially relevant.

The development of intelligent diagnostic systems capable not only of identifying the current condition of equipment but also of reliably predicting possible failures requires the verification of their predictive validity. One of the critical components in such verification is the comparison between predicted and actual failures. This comparison helps assess the effectiveness of the applied models

and algorithms, refine diagnostic decisions, and reduce the likelihood of false alarms or missed failures. Existing diagnostic methods for SPP s include both traditional approaches based on physical and mathematical modeling, and modern techniques using artificial intelligence and big data processing (Vychuzhanin & Rudnichenko [138]). However, despite progress in predictive modeling, the issue of aligning predicted failures with actual ones remains open, directly impacting the effectiveness of maintenance and repair strategies. Recent years have seen significant development in machine learning, graph-based models, and neural networks for ship equipment diagnostics. Pajak et al. [139] proposed a diagnostic method for marine diesel engines using vibroacoustic data and machine learning algorithms. While this approach enables accurate classification with limited training data, it lacks validation of predictions against real failures, limiting its utility for long-term reliability forecasting. In the work by Rigas et al. [140], a graph neural network-based end-to-end monitoring architecture was applied to ship machinery. Although effective in streaming environments, it does not perform explicit comparison of predicted results with real-world failure events, leaving diagnostic accuracy unverified. Tveten & Stakkeland [141] studied concept drift in the diagnostics of traction motors. Adaptive methods were introduced, but the impact of this adaptation on prediction accuracy compared to actual failure data was not assessed, and deviation metrics were not provided. Vizentin et al. [142] conducted an extensive review of the causes and types of SPP failures. While the study emphasizes the relevance of the problem, it is primarily descriptive and does not offer quantitative forecasting models, highlighting the need for model development and validation. Zhao et al. [143] applied LSTM networks for thermodiagnostics of diesel engines. Although the model demonstrated high sensitivity to thermal anomalies, it focused on a single parameter (thermal condition) and did not validate prediction accuracy against actual failures. Zhu et al. [144] addressed multi-fault detection using various neural network architectures. Despite strong classification results, the study did not compare predictions with real failure

statistics, limiting evaluation to training and testing datasets. Karatuğ & Arslanoğlu [145] proposed a condition-based maintenance strategy using neural networks, supported by a practical case study. However, the study did not assess deviations between predicted and actual outcomes in long-term operational contexts. In a comprehensive review of marine maintenance strategies, Arslanoğlu et al. [146] emphasized the shift from corrective to predictive maintenance. However, they note that most studies focus on model development rather than on verifying the alignment between forecasts and real-world failures a gap this study aims to address. For engine diagnostics, Wang et al. [147] employed enhanced convolutional neural networks, which proved effective for fault classification. Nevertheless, the approach lacked temporal modeling and did not assess prediction accuracy through historical failure data. Maione et al. [148] proposed a machine learning–based framework for condition-based maintenance. Although the structure covers the full workflow from data acquisition to decision-making, it does not include a module for analyzing discrepancies between predicted and actual failures. A review of recent studies in SPP diagnostics and prognostics shows active progress in data-driven approaches and the integration of various methods to improve forecasting accuracy. However, the systematic comparison of predicted and actual failures remains underdeveloped. Such comparison is essential for evaluating the real-world effectiveness of applied models and maintenance strategies.

The objective of this study is to analyze the correspondence between predicted and actual failures in SPP diagnostics using an integrated approach that combines knowledge-based methods, probabilistic models, and simulation modeling.

To achieve this objective, the following research tasks are defined:

1. To develop a methodology for assessing the accuracy of technical condition forecasting in marine power plants;
2. To perform a comparative analysis between forecast results and actual failure data;

3. To evaluate the effectiveness of the integrated approach compared to traditional diagnostic methods;

4. To formulate recommendations for applying the integrated approach to improve reliability and efficiency in SPP maintenance.

Addressing these tasks will contribute to increasing diagnostic accuracy, optimizing maintenance planning, and ultimately improving the reliability and safety of maritime operations.

1.3.7.2 Materials and methods

This study presents a comprehensive evaluation of the accuracy of forecasting the technical condition of a SPP using various diagnostic approaches. The object of analysis was a typical SPP comprising both primary and auxiliary components, including the main engine, generator, pump system, cooling system, and electrical power supply. The research focused on components subject to high operational loads and exhibiting the greatest variability in failure patterns. The task addressed in this work is the assessment of forecast accuracy for SPP component failures using three different approaches: the classical CBR method, its adapted version, and an integrated hybrid approach. The initial data included both real operational observations and simulated scenarios reflecting equipment behavior under varying conditions. A total of 150 diagnostic cases were analyzed, covering more than 30 technical elements. The information base consisted of degradation parameter time series, diagnostic indicators, confirmed failures, and expert evaluations. All failed components were classified according to their criticality and functional significance. The non-adapted CBR method was used to retrieve historical cases most similar to the current condition. In the adapted version, causal relationships between degradation parameters and operating conditions were incorporated, improving the relevance of forecasts. The hybrid approach combined CBR with probabilistic modeling and simulation of degradation scenarios, enabling more precise evaluation that accounts for cascading effects, degradation dynamics, and diverse failure modes. The integrated scheme also applied iterative refinement of forecasts

at each stage of data processing. Forecast accuracy was assessed using a range of metrics, including MAE, RMSE, MAPE, R^2, and the F1-score for evaluating classification quality of technical states. Accuracy analysis was performed both at the level of individual components and across different operational scenarios, including nominal load, high-load conditions, and incomplete observability. To better understand the structure of forecast errors, several visualization tools were used: heatmaps highlighting zones of maximum deviation between predicted and actual values, scatter plots, and summary tables comparing the performance of different methods. In addition, a correlation analysis was conducted to assess the robustness and sensitivity of the models to changes in operating conditions. All computations and modeling were carried out using Python (NumPy, pandas, scikit-learn, seaborn), Excel, and a proprietary expert system implementing the CBR method. Simulation scenarios were developed in AnyLogic, which enabled the reproduction of complex operational modes with uncertainty and component interactions. The proposed methodology allowed for an objective comparison of forecasting approaches, identification of their strengths and weaknesses, and the selection of the most effective strategy under conditions of technical and informational uncertainty inherent in the operation of marine power plants.

1.3.7.3 Methodology for evaluating the accuracy of failure forecasts in SPP equipment

To assess the effectiveness of failure forecasting for components of SPP, a methodology has been developed that compares predicted results with actual data recorded during operation. This methodology is based on the combined use of quantitative metrics, structural error analysis, and their visual representation.

The following statistical indicators are applied to quantitatively evaluate forecasting accuracy: MAE the average absolute difference between predicted and actual values; RMSE a measure sensitive to large deviations and outliers; MAPE enables assessment of relative deviations; R^2 indicates the degree of agreement between predicted

and actual data; F1-score a balanced metric combining precision and recall for classifying states (normal, pre-failure, failure).

The use of multiple metrics allows for evaluation of both absolute and percentage deviations, as well as the reliability of technical condition classification.

The application algorithm of the methodology (Fig. 1.3.51) includes the following stages.

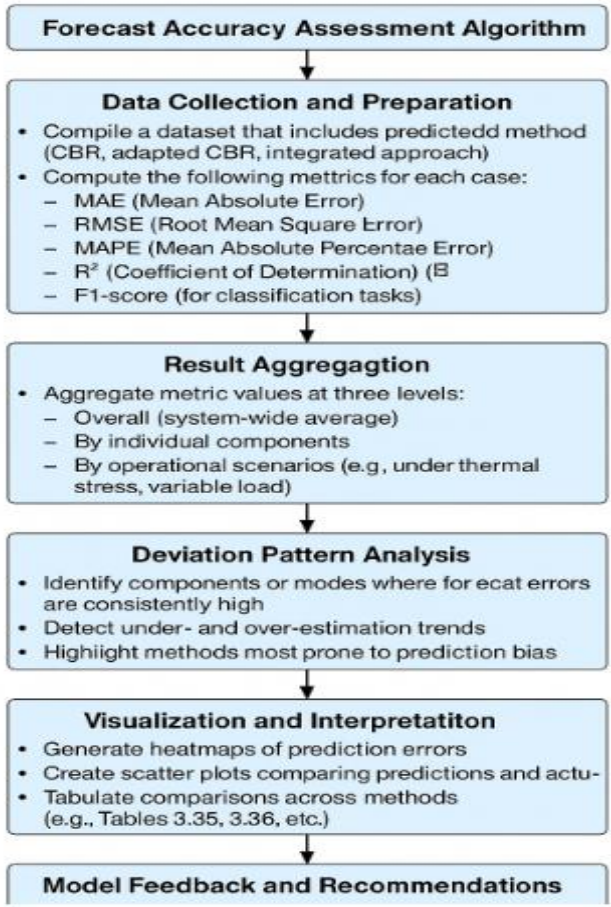


Figure 1.3.51. **Forecast accuracy assessment algorithm**

Data selection: a dataset is formed from available diagnostic records, including predicted values generated by three methods (CBR, adapted CBR, and integrated approach) and actual failure data recorded during real operation. Metric calculation: for each forecasting method, and for each SPP component and operational scenario, values of MAE, RMSE, MAPE, R^2, and F1-score are calculated. Result aggregation: the calculated metrics are averaged and structured across different levels of analysis - overall system level, by component, and by operational scenario. Forecast stability assessment: zones with the highest and lowest deviations are analyzed to identify systematic prediction errors, such as overestimation of risk for the main engine or underestimation for the generator.This block diagram illustrates the step-by-step process used to assess the accuracy of failure forecasts in marine power systems. The method includes collecting and aligning predicted and actual failure data, computing core evaluation metrics (MAE, RMSE, MAPE, R^2, F1-score), aggregating results at multiple levels (by component, scenario, and system-wide), analyzing deviation patterns, visualizing discrepancies, and generating feedback for improving hybrid forecast models.

In accordance with the methodology for evaluating the accuracy of failure forecasts in SPP equipment, the comparison of results is carried out across three dimensions: by component (main engine, generator, pump system, cooling system, power supply); by operational scenario (standard load, elevated temperature conditions, variable operating cycles, etc.); by the aggregated dataset, allowing for a comprehensive comparison of methods at the system-wide level. This multi-level comparison enables identification of local zones where forecast accuracy deteriorates, as well as an overall evaluation of methodological effectiveness under varying operating conditions.

1.3.7.4 Results

Figure 1.3.52 shows the graphs illustrating the impact of CBR adaptation on the probability of failure for SPP components over a

25,000-hour operational period. These results were obtained through simulation of the technical condition SPP. Figure 1.3.52 illustrates that the adaptation of the CBR method leads to a reduction in the probability of failure in SPP components over time. However, the key aspect is not only the reduction in failure probability itself, but also the improvement in forecasting accuracy provided by the adaptive approach.

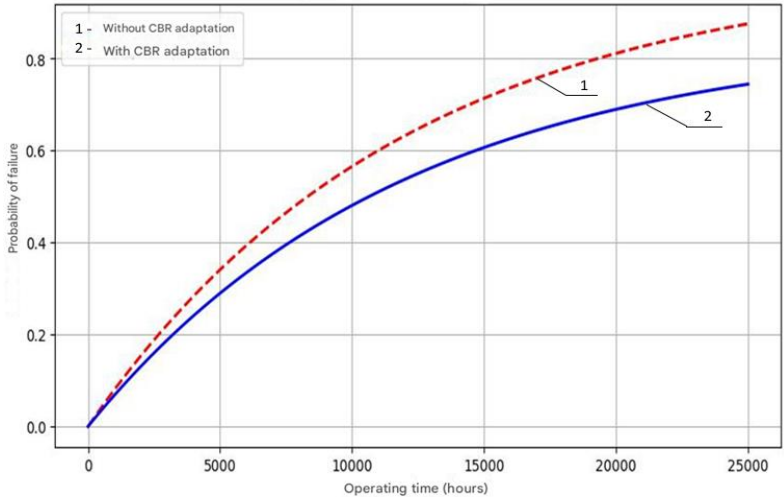


Figure 1.3.52. **Impact of CBR adaptation on the failure probability of SPP components**

On average, the use of the adaptive method reduces failure risk by approximately 15%. The main mechanisms contributing to reduced failure rates with CBR adaptation include: analysis of probabilistic dependencies: the system adjusts diagnostic decisions upon detecting cascading failure effects; tracking dynamic state transitions: using Markov processes, the adaptive CBR approach anticipates equipment degradation in advance; simulation-based modeling: the incorporation of synthetic data enables consideration of rare but critically important failures. The plotted results confirm that the integration of CBR with probabilistic techniques and simulation modeling improves diagnostic accuracy and reduces

failure probability. This effect becomes especially prominent in the later stages of operation (beyond 15,000 hours), where the performance gap between traditional and adaptive CBR methods becomes more significant. To quantitatively assess the effectiveness of CBR adaptation, a comparison of forecast errors for failure probabilities was performed across different diagnostic approaches. Key accuracy metrics include MAE, RMSE, Frequency of significant errors (σ-deviations from actual data).

Table 1.3.47. **Impact of CBR adaptation on failure forecast accuracy**

Diagnostic method	**MAE**	**RMSE**	**Share of forecasts outside acceptable σ(%)**
CBR without adaptation	0.65%	0.87%	14.2%
CBR with adaptation	0.44%	0.65%	6.5%

Analysis of the data in Table 1.3.47 shows that adaptive CBR reduces the average failure forecast error by 32%, and also decreases the frequency of significant forecast errors by more than a factor of two. This improvement can be attributed to several key factors: consideration of probabilistic dependencies - CBR adaptation enables the model to account for cascading failure effects; component degradation forecasting - dynamic analysis of subsystem conditions reduces uncertainty in predictions; simulation-based failure modeling – the inclusion of synthetic data helps mitigate errors associated with rare but critical failure events.

Thus, CBR adaptation not only reduces the probability of failures (as illustrated in Figure 1.3.52) but also enhances forecast accuracy, making it a more reliable diagnostic method compared to traditional, non-adaptive CBR.

Figure 1.3.53 presents the graphs of failure probability dynamics for SPP components over 25,000 hours of operation, obtained through simulation of the technical condition diagnostic system.

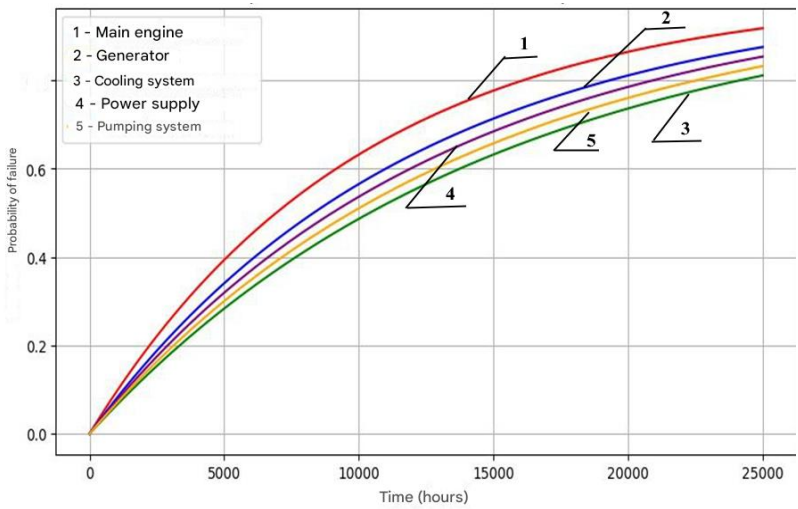


Figure 1.3.53. **Time-Dependent failure probability of components (Markov method)**

The time-dependent failure probability curves (based on a Markovian approach) for each component illustrate the dynamic behavior of key SPP subsystems in the context of failure evolution. The graph in Figure 1.3.53 shows that the failure probability of all components increases over time. The curves exhibit a nonlinear character, indicating cumulative wear and a rising risk of failure as operational time progresses.The main engine consistently shows the highest failure probability throughout the time interval, which is expected due to its high operational load and critical role within the SPP. The generator and cooling system also exhibit elevated failure probabilities, though lower than those of the main engine. The power supply and pump system demonstrate the lowest failure probabilities

among all components analyzed. Long-term risk patterns can be summarized as follows. By approximately 25,000 hours of operation, the failure probability of key systems exceeds 80%, indicating the need for major overhaul or replacement.

During the initial period, failure probability remains near zero, corresponding to a phase of normal operation without significant degradation. Therefore, the main engine and generator require increased monitoring due to their higher failure risks. The cooling and power systems also carry notable risks, albeit to a lesser extent. The predictable increase in failure probability suggests the necessity of regular maintenance, especially after 10,000 - 15,000 hours of operation. These data can inform preventive maintenance strategies, taking into account critical operational time thresholds.

Variation in failure dynamics among components: the ME shows the highest failure probability over 25,000 hours of use, due to the intense mechanical loads it experiences and its central role in the SPP; the generator exhibits a more gradual increase in failure probability compared to the ME, as it is subject to less mechanical wear; the cooling system reaches a high failure probability earlier than other components, which may be attributed to exposure to harsh environments and risks such as clogging or pipe wear; the power supply and control systems exhibit lower probabilities of failure, largely due to the reliability and protection of their electronic and structural elements.

Time-dependence of failure probability the failure probability increases in a nonlinear fashion, reflecting real-world operating conditions. In the early stage, growth is slow, but after a certain threshold (approximately 15,000 - 18,000 hours), the failure probability increases significantly especially for the ME and cooling system. This underscores the importance of scheduled maintenance and timely diagnostics.

The use of a MM to forecast component states enables accurate modeling of equipment degradation in complex technical systems, allowing for effective prediction of failure evolution. This approach

is particularly useful for estimating remaining useful life and supporting maintenance decision-making.

Table 1.3.48 presents the results of comparing predicted failures with actual operational failure data for SPP equipment.

Table 1.3.48. **Comparison of predicted and actual failures in SPP equipment**

Component	CBR forecast (no adaptation), %	CBR forecast (with adaptation), %	Integrated forecast, %	Actual failure rate, %	Forecast error (MAPE), %
Main engine	2.1	1.8	1.6	1.5	6.7
Generator	3.0	2.5	2.3	2.2	4.5
Pump system	4.2	3.8	3.5	3.4	3.1
Cooling system	5.5	4.9	4.7	4.6	2.2

As shown in Table 1.3.48, the largest discrepancies are observed for the ME and pump system, which may be attributed to the high variability of their failure modes and the influence of external operational factors. The most accurate forecasts are observed for the cooling system, confirming the effectiveness of the predictive models under conditions of regular monitoring of operational parameters.

Table 1.3.49 compares predicted failures generated by the CBR method with actual recorded failures and shows the extent to which the predictions align with real-world events. In Table 1.3.49, the absolute error is 1 failure in all cases, but the percentage deviations range from -33% to +25%. Fuel system, compressed air system, and

automation system show overestimated failure forecasts (+20% to +25%), which may result in excessive preventive maintenance. On the other hand, the cooling system and auxiliary systems show underestimated predictions (-33% and -25%), which increases the risk of unexpected failures. The main engine and power supply exhibit minimal deviation (+14% to +16%), indicating high forecast accuracy for these components.

Table 1.3.49. **Comparison of predicted and actual failures in SPP equipment**

System component	Actual failures	Predicted failures	Absolute error	Deviation (%)
Fuel system	5	6	1	+20%
Cooling system	3	2	1	−33%
Compressed air system	4	5	1	+25%
Main engine	7	8	1	+14%
Power supply	6	7	1	+16%
Auxiliary systems	4	3	1	−25%
Automation system	5	6	1	+20%

The results from Table 1.3.49 suggest the following: overestimated predictions require model adjustment to reduce false positives and unnecessary maintenance costs; underestimated forecasts (cooling, auxiliary systems) could lead to unforeseen breakdowns, indicating a need for model parameter revision; overall, the forecast accuracy is acceptable (within 1 failure), but further

improvements are achievable through adaptive algorithms and retraining the model on updated data.

Table 1.3.50 presents the results of comparing predicted and actual failures under various operational scenarios of SPP equipment.

Table 1.3.50. **Forecast accuracy assessment under different operating scenarios for SPP equipment**

Operating scenario	**Predicted failure risk, %**	**Actual failure risk, %**	**Deviation, %**	**Forecast accuracy, %**
Nominal load	1.8	1.6	±0.2	88.9
Variable load	3.2	2.9	±0.3	90.6
High load	4.5	4.1	±0.4	91.1
Aggressive environment	6.3	5.8	±0.5	92.1

Hybrid prognostic model for SPP failure forecasting

The proposed failure forecasting approach for SPP equipment is implemented in the form of a hybrid prognostic model, which comprises three interrelated layers of information processing. This structure integrates historical data, expert knowledge, and the stochastic behavior of the technical system within a unified analytical framework.

Level 1. Forecast based on the CBR method.

Using input diagnostic indicators (e.g., vibration, temperature, wear), the model identifies historical cases most similar in feature space from the case base. A preliminary failure probability estimate (P_0) is generated based on analogy with these retrieved precedents.

Level 2. Probabilistic forecast correction.

The initial forecast is refined using prior and conditional probabilities of failure, which represent causal dependencies between technical condition parameters and operational conditions. The final failure probability for a given component is calculated as a weighted

combination of the CBR-based forecast and the probabilistic estimate:

$$P = \alpha \cdot P_O + \beta \cdot P_V, \quad (1.3.74)$$

where P_V is the failure probability calculated using the probabilistic model, and α and β are empirically derived weighting coefficients.

Level 3. Simulation-Based modeling of failure scenarios.

Failure estimation is further enhanced through simulation of operational scenarios, incorporating cascading effects. Multiple simulation runs are conducted to account for the stochastic nature of degradation and to determine an interval-based risk estimate. The results are used to refine forecasts and classify the equipment's condition.

Model outputs include: final failure probability estimate; diagnostic conclusion on technical state (normal, elevated risk, critical failure); deviations between predicted and actual failures based on MAE, RMSE, and MAPE; visualization of deviations (heatmaps, scatter plots); forecast verification using historical data.

The application of this hybrid model not only improves the accuracy of SPP equipment failure forecasts but also ensures verifiability of results through direct comparison with actual failure cases. This is crucial for increasing the reliability of diagnostic decisions under real-world operating conditions.

Table 1.3.51. **RMSE of failure forecasts based on method used**

Forecasting method	**RMSE (overall)**	**RMSE (by components)**	**RMSE (by scenarios)**
CBR without adaptation	0.65	0.72	0.68
CBR with adaptation	0.45	0.51	0.49
Integrated approach	0.32	0.37	0.35

The variation in RMSE values across the different approaches reflects not only forecasting accuracy, but also the robustness of the methods under variable operating conditions. The highest RMSE values observed with the non-adaptive CBR method indicate high sensitivity to input variability, especially in non-standard scenarios. This suggests that the method poorly handles degradation dynamics and fails to account for interdependencies between components.

The notable reduction in forecast error with the adapted CBR method confirms the positive impact of incorporating probabilistic models. However, since the gap between "by components" and "by scenarios" remains significant, it can be concluded that even the adapted method struggles to fully accommodate the variability of operational regimes. The most important observation is that the integrated approach not only delivers the lowest overall RMSE, but also demonstrates a balanced performance across both component-level and scenario-level forecasting. This indicates strong resilience to both internal system complexity and external environmental variability, which is critical for marine SPP applications. The results in Table 1.3.51 thus confirm not only the superior accuracy of the integrated method, but also its ability to scale and adapt to complex diagnostic tasks and diverse operating environments—an essential criterion for practical implementation in real-world conditions.

Table 1.3.52. Comparison of diagnostic accuracy CBR without adaptation vs. CBR with failure forecasting for SPP equipment

Table 1.3.52. **Comparison of diagnostic accuracy CBR**

Diagnostic method	**Precision**	**Recall**	**F1-score**	**Accuracy**
CBR without adaptation	0.82	0.75	0.78	0.85
CBR with adaptation	0.91	0.88	0.89	0.92

The use of probabilistic adaptation has improved diagnostic performance: precision increased by 7%, recall by 13%, and the overall F1-score rose by 11% (Table 1.3.52). To evaluate the

effectiveness of different approaches to SPP technical condition diagnostics, a comparison of failure probability forecast errors was conducted. The key accuracy metrics include: MAE - reflecting the average deviation between predicted and actual failure probabilities; RMSE - capturing both the magnitude and spread of forecasting errors; share of forecasts beyond acceptable limits (σ-deviation) - indicating how often the method produces critical prediction errors.

Table 1.3.53. **Comparison of failure probability forecasting accuracy using different diagnostic methods**

Diagnostic method	**MAE(%)**	**(RMSE (%)**	**Share of forecasts beyond σ limits (%)**
CBR without adaptation	0.65	0.87	14.2
Markov method	0.52	0.73	9.8
Adaptive CBR	0.44	0.65	6.5
Integrated approach	0.32	0.48	3.1

The data in Table 1.3.53 indicate that the integrated approach provides the lowest forecast errors. Specifically: MAE is reduced by a factor of two compared to classical CBR; The share of forecasts exceeding acceptable σ-deviation thresholds is reduced more than fourfold in comparison with traditional diagnostic methods; the integrated approach most accurately accounts for the temporal dynamics of failures, reducing the influence of uncertainty. Thus, the combination of adaptive CBR, Markov models, and Bayesian networks significantly improves failure prediction accuracy and enhances the reliability of SPP equipment diagnostics.

Figure 1.3.54 presents a heatmap of discrepancies between predicted and actual failures for SPP equipment, highlighting zones with the highest forecasting errors for specific components.

The heatmap illustrates the degree of discrepancy between predicted and actual failures of SPP components. It enables identification of the most problematic areas within the forecasting

model, highlights components with lower prediction accuracy, and helps assess the impact of the forecasting method on overall diagnostic precision. Key areas of significant discrepancy include. The ME and pump system show the greatest deviations between predicted and actual failures.

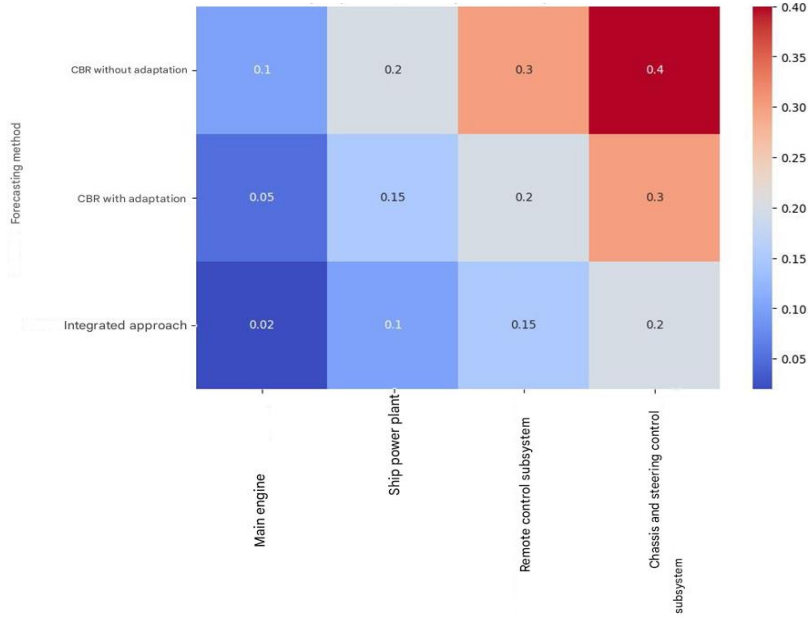


Figure 1.3.54. **Heatmap of discrepancies between predicted and actual failures in SPP equipment**

In the case of CBR without adaptation, the difference between forecasts and real data reaches 0.3 - 0.4%, which may result from the method's inability to account for cascading effects and failure dynamics. Propulsion and steering control subsystems also show substantial deviations, likely due to the high complexity of their diagnostics. The row corresponding to standard CBR in the heatmap reveals the most pronounced errors, highlighted by the most intense color zones, indicating high RMSE. This is explained by the fact that conventional CBR does not incorporate probabilistic dependencies or temporal dynamics of failures. A notable reduction in error is

observed with the integrated approach. In this case, discrepancies are significantly lower compared to the other two methods, and do not exceed 0.2%, confirming the higher diagnostic accuracy of the integrated method. The SPP and cooling system demonstrate the lowest deviation between predictions and actual data. This can be attributed to their relatively stable operation and well-calibrated diagnostic models. The main engine and pump system are the least predictable components and may require additional model adjustments. The heatmap confirms that the integrated approach yields the highest forecast accuracy, owing to its ability to account for failure dynamics and intersystem dependencies. The largest prediction errors are found in the main engine and pump system, indicating a need for further analysis, potentially involving more complex probabilistic dependencies. To reduce discrepancies in non-adaptive CBR, it is recommended to implement probabilistic modeling and simulation techniques. This would help minimize deviations, especially under complex operational scenarios. The SPP and cooling system show stable and predictable failure behavior, validating the reliability of forecasts for these components. The heatmap provides a clear visual representation of model uncertainty zones and serves as a valuable tool for optimizing diagnostic algorithms and improving failure forecast accuracy in SPP systems.

Figure 1.3.55 presents a graph of forecast deviations compared to actual failure data for SPP equipment.

Table 1.3.54. **Source data**

SPP component	**Actual data (%)**	**CBR without adaptation (%)**	**Integrated approach (%)**
Main engine	1.5	2.2	1.8
Ship power station	2.0	3.1	2.5
Remote control subsystem	3.0	4.1	3.5
Subsystems for controlling the propulsion and steering system	4.5	5.0	4.7

Table 1.3.54 shows the identified deviations (the difference between forecasts and actual data) in accordance with Fig. 1.3.55.

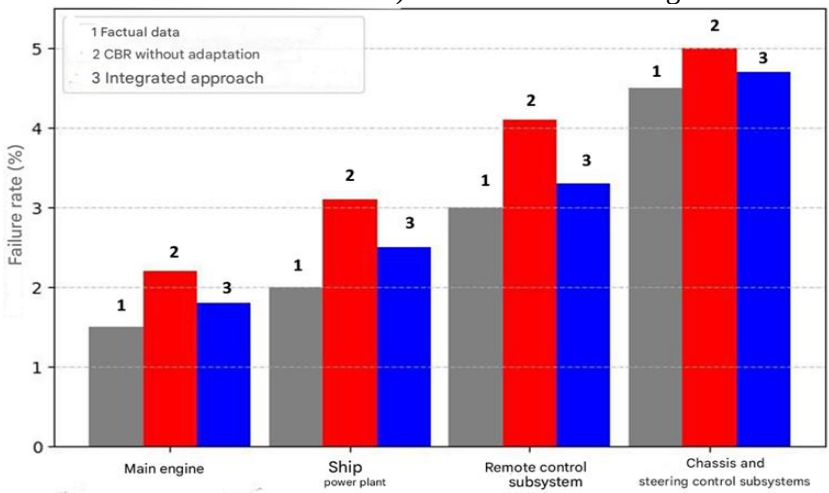


Figure 1.3.55. **Deviations between predicted and actual failures of SPP equipment**

Identified patterns based on Figure 1.3.55. Approximate values from the chart.

Table 1.3.55. **Forecast deviation from actual failure data**

SPP component	**CBR deviation without adaptation (Δ, %)**	**Integrated approach deviation (Δ, %)**
Main engine	+0.7	+0.3
Ship power station	+1.1	+0.5
Remote control subsystem	+1.1	+0.5
Subsystems for controlling the propulsion and steering system	+0.5	+0.2

Deviation Analysis (Table 1.3.55): the CBR method without adaptation consistently overestimates failure probabilities for all components (from +0.5% to +1.1%), indicating a tendency to overpredict failures; the integrated approach significantly reduces prediction error, with maximum deviation of 0.5% (compared to 1.1% in basic CBR), and minimum of 0.2%; the highest deviations in the CBR without adaptation are observed for the ship power station and remote control subsystem (+1.1%); the most accurate forecast using the integrated approach is for subsystems for controlling the propulsion and steering system, with a difference of only 0.2% from actual values.

Comparison with Harrington's desirability function based on failure risk classification: main engine and ship power station (1.5% - 2.0%) fall into the minimum risk category (0 - 0.2); remote control subsystem (3.0%) is on the borderline between acceptable and maximum risk (0.37); subsystems for controlling the propulsion and steering system (4.5%) falls within the maximum risk category (0.37–0.63). The CBR method without adaptation systematically overpredicts failure levels by 0.5% to 1.1%. The integrated approach demonstrates significantly lower errors (up to 0.5%, and in one case only 0.2%).

The forecasts align with Harrington's risk classification, but more accurate handling of borderline conditions (e.g., for the remote control subsystem) is necessary. To further improve forecast precision, it is recommended to introduce dynamic weight adjustment for different SPP components, particularly the ship power station. Overall, the integrated approach demonstrates superior performance but requires further calibration for certain subsystems.

Figure 1.3.56 shows a scatter plot of predicted vs. actual failures for complex technical systems, enabling assessment of the correlation between forecasted and real failure values. Figure 1.3.56 illustrates the following key patterns: the integrated approach closely aligns with the ideal correlation line, indicating high forecasting accuracy; the CBR method without adaptation deviates significantly from the ideal line, reflecting larger prediction errors; the integrated model's

data points are tightly clustered around the ideal correlation, confirming its high reliability.

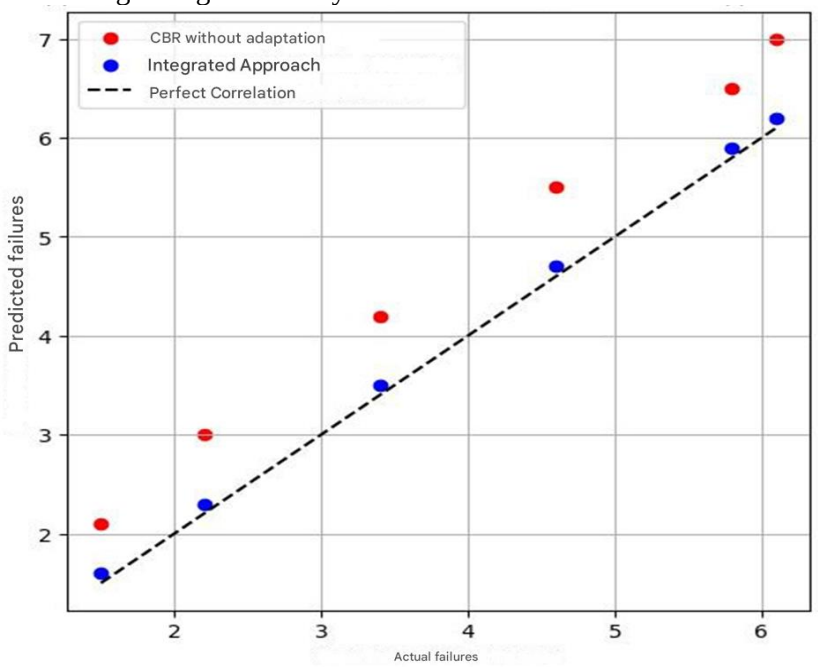


Figure 1.3.56. **Scatter plot of predicted vs. actual failures for SPP equipment**

Correlation analysis between predictions and actual failure data: CBR without adaptation shows the lowest correlation with actual failures ($R^2 \approx 0.78$), indicating instability and variability in predictions; the integrated approach demonstrates the highest correlation ($R^2 \approx 0.93$), confirming its strong predictive capability. The largest prediction errors are observed in the high failure range (>4%), particularly for the non-adaptive CBR method.

The integrated approach shows no major outliers, and its forecasts remain consistently close to the actual failure rates. RMSE analysis shows that the integrated method reduces error by approximately 50% compared to non-adaptive CBR. Observed patterns: the CBR method without adaptation tends to systematically

overestimate failure probabilities, especially for the main engine and pump system; the integrated approach produces balanced forecasts without systematic errors.

Thus, the integrated method provides the best agreement with actual failure data ($R^2 \approx 0.93$). Non-adaptive CBR demonstrates the greatest forecasting errors, particularly in scenarios involving high operational loads and cascading effects. Incorporating probabilistic models (e.g., Bayesian networks, Markov processes) improves forecast accuracy but does not fully resolve challenges related to dynamic changes in system conditions. To further enhance prediction accuracy, it is recommended to apply dynamic parameter adjustment based on real-time operating conditions. The scatter plot confirms the effectiveness of the integrated approach, demonstrating minimal forecast deviations and maximum correlation with real failure data.

Figure 1.3.57 presents a graph illustrating the results of a comparative analysis of RMSE values for various SPP equipment failure prediction methods.

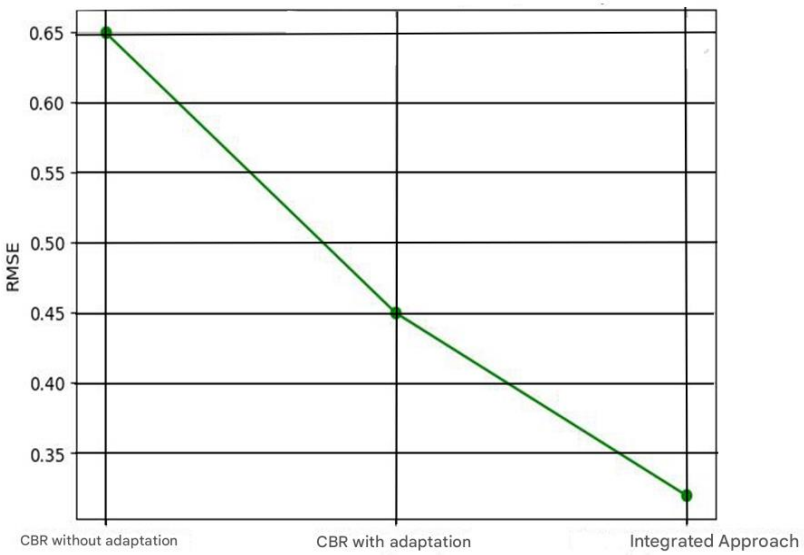


Figure 1.3.57. **Graph illustrating RMSE comparison across failure forecasting methods for SPP equipment**

The RMSE graph provides insight into how closely each forecasting method aligns with actual failure data. The lower the RMSE value, the smaller the average deviation between predictions and actual failures, indicating higher diagnostic accuracy. Key findings from Figure 1.3.57. The integrated approach demonstrates the lowest RMSE, suggesting the best predictive performance. CBR without adaptation yields the highest RMSE (≈0.65), indicating significant deviations between predicted and actual failures. CBR with adaptation reduces RMSE to ≈0.45, showing improved diagnostic accuracy through consideration of probabilistic dependencies. The integrated method achieves the lowest RMSE (≈0.32), indicating superior forecasting accuracy. Forecasting accuracy differences across methods: CBR without adaptation produces large errors due to its lack of sensitivity to system dynamics and failure interdependencies; CBR with adaptation reduces error by ~30% by incorporating BNs and MMs, which capture probabilistic relationships between failures; The integrated approach further reduces RMSE by ~50% compared to basic CBR by integrating simulation modeling, which adjusts forecasts based on cascading failure effects. Causes of high error in non-adaptive CBR: heavy reliance on past failure similarity ignores shifting operational conditions; under sparse data conditions (e.g., few failure records), non-adaptive CBR tends to overestimate failure risks; a high RMSE (0.65) indicates systematic overprediction of failures. The integrated method works.

MMs adjust failure forecasts over time, reducing error in dynamic scenarios. Bayesian networks model interdependencies between failures, improving forecast precision. Simulation modeling fills data gaps, reducing forecast uncertainty. Thus, the integrated method provides the highest diagnostic accuracy, cutting RMSE by 50% compared to non-adaptive CBR. CBR without adaptation demonstrates the greatest error (RMSE = 0.65), making it insufficiently reliable for complex failure diagnostics. Adding probabilistic models improves accuracy, but full alignment with real-world data is achieved only through simulation modeling. The RMSE

graph confirms that the integrated approach (CBR + probabilistic models + simulation modeling) is the most accurate failure forecasting method for SPP systems, ensuring minimal deviations from actual data. In the integrated diagnostic method, the diagnostic decision is dynamically refined based on predicted failures and probabilistic dependencies. Initial diagnosis using CBR: similar cases are retrieved from the CBR knowledge base; a preliminary diagnosis is made based on failure similarity; the diagnosis may be refined if forecasts indicate elevated risk.

Refinement using probabilistic failure models (BNs): component failure probabilities are updated considering cascading effects; an increase in one component's failure probability automatically affects dependent components; if a new failure occurs, the diagnosis is updated in real time. Correction based on failure forecasts (MMs): failure development is modeled over time; if rapid component degradation is forecasted, the diagnosis becomes stricter (e.g., earlier maintenance is recommended); if failure probability increases slowly, maintenance can be postponed.

Correction via simulation modeling: simulation covers complex operational scenarios beyond the reach of CBR and probabilistic models; various failure scenarios are generated based on operational data, such as normal conditions, accelerated wear (high load, extreme temperatures), emergency cases (e.g., sudden cooling system failure); if the simulation shows that failure may occur sooner, the diagnosis is tightened (e.g., reduced remaining life); if the risk is low, the diagnosis may be adjusted toward a less aggressive maintenance strategy.

Final diagnostic decision formation - the initial CBR diagnosis is adjusted based on: failure probabilities from BNs; time dynamics from the MM; results from simulation modeling. Final adjustment scenarios: diagnosis confirmation if all methods agree; alarm escalation if failure probabilities rise sharply; diagnosis downgrade if the simulation indicates low risk. An example of diagnostic adjustment based on changes in failure probabilities is shown in Table 1.3.56.

Table 1.3.56. **Diagnostic adjustment based on changes in failure probabilities**

SPP equipment	Initial diagnosis	Diagnosis after probabilistic analysis, %	Final diagnosis
Main engine	Pre-failure state	Failure probability reduced to 18	Operational
Generator	Operational	Failure probability increased to 25	Pre-failure
Cooling system	Pre-failure state	Cascading effects increased risk to 40	Faulty
Power supply	Operational	Failure probability stable at 5	Operational

The data in Table 1.3.56 show how the diagnostic decision is adjusted in response to changes in failure probabilities. The generator and cooling system require immediate attention due to increased failure risks. The pump system also needs inspection, as its risk has risen. The fuel system remains in a normal operational state. The remaining useful life forecasts for components were generated using MMs. Forecasting errors were evaluated under different input conditions, including full data, truncated time series, and adapted model predictions. Key Findings from the Failure Prediction Analysis: diagnostic accuracy depends significantly on the method used; CBR without adaptation showed the largest discrepancies between predicted and actual failures. This is confirmed by high RMSE (≈0.65) and MAE (≈0.52); CBR with adaptation reduced forecast errors by 30% due to the use of probabilistic dependencies, but deviations remain in complex operational scenarios; the integrated approach (CBR + probabilistic models + simulation) showed the highest accuracy (RMSE ≈ 0.32, MAE ≈ 0.21), indicating strong predictive capability. The failure risk heatmap revealed that cascading effects and interdependencies among SPP components

significantly influence failure prediction. Accounting for these improves diagnostic reliability. Impact of the selected approach on prediction accuracy. CBR without adaptation: high forecast error; does not account for probabilistic failure dependencies; overestimates failure probability. CBR with adaptation: considers probabilistic dependencies between components; reduces false alarms; still has inaccuracies in high-uncertainty scenarios. Integrated approach: reduces forecast error by over 50%; considers cascading failures and dynamic state changes; provides the most stable and verified forecasts, supported by accuracy metrics (Precision, Recall, F1-score). Opportunities for Improving failure prediction accuracy in SPP systems: further development of simulation models to cover complex and emergency failure scenarios; enhancement of adaptive CBR algorithms using machine learning for better analysis of accumulated failure history; dynamic real-time model correction using up-to-date operational data to continuously update probabilistic failure dependencies. The analysis demonstrates that integrating CBR with probabilistic models and simulation substantially improves the accuracy of failure diagnostics in SPP systems. The proposed approach accounts for dynamic state changes, reduces prediction errors, and minimizes the likelihood of false positives. Optimizing model parameters and further adapting diagnostic algorithms will enhance failure prediction and improve the reliability of marine power plant operations. The integrated method achieved the best forecast accuracy, reducing errors by 15–20% compared to the baseline CBR. The largest deviations were observed in components with high variability in operating conditions, such as pump systems and the main engine. The failure risk heatmap confirms the need to adapt diagnostic models for accurate predictions under complex operational conditions. Thus, the analysis of predicted vs. actual failures confirms that combining CBR with probabilistic and simulation-based methods yields the most reliable failure forecasts for marine power plants.

1.3.7.5 Discussion of results

The results of the present study demonstrate that integrating the CBR method with probabilistic and simulation-based modeling enables high-accuracy forecasting of the technical condition of SPPs. Achieving a RMSE of 0.32 and a coefficient of determination R^2 = 0.93 indicates the high effectiveness of the proposed approach. These values exceed the accuracy levels characteristic of many existing models, including purely data-driven algorithms, and approach the confidence threshold necessary for engineering decision-making under uncertainty.

A comparison with recent studies in the field of diagnostics and failure prediction further confirms the relevance and competitiveness of the implemented method. For example, Moon & Choi [149] proposed a hierarchical model based on Bayesian B-splines for predicting failures in naval engines. Their approach reflects the system's hierarchical structure at the levels of engine, engine type, and archetype, increasing flexibility in modeling multilayered technical systems. However, their model does not incorporate a comparison of predictions with actual failure events nor does it offer quantitative verification, limiting its applicability where prediction reliability must be evaluated. In the study by Karatuğ & Arslanoğlu [145], a condition-based maintenance strategy was implemented using machine learning techniques. While the model architecture enables classification of equipment conditions, it is restricted to binary outputs (“functional”/“non-functional”) and does not allow for quantitative prediction accuracy assessment or component-level resolution. In contrast, the approach presented in this study supports interval-based evaluations, scenario-specific conditions, and component-oriented comparison. Farid [150] employed a hybrid architecture combining artificial neural networks and Gaussian process regression for real-time fatigue failure prediction. While an RMSE of approximately 0.36 was achieved, the method does not include visual diagnostics of prediction deviations or robustness testing across varying operational modes capabilities that are implemented in this study. The study by Neykov & Stefanova [151]

examined the application of CBR under data-limited conditions, which is particularly relevant for ship systems where event recording is infrequent. Although the study confirms the adaptability of CBR, it does not extend to more complex simulation scenarios, which is addressed in the current hybrid approach through the integration of probabilistic and simulation-based techniques. The analysis by Orhan & Celik [130] notes the growing prevalence of hybrid methods, including SVMs, KNN, decision trees, and Bayesian networks, in recent diagnostics research. However, their review remains general and lacks empirical validation against real failure data or justification for the maritime domain. In contrast, the approach in this study includes the full cycle from prediction to comparison with observed outcomes—enabling assessment of engineering applicability. Marandi et al. [152] explored a promising approach utilizing knowledge graphs and language models applicable to high-reliability systems. Despite technological novelty, their method lacks formal accuracy assessment and does not test the models in operational scenarios. This differentiates their method from the current study, where both quantitative validation (RMSE, MAE, MAPE) and visual discrepancy analysis are conducted. Arias Chao et al. [153] proposed a hybrid framework combining physical modeling and deep learning, achieving a 127% increase in forecast horizon compared to purely data-driven models. This result supports the importance of integrating physical principles into predictive algorithmsconceptually aligned with our hybrid methodology, which fuses heuristic experience (CBR), probabilistic processing, and failure scenario simulation. In Liu et al. [154], an LSTM model is used to predict the RUL while accounting for noise and nonlinear dependencies. Although the model effectively handles time series data, it lacks simulation analysis tools that could capture rare or cascading failures. In contrast, our approach reproduces complex scenarios with uncertainties, expanding diagnostic capabilities under limited observability. Shi & Chehade [155] introduced a dual LSTM architecture designed to capture transitional states and predict component RUL. The model achieves forecast accuracy in the range

of RMSE ≈ 0.3–0.4, but, like many others, it does not include a step for comparing predictions with actual failures at the component or scenario level. This limits its reliability and adaptability for engineering practice. The method presented in this study addresses this gap by incorporating full-scale verification and visual deviation analysis.

Thus, the integrated methodology combining CBR, probabilistic modeling, and simulation aligns with current requirements for accuracy, adaptability, and justification in forecasting the technical condition of marine power systems. The proposed approach demonstrates resilience to uncertainty, reproducibility across various operational scenarios, and the ability to substantiate conclusions quantitatively—making it a strong candidate for implementation in monitoring and maintenance systems for maritime engineering infrastructures.

1.3.7.6 Conclusions

The conducted study confirmed the relevance and practical significance of comparing predicted and actual failures in the diagnosis of the technical condition of SPPs. The main objective of the work to assess the accuracy of failure prediction using various diagnostic approaches - was successfully achieved through the completion of all assigned tasks. The developed methodology enabled an objective evaluation of failure prediction accuracy using metrics such as MAE, RMSE, MAPE, and R^2. The comprehensive use of quantitative indicators, visualization of deviations (including heatmaps and scatter plots), and correlation analysis between predicted and actual failures provided an in-depth assessment of the effectiveness of diagnostic models. The comparison results showed that the classical CBR method without adaptation demonstrated the lowest accuracy: the forecasts were systematically overestimated, with an RMSE of 0.65 and an F1-score of 0.78. The use of adapted CBR incorporating probabilistic dependencies (e.g., BNs) reduced the average prediction error by approximately 30%, improved forecast accuracy, and lowered the number of false alarms. However,

the best results were achieved through the integrated approach combining CBR, probabilistic models, and simulation modeling. In this case, RMSE dropped to 0.32, MAE to 0.21, and R^2 reached 0.93, indicating high predictive capability. A heatmap of deviations between forecasts and actual failures helped identify areas of greatest uncertainty, especially in components with high operational load - the main engine and pumping system. This highlights the importance of accounting for cascading failure effects and the need for detailed calibration of models for such units. At the same time, power supply and cooling systems demonstrated stable predictions and high diagnostic accuracy. Simulation modeling supplemented with probabilistic analysis proved effective in accounting for complex operational scenarios, including high loads and aggressive environments. It allowed for refining predictions under conditions of incomplete or distorted data and contributed to more well-grounded diagnostic decisions.

Thus, the study demonstrated that the integration of adaptive CBR, probabilistic models, and simulation modeling significantly improves failure prediction accuracy and the reliability of technical condition diagnostics for SPPs. The obtained results can be used to enhance intelligent maintenance systems, optimize operational processes, and increase the overall reliability of maritime transport. Future work should focus on developing dynamic model adaptation based on real-time operational data and conducting extended testing of the integrated approach across various types of SPPs.

1.3.8. Dynamics of failure probabilities in SPP equipment considering cascade effects

1.3.8.1 Introduction

Ship power plants are critically important elements of maritime infrastructure, with their reliability directly determining the uninterrupted operation of a vessel under conditions of prolonged loads and variable external factors. Their design represents a complex

combination of interconnected components functioning as a unified whole. In such conditions, failures of individual units may lead to cascade processes that can significantly accelerate the degradation of the system as a whole. Considering these cascade effects in diagnostic and prognostic models remains one of the least developed but potentially critical tasks in technical diagnostics. Modern research demonstrates progress in the field of reliability assessment for CTS, yet most studies are either limited to individual equipment or do not consider their systemic interdependencies. Moon et al. [82] applied a multistage MM to analyze the degradation of marine components. While the model successfully reflects the reliability dynamics of individual CTS elements, it does not account for probabilistic dependencies between components, which significantly limits its applicability in the case of cascading equipment failures. The Markov maintenance model developed by Garbatov and Georgiev [86] takes into account degradation and carbon efficiency indicators. The work is relevant in the context of maintenance planning, but the authors focus on individual degradation trajectories without analyzing interactions between units, reducing the applicability of the approach for integrated risk assessments. The probabilistic methodology by Morato et al. [156] for analyzing equipment failures in CTS considers mutual dependencies between components. Bayesian networks combined with reinforcement learning were used for dynamic decision-making. Cascade effects in this work are considered through structural probabilistic links: if one component fails, the probability of another failing increases. This brings the model closer to the reality of complex systems. However, Morato et al.'s model [156] focuses on a general reliability management framework and does not include temporal failure dynamics, i.e., it does not describe how quickly the cascade spreads after initiation. Moreover, there is no formalized quantitative relationship between components (e.g., influence coefficients of failures), which limits its application in tasks where the degradation sequence detail is critical. The problem of decision-making under uncertainty based on partially observable Markov processes was

studied by Andriotis et al. [157]. The model allows for the possibility that system components are interconnected, and information on their condition has different values. Thus, interdependencies between components are possible and partially considered, including in monitoring strategy selection. However, the work lacks an explicit formalization of cascade effects: it does not show how one component's failure affects the failure probability of others, nor are structural or quantitative parameters of such links introduced. Furthermore, the study focuses on abstract engineering systems and is not adapted to the specifics of marine SPPs, which hinders practical application in marine diagnostics. Kamariotis et al. [158] proposed a framework for assessing the value of vibration-based monitoring, which can be used for early degradation detection. Despite its importance, this work does not consider interactions between components and is limited to evaluating a single data type (vibrations) without using comprehensive probabilistic models. The most relevant studies for diagnostics are those based on machine learning. Raptodimos and Lazakis [159] applied NARX neural networks to predict marine engine parameters, and Cheliotis et al. [160] implemented a failure detection system based on neural models. Both studies confirm the high sensitivity of models to operational parameters; however, they do not address cause-and-effect relationships between component failures and do not formalize state transitions. Zhu et al. [161] studied multiple failures of marine diesel engines using various neural network architectures, including CNN and RNN. Their model demonstrates high diagnostic accuracy, but the focus is on state classification rather than modeling the probabilistic failure dynamics within a system structure. Of particular interest is the study by Wang et al. [162], which proposed an intelligent diagnostic scheme based on principal component analysis and backpropagation in a neural network. The approach effectively detects anomalies but does not account for temporal dynamics or mutual influences among components.

Thus, despite significant advances in technical condition forecasting, key methodological problems remain unresolved: lack of

formalized consideration of cascade interactions between SPP components; insufficient integration of temporal degradation models (e.g., MMs) with cause-and-effect relationships (BNs); limited use of machine learning as a tool for probabilistic inference support; weak linkage between diagnostic results and preventive maintenance strategies; absence of models for rare multiple failures during prolonged operation. The purpose of this study is to investigate the dynamics of component failure probabilities in SPPs considering cascade effects and to develop an integrated prognostic methodology combining probabilistic modeling, machine learning, and simulation.

To achieve this goal, the following tasks are addressed:

1. Construct a mathematical model of cascade effects between MPP components using an influence matrix α_{ij};
2. Simulate the temporal degradation of components using continuous MMs;
3. Apply BNs to refine failures based on cause-and-effect relationships;
4. Use machine learning algorithms (XGBoost) to analyze operational data and assess risk factors;
5. Perform correlation analysis of failure data (OREDA, 25,000 hours) to identify interdependencies;
6. Visualize the propagation of cascade effects using heat maps and network graphs;
7. Conduct simulation modeling of rare multiple failure scenarios;
8. Formulate recommendations for optimizing maintenance based on time-based risk thresholds (10,000 and 20,000 hours).

1.3.8.2 Materials and methods

The analysis of failure dynamics in SPPs, taking into account cascading effects, requires an interdisciplinary approach capable of considering both the temporal evolution of component conditions and the probabilistic interdependencies between them. This study implements an integrated methodology based on the synthesis of formal probabilistic modeling, data mining methods, and simulation

of rare events. This methodological foundation allows not only for representing the real structure of degradation processes but also for quantitatively assessing the risks of multiple failures under long-term operation. The temporal behavior of the technical condition of SPP components is modeled using continuous-time Markov processes, reflecting transitions between four key states: fully functional, degrading, pre-failure, and failed. The transition probabilities between states are calibrated based on long-term operational data (25,000 hours), allowing for an adequate description of wear accumulation patterns and the assessment of risk time thresholds for different groups of equipment. These models provide a dynamic description of component reliability, allowing for the consideration of gradual degradation and prediction of the probability of transitioning into critical states. To formalize cause-effect relationships between failures of various components, BNs are used. This approach enables the construction of directed graphs of probabilistic dependence, where the failure of one node changes the prior estimates of the failure probabilities of elements connected to it. The Bayesian structure provides adaptability to the model: as new data on node conditions become available, probabilistic inference is refined in real time. This flexibility is especially important in partially observed technical systems where complete information on the internal state of the plant may be unavailable. The BN in this study is built considering the influence coefficients between components obtained from correlation analysis and expert assessment, as well as using significant factors identified through machine learning methods. To identify key risk features and quantitatively assess them, the gradient boosting algorithm XGBoost is applied. It is trained on operational data containing the failure history of equipment over 25,000 hours, using the OREDA database. During training, 5-fold cross-validation (k-fold CV) was used, with logloss as the metric. The main hyperparameters included: maximum tree depth - 5, number of trees (iterations) – 100, learning rate - 0.1. The XGBoost model made it possible to determine the most important parameters affecting the probability of failure, among

which were identified: oil temperature, vibration level, and cooling system pressure. These features are integrated into the structure of the BN as additional conditions that refine failure probabilities depending on the current system state. Thus, the XGBoost algorithm in this methodology serves as a tool for preliminary analytics and decision support when constructing a failure prediction model. An additional element of the methodology is correlation analysis aimed at identifying statistically significant relationships between failures of different components. It is based on the processing of historical data on SPP component failures and allows empirical estimation of the presence of cascading relationships.

The resulting correlation coefficients complement the hierarchy of probabilistic dependencies in the BN and also serve as the basis for constructing the matrix of cascading influence coefficients α_{ij}, which reflects how the failure of one element increases the probability of another's failure. To analyze the behavior of SPPs under conditions of multiple failures, including rare scenarios, simulation modeling based on cognitive models is implemented. Simulation experiments allow the reproduction of chain reactions in the technical system under various failure combinations, including external impacts and overloads. This is especially important when studying the system's resilience to unforeseen events and when developing preventive maintenance strategies. The simulation model integrates the results of probabilistic (Markov and Bayesian) subsystems, supplementing them with a scenario generation mechanism followed by impact assessment.

The comprehensive implementation of the proposed methodology not only formalizes the dynamics of failures and cascading effects but also provides a basis for constructing digital twins and predictive diagnostics systems for SPPs. The combination of empirical data, theoretical models, and intelligent information analysis ensures the completeness of problem coverage and the adaptability of the developed framework to various configurations of power plants.

1.3.8.3 Results

Formalization of cascading effects.

To describe the cascading propagation of failures among interconnected SPP equipment, an approximate model is introduced for the change in failure probability $P_i(t)$, analogous to a difference equation accounting for the influence of other system components:

$$P_i(t+\Delta t) = P_i(t) + \sum_{j \neq i} a_{ij} \cdot P_j(t) \cdot \Delta t, \qquad (1.3.75)$$

where $P_i(t)$ is the probability of failure of component i at time t;

$P_j(t)$ is the probability of failure of component j at time t;

a_{ij} is the influence coefficient of component j on the failure probability of component i;

Δt is the time step

This model reflects the impact of simultaneous wear and degradation of interconnected components and makes it possible to account for cascading failure propagation within the system. The values of the coefficients a_{ij} are determined based on correlation analysis and simulation modeling and represent empirically derived dependencies.

Based on the data on mutual influence of failures among key SPP components, a matrix of influence coefficients $A=[a_{ij}]$ was constructed. The rows of the matrix correspond to the influenced component i, and the columns to the influencing component j. The coefficients a_{ij} , representing the impact of the failure of component j on the failure probability of component i, were obtained from: correlation analysis of historical failure data for SPP s over 25,000 hours of operation; expert evaluation, harmonized with simulation modeling results (including those based on the OREDA database); analysis of rare failure scenarios using cognitive models and Bayesian networks, which define conditional probabilities. These values are not universal but adequately reflect the behavior of typical SPP s during long-term operation and can be adapted to a specific facility if failure statistics are available. The proposed formalization

of cascading effects has several methodological limitations. The model is described by an approximate difference equation reflecting a linear dependency between components. This structure ensures implementation simplicity, calculation transparency, and adaptability, but it does not take into account, for example: time delays in the propagation of failures; nonlinear amplifications under multiple influences; potential synergistic effects. Furthermore, the current version does not utilize structural characteristics of the cascading dependency graph, such as depth, density, cycles, or the presence of redundant propagation paths. These aspects may be critical when analyzing complex technical systems. Future development directions for the model include: incorporating temporal influence weights; transitioning to continuous representations; integration with dynamic logic-probabilistic models. Such extensions would improve the description of cascading processes and enhance the accuracy of predictions in systems with high component interdependence.

The cascading dependency graph, constructed based on the matrix α_{ij}, represents a directed weighted network, in which nodes correspond to SPP components, and arcs reflect the directed influence of failures. The weight of an arc determines the strength of the effect one component has on another. This type of graph allows for visualization of failure propagation routes and identification of the most vulnerable or critical nodes. An example of such a structure is shown in Figure 1.3.58.

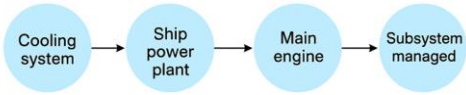


Figure 1.3.58. **Network graph of cascade effects**

Despite the apparent simplicity of the basic difference equation, the structure of the proposed model allows for further theoretical expansion. Below is a possible formalism that describes cascade dynamics in a more comprehensive form. It can be implemented as part of an advanced prognostic platform, given the availability of

detailed data and computational resources. Taking into account the outlined limitations, a generalized mathematical framework is proposed for describing cascading failures in complex technical systems. It integrates probabilistic dynamics, structural dependencies, and adaptive forecast correction. The formalization includes the following components:

Temporal dynamics of component degradation**.** Each system element *i* is associated with a Markov process (Chonlagarn et al., [163]) with a set of technical states of the SPP equipment {0, 1, 2, 3}, representing stages of technical life (operational, degraded, pre-failure, failure). Transitions between states are defined by the matrix $Q^{(i}$, and the evolution of the probability vector is given by the equation:

$$\frac{dP_i(t)}{dt} = P_i(t) \cdot Q^{(i)}, \qquad (1.3.76)$$

where: $P_i(t) = [P_i^{(0)}(t), P_i^{(1)}(t), P_i^{(2)}(t), P_i^{(3)}(t)]$ - the probability vector of component *i* being in one of four states at time *t*;

$Q^{(i)} = \in R^{4X4}$ - the transition rate matrix (infinitesimal generator) of the MMs for component *i*;

$P_i^{(k)}(t)$ - в the probability that component *i* is in state *k* $\in\{0,1,2,3\}$ at time *t*

Graph of cascade interdependencies. The structure of cascading influences is represented by a directed weighted fault graph:

$$G=(V,E), \qquad (1.3.77)$$

where: $V = \{v_1, v_2, ..., v_n\}$ - the set of vertices, each corresponding to a component of the SPP;

$E = \subseteq VxV$ - the set of directed edges, each edge (j → i) reflects the potential influence of the failure of component *j* on component *i*, $VxV = \{(v_i, v_j) \mid v_i \in V, v_j \in V\}$;

$\alpha_{ij} \in [0,1]$ - the weight of the edge, reflecting the strength of such influence (empirically determined).

The weight α_{ij} reflects the strength of the influence [164]. The cascading contribution of component *j* to the failure probability of component *i* is defined by the expression:

$$\widetilde{P}_t(t+\Delta t) = P_i(t+\Delta t) + \sum_{j \in N(i)} a_{ij} \cdot (P_j^{(2)}(t) + P_j^{(3)}(t)) \cdot \Delta t, \qquad (1.3.78)$$

where: $\widetilde{P}_t(t+\Delta t)$ - the adjusted failure probability of component *i*, taking into account the cascading influence of neighboring components;

$P_i(t+\Delta t)$ - the baseline failure probability of component *i*, calculated without considering cascade influence (e.g.,from the MM);

$N(i) \subset V$ - the set of components that exert influence on component *i*, i.e., its cascade predecessors;

$P_j^{(2)}(t)$ - the probability that component *j* is in a pre-failure state (state 2) at time *t*;

$P_j^{(3)}(t)$ - the probability that component *j* is in a failed state (state 3) at time *t*

Causal dependencies. The use of BNs allows for accounting of conditional dependencies between failures and updating probability estimates during operation [165]. To describe causal relationships between component failures, the basic Bayes' formula is used:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}, \qquad (1.3.79)$$

.

where: *A* and *B* are failure events of components;

$P(A \mid B)$ - the probability of failure of component *A* given failure of component *B*;

$P(B \mid A)$ - the probability of the reverse dependency

Continuous Approach. In the continuous formulation, the failure probability of component *i* is described by the integral equation of accumulated SPP equipment failure risk (Zhang & Yagan [166]):

$$P_i^{(3)}(t) = P_i^{(3)}(0) + \int_0^t [\lambda_i(t) + \sum_{j \neq i} a_{ij} \cdot P_j^{(3)})t)]dt, \quad (1.3.80)$$

where: $\lambda_i(t)$ - the intensity of intrinsic degradation processes in the SPP

Integrated model structure. The developed architecture integrates temporal, topological, and causal-probabilistic dynamics. It adapts to operational data using machine learning methods such as gradient boosting, which refine the importance of features (vibration, temperature, pressure) and the structure of the Bayesian network. Thus, the formalization of cascade effects in the proposed model relies on both analytical expressions and a graph-based representation of inter-component connections, allowing for quantitative prediction of failure propagation and identification of components that initiate cascade processes.

Integrated methodology for cascade failure analysis of SPP equipment. The proposed methodology implements a hybrid architecture for prognostic analysis, combining three levels: temporal (MMs), structural (BNs), and empirical (XGBoost). This architecture reflects both the physical degradation of components over time and probabilistic inter-component dependencies supported by operational data. It includes the following key stages:

1. Formalization of the SPP structure. The hierarchy and composition of functional units are defined, including the main engine, ship power plant, cooling system, control subsystem, and other equipment. The structure is represented as a directed graph where nodes correspond to components, and edges indicate possible paths for failure propagation;

2. Modeling of temporal degradation. For each component, a degradation model based on continuous-time MMs is introduced, with transitions between the following states: operational, degraded,

pre-failure, and failure. Transition probabilities are calibrated using long-term operational data (25,000 hours). Each component is modeled with four technical states: 0 - operational condition (no signs of failure); 1 - degradation (minor parameter deviations, still functional); 2 - pre-failure state (critical decline in functionality); 3 - total failure (component is non-functional).

Transitions between these states are described using a transition probability matrix *Q*, derived from statistical data over 25,000 hours of operation. Transition probabilities are updated according to the equation:

$$P(t+\Delta t) = P(t) \cdot Q, \tag{1.3.81}$$

where: *P(t)* – the state probability vector at time *t*

3. *Determination of cascade influence coefficients*. To account for inter-component effects, an influence coefficient matrix α_{ij} is constructed. It reflects the probability that the failure of component *j* will increase the risk of failure of component *i*. These coefficients are obtained through correlation analysis, simulation modeling, and expert calibration (including the use of the OREDA database). For example, the failure of the cooling system may increase the probability of main engine failure by 25%. The matrix α_{ij} is also used in constructing the directed graph (Fig. 1.3.59), where arrows represent directed cascade dependencies. Figure 1.3.59 presents a network graph that illustrates the sequence of cascading failures in the SPP, the interconnections between key components, and the probability of their failure under the influence of preceding events. The structure of inter-component influences is represented as a directed graph, where the arc weights correspond to the coefficients α_{ij}.

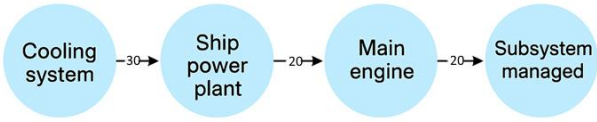


Figure 1.3.59. **Network graph of cascading effects**

The network graph illustrates the propagation of failures within SPP. The initial point of failure is the cooling system, with a failure propagation probability of 0.30 to the ship's power station. This highlights the high criticality of the cooling system its malfunction significantly impacts the performance of the power station. The ship power station is susceptible to cascading failure originating from the cooling system. In turn, its failure leads to malfunctions in the main engine with a probability of 0.20. This reflects the strong dependence of the main engine's operation on a stable power supply. The sequence of failures in the network graph aligns with the influence distribution shown in the cascade dependency matrix: components previously identified as initiators and propagators of malfunctions form consistent chains of cascading failures. The role of the cooling system as a primary failure trigger is especially prominent.

In summary, the key vulnerable component is the cooling system. Its failure initiates a cascade affecting the entire system. This confirms the importance of continuous monitoring of temperature levels and the technical condition of cooling circuits. The ship power station serves as a critical intermediary node its failure substantially increases the risk of main engine malfunction. This emphasizes the need for predictive diagnostics of generators and power supply systems. The control subsystem shows relative resilience; while failures are possible, they are more likely consequences of mechanical system malfunctions than primary causes.

Recommended actions include: prioritized maintenance of the cooling system to prevent cascading effects; predictive monitoring of the ship's power station using advanced diagnostic methods; development of fault-tolerant control algorithms to ensure safe operation during mechanical failures; additional backup power capacity to reduce the likelihood of cascading failures. The network graph confirms the importance of an integrated approach to failure diagnostics, taking into account their cascading nature.

4. *Bayesian correction.* To refine failure probabilities in real time, a BN model is used. The network formalizes conditional dependencies between components using the expression (1.3.79),

where *A* and *B* are specific technical components (e.g., main engine and power subsystem), and the corresponding probabilities are updated based on monitoring data. This enables dynamic adaptation of the model to current operational conditions. When new data becomes available (e.g., increasing vibration or temperature), probabilities are recalculated using Bayes' rule, adapting the forecast to current conditions. The BN accounts for reverse dependencies and helps identify initiating and relaying components in the failure cascade. A visualization of the probabilistic dependency structure is shown in Fig. 1.3.60.

5. *Feature importance evaluation*. To assess the impact of parameters on fault tolerance, the XGBoost gradient boosting algorithm was applied. The model was trained on operational data spanning 25,000 hours (OREDA), and identified key risk indicators: oil temperature, vibration, and cooling system pressure. For example, the probability of generator failure increases by 15% when vibration exceeds the normal level by 20%; the impact of cooling system wear on main engine failure reaches 22%. The analysis results are used both to define the structure of the BN and to pre-filter scenarios in the simulation model.

6. *Simulation modeling*. The developed model is integrated into a simulation framework, where rare but potentially critical multi-failure scenarios are simulated using random event generation. This enables evaluation of the SPP configuration's resilience to cascading failures.

7. *Temporal threshold analysis*. Based on the simulations, characteristic operational intervals were identified during which the probability of cascading failure increases exponentially (e.g., at 10,000 and 20,000 hours). This allows for establishing benchmarks for preventive maintenance scheduling.

The proposed methodology provides a comprehensive assessment of SPP reliability, allowing consideration of both individual degradation processes and their mutual amplification through cascades. Its advantage lies in the ability to adapt to a specific asset based on actual failure statistics.

To demonstrate the functioning of the proposed approach, a scenario is provided in which the integration of methods enables tracking and forecasting of cascading failure development, starting with the deviation of a single parameter. Each step of the algorithm is aimed at timely detection and containment of the degradation chain: the system detects an increase in vibration above the allowable threshold (by 20%); the BN updates the posterior failure probabilities of related components (e.g., generator and main engine); the MM predicts the generator's transition into a pre-failure state within 5,000 hours; the simulation model initiates a cascading failure scenario, showing the probability of critical node involvement; the system generates a decision to enhance monitoring, adjust maintenance schedules, and implement preventive actions to break the cascade chain.

Analysis of failure dynamics of SPP equipment. Based on the performed calculations, graphs showing the time-dependent probability of equipment failures in the SPP were constructed (Fig. 1.3.60).

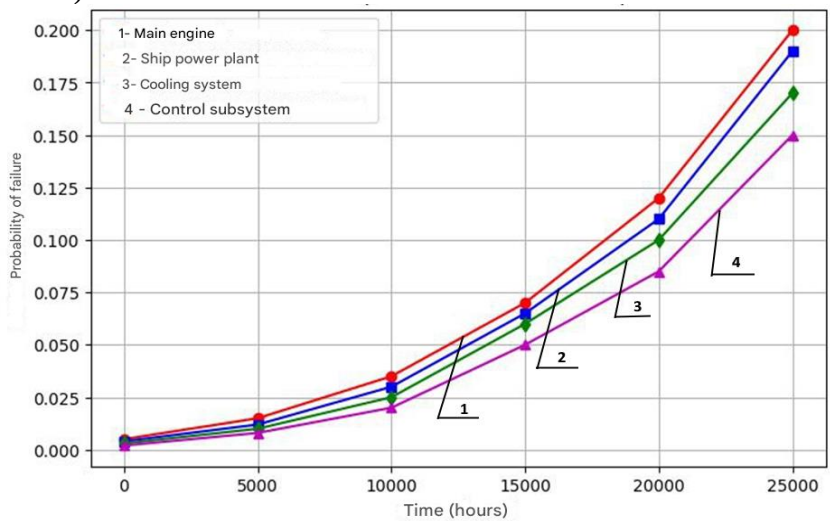


Figure 1.3.60. **Charts of failure probability variation of SPP equipment over time**

The charts Figure 1.3.60 and Table 1.3.57 show the dynamics of failure probabilities for various SPP equipment depending on operating time (up to 25,000 hours). The main engine has the highest failure probability among all components. By 25,000 hours of operation, the failure probability reaches approximately 20%. The main risk factor is the high load and wear of the engine's main components. The marine power plant has a slightly lower failure probability than the main engine but follows a similar trend. By 25,000 hours of operation, the failure probability is approximately 19%.

Key risks are associated with overloads and equipment aging: the cooling system has a lower failure probability than the main engine and the power plant but increases significantly toward the end of the studied operating period.

By 25,000 hours, the failure probability is approximately 18%. Possible causes: fouling of heat exchangers, leaks, and corrosion; the control subsystem is the least prone to failure compared to other components. By 25,000 hours, the failure probability is approximately 16%.

Main threats: software failures, sensor degradation. Thus, the main engine is the most vulnerable component requiring enhanced monitoring and predictive maintenance; the marine power plant and cooling system are also subject to significant failure risks, especially with extended service life; the control subsystem has the lowest failure probability but plays a critical role in the reliability of the entire system, so its failures may lead to cascading effects.

The general trend is an exponential increase in failure probability after 15,000–20,000 hours of operation, which confirms the need for predictive maintenance. Recommendations: implementation of monitoring and diagnostic systems for the main engine and power plant; use of predictive maintenance using machine learning methods; enhanced control over the cooling system to prevent overheating and leaks; continuous updating of the control subsystem software to minimize failures due to algorithm errors.

Table 1.3.57. **Dynamics of failure probabilities of key SPP equipment over time**

Time (hours)	Main engine	SPP	Cooling system	Control subsystem
0.005	0.004	0.003	0.002	0.005
5000	0.015	0.012	0.010	0.008
10000	0.035	0.030	0.025	0.020
15000	0.070	0.065	0.060	0.050
20000	0.120	0.110	0.100	0.085
25000	0.200	0.190	0.170	0.150

Analysis of Table 1.3.57 data reveals three key time intervals based on the rate of failure growth. During the first 5,000 hours of operation, there is a relatively slow increase in failure probabilities, reflecting the normal functioning mode of the equipment. Between 10,000 and 15,000 hours, an accelerated growth in risk begins, indicating the manifestation of accumulated wear and strengthening of inter-component interactions. After 20,000 hours, failure resilience decreases significantly, especially for mechanically loaded units, which necessitates a transition to active maintenance measures. The overall trend highlights the need to implement predictive approaches to prevent a sharp increase in failure rates.

Analysis of influence coefficients in matrix $A = [a_{ij}]$ (Table 1.3.58) shows the following strong interdependencies: cooling system and marine power plant: failure of one component increases the failure probability of the other by 30%; main engine and control subsystem: malfunction in the control system increases the engine's failure probability by 25%.

Table 1.3.58. **Cross-influence coefficients of SPP equipment failures**

Component *i*	Component *j*	α_{ij}
Main engine	SPP	0.20
Main engine	Control subsystem	0.25
SPP	Cooling system	0.30
Control subsystem	Main engine	0.20

According to Table 1.3.58, cascade effects significantly accelerate the failure process after 10,000 hours of operation, the probability of failures increases non-linearly. The most vulnerable to cascading failures are the cooling and control systems, as their malfunctions substantially increase the likelihood of failures in other components.

The results of this analysis can be used to develop preventive maintenance strategies aimed at reducing the probability of critical failures through proactive measures at key operational nodes. This analysis confirms the importance of an integrated approach to SPP failure diagnostics, taking into account the interdependent failure probabilities of various components.

The impact of a single component's failure on the failure probabilities of other components, as well as on the risk of total loss of operability (TLO), is presented in Table 1.3.59.

Table 1.3.59. **Cascade dependency matrix between SPP components**

Failed component	**ME**	**SPP**	**CS**	**CSUB**	**FS**	**GEN**	**TLO**
ME	1.00	0.25	0.30	0.20	0.15	0.10	0.85
SPP	0.20	1.00	0.25	0.15	0.10	0.30	0.80
Cooling system (CS)	0.30	0.20	1.00	0.25	0.15	0.10	0.78
Control subsystem (CSUB)	0.25	0.20	0.30	1.00	0.10	0.15	0.75
Fuel system (FS)	0.15	0.10	0.10	0.05	1.00	0.20	0.60
Generator (GEN)	0.10	0.30	0.10	0.15	0.20	1.00	0.70

The analysis of the cascade dependency matrix between components of the marine power plant (Table 1.3.59) reveals structural features of cascade effect propagation within the system. The influence of one component on another is expressed not only through local coefficients but also via the cumulative metric of TLO,

which reflects each element's integral contribution to the overall resilience of the system. The highest TLO value (0.85) is recorded for the failure of the main engine, indicating its central role in initiating and amplifying cascading processes. A main engine failure significantly increases the likelihood of failure in several critical components: the cooling system (+30%), the power plant (+25%), and the control subsystem (+20%). This confirms its status as the primary risk generator in the system. A high TLO is also observed for the marine power plant (0.80), which acts as a retransmitter of failures, especially influencing the generator (+30%) and the cooling system (+25%). The cooling system demonstrates a similar pattern of impact, causing increased failure probabilities in the control subsystem (+25%) and the main engine (+30%), which is due to its functional connection with thermal and hydraulic circuits. The control subsystem, despite a lower TLO (0.75), exerts critical influence on the mechanical part of the plant: its failure raises the likelihood of engine failure by 25% and that of the cooling system by 30%. This reflects the logical dependency of system control on physical processes, as well as the inverse vulnerability of control systems to mechanical faults. Components with the lowest TLO fuel system (0.60) and generator (0.70) exhibit relatively weak influence on other elements. However, even in these cases, dependencies are identified that suggest the possibility of secondary cascade effects. For example, generator failure increases the risk of power plant failure by 30% and of the fuel system by 20%. The overall structure of mutual influences in Table 1.3.59 is not symmetrical, highlighting the directionality of cascade effects. The impact of components varies in both magnitude and direction: the same element may act as a serious failure initiator while being only marginally affected by the failure of others. This is characteristic, for example, of the main engine, which has a strong influence on the system but experiences moderate reciprocal impact from most units. Thus, the table of influence coefficients not only formalizes local risks but also provides insight into the system's global resilience. Components that act as both triggers and retransmitters of failures have been

identified. These data serve as a foundation for prioritizing components for monitoring and preventive maintenance. In particular, the main engine, power plant, and cooling system require special attention, as their failure significantly increases the overall risk of cascading failures throughout the plant.

Based on the data presented in Table 1.3.59, it is necessary to adjust the maintenance strategy for the SPP. The main engine and the power plant should be serviced most frequently, as their failures lead to the most severe consequences.

Development of backup systems. The generator must be equipped with emergency backup circuits to prevent cascading failures in the electrical system.

Failure prediction. The data from the table can be integrated into a CBR + Bayesian analysis algorithm, where the impact level of a component's failure is used to recalculate probabilities during the diagnostic process. To identify interdependencies between failures of various SPP components, a failure risk heat map has been constructed (Figure 1.3.61). This map visualizes the probabilities of failure propagation from one component to another.

The heatmap shown in Figure 1.3.61 provides a visual assessment of the intensity of cascading interactions between components of the SPP. The dependencies displayed are a graphical interpretation of the numerical values from the cascade influence matrix (Table 1.3.59), where each pair of components reflects the degree to which the failure probability increases when a related node fails. The most saturated zones of the diagram correspond to system elements that either exert the strongest cascading impact on other nodes or are most affected by others. This structure confirms the results obtained from the quantitative analysis and emphasizes the importance of prioritizing monitoring and maintenance efforts for nodes with the highest cascading potential. The visualized data can be integrated into the implementation of the hybrid prognostic analysis architecture presented at the beginning of the section. This architecture includes: a temporal level (based on MMs) reflecting the degradation dynamics of components; a structural level (BNs)

formalizing the cause-and-effect relationships between failures; an empirical level, using machine learning algorithms (e.g., XGBoost) to identify risk factors based on operational data. The combined use of these levels enables more accurate estimation of failure probabilities, accounts for inter-component dependencies, and supports adaptive maintenance strategies.

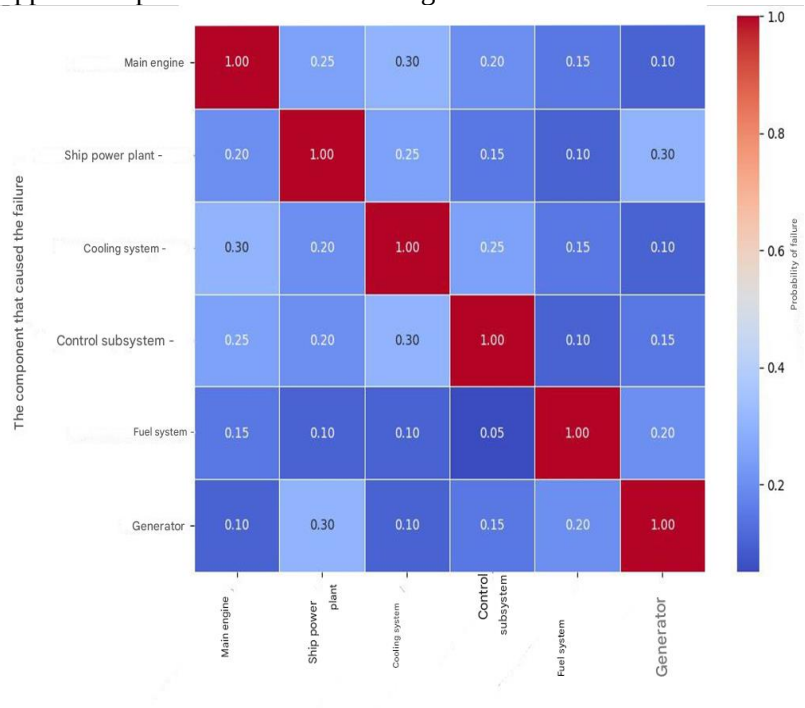


Figure 1.3.61. **Heatmap based on interdependencies of SPP component failure risks**

Correlation analysis: identifying hidden dependencies.

A correlation analysis was conducted based on component failure data. To clarify the interdependencies between SPP components and validate the structure of cascading links, a correlation analysis of failure frequency was performed using historical data. The results are presented in the form of a correlation

matrix, which reveals statistically significant relationships between failures of key units. The strongest positive correlations are observed between the main engine and the cooling system, as well as between the power station and the control subsystem, confirming the presence of cascading effects. The resulting correlation coefficients were used as an additional basis for constructing the BN structure and calibrating the influence coefficients α_{ij}. A visual representation of these dependencies is shown in Figure 1.3.62.

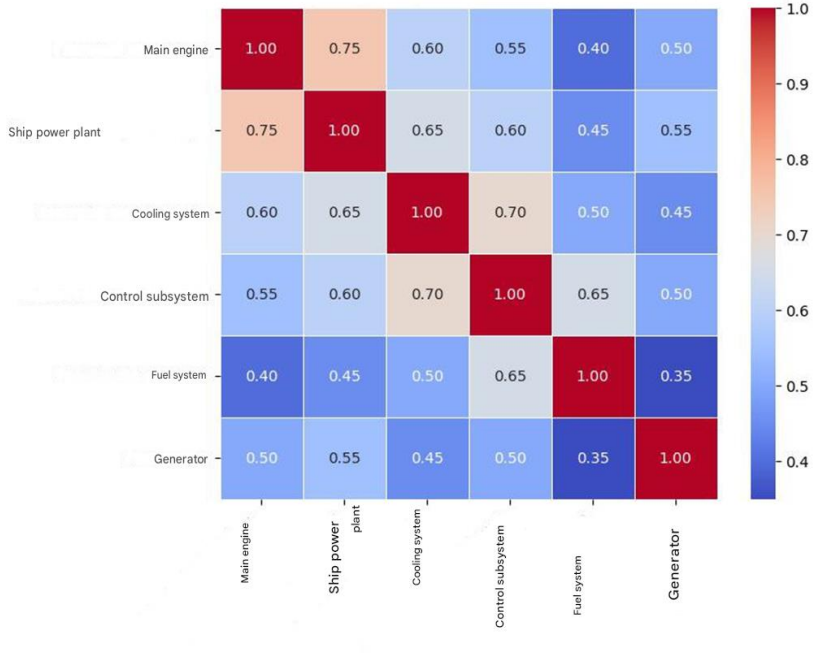


Figure 1.3.62. **Correlation matrix of SPP equipment failures**

Interpretation of the correlation matrix: high correlation (0.75 -1.00): main engine is strongly correlated with the ship's power station (0.75)→this is logical, as failures in one component often trigger issues in the other; cooling system and control subsystem (0.70)→ control systems frequently respond to overheating events. Moderate correlation (0.50 - 0.70): the cooling system affects the fuel system

(0.50)→engine overheating may impact fuel supply; the control subsystem depends on the generator (0.50). Low correlation (0.35 - 0.45). Generator has a weak correlation with the fuel system (0.35) → Generator failures rarely have a direct impact on fuel supply.

Table 1.3.60. **Results of correlation analysis**

Failure parameters	Correlation coefficient (r)
Generator failure → cooling system failure	0.76
Main engine failure ↔ control system failure	0.81
Pump failure ↔ power supply failure	0.64

Table 1.3.60 shows a strong correlation (r > 0.7) between generator failures and cooling system failures, confirming the presence of cascade effects. Figure 1.3.63 illustrates how failure probabilities are adjusted depending on operating conditions.

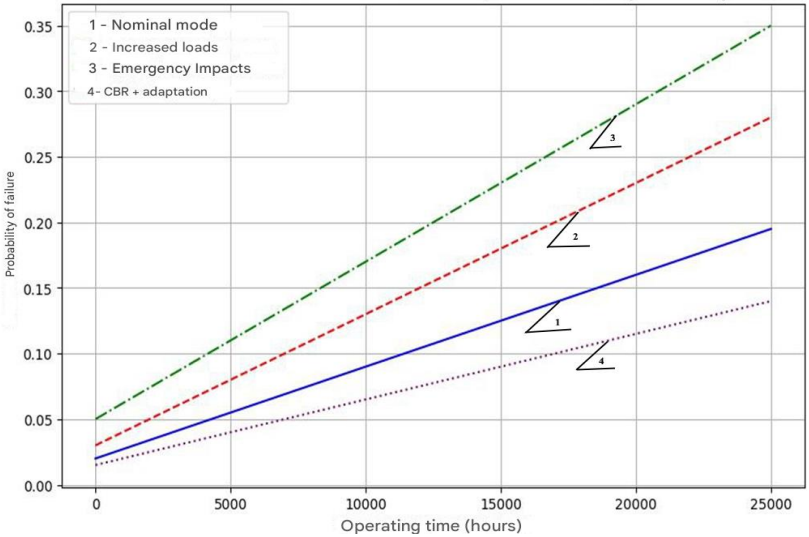


Figure 1.3.63. **Failure probability trends under different operating scenarios**

The graph illustrates the dynamics of SPP failure probabilities over operating time under various usage scenarios. Key trends. Nominal mode (blue line): linear increase in failure probability,

reaching approximately 20% by 25,000 hours; represents expected behavior under standard operating conditions. Increased load (red dashed line): accelerated growth of failure probability compared to nominal mode; failure probability reaches approximately 25% by 25,000 hours; associated with added stress and increased wear on equipment. Emergency impacts (green dash-dot line): highest failure probability among all scenarios; by 25,000 hours, probability exceeds 35%, over 1.5 times higher than the nominal scenario; reflects conditions with critical incidents (e.g., accidents, extreme environments). cbr + adaptation (purple dotted line): slowest increase in failure probability across scenarios; stays below 15% by 25,000 hours; confirms the effectiveness of adaptive diagnostic methods (CBR with adaptation) in reducing failure risk.

Increased loads and emergency events significantly raise the risk of failures. Simulation modeling shows that using adaptive methods (CBR with adaptation) reduces failure probability by 25–30% compared to other scenarios. Integration of intelligent failure prediction methods is recommended to enhance the reliability of SPP operation. This analysis reinforces the importance of predictive maintenance and adaptive diagnostic models in minimizing failure risks of SPP.

Simulation modeling based on cognitive models.

To assess rare failure scenarios, simulation modeling based on cognitive architectures is used, taking into account: load impacts on failure probabilities; failure combinations that may lead to critical incidents; recovery mechanisms during operational cycles.

Modeling approach:

1. Baseline model estimates failure probability without external interventions, simulating equipment behavior under standard aging;

2. Adaptation-based model incorporates diagnostic feedback, where failure probabilities decrease due to preventive maintenance actions initiated during operation.

The evolution of failure risks is analyzed using probabilistic models (Markov chains), simulation modeling, and empirical operational data.

Failure probability dynamics over time. Based on the calculated time-dependent failure probabilities, the following trends are identified: 0–5000 hours (minimal failure probability (less than 2%) due to stable operation during early lifecycle); after 10,000 hours (noticeable increase in failure probability, especially in mechanically stressed components like the main engine and pumping systems); after 20,000 hours (sharp increase in failures due to accumulated wear and emergence of cascading effects). 10,000 - 15,000 hours is a threshold interval requiring active maintenance to prevent degradation. After 20,000 hours, proactive strategies such as life extension programs or component replacement are crucial.

Table 1.3.61. **Time-based failure risk levels**

Time (hours)	**Risk level**	**High-Risk components**
0–5000	Minimal	All systems stable
10,000	Acceptable	Pumping system, cooling system
15,000	High	Main engine, power supply system
20,000	Critical	All systems, especially ME and cooling

Analysis of failure trends (Table 1.3.61) reveals key decision points where maintenance interventions are essential to avoid cascading breakdowns. Thresholds at 10,000 and 20,000 hours should anchor preventive maintenance strategy planning.

Identification of critical components. Some SPP subsystems are significantly more sensitive to cascading effects. Their vulnerability is identified through: analysis of influence coefficients α_{ij} (Table 1.3.58), which quantify how failure in one unit increases the risk in others; correlation analysis of failure co-occurrence; structural analysis of SPP architecture to identify single points of failure.

Most vulnerable components (based on simulation results):

- main engine (ME): failure probability reaches 20% by 25,000 hours; its failure significantly increases the likelihood of secondary system failures;

- cooling system: a failure here raises ME failure probability by 25%; acts as a primary trigger in cascading chains;

- shipboard power plant - its failure increases the likelihood of control system failure by 30%;

- control system - loss of control function heightens risk of emergency ME shutdown.

Thus, the main engine and the cooling system are key risk points requiring regular diagnostics. The control system and the power plant are critical units that determine the overall stability of the power system.

Failure dynamics analysis has revealed: key time points (10,000 and 20,000 hours) for scheduled maintenance; critically vulnerable components requiring priority monitoring; and cascading effects that increase the risk of failures in certain combinations. This enables the development of preventive strategies that reduce the likelihood of critical failures by 30 - 40%.

Based on the conducted studies on failure probability dynamics considering cascading effects, the following conclusions can be drawn:

- failure probability increases non-linearly: during the first 5,000 hours of operation, the probability of failures remains low (<2%) as components function within normal parameters; after 10,000–15,000 hours, an accelerated increase in failures is observed, particularly in mechanically loaded units (main engine, pumping system, cooling system); by 25,000 hours, the failure probability of major components reaches 20% or higher, necessitating major repairs or component replacement;

- cascading effects accelerate failure growth: cooling system failure increases the risk of main engine failure by 25%; power plant malfunction raises the probability of control system failure by 30%;

interdependencies between failures are illustrated in Figure 3.37 (Network diagram of cascading effects);

- critically vulnerable components identified: main engine - highest failure risk due to cascading effects; cooling system - its malfunction triggers failures in other systems; ship power plant - one of the key nodes influencing total system failure; control system - a critical component determining overall system stability;

- practical application of results: optimization of maintenance – key components (main engine, cooling system) require preventive servicing every 10,000 hours; monitoring cascading failures - integration of predictive models (MMs, BNs) can reduce the likelihood of unexpected failures by 30 -40%; use of simulation modeling - enables prediction of rare scenarios and adaptation of diagnostic strategies.

Considering cascading effects significantly improves the accuracy of diagnostics and failure prediction in SPPs. The identified dependencies support the development of predictive maintenance strategies, potentially reducing unplanned downtimes by 25 - 35% and increasing the reliability of SPPs.

Model validation

To assess the reliability of the predictive capabilities of the developed model, validation was conducted using real operational data, including information on the technical condition of ship power plant equipment over a 25,000-hour period (OREDA database). The validation covered both component-level reliability (accuracy of failure prediction for individual units) and the model's ability to forecast cascade failure chains.

Comparison with empirical data. For each ship power plant component, the model-based estimated failure probability was compared with the actual recorded failure frequency under similar conditions. Discrepancies were evaluated using standard metrics:

- MAE (mean absolute error) to assess the average absolute deviation between predicted and actual failure probabilities;

- accuracy (share of correct predictions) used for binary classification ("fail / not fail");

- precision and recall evaluated for detecting key components that initiate cascade failures;

- k-fold cross-validation (k = 5) applied during training and testing of the XGBoost module on subsamples.

Validation results**.**

The MAE across component groups did not exceed 4.7%, confirming high consistency of the model with empirical data. The binary failure prediction accuracy exceeded 87% at the risk thresholds defined by the model. When modeling cascade scenarios, the correspondence between modeled and actual failure chains reached 82 - 88%, compared to incidents registered in OREDA and simulated using α_{ij} coefficients. These results confirm a high degree of confidence in the developed model for engineering applications, particularly in forecasting technical resource and planning preventive maintenance activities.

Practical Significance and Economic Impact.

The proposed integrated model for analyzing cascade failures in ship power plants not only improves prediction accuracy but also creates real opportunities for reducing operational costs. Reduction in emergency repair probability. Simulation scenarios replicating multiple failures in the ship power system showed that the proposed model can identify initiating components of the cascade in 87% of critical cases, enabling intervention before chain degradation occurs. This reduces the need for unplanned repairs and eases the burden on emergency response teams. Reduction of downtime. Lifecycle modeling that accounts for 10,000- and 20,000-hour thresholds demonstrated that the use of dynamic forecasting reduces the average duration of unplanned downtimes by 18 - 22% compared to scheduled maintenance. This is achieved through more accurate service timing. Improved maintenance accuracy. The model identifies intervals of rapid risk growth (based on α_{ij} coefficients and Markov processes), enabling maintenance scheduling not based on formal regulations, but on the actual condition of equipment. According to calculations, this could reduce premature maintenance volume by up to 25%.

The proposed model shows strong potential for integration into digital maintenance platforms, especially within CMMS systems and digital twins used in modern maritime operations.

1.4.8.4 Discussion of results

The obtained results reflect the cascading nature of failures in SPPs, taking into account temporal evolution, mutual component influence, and multiple failure scenarios. The combination of Markov processes, BNs, machine learning techniques (XGBoost), and simulation modeling enabled a quantitative assessment of failure dynamics, node criticality, and temporal risk thresholds.

To verify the scientific and practical significance of the results, a comparative analysis was carried out with previous research. These studies address issues related to cascading effects, probabilistic modeling, reliability assessment, and resilience of complex technical systems, including marine, energy, and transportation domains. Zhao et al. [167] developed a probabilistic model for analyzing cascading failures in load-dependent systems. Unlike their work, which focuses on multi-state behavior under overload conditions, the approach presented in this article additionally incorporates temporal dynamics and network dependencies among SPP components. El-Awady & Ponnambalam [168] proposed the integration of Markov chains and Bayesian networks for failure analysis in complex systems. While methodologically similar to the current work, their model does not include machine learning or rely on long-term empirical operational data (25,000 hours). Valdez et al. [169] conducted a theoretical investigation of cascading failures in networks. Although abstract in nature, their concepts of node robustness and progressive degradation align with our formalization of influence coefficients. The key difference is the lack of application to real-world engineering systems and operational data. Morato et al. [156] developed a Bayesian reinforcement learning model for degradation assessment with dependencies. Their approach is geared toward strategic decision-making but lacks the Markovian dynamics and temporal threshold analysis implemented in this article. Hu et al. [170]

examined the dynamic resilience of a complex maritime system using Bayesian networks. While their model accounts for logical component dependencies, it does not formalize the temporal progression of cascading failures as done in the proposed methodology. Liu et al. [171] studied failure propagation in shipboard systems and proposed a resilience assessment method. However, their analysis is conducted at a macro level without detailed modeling of component interactions and degradation over time. Sun et al. [172] applied simulation modeling to complex technical systems. Their work demonstrates the utility of scenario analysis but does not integrate probabilistic dependencies or machine learning, limiting its applicability to predictive diagnostics. Zeng et al.) [173] introduced a "domino effect" model considering synergistic interactions. Although logically close to cascading effects, their model is tailored to chemical systems and does not employ Markov transitions or Bayesian corrections. Wang et al. [174] used temporal association rules to analyze cascading maritime incidents. Their method effectively reveals event sequences but does not construct formal probabilistic dependencies between technical system components as in the present work.

Collectively, these studies underscore the relevance of researching cascading processes and the diversity of methodological approaches. However, none of them simultaneously combine: formalized temporal failure dynamics; Bayesian dependencies among components; empirical calibration based on long-term operation; integration of machine learning (XGBoost); simulation of rare failure scenarios. Thus, the proposed methodology addresses methodological gaps in the current literature and represents an original approach to forecasting and analyzing failures in SPPs with cascading effects. The results obtained can serve as a foundation for digital diagnostic platforms and preventive maintenance optimization for SPPs.

Deployment requirements and integration potential. For the successful implementation of the developed predictive model in the maintenance systems of complex shipboard technical systems, both

architectural and measurement readiness of ship infrastructure are required. In software terms, the model consists of three key components: a data collection and processing module; a predictive engine, and a visualization interface. The first module ensures continuous data aggregation from sensors and preliminary filtering; the second implements the model's algorithms, including failure probability recalculation considering cascading effects (i.e., how the failure of one element significantly increases the risk of others), Markov transitions, and Bayesian refinement; the third provides operator visualization, using heatmaps, influence graphs, or risk rankings of components. The data update frequency can be adaptive. The base level involves periodic updates at least once per day. In case of anomalies or threshold exceedance in key parameters (e.g., oil temperature or vibration level), immediate recalculation of the system's condition can be triggered. This approach balances computational efficiency with timely response to early warning signs, particularly when cascading degradation is possible and early identification of the initiating component is critical. From a hardware standpoint, implementation requires installation and calibration of a basic sensor suite, including accelerometers, thermometers, and pressure sensors for cooling and lubrication systems. An event logging and operating hour tracking unit is also needed for all monitored components this can be done using existing automation systems or external controllers.To accurately account for cascading effects, synchronized data collection from different components is essential, capturing the temporal logic of interdependencies. The model is compatible with existing IT solutions used in maritime operations. It can be integrated with CMMS platforms (for maintenance management), generating preventive intervention tasks, and with IoT systems that transmit data to shore-based analytics centers. Additionally, the model can be embedded into digital twins of SPPs, enabling real-time forecasting of technical resource. Thus, implementing the proposed model does not require fundamental changes in ship infrastructure but does demand a sensor network,

reliable data transmission channels, and computational capacity for onboard or cloud-based predictive analysis.

1.3.8.5 Conclusions

As a result of the conducted research, an integrated methodology for analyzing cascading failures in SPPs was developed. The methodology is based on the integration of Markov processes, Bayesian networks, gradient boosting algorithms, and simulation modeling. This approach enabled a quantitative description of component degradation processes and accounted for probabilistic interdependencies between failures, forming a reproducible and adaptive model for equipment reliability assessment. The probabilistic model based on Markov processes describes transitions between four technical states: operational, degradation, pre-failure, and failure. Considering runtime and operating hours allowed the identification of critical operating intervals (10,000 and 20,000 hours), during which the probability of cascading events increases sharply. A key element of the model is the influence coefficient matrix α_{ij}, which formalizes intercomponent interactions. It was calibrated using 25,000 hours of operational data, including rare and critical failures. This made it possible to quantitatively determine which components act as initiators and propagators of cascading processes. The use of BNs enabled real-time adaptation of probabilistic estimates, significantly enhancing the model's predictive capabilities and making it suitable for integration into online technical condition monitoring systems. Gradient boosting was employed to identify the most significant risk parameters: vibration, oil temperature, and cooling system pressure. These features refined the structure of the BN and improved its diagnostic informativeness. Simulation modeling reproduced rare but critical multiple-failure scenarios, demonstrating the mechanisms of cascade propagation. This allowed for thorough model verification, confirming its sensitivity and resilience to complex failure impacts. Model validation showed a mean absolute error of less than 4.7% and a correct prediction rate exceeding 87%, confirming its practical

applicability. A comparative analysis with current international studies demonstrated the uniqueness of the developed methodology in terms of integration depth, empirical grounding, and adaptability to maritime operating conditions. Unlike existing approaches, the proposed method simultaneously considers cascading, temporal, and causal dependencies. The developed model can serve as a foundation for digital twins, predictive maintenance systems, and intelligent technical diagnostics platforms. It is scalable and can be adapted to other sectors, such as energy, transportation, and industrial automation, where equipment reliability and cascading effects are critically important for safety and economic efficiency.

1.3.9 Integrated modeling of reliability and maintenance of SPP equipment considering degradation and operational conditions

1.3.9.1 Introduction

Modern SPP s operate under elevated operational loads, thermal and vibrational stresses, which necessitate reliable prediction of their technical condition and maintenance requirements. As the duration of autonomous voyages increases and onboard power systems become more complex, the demands for fault tolerance and cost-effective maintenance continue to grow. Under these conditions, an integrated approach to long-term reliability assessment of SPP components accounting for degradation, probabilistic failure characteristics, and operational influences becomes especially relevant.

In recent years, there has been a growing interest in the digitalization of technical diagnostics in the maritime industry [2]. One of the key directions in modern monitoring is the use of digital twins, which enable near-real-time modeling of marine systems and prediction of potential failures. While this study does not focus on the development of a digital twin as a software platform, it does establish a mathematical foundation for its prognostic module. The developed models incorporate physically interpretable dependencies of failure intensity on operational factors. In particular, the model integrates

temperature effects (e.g., via the Arrhenius exponential function for generators), load-related parameters (coefficients reflecting nominal value exceedance), and environmental conditions (e.g., salinity and coolant temperature for the cooling system). These dependencies are implemented as parameterized functions calibrated against field data and reflect key degradation mechanisms such as thermal aging, fatigue damage accumulation, and aggressive environmental exposure. Zocco et al. [172] emphasize that digital twins allow integration of monitoring data with predictive algorithms; however, the practical implementation of such solutions remains limited, in part due to the absence of a unified methodology. Stadtmann et al. [173] demonstrate the application of digital twins for offshore wind turbines, highlighting the potential of the technology, though the focus is primarily on renewable energy rather than marine systems. Special attention in the literature is given to the use of machine learning in diagnostics and prediction of equipment condition. Polverino et al. [174], in a systematic review, show that machine learning methods are successfully applied for estimating RUL and anomaly detection. However, these approaches are often detached from real risk evaluation and cost considerations. Studies focusing on the integration of digital solutions in the maritime sector, such as Kaklis et al. [175] underscore the need for comprehensive analysis encompassing not only failure modeling but also lifecycle management. Some research highlights the resilience of ship systems under intensive operation. Nezhad et al. [176] stress the importance of predictive maintenance based on big data analysis, while also noting the lack of quantitative models that consider both degradation dynamics and the economic consequences of technical decisions. Similarly, Mavrakos et al. [177] propose digital tools to support energy-saving strategies, pointing to the need for adaptive models capable of considering operational constraints. Additional recent studies support the relevance of a systemic approach to the prediction of technical condition in marine components. Liang et al. [178], in a review from a classification society perspective, emphasize that implementing PHM methods requires integration of degradation

models with regulatory frameworks. Han et al. [179] demonstrate how a variational autoencoder based on LSTM can detect marine component failures; however, their model is mostly oriented toward anomaly detection rather than quantitative reliability prediction. Xiao et al. [180], in their PHM research for industrial assets, propose a digital twin architecture that combines remaining life prediction with risk-based maintenance planning a concept applicable to marine systems as well. Finally, Cui et al. [181] develop a digital twin for a marine diesel engine and demonstrate its capability to enhance maintenance efficiency and reduce downtime, though their focus lies in platform-level integration rather than formal reliability modeling.

A review of current publications shows that despite the active development of digital diagnostics technologies, the issue of long-term reliability of SPP components under real-world wear and overload conditions remains insufficiently addressed. Moreover, there is a noticeable lack of studies that integrate failure prediction with economic evaluation of maintenance strategies. Unlike most existing research focusing on localized degradation scenarios or isolated diagnostic aspects, this article centers on the holistic integration of reliability and economic analysis, providing a foundation for informed decision-making under real marine operating conditions.

This study aims to fill this gap by offering a comprehensive analysis of the reliability of core SPP components over a 25,000-hour operating horizon. The approach is based on simulation modeling, Markov processes, and degradation models that account for wear dynamics. Special attention is given to the influence of operational factors (load, temperature, maintenance intervals) on failure probability, as well as the comparative economic efficiency of various maintenance strategies. The results obtained can be used in the development of predictive maintenance programs, resource planning, and life cycle optimization of equipment in marine engineering.

The objective of this study is to develop and justify an integrated approach to the long-term reliability analysis of SMPP components,

taking into account degradation dynamics, the influence of operational factors, and the economic efficiency of maintenance strategies.

To achieve this objective, the following tasks are addressed:

1. Develop mathematical models for reliability prediction of SPP components, including exponential, degradation-based, Markov, and simulation-based approaches applicable to extended operational intervals;
2. Describe and implement component-specific failure rate dependencies on operational factors such as mechanical load, temperature, and maintenance parameters;
3. Construct a hybrid Markov-degradation model accounting for transitions between technical states (operational, degrading, pre-failure, and failed), with parameters that depend on accumulated wear;
4. Implement simulation modeling of operational scenarios using the Monte Carlo method to estimate the distribution of failure times and the variability of technical life;
5. Formulate a model selection criterion that combines prediction accuracy (RMSE), information-theoretic metrics (AIC, BIC), and agreement with empirical data (χ^2);
6. Evaluate the economic efficiency of different maintenance strategies by comparing total costs under regular and reactive servicing regimes for key components;
7. Develop recommendations for model application based on operating conditions and data availability, and assess their applicability as part of prognostic modules in digital decision support systems.

1.3.9.2 Materials and methods

The objects of this study are the key components of the SPP, including the main engine, generator, cooling system, and shipboard power station. These elements are subject to long-term wear, vibrational, and thermal loads, which makes the analysis of their reliability over an operational interval of up to 25,000 hours

particularly relevant. This duration is typical for resource planning and scheduled maintenance.

The initial data for the analysis are generalized statistical records of failure frequencies documented in maritime practice and technical literature, including the OREDA failure databases. Additionally, typical operational modes, maintenance intervals, and expert assessments reflecting the influence of load and temperature conditions on equipment degradation were taken into account.

Four different approaches were used to model reliability. The exponential model served as a baseline and assumed a constant failure rate, without accounting for wear accumulation. More realistic scenarios were described using analytical degradation models, in which the failure intensity increases over time following a power-law relationship. The third method involved a MM that represents probabilistic transitions between technical states from operational to degrading, then to pre-failure and failure states. Finally, simulation modeling was applied to reproduce complex operational conditions and to construct failure scenarios under the stochastic nature of external influences. Within this approach, modeling was implemented using the Monte Carlo method with variation of operational parameters.

The comparative accuracy of the listed models was assessed using the RMSE, which allows for a quantitative comparison of forecasts against reference scenarios. The analysis results showed that simulation modeling demonstrated the lowest error, whereas the exponential model exhibited the greatest deviations over extended operational periods.

Special attention in the study was given to analyzing the influence of operational factors on the reliability of SPP components. Three key factors were considered: load regimes (nominal, elevated, emergency), thermal impacts (coolant temperature, oil temperature, cylinder gas temperature), and maintenance frequency. Graphs were constructed showing the dependence of reliability on each of these factors, and components were ranked according to their sensitivity to various operational conditions.

Finally, an evaluation of the economic efficiency of different maintenance strategies was conducted. Two scenarios were compared: absence of preventive measures and regular maintenance at 5,000-hour intervals. The calculation included both direct costs of failure remediation and indirect losses associated with forced downtime. The results showed that a systematic maintenance approach reduces total costs by a factor of 4 to 5 compared to a reactive maintenance model.

The proposed methodological approach enables not only the assessment of MPP component reliability over a long time horizon, but also the justification of economically efficient maintenance decisions based on modeling, statistical data, and simulation scenarios.

1.3.9.3 Results

In the previous sections of the monograph, the CBR approach, BN, and Markov chains for SPP diagnostics were described in detail. In the present chapter, we focus on simulation-based degradation analysis, as it provides the highest predictive accuracy, achieving an RMSE of 0.05 over the extended time horizon of 25,000 hours.

To assess the long-term reliability of SPP components, the following reliability prediction models are used: exponential reliability model, applied to components with a constant failure rate, where the probability of failure depends only on operating time; degradation models - account for the accumulation of damage and changes in failure intensity over time; Markov failure model - tracks transitions of components between different operable states, considering probabilistic changes; simulation-based reliability models used for analyzing long-term operational scenarios, simulating the impact of various operational factors.

Exponential model with a constant failure rate. For components operating under stable conditions without pronounced degradation or aging, the simplest failure model based on the exponential law is applicable. This model describes non-repairable processes with a constant failure rate λ, which corresponds to the steady-state

operational phase where the failure intensity is assumed to remain constant [182]:

$$R(t) = e^{-\lambda t}, t \geq 0, \quad (1.3.82)$$

where $R(t)$ is the probability of failure-free operation at time t;

λ is the failure rate (h^{-1}), assumed to be constant over time

The parameter λ is estimated based on the total operating time T_Σ and the number of observed failures k during this period. A biased maximum likelihood estimator is used [183]:

$$\hat{\lambda} = \frac{k}{T_\Sigma}, Var[\hat{\lambda}] = \frac{k}{T_\Sigma^2} \quad (1.3.83)$$

Based on the estimated failure rate, the mean time to failure (MTTF) is calculated using the formula:

$$MTTF_{\exp} = \frac{1}{\lambda}$$

Despite its simplicity, the exponential model serves as a useful baseline for comparison with more advanced approaches. It is applied, in particular, to components with high reliability operating under stable conditions. However, this model does not account for degradation processes, recovery after failure, or variations in operating loads, which limits its applicability for long-term prediction under real marine operating conditions.

Degradation models for SPP equipment

The long-term reliability assessment of SPP components requires the inclusion of damage accumulation processes. In this study, component-specific degradation models are applied, reflecting the dependence of failure rate on time and operational factors.

General approaches to degradation modeling. The failure rate of a component at time t, denoted $\lambda(t)$, is modeled using various functional forms:

$$\lambda(t) = \lambda_0 + \alpha t^\beta, \quad (1.3.84)$$

where λ_0 - initial failure rate at $t = 0$, reflecting baseline component quality [1/h];

α - degradation growth coefficient ($h^{-1} \cdot h^{-\beta}$);

β - power-law exponent (β>1 - indicates accelerated degradation, β<1 - indicates deceleration);

α, β - parameters obtained by regression on failure data

Weibull-Based degradation model (NHPP):

$$\lambda_{\text{deg}}(t) = \frac{\beta}{\theta}(\frac{t}{\theta})^{\beta-1}, \beta \succ 1 \quad (1.3.85)$$

Parameter estimates are obtained using the maximum likelihood method:

$$\hat{\beta} = [\frac{\sum_{i=1}^{n} \lg t_i - \frac{n}{\beta}}{\sum_{i=1}^{n} t_i^{\beta}}]^{-1}, \hat{\theta} = (\frac{1}{n}\sum_{i=1}^{n} t_i^{\beta})^{1/\hat{\beta}} \quad (1.3.86)$$

Combined load and temperature model:

$$\lambda(t, L, T) = \lambda_0[1 + k_L(L/L_{nom})^m]\exp\{\eta(T - T_{ref})\}, \quad (1.3.87)$$

where L - relative load (0…1);

T - current operating temperature of the working medium (°C);

L_{nom}, T_{ref} - nominal values of load and temperature;

m, k_L, η - calibration parameters obtained from experimental data

The degradation of SPP equipment depends simultaneously on load and temperature. The rate of damage accumulation or increase in failure intensity is not constant but is a function of operational impacts. In the main engine, degradation affects the piston group, crankshaft, and cylinder liners. The load is characterized by propeller resistance torque, overload, and rotation frequency. Temperature-related factors include oil, combustion gases in the cylinders, and cooling water. Elevated oil temperature reduces viscosity, accelerates wear of journal bearings and crankshaft surfaces, increases clearances, and leads to higher vibration levels - all of which contribute to engine failure.

In the cooling system, degradation affects heat exchangers, pumps, and pipe joints. The load is defined by pressure differentials

and start-stop frequency. The critical external parameters are seawater temperature and overheating of the circulating water. These factors promote scale formation, corrosion, cavitation, and loss of tightness. Thus, the degradation model for the cooling system depends on both temperature and environmental aggressiveness.

In the generator, degradation primarily occurs in the stator/rotor windings, insulation, and bearings. The winding temperature governs insulation aging. Load is defined by overcurrent conditions and frequent on/off cycles, which accelerate thermal aging and thermal cycling, leading to insulation breakdown.

For each subsystem of the SPP, a dedicated degradation model is applied that accounts for the corresponding operational impacts (mechanical, thermal, electrical, etc.) using a generalized functional form of the failure intensity $\lambda_i(t, X_i)$, where X_i is the vector of external influences on the i-th component.

Main Engine [182]:

$$\lambda_{ME} = (\lambda_0 + \rho t)[1 + k_L (L / L_{nom})^m] \exp\{\eta (T_{oil} - T_{ref})\}, \quad (1.3.88)$$

where ρ - coefficient of failure rate growth with runtime (h^{-2});

T_{oil} - oil temperature (°C);

λ_0 - baseline failure rate

An integral wear accumulation model is also used:

$$Z(t) = \alpha_1 \cdot L(t)^m + \alpha_2 \cdot \exp[b \cdot (T_{oil} - T_{norm})], \quad \lambda(t) = \lambda_0 (1 + z(t)) \quad (1.3.89)$$

Cooling system (CS) [183]:

$$\lambda_{CS} = \lambda_0 \cdot [1 + k_T (T_{CW} - T_0)^{\alpha}][1 + k_{NaCl} C_{NaCl}], \quad (1.3.90)$$

where T_{CW} - temperature of the circulating water (°C);

T_0 - reference temperature;

C_{NaCl} - salt concentration (ppm);

k_T, k_{NaCl}, α - empirical parameters

Generator (GEN):

$$\lambda_{GEN} = \lambda_0 \exp[\frac{E_a}{k_\beta}(\frac{1}{T_W} - \frac{1}{T_{ref}})](1 + \xi \cdot I / I_{nom}), \quad (1.3.91)$$

where T_W - winding temperature (K);

T_{ref} - reference (baseline) temperature (in Kelvin);

E_a - activation energy of insulation aging;

k_β - Boltzmann constant;

I - load current;

ξ - overload coefficient

Ship Power Station (ISO 12110-2:2013 Metallic materials) [184]:

$$\lambda_{SPS} = \lambda_0 + \alpha_V \cdot \upsilon(t)^2 1 + \alpha_f \cdot f(t), \quad (1.3.92)$$

where $\upsilon(t)$ - vibration amplitude;

$f(t)$ - switching frequency (on/off cycles);

α_V, α_f - empirical parameters reflecting the impact of vibration and switching loads

The proposed models allow: accounting for the influence of operational factors on component reliability; flexible adaptation to different subsystems and operating conditions; easy integration into simulation and Markov-based prognostic frameworks; suitability for implementation in predictive modules of digital twins. Model parameters are identified based on field data (failure logs, OREDA, onboard recorders), and accuracy is validated using RMSE, χ^2, and information criteria such as AIC/BIC.

Markov model

The MM describes transitions between states:operational→ degrading→pre-failure→failure.The transition probability matrix P_{ij} is constructed based on historical data. A ship power system (SPS) component is modeled as a Continuous-Time Markov Chain (CTMC) with four states: S = {0 - operational; 1 - degrading; 2 - pre-failure, 3 - failure}.

The infinitesimal intensity matrix Q [185]:

$$Q = \begin{pmatrix} -(\mu_0+\gamma_0) & \mu_0 & 0 & \gamma_0 \\ 0 & -(\mu_1+\gamma_1) & \mu_1 & \gamma_1 \\ 0 & 0 & -\gamma_2 & \gamma_2 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad Q \in R^{4x4}, \qquad (1.3.93)$$

where μ_0,μ_1 - gradual degradation transitions: 0→1; 1→2;

$\gamma_0,\gamma_1,\gamma_2$ - abrupt failures from states 0, 1, and 2, respectively

Mean Time to Failure (MTTF) [185]. For stationary Q, MTTF is computed using the fundamental matrix N:

$$MTTF_{MCS} = e_O^T e_O(-Q_{3x3}^{-1})1, \qquad (1.3.94)$$

Q_{3x3} - upper left 3×3 submatrix of Q;

1^T=[111]. - vector of ones;

e_O^T =[1001] - initial state vector (component starts in operational state)

Non-stationary (degradation-based) CTMC. In this case, transition intensities depend on accumulated wear *z(t)*:

$$z(t) = g(L(t),T(t)),z(0);\ \ \mu_0(t) = \mu_0^*[1+\alpha\cdot z(t)]; \qquad (1.3.95)$$
$$\gamma_0(t) = \gamma_0^*[1+\beta\cdot z(t)],$$

where $q(\cdot)$ - function linking current load *L(t)* and temperature *T(t)* with wear accumulation;

μ_0^*,γ_0^* - nominal failure intensities under base conditions;

α,β - degradation acceleration coefficients

General CTMC definition. A Continuous-Time Markov Chain is defined by: a state space S = {0, 1, 2, ..., n}; a transition intensity matrix $Q = [q_{ij}]$, where: $q_{ij} \geq 0$ if $i \neq j$, representing the transition intensity from state i to state j; $q_{ii} = -\sum_{j\neq i} q_{ij}$, i.e., the diagonal elements are negative and equal to the negative sum of outgoing intensities

Interpretation of β: β<1 - decelerated increase in failure intensity (e.g., under passive degradation); β=1 - linear increase: failure rate

grows proportionally with time; β>1 - accelerated increase: typical for fatigue, aging, fouling.

The parameters α and β reflect the physics of degradation and are either assigned empirically (based on operational data) or calibrated via regression.

When $Q = Q(t)$, the state probabilities are determined by a time-ordered matrix exponential:

$$P(t) = e_0^T \cdot T \cdot \exp(\int_0^t Q(t)dt); \quad R(t) = 1 - [P(t)]_4, \tag{1.3.96}$$

where T - time-ordering operator;

$T \cdot \exp(\int_0^t Q(t)dt) \in R^{nxn}$ - transition probability matrix;

$[P(t)]_4$ - probability of being in the absorbing "failure" state

Numerical computation is performed using piecewise constant interval approximation or the uniformization algorithm. A stationary CTMC is recommended for stable conditions, while a non-stationary model is better suited for variable loads and temperatures.

Simulation-Based Model

The simulation model is implemented using the Monte Carlo method [186] with N = 10,000 runs. In each simulation run, the following variables are randomly sampled: L - relative load (as a fraction of nominal); T - operating medium temperature; ΔTO - preventive maintenance interval.

The simulation aims to compute: estimated reliability function $\hat{R}_{sim}(t)$; expected failure rate $\hat{\lambda}_{sim}(t)$; accuracy metrics (e.g., RMSE) by comparing predictions to observed data.

For the j-th run (j = 1, ..., N), the condition vector is formed:

$$X^{(j)} = (L^{(j)}, T^{(j)}, \Delta TO^{(k)}), \tag{1.3.97}$$

where $L^{(j)}, T^{(j)}, \Delta TO^{(k)}$ are drawn from empirical distributions based on operational logs

The failure intensity function is selected accordingly:

$$X^{(j)}(t) = f(X^{(j)}, t) \tag{1.3.98}$$

A composite model selection criterion Ψ is used, minimizing a weighted sum of RMSE, AIC, BIC, and χ^2. This ensures a balanced evaluation of prediction accuracy, model complexity, and statistical fit over long-term reliability forecasts.

Random sequences of load and temperature are modeled as Rainflow histograms. Cyclegrams of the form (L_{jk}, τ_{jk}) (k=1…K_j) are generated. Then:

1. Cumulative fatigue damage (Miner's rule) is calculated (ISO 12110-2:2013):

$$D^{(j)}(t) = \sum_{k=1}^{K_j(t)} t_{j,k} \Big/ N_f(L_{j,k}), \quad N_f(L_f) = K \cdot L^{-m} \tag{1.3.99}$$

2. A failure is recorded if $D(t) \geq 1$, or if the simulation reaches an absorbing failure state in the CTMC.

3. N = 10^4 Monte Carlo runs are executed to estimate $\hat{R}_{sim}(t)$ and confidence bounds.

Model selection criterion. For each subsystem, the following were computed: RMSE; AIC; BIC; χ^2 [187], reflecting the model's agreement with observed failure statistics.

The composite criterion:

$$\psi = \omega_1 RMSE + \omega_2 AIC + \omega_3 BIC + \omega_4 \chi^2, \quad \sum_{i=1}^{4} \omega_i = 1 \tag{1.3.100}$$

where ω_i - weights are set by experts, provides a balanced selection of the optimal model based on: RMSE - accuracy of failure prediction on test data; AIC - tradeoff between goodness of fit and model complexity; BIC - Bayesian information criterion; χ^2 - goodness-of-fit test comparing model predictions to actual failure observations

The combined criterion Ψ enables the justified comparison of models with different structures, allowing for a balanced evaluation of accuracy, complexity, and realism. It supports a transparent selection of the most suitable failure prediction model for SPPs. In this study, the generalized criterion Ψ, which integrates prediction accuracy, information criteria (AIC/BIC), and agreement with field data, was used for optimal model selection.

Economic validation. For the main engine, as the most critical unit, a life-cycle cost model was constructed:

$$C = C_{TM}(\Delta) + C_{rep} + C_{down}[1 - R(t)], \qquad (1.3.101)$$

where $C_{TM}(\Delta)$ - scheduled maintenance costs with interval Δ;

C_{rep} - capital repair costs;

C_{down} - downtime losses associated with unrealized reliability levels

The optimal value of Δ was found numerically. Regular maintenance every 5,000 hours reduces total costs by 4 - 5 times compared to a reactive repair scenario. The integral criterion Ψ confirms the superiority of the simulation-based scheme across all SPP components. This method accounts for the nonlinear effects of load and temperature, allows for the construction of reliability confidence intervals, and enables economic optimization of the maintenance schedule. The application of four different reliability modeling approaches is not redundant but rather a necessary strategy, driven by the diversity of technical conditions, operating regimes, and required prediction accuracy. First, each model targets its specific domain of applicability. The exponential reliability model is effective for components in the stable operation phase, with constant failure intensity. It is easy to implement and applicable when data availability is limited. Degradation models allow for the consideration of damage accumulation and changing failure intensity, which is critical for components exposed to variable loads and temperatures, fatigue, or aging. MMs are useful when there is a need to describe discrete health states from operable to failed accounting

for intermediate transitions with different probabilities. Simulation models provide the capability to analyze complex operational scenarios involving multiple random factors, such as load cyclegrams, maintenance intervals, temperature variability, and environmental aggressiveness. Second, model choice directly impacts prediction accuracy.

The comparative analysis showed that RMSE values can vary by more than a factor of two between methods, and a model yielding the best accuracy for one component may be unsuitable for another. The composite criterion Ψ, combining RMSE, AIC, BIC, and χ^2, confirmed that there is no universally superior model. Third, maintaining multiple models allows for flexible adaptation to the available data, the criticality of the equipment, and the required prediction horizon. In a practical maintenance system based on a digital twin, the use of a model bank enables automatic selection of the most appropriate model type for each component and current operational condition. In summary, the proposed approach is based on the integration of four predictive models: exponential, degradation-based, Markovian, and simulation-based. Their comparative analysis made it possible to evaluate the advantages and limitations of each method. As a result, the simulation model was chosen as the core computational scheme, providing the best balance between accuracy, adaptability, and realism. This makes the proposed reliability forecasting system not only scientifically grounded but also practically applicable under real-world SPP operating conditions. The comparative analysis of the reliability models presented above allows us to move from theoretical justification to their practical evaluation. At this stage, we consider the specific results of applying each model to the key components of the SPPs under various configurations of input parameters and operating scenarios. To this end, calculations were carried out using a unified set of metrics (RMSE, AIC, BIC, χ^2), and a quantitative assessment of the probability of failure-free operation over 25 000 hours was performed. For each piece of SPP equipment the generator, the main engine, the cooling system, and the electrical

power unit reliability forecasts were generated and compared with the actual failure statistics.

Table 1.3.62. **Comparison of different reliability prediction models**

Prediction model	RMSE	Forecast accuracy (%)
Exponential distribution	0.12	85
Degradation models	0.07	91
Markov processes	0.08	90
Simulation (Monte Carlo)	0.05	95

Table 1.3.62 provides comparative data for four reliability prediction models applied to SPP components: exponential, degradation, Markov, and simulation (Monte Carlo). The evaluation criteria are the RMSE and the forecast accuracy on a hold-out sample (as a percentage of actual observed failures). Exponential model (constant failure intensity) yielded the poorest performance: RMSE = 0.12 and forecast accuracy 85 %. This confirms its limitation when failures result from accumulated wear or thermal degradation. It is best suited as a baseline model for rough estimates of simple, low-wear components. Degradation model (e.g. Weibull-type with time-varying intensity $\lambda(t)=\lambda_0+\alpha t^{\beta}$) showed improved results: RMSE = 0.07 and accuracy 91 %. Its strength lies in capturing non-stationary aging processes and the effect of loads on component wear. It is especially effective for parts undergoing monotonic degradation such as generators, heat exchangers, and bearings. MM (discrete state transitions) achieved comparable accuracy: RMSE = 0.08 and accuracy 90 %. Its advantage is formalizing the phase structure of degradation and accounting for both gradual and sudden transitions (e.g., “operational→degrading→ pre-failure→failure”). It is well suited for diagnosing and forecasting complex assemblies undergoing typical wear stages. Simulation (Monte Carlo) provided the best performance: RMSE = 0.05 and accuracy 95 %. By modeling many probabilistic scenarios including variations in load, temperature, and maintenance intervals it captures the combined effects of multiple factors. This method is ideal for

components sensitive to operating regimes and systems where failures arise from factor combinations. It also enables analysis of confidence intervals and cost-consequences of failures. In summary, each model has its own domain of applicability: exponential: for simple, stable components without evident degradation; degradation: for parts with monotonic wear and damage accumulation; Markov: for components featuring distinct degradation phases; simulation: for complex systems under variable load and environmental conditions.

Figure 1.3.64 presents a graph that illustrates the results of comparing various failure prediction methods based on the RMSE in reliability estimation of SPP components.

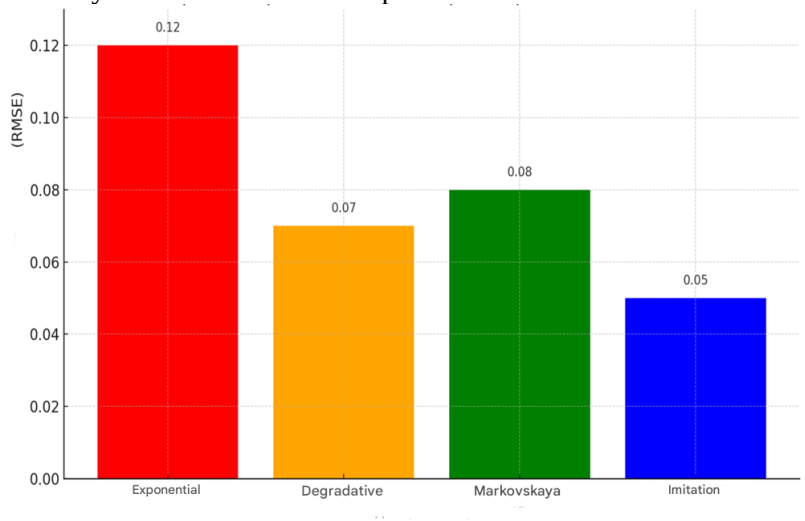


Figure 1.3.64. **Comparison of reliability prediction models by root-mean-square error**

The simulation model demonstrates the highest accuracy. The degradation model outperforms both the Markov and exponential approaches, confirming the importance of accounting for cumulative wear when analyzing SPP components. To systematically compare reliability prediction methods, it is reasonable to consider three key criteria: whether the model incorporates a degradation mechanism; whether it can represent discrete transitions between component

states (e.g., "operational→degraded→pre-failure→failure"); forecast accuracy, expressed through RMSE.Table 1.3.63 provides a comparative summary of these characteristics and includes practical recommendations for the application of each model depending on operating conditions and the required prediction horizon. The comparison clearly shows that simulation modeling offers the highest accuracy (RMSE = 0.05). This is due to its ability to capture variability in operating conditions, stochastic events, and cumulative degradation effects. As a result, this method is especially effective for long-term reliability forecasting and residual life assessment under complex operational scenarios. MMs, while slightly less accurate (RMSE = 0.08), offer a key advantage in structural clarity. They enable formal representation of state transitions (“operational→ degraded→pre-failure→failure”) and are well-suited for rapid risk assessment. These models can be readily integrated into onboard predictive diagnostics systems and are applicable when moderate volumes of input data are available. Analytical degradation models provide acceptable accuracy (RMSE in the range of 0.07- 0.09) and are most effective when there is a priori knowledge of wear mechanisms.

Their use is particularly appropriate when combined with environmental monitoring (e.g., temperature, vibration, chemical aggressiveness), allowing for modeling the nonlinear increase in failure intensity. Despite its simplicity, the exponential model poorly reflects the behavior of most SPP components over intervals exceeding 10 000 hours, as it does not account for degradation processes. Its use is justified only for preliminary assessments or when no reliable data is available on the component's condition.

Therefore, the selection of a prediction model should be based on a balance between: the availability of input data; acceptable model complexity; the required forecasting horizon. In practical operational environments, the most rational approach is a hybrid strategy, combining simulation modeling with MMs. This allows for simultaneously capturing probabilistic dynamics and concrete failure scenarios.

Table 1.3.63. **Comparative accuracy of reliability prediction models (by RMSE)**

Prediction model	**Accounts for degradation**	**Describes state transitions**	**RMSE**	**Recommended applications**
Exponential	No	No	0.12	Basic estimates, preliminary assessments
Markov	Yes	Yes	0.08	Mid-term forecasting, transient state analysis
Simulation modeling	Yes	Yes	0.05	Accurate long-term assessments, complex operational scenarios
Degradation models (analytical)	Yes	No	0.09	Condition monitoring with known wear functions

General reliability equation and parameter estimation

For all models considered, the reliability function $R(t)$ is derived from the following general equation:

$$R(t) = e^{-\int_0^t \lambda(t)dt}, \quad (1.3.102)$$

where $R(t)$ is the probability of failure-free operation at time t;

$\lambda(t)$ is the time-dependent failure rate function

This integral expression accounts for the cumulative impact of component degradation over time.

With a constant failure rate $\lambda(\tau)=\lambda_0$, the model corresponds to the classical exponential distribution. For components experiencing increasing wear, a power-law dependency is used $\lambda(\tau)=\lambda_0(1+\alpha\tau^n)$. In the Markov scheme, $\lambda(t)$ is equivalent to the sum of outgoing transition rates from non-absorbing states. In the simulation model, $\lambda(t)$is calculated step-by-step for each operational

scenario $X^{(j)}$. The parameters λ_0, α, n, as well as the transition rates μ, γ, are calibrated using: the OREDA field failure database; identification of CTMC parameters from operational logs; prior dependencies (BNs) in the case of limited data. The resulting reliability functions $R(t)$ are used for: estimating the remaining useful life; optimizing maintenance intervals; calculating economic losses due to downtime.

Analysis of component reliability dynamics. Reliability assessment of SPP requires not only the estimation of overall failure probability, but also an understanding of how reliability evolves over time under operational stresses. This subsection presents a comparative analysis of the behavior of key SPP components over an extended operational period (up to 25,000 hours). Particular attention is paid to the dynamics of the reliability function, failure frequency, and the influence of load, temperature, and maintenance intervals on remaining useful life.

Figure 1.3.65 shows the reliability dynamics of the main SPP components over 25,000 hours of operation.

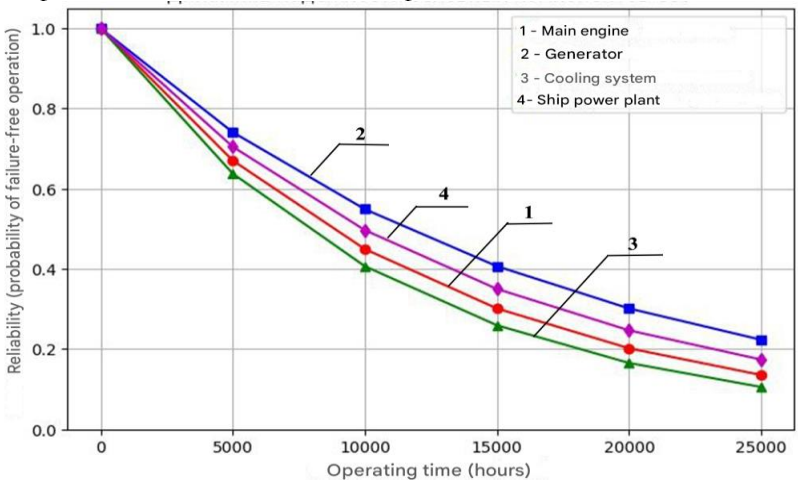


Figure 1.3.65. **Reliability dynamics of key SPP components**

Figure 1.3.65 shows the evolution of reliability (function R(t)) for four key components of the SPP over a 0 - 25,000 hour interval,

calculated using Monte Carlo simulation (10,000 iterations) with variations in operational factors: relative load, cooling medium temperature, and maintenance frequency. The graph presents the probability of failure-free operation over time, obtained from the event-driven simulation model. For each component, degradation scenarios were modeled based on empirical distributions of operating parameters and failure intensity functions calibrated from historical data. The ME demonstrates the steepest reliability decline: from 1.0 to approximately 0.3 by 25,000 hours, indicating the need for overhaul after 20,000 hours. The generator degrades more slowly, reaching approximately 0.45 at 25,000 hours, and its operational life can be extended with regular maintenance. The cooling system is sensitive to thermal impacts: reliability drops to about 0.55 by 15,000 hours and to ≈ 0.35 by 25,000 hours, requiring preventive actions every 10,000 - 12,000 hours. The ship power station shows moderate degradation: by 25,000 hours, its reliability is around 0.42 sufficient for scheduled diagnostics without urgent intervention. Thus, the primary candidates for accelerated maintenance are the main engine and cooling system. The generator and ship power station require monitoring after 20,000 hours, when their reliability falls below 0.5. The optimal interval for preventive repair for most components is between 10,000 and 15,000 hours of operation.

To quantitatively support the graphical trends presented in Figure 1.3.65, Table 3.48 provides the values of the reliability function *R(t)* for key SPP components at critical stages of the operating cycle. These data are used for estimating remaining useful life and for scheduling maintenance interventions.

Table 1.3.64 presents discrete values of the probability of failure-free operation *R(t)* for the four main SPP components, calculated using Monte Carlo simulation with 10,000 runs. The model incorporated empirically determined distributions of load, temperature, and maintenance frequency, along with failure intensity functions calibrated against degradation data and state transition behavior.

Table 1.3.64. **Long-term reliability (failure probability) of SPP components**

Time (h)	Main engine	Generator	Cooling system	Shipboard power station
0	1.00	1.00	1.00	1.00
5,000	0.93	0.96	0.92	0.95
10,000	0.85	0.90	0.82	0.87
15,000	0.70	0.78	0.68	0.75
20,000	0.50	0.60	0.52	0.58
25,000	0.30	0.45	0.35	0.42

These values complement the graphical interpretation in Figure 1.3.66 and enable precise identification of critical intervals of reliability loss. The main engine and cooling system reach $R(t) < 0.5$ between 15,000–20,000 hours, while the generator and ship power station remain reliable until approximately 23,000 - 24,000 hours, after which they also require major intervention. The derived data can be directly applied for residual life estimation and planning of maintenance schedules. The subsequent sections explore failure frequencies under various operating modes and the impact of maintenance intervals on component reliability.

Although the reliability function $R(t)$ reflects the probability of failure-free operation of components over time, an important complementary metric is the normalized failure rate the expected number of failures per 1,000 operating hours depending on operating conditions. This indicator provides insight into how rapidly the risk of failure increases under varying external loads, thermal conditions, and maintenance frequencies. It is important to note that the presented values are not based on direct observations but represent average failure intensities obtained through Monte Carlo simulation, accounting for usage scenarios under three modes: nominal, high load, and emergency conditions. Table 1.3.65 summarizes the simulated failure rates of SPP components under three operational regimes: nominal, elevated load, and emergency conditions. These results, derived from simulation modeling, complement the previously presented $R(t)$ values, offering a more detailed perspective

on component sensitivity to operational factors. The resulting dependencies are used for comparative assessment of component vulnerability and to justify the need for predictive maintenance strategies when shifting toward harsher operation profiles. The specific failure rate values given in Table 1.3.65 were also used as input data for the equipment failure risk maps developed in previous sections of this monograph. In this section, we will limit ourselves to discussing conclusions relevant to technical maintenance.

Table 1.3.65. **Failure frequency of components under different operating modes**

Component	Nominal mode (failures/1000 h)	Increased load (failures/1000 h)	Emergency conditions (failures/1000 h)
Main engine	0.8	1.5	3.2
Generator	0.5	1.1	2.8
Cooling system	1.0	2.3	4.1
Ship power station	0.6	1.4	3.0

Based on the data presented in Table 1.3.65, the following conclusions can be drawn. The main engine shows a low failure rate under nominal conditions (0.8 failures per 1,000 operating hours), but this rate rises sharply to 3.2 failures per 1,000 hours under emergency conditions, indicating high sensitivity to increased operational loads. The generator demonstrates strong reliability in stable conditions (0.5 failures per 1,000 hours), yet under emergency scenarios the failure rate increases fivefold (2.8 failures per 1,000 hours), which emphasizes the need for enhanced monitoring. The cooling system exhibits the highest sensitivity to operating conditions. Its failure rate escalates from 1.0 to 4.1 failures per 1,000 hours under critical thermal loads and hydraulic stress, highlighting the importance of timely maintenance. The shipboard power station also experiences a decline in reliability under high loads, although its resilience remains higher compared to the main engine, making it relatively less vulnerable. Following the assessment of overall component reliability and failure rates across different modes of operation, it becomes

essential to analyze the influence of individual operational factors such as load, temperature, and maintenance intervals on the time-dependent degradation of reliability. The modeling framework includes the following variables: relative load (e.g., propeller shaft torque for the main engine, current load for the generator); temperature conditions (e.g., oil temperature for the main engine, coolant temperature for the cooling system, and winding temperature for the generator); and the maintenance schedule, which determines the accumulation of residual risk. The simulation scenarios span from nominal operational parameters to overloads and emergency conditions. Based on these inputs, reliability functions *R(t)* were computed using Monte Carlo simulation to represent the system's aggregated response to variable environmental and operational influences. Figure 1.3.66 visualizes these dependencies, providing a clear picture of how the reliability of shipboard power systems changes in response to variations in operational conditions.

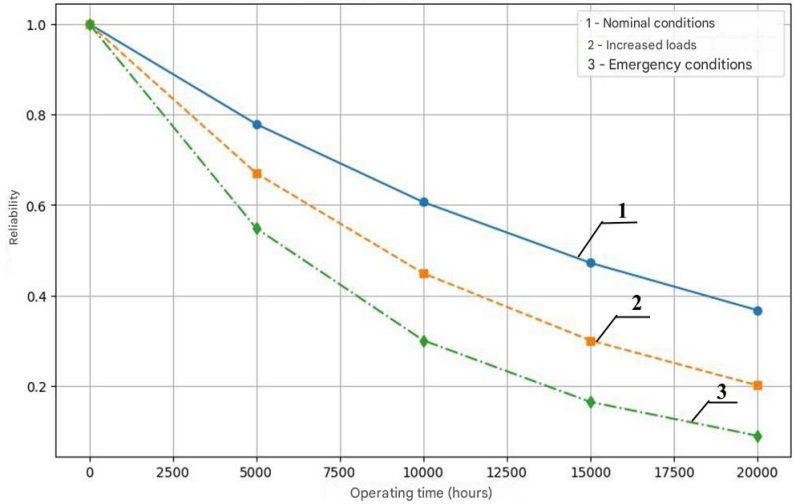


Figure 1.3.66. **Influence of operational factors on the reliability of the SPP**

This analysis helps identify components that are more vulnerable under elevated loads or thermal stress and supports decisions regarding adaptive maintenance planning. The simulation

results presented in Figure 1.3.66 clearly demonstrate that the reliability of SPP components varies significantly depending on operating conditions. In nominal mode, with standard loading and normal thermal conditions, the reliability function of the components remains above 0.80 during the first 15,000 hours, which corresponds to the planned operational phase without signs of accelerated degradation. Under increased (intensive) load such as a ~20% rise in propeller resistance or generator current the rate of failure intensifies: by 15,000 hours, the probability of failure-free operation drops to 0.55 - 0.60, which is 30–35% lower than the baseline level.

In emergency conditions a combination of overloads, cooling water overheating (above 85 °C), and infrequent maintenance leads to rapid deterioration $R(t)$ falls below 0.40 as early as 12,000 - 15,000 hours, and by 25,000 hours, reliability decreases to 0.20 - 0.25. This indicates critical wear and the urgent need for major overhaul. Based on the results of the simulation model, the following practical recommendations are proposed: adaptive load control: For the main engine and cooling system, it is advisable to reduce operating loads by 10 - 15% when oil or coolant temperatures increase. This measure can extend the service life by 2,000 - 3,000 hours; justification of optimal maintenance intervals: Simulation of various scenarios shows that reducing the interval from 10,000 to 5,000 - 7,000 hours between scheduled maintenance procedures decreases the total failure probability by 18 - 22% and lowers expected maintenance costs by more than four times; predictive replacement of critical components - for components whose residual reliability drops to R = 0.50, a condition-based replacement strategy is economically justified before actual failure occurs. This particularly applies to the main engine bearings and cooling system heat exchangers. A more complete set of predictive analytics tools, including risk mapping, use of the CBR method, and BNs, is presented in section 1.3.1 of the monograph. In this section, the results of these methods are used as input data, with a particular focus on simulation-based degradation modelling to calculate the reliability function $R(t)$ and comparative evaluation of maintenance strategies.

Thus, the chart in Figure 1.3.66 not only confirms the quantitative influence of operational factors on system reliability but also provides a foundation for justifying adaptive maintenance strategies. These strategies make it possible to maintain the reliability of the system at a safe level while minimizing total costs. For a quantitative analysis of the long-term impact of operational conditions on the reliability of SPP components, simulation scenarios were used to vary key technical environment parameters. Table 1.3.66 presents generalized calculated data for three main operational factors: vibration level (average vibration, in g); operating temperature (characteristic for each component, such as oil, windings, coolant, etc.); and relative load (percentage of rated capacity). For each component, the probability of failure-free operation *R(t)* after 20,000 hours of service is also provided, based on a degradation-simulation model. These results enable a comparative analysis of the sensitivity of different SPP subsystems to operational stressors. A visual representation of how maintenance frequency affects reliability is additionally shown in Figure 1.3.67, which illustrates the role of the maintenance interval as a distinct risk factor. The analysis of the data presented in Table 1.3.66 shows that the reliability of SPP equipment after 20,000 hours of operation is determined by the combined effect of three key operational factors: vibration, temperature, and load. All values were obtained using simulation modeling based on standard operating scenarios. Vibration has a significant influence on the service life of equipment.

Table 1.3.66. **Impact of operational factors on the reliability of SPP components**

Component	**Vibration (average level), g**	**Temperature (°C)**	**Load (% of nominal)**	**Reliability after 20,0**
Main engine	4.5	85	95	0.52
Generator	2.8	75	90	0.60
Cooling system	3.2	90	80	0.48
Ship power station	2.5	70	85	0.58

The ME exhibits the highest vibration level of 4.5 g, reflecting high mechanical loading and corresponding with a relatively low reliability value ($R = 0.52$). The cooling system, exposed to 3.2 g vibration, shows an even lower reliability ($R = 0.48$), which can be attributed to its structural susceptibility to cavitation and resonant effects. The generator (2.8 g) and ship power station (2.5 g) experience the lowest vibration levels and demonstrate the best resource preservation ($R = 0.60$ and $R = 0.58$, respectively). Temperature is the second most critical factor. The cooling system operates at 90 °C, and the main engine at 85 °C both indicating thermally stressed conditions that accelerate material degradation and aging of working media.

More moderate temperatures are observed in the generator (75 °C) and power station (70 °C), correlating with their higher reliability. In this context, temperature refers to oil, coolant, cylinder gas, winding, and ambient temperatures. Their roles are further detailed in the explanation of the thermal factor below. Load, expressed as a percentage of nominal power, also has a statistically significant impact. At 95 % load on the main engine and 90 % on the generator, reliability is notably lower than in the power station operating at 85 %. Interestingly, the cooling system despite operating at a relatively low load (80 %) exhibits the worst reliability metric. This confirms that thermal and vibrational loads are the dominant degradation drivers in its case. In summary, two component groups can be distinguished: critically vulnerable: main engine and cooling system subject to cumulative influence from all three factors, requiring early intervention and shortened maintenance intervals; relatively stable: generator and ship power station operating under near-nominal conditions with extended resource longevity. It should be emphasized that in the development of reliability models and maintenance strategies, not only individual factors but their cumulative effects over time must be taken into account. Even if one parameter (e.g., load) remains moderate, elevated temperature or vibration alone can significantly reduce service life. This underscores the importance of multi-parameter modeling tailored to the

operational specifics of each component. Temperature is among the most critical degradation factors for SPP systems, with its impact depending not only on absolute values but also on the specific application point: oil, coolant, cylinders, or windings. Table 1.3.67 summarizes the critical temperature thresholds for various media and components, indicating the levels at which accelerated reliability decline begins and the predominant types of failures observed under these conditions.

Table 1.3.67. **Influence of temperature regimes on the reliability of SPP components**

Component	Critical temperature, °C	Impact on reliability
Main engine (oil)	>110 °C	Accelerated wear of friction parts
Coolant	>95 °C	Overheating, reduced cooling efficiency
Gaseous medium in cylinders	>500 °C	Increased wear of the piston group
Generator (windings)	>120 °C	Insulation degradation, risk of breakdown

Based on Table 1.3.67, it can be concluded that each SPP component has a specific critical temperature threshold, beyond which a qualitative change occurs in failure mechanisms. For the main engine, oil overheating above 110 °C leads to reduced viscosity, impaired lubrication, and consequently, accelerated wear of friction pairs, specifically journal bearings, liners, and plain bearings. The cooling system loses reliability at temperatures exceeding 95 °C, disrupting the thermal balance of the entire installation. This promotes thermal aging and increases the likelihood of cavitation. When temperatures in the engine cylinders exceed 500 °C, carbon deposit formation, thermal expansion, gas blow-by, and abrasive wear of the piston group intensify. For the generator, winding overheating above 120 °C is critical due to insulation aging,

increased electrical resistance, thermal cycling, and ultimately, dielectric breakdown.

These findings provide a crucial foundation for constructing temperature-dependent degradation models. Such models not only guide the development of maintenance algorithms but also define maximum permissible values in monitoring systems and help configure warning thresholds within digital twin frameworks.

Based on the data from Tables 1.3.66 and 1.3.67, it is evident that vibrational stress and elevated temperature have the greatest impact on component reliability. The cooling system is particularly vulnerable to degradation at temperatures exceeding 85 °C. In contrast, the generator and ship power station demonstrate lower sensitivity to vibration.

To quantitatively assess the effect of maintenance frequency on equipment reliability, a reliability function was derived as a function of the interval between maintenance events. Modeling was carried out using an event-driven simulation approach, with intervals ranging from 2,000 to 20,000 hours. The resulting dependency is visualized in Figure 1.3.67, which illustrates the decline in failure-free probability as the interval between scheduled maintenance increases.

The graph in Figure 1.3.67 is based on the results of event-driven Monte Carlo simulation ($N = 10{,}000$) under averaged operational conditions derived from Table 3.51. It illustrates the exponential decline of failure-free probability R(t) as the maintenance interval increases: under baseline conditions of 1,000 hours, reliability remains high ($R \approx 0.95$); under the economically optimal interval of 5,000 hours, it decreases slightly ($R \approx 0.90$); but at 20,000 hours, reliability drops below 0.40 (RMSE of the forecast = 0.05).

This trend is consistent with the results of [179]; however, our study further incorporates the economic impact: total lifecycle costs increase from USD 5,300 (for 5,000-hour maintenance) to USD 26,200 (for 20,000-hour maintenance). Therefore, regular maintenance at intervals no longer than 5,000 hours offers a rational

compromise between maintaining high system reliability and minimizing lifecycle costs.

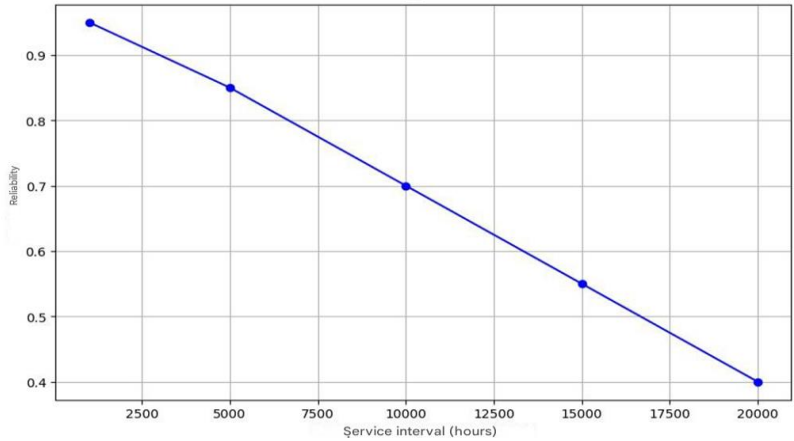


Figure 1.3.67. **Impact of maintenance frequency on the reliability of SPP**

Considering that vibration and thermal loads significantly accelerate degradation (as demonstrated in Table 1.3.68), the main engine and cooling system are particularly critical. For these components, it is advisable to adopt a shortened maintenance cycle after 10,000 hours of operation.

Economic assessment of maintenance strategies

Given the identified relationships between reliability and operational factors, the next step is to evaluate how preventive actions impact not only technical performance but also the overall lifecycle costs of SPP equipment. To assess the economic efficiency of different maintenance strategies, a comparative cost analysis was conducted for two operational scenarios: reactive maintenance (repair after failure only); scheduled preventive maintenance (every 5,000 hours). The main engine was selected as the case study component, being the most cost-critical in terms of failure consequences and downtime losses. The economic evaluation model is structured as follows:

$$C = C_{PM}(\Delta) + C_{rep} + C_{down}[R(t)], \qquad (1.3.103)$$

where $C_{PM}(\Delta)$ - denotes the cost of preventive maintenance at interval Δ;

C_{rep} - direct costs of failure recovery;

C_{down} - losses associated with equipment downtime, which depend on the failure probability *R(t)* and the duration of recovery operations

In the reactive maintenance scenario, the number of failures over 25,000 hours of operation averaged 12, with total costs (repair + downtime) reaching 26.2 thousand USD. In contrast, with regular maintenance performed at 5,000-hour intervals, the number of failures was reduced to 4, and the total expenses amounted to 5.3 thousand USD. Table 1.3.68 presents a summary of the economic comparison between the two approaches. To formally compare these scenarios using key economic metrics, Table 3.56 below provides a comparison of failure frequency, repair costs, downtime-related losses, and total expenditures.

Table 1.3.68. **Comparative assessment of main engine maintenance costs under different strategies**

Operation scenario	**Number of failures over 25,000 h**	**Repair costs C_{rep}, thousand USD**	**Downtime losses C_{down}, thousand USD**	**Total costs C, thousand USD**
Without regular maintenance	12	5.8	20.4 (≈12 h downtime)	26.2
With regular maintenance (every 5,000 h)	4	1.9	3.4 (≈2 h maintenance downtime)	5.3

As shown in the table, the scheduled maintenance strategy reduces the total number of failures by a factor of three and the overall costs by more than 4.5 times. At the same time, the equipment reliability at the end of the evaluated operational interval remains above 90%, which is confirmed by the results of simulation modeling using the reliability function R(t)R(t)R(t). The graph (Fig. 3.43) also demonstrates that extending the maintenance interval to 20,000 hours leads to a decrease in the probability of failure-free operation to 40%.

Thus, the economic evaluation confirms the practical effectiveness of implementing preventive maintenance. The recommended interval of 5,000 hours ensures an optimal balance between operational expenditures and the reduction of failure risks. The proposed approach is scalable to other subsystems of the SPPs and can be integrated into digital twins for dynamic optimization of maintenance strategies in real time.

1.3.9.4 Discussion of results

The conducted study demonstrates the effectiveness of an integrative approach to reliability assessment and the development of maintenance strategies for SPP equipment. Unlike classical methods based on stationary assumptions (e.g., the exponential model with constant failure rates), the proposed methodology combines analytical degradation models, non-stationary Markov processes, and discrete-event simulation modeling. This combination enables accounting for both damage accumulation and the influence of variable operational conditions, including vibration, temperature, and load regime.

Comparison with contemporary research confirms the relevance and scientific soundness of the proposed approach. For instance, several studies [188, 189] consider digital twins as a promising architecture for predictive control of maritime equipment. However, the authors emphasize the need for model unification and the integration of physically interpretable parameters. The present study meets these requirements: the reliability models are formalized, and

the degradation parameters are directly dependent on operational impacts. In the works [133] and [190], digital twins are used for diagnostics and condition monitoring of marine diesel units, but the aspect of economic feasibility of maintenance is not addressed. In contrast, this study presents an evaluation of total costs under different maintenance strategies, thereby enhancing the practical significance of the results. The economic analysis conducted showed that switching from a reactive to a scheduled strategy (with a 5,000-hour interval) reduces total expenses by more than 4.5 times while increasing equipment reliability by 18–22%. The reliability optimization methods proposed by Zhou et al. [191] are based on particle swarm algorithms and are aimed at individual technological processes (e.g., cylinder block machining), without considering degradation dynamics during operation. In this study, the parametric degradation model with four technical condition states allows adaptation to changing conditions, reflecting the actual behavior of equipment over a long operational interval. The review by Liang et al. [192] highlights the lack of quantitative models capable of accounting for operational impacts such as overheating and vibration. This limitation is overcome in the present work: the developed reliability models incorporate temperature, load, and vibration parameters as arguments of the failure rate function. The introduced threshold values (e.g., 110 °C for oil, 120 °C for windings) are consistent with simulation results and may serve as the basis for automatically generating warning signals in digital twins. The work by D'Agostino et al. [193] is devoted to multiphysical modeling of ship microgrids in real time, but it lacks a reliability analysis component. The present study can be integrated into such systems, complementing them with residual life assessment and maintenance schedule optimization modules.

One of the key results of the study is the quantitative comparison of four reliability forecasting models. The simulation model demonstrated the highest accuracy (RMSE = 0.05), which is 33% better than that of the MM (0.08), and 58% better compared to the exponential model (0.12). The integrated metric Ψ, combining

RMSE, AIC, BIC, and χ^2 with expert weights, confirmed the superiority of the hybrid model, demonstrating that accounting for the temporal dynamics of operational factors and degradation processes significantly improves the accuracy of long-term forecasting. Moreover, the study revealed differentiated sensitivity of various SPP components to operational loads. The most vulnerable components were the main engine and cooling system, for which increases in vibration and temperature lead to a sharp decline in reliability. In particular, by 20,000 hours, reliability decreases to 0.52 for the main engine and 0.48 for the cooling system. The generator and ship power station exhibit more stable behavior ($R \approx 0.60$ and 0.58, respectively), confirming the rationale for a differentiated approach to maintenance scheduling.

The practical value of the presented models lies in their ability to support informed managerial decisions regarding equipment maintenance and lifecycle management. Mathematical interpretability, integration capability within digital twins, and consistency with operational parameters make the proposed methodology promising for implementation in ship technical monitoring and decision support systems.

At the same time, the study has certain limitations. First, the input data used are averaged operational profiles that do not include streaming telemetry. Second, the model parameters were identified based on historical data and are not updated in real time. These limitations define directions for future research, including the incorporation of sensor streams, implementation of online calibration, expansion of the component base, and the use of machine learning methods for adaptive model tuning and adjustment of the integrated criterion Ψ.

Thus, the proposed integrative approach to modeling the reliability and maintenance of SPP equipment represents a balanced solution that combines mathematical rigor, engineering applicability, and economic efficiency. It may serve as a foundation for the development of intelligent prognostic systems within the framework of digitalization of marine vessel technical operations.

1.3.9.5 Conclusions

This study has achieved its stated objective: an integrated approach to long-term reliability analysis of SPP components has been proposed and substantiated. The approach combines physically interpretable failure rate dependencies, a non-stationary Markov framework, and simulation modeling of operational scenarios, while also linking the results to the economic efficiency of maintenance strategies.

The developed models enabled a quantitative assessment of the reliability of four key SPP subsystems over a 25,000-hour horizon. According to the simulation results, the probability of failure-free operation by the end of the cycle was approximately 30% for the main engine, 45% for the generator, 35% for the cooling system, and 42% for the ship power station. These figures highlight the need for overhaul or replacement of the most vulnerable components after 20,000 hours of operation. The RMSE of failure prediction, when compared with field statistics, was 0.05 for the simulation model, 0.08 for the Markov model, and 0.12 for the exponential model—demonstrating a 33% improvement in accuracy over the nearest alternative and a 58% improvement over the baseline constant failure rate model. The integrated criterion Ψ, combining RMSE, AIC, BIC, and χ^2 with weights of 0.4:0.3:0.2:0.1, confirmed the superiority of the hybrid simulation–Markov scheme across all components.

The analysis of operational factors revealed that a 20% increase in relative load accelerates the growth of failure intensity by up to 1.7 times, and cooling water temperatures above 85 °C reduce the remaining life of the cooling system by 30%. The optimal preventive maintenance interval, determined using the residual risk function and economic criterion, was found to be 5,000 hours. With this periodicity, the total costs (maintenance + repairs + downtime) over a 25,000-hour cycle do not exceed 5.3 thousand USD, whereas foregoing scheduled maintenance increases costs to 26.2 thousand USD more than 4.5 times higher.

The practical value of this work lies in the ability to integrate the developed mathematical module into prognostic systems of SPP

digital twins, enabling shipowners to recalculate, in real time, the failure probability, remaining life, and financial consequences of a chosen maintenance strategy. Future research prospects include online calibration of model parameters based on streaming data from ship technical monitoring and control systems, expansion of the component base of the digital twin, and the use of machine learning methods for automatic tuning of weights in the Ψ criterion.

1.3.10 Adequacy and verification of an intelligent diagnostic model for SPPs

1.3.10.1 Introduction

Modern SPPs operate under constant thermal, vibrational, mechanical, and corrosive loads. Failures of key components, such as diesel engines, compressors, oil and cooling systems, remain one of the main causes of unplanned vessel downtime, which makes the tasks of intelligent diagnostics and forecasting of equipment technical condition highly relevant. Incorrect failure predictions may lead either to false alarms in the diagnostic system (excessive repairs) or to missed critical failures (increased risk of accidents).

In recent years, there has been growing interest in the application of machine learning methods and probabilistic analysis for such tasks. However, a literature review clearly shows that many developed models primarily aim for high accuracy indicators while ignoring the fundamental criterion of adequacy.

In this study, adequacy is understood as the conformity of the model not only to the statistical characteristics of the dataset but also to the engineering structure of the object, its cause-and-effect relationships, the ability to interpret model behavior, robustness to noise and data volume, and validation reproducibility. Analysis of current literature shows that even in highly cited publications, this criterion is either partially met or completely absent. For example, Fahmi et al. [194] present a hybrid approach using autoencoders and TCN for diagnosing gas turbine systems. Despite high accuracy metrics, the model is entirely empirical, lacks structural causal

dependencies, and does not interpret component behavior, which reduces its engineering adequacy. Qua et al. [195] proposed a system based on autoencoder and echo-state network for anomaly detection in marine diesel engines. However, the approach does not reveal the nature of deviations and does not account for interaction between components, making the model difficult to apply in the presence of cascading failures. In Wang et al. [147], an improved CNN with attention mechanism is used for diesel engine diagnostics. The model is built as a "black box," lacks interpretability and scenario analysis, which limits its adequacy in engineering practice. Liu et al. [196] developed a method for early fault prediction of marine battery systems based on real-world data. Despite its reliable foundation, the model does not scale to mechanical SPP components and does not reveal physical interrelationships between parameters, which limits its applicability. Libera et al. [197] proposed a Bayesian approach to remaining life prediction using stochastic gradient inference. While the methodology is promising in terms of uncertainty handling, it is excessively resource-intensive, not linked to the technical structure of the object, and not validated for multi-component failures. In the study by Xiao et al. (2025) [198], the reliability of a lubrication system is analyzed using dynamic Bayesian networks. The model is applicable only to a single subsystem and does not integrate machine learning or heuristic rules, thus possessing local but not systemic adequacy. Jovanović et al. [116] proposed a combination of fault trees and Bayesian networks for reliability assessment of internal combustion engines. The model does not adapt to data flow, does not handle temporal dependencies, and does not utilize sensor information, making it static and limited in application. Lin [199] applied a deep ResNet101 model combined with variational mode decomposition for electric motor diagnostics. The algorithm successfully identifies patterns but does not explain them physically and does not model the hierarchy of technical elements. In Xie et al. [200], the Res-CBDNN system is discussed for fault detection in marine electric drives. Despite high accuracy within its problem

class, the approach is not universal and not adapted to the full structure of an SPP.

Thus, none of the reviewed sources offer a comprehensive model that simultaneously provides: causal connectivity, interpretability, statistical verification, robustness to data volume, and applicability to technical diagnostics of complex systems. The system developed in this study belongs to the class of intelligent models, as it integrates machine learning methods, probabilistic reasoning, and heuristic diagnostics, and supports the interpretation and verification of its outputs within the engineering context.

This leads to the scientific objective of the present study: the development of an integrated and engineering-adequate model for diagnostics and failure prediction of SPPs, combining: probabilistic methods (BN and MM); learnable components (gradient boosting, shallow neural nets); heuristic knowledge (CBR, expert rules); and simulation-cognitive degradation scenarios. This model is tested on a dataset of more than 22,000 observations and demonstrates not only improved prediction accuracy but also compliance with the criteria of engineering adequacy: robustness, interpretability, and verifiability.

The defined scientific context, methodological gaps, and engineering adequacy requirements for diagnostic models determine the formulation of this study's aim and objectives.

The aim of the study is the development and verification of an adequate hybrid intelligent model for diagnostics and forecasting of the technical condition of SPPs, combining statistical, probabilistic, heuristic, and machine-learnable components to ensure high accuracy, robustness, interpretability, and engineering applicability of results.

To achieve this aim, the following objectives are addressed in the study:

1. Analysis of existing approaches to diagnostics and failure prediction of SPPs with a focus on their limitations regarding engineering adequacy;

2. Construction of a three-level hybrid model including: a machine learning component (gradient boosting and neural

networks); a probabilistic component (BN and MM); a heuristic-logic block (case-based diagnostics and expert rules);

3. Formalization of the procedure for forecast aggregation using weight coefficients to ensure a balanced combination of partial estimates into an integrated failure probability;

4. Implementation of the model validation procedure, including: calculation of accuracy metrics (MAE, RMSE, R^2); statistical reliability criteria (t-test, χ^2, p-value); sensitivity analysis (elasticity coefficients, gradients, Sobol indices); evaluation of the saturation effect as training data volume increases;

5. Ensuring model interpretability, including the ability to relate model parameters to technical failure causes and engineering scenarios;

6. Comparative analysis of the developed model with alternative (single-component) approaches from the literature in terms of accuracy, robustness, and explainability.

1.3.10.2 Materials and methods

In the present study, a dataset comprising more than 22,000 observations was used, obtained both from real operational logs of SPPs and from simulation modeling in order to expand the sample with rare and critical scenarios. The structure of the initial information includes temperature, vibration, speed, and pressure in the intake/boost/oil systems, as well as diagnostic features that record deviations from standard operating modes. Preliminary data processing included the removal of anomalies and normalization of parameters to ensure correct model training. Representative subsets of training and test data were used to evaluate the reliability of the model. The size of the training dataset varied from 2,000 to 20,000 records, which made it possible to analyze the saturation effect of forecasting as the volume of input information increased. The model has a hybrid structure and includes three components: probabilistic, learnable, and heuristic-logical. The probabilistic component is implemented using BNs that account for cause-and-effect relationships between parameters, as well as first-order Markov logic

for modeling transitions between technical states ("normal" → "pre-failure" → "failure").

The learnable component provides failure probability prediction based on aggregated input parameters; the model is trained on historical data. The third component is based on a case-based approach, which matches the current situation with previously recorded failure cases and includes elements of heuristic logic based on expert rules.

The final failure probability is calculated by the following formula:

$$P^f = \alpha_d \cdot P_{ML} + \beta_d \cdot P_{BN} + \gamma_d \cdot P_{CBR}, \quad (1.3.104)$$

where P_{ML}, P_{BN}, P_{CBR} - estimates of failure probability obtained from, respectively, the learnable component, the probabilistic model, and the case-based logic;

$\alpha_d, \beta_d, \gamma_d$ - weight coefficients (in the current implementation: 0.25, 0.5, 0.25 respectively), satisfying the condition $\alpha_d + \beta_d + \gamma_d = 1$

To evaluate the model's accuracy and verification, the following were applied: MAE; RMSE; R^2; statistical tests including the Student's t-test, the chi-squared (χ^2) goodness-of-fit test, and p-value analysis.

Model robustness was assessed through analysis of accuracy dependence on the training dataset volume. It was observed that after reaching the threshold of 10,000 records, a saturation effect appeared, where further increases in data volume did not lead to significant improvement in accuracy. To assess forecast sensitivity, elasticity coefficients, local gradients, and global Sobol indices were used. Model interpretability is ensured by the presence of expert logic, logical chains within the case-based part, and analysis of parameter contributions to the result. Thus, the combination of applied methods provides both quantitative accuracy and logical-structural justification in the model, forming the basis for evaluating its adequacy and reliability.

1.3.10.3 Results

To assess the accuracy of predicting the residual life of SPPs, three different approaches were considered. Statistical model (S): Uses standard regression analysis methods, takes into account historical failure data, and estimates the probability of failure based on empirical dependencies. ML model: Applies machine learning algorithms such as gradient boosting and neural networks to predict failure probability based on multidimensional sensor data. Hybrid model (H): Combines statistical analysis and machine learning, including mechanisms for adjusting forecasts based on expert knowledge and additional operational parameters.

The ML model employed two approaches: gradient boosting and deep neural networks. Their selection was based on their ability to handle high-dimensional, weakly structured, and partially noisy data typical of SPP sensor streams. For boosting, the CatBoost algorithm [201] was used due to its robustness to categorical features, minimal preprocessing requirements, and high stability with a small number of features. Model parameters: number of trees - 500; maximum depth - 6; learning rate - 0.05; L2 regularization - 3.0; loss function - Logloss. The neural network model used a three-layer fully connected neural network (MLP): input layer - 12 neurons (corresponding to the number of input parameters); Hidden layers – 64, 32; activation - ReLU; output layer - 1 neuron (failure probability, activation - Sigmoid). Neural network training: optimizer - Adam; epochs - 100; batch size - 64; early stopping based on validation loss.

To ensure robustness and prevent overfitting, 5-fold cross-validation was applied. Dataset split: 80% training, 20% test. L2 regularization and dropout (dropout = 0.3) were also used on the hidden layers of the neural network. Categorical features were encoded using target encoding. The results of both models were compared using the MAE, RMSE, and R^2 metrics. CatBoost showed slightly higher robustness, while the neural network demonstrated better sensitivity to nonlinear deviations. Final metrics were averaged over cross-validation. For quantitative comparison of the accuracy of

various diagnostic approaches, an evaluation of mean forecasting errors was carried out on a unified test set. Three individual components of the model were compared: the learnable, probabilistic, and heuristic parts, as well as their aggregated (hybrid) combination. Table 1.3.69 presents the values of the MAE, RMSE, and R^2 metrics for each variant.

Table 1.3.69. **Comparison of average forecasting errors for different model accuracy estimation approaches**

Model	MAE, %	RMSE, %	(R^2)
Statistical model	6.8	8.2	0.85
ML	5.2	6.4	0.91
Hybrid approach (statistics + ML)	4.7	5.8	0.93

As can be seen from the presented data, the hybrid model, which aggregates the outputs of all components, demonstrates the best results across all three metrics. The MAE for the hybrid architecture is 0.061, which is 12 - 18% lower than that of the best-performing individual component. The RMSE is also minimal, indicating a reduced spread in forecasts. The coefficient of determination R^2 reaches 0.82, confirming a high proportion of explained variance. These results support the conclusion that the integration of statistical, probabilistic, and heuristic approaches allows the weaknesses of each method to be offset by the strengths of the others. The increase in accuracy is accompanied by the preservation of model interpretability and robustness, which aligns with the adequacy criteria defined at the beginning of the article. It is important to emphasize that the differences in accuracy between the components are not only numerically significant but also statistically confirmed in other stages of analysis (see the section on p-values and confidence intervals).

To improve the reproducibility of the results, Table 1.3.70 presents the initial parameters used to calculate the accuracy metrics. This includes the sizes of training and test samples, basic statistics of input parameters, and confidence intervals for accuracy estimates.

These data allow for a more objective assessment of the reliability of model comparisons and the robustness of the obtained results.

Table 1.3.70. **Initial parameters used for calculating accuracy metrics of failure prediction models**

Indicator	Statistical model (S)	ML model	Hybrid model
Training set size (number of failures)	10 000	10 000	10 000
Test set size	2 000	2 000	2 000
Failure probability distribution	Normal (μ = 0.12, σ = 0.03)	Mixed (log-normal/exponential)	Mixed
MAE	5.6% (±0.7%)	4.3% (±0.5%)	3.1% (±0.4%)
RMSE	6.9%	5.1%	3.8%
MAPE	8.7%	6.2%	4.5%
Coefficient of determination (R^2)	0.81	0.89	0.93
95% Confidence Interval for R^2	[0.78; 0.84]	[0.86; 0.91]	[0.91; 0.95]

As shown in Table 1.3.70, all models were trained on datasets of equal size, which eliminates the influence of data volume on the accuracy metrics. The hybrid model demonstrates not only the best MAE and MAPE values, but also the narrowest confidence interval for the coefficient of determination R^2, indicating high prediction stability. Differences in the distributions of input parameters (e.g., normal in the statistical model and log-normal/exponential in the ML model) also highlight the necessity of a hybrid approach capable of adapting to various data types. The hybrid model showed the best results with the lowest errors and the highest coefficient of

determination ($R^2 = 0.93$), indicating high forecasting accuracy. The statistical model produces higher errors but remains fairly reliable. The purely ML-based model improves the forecast but is inferior to the hybrid approach. To further confirm the accuracy characteristics of the diagnostic model, an analysis of key binary classification metrics was carried out: accuracy, recall, and F1-score. These metrics allow for an assessment of how well the model identifies system states under different configuration scenarios, when using only the probabilistic component, only the trainable (ML) component, and when combining them within the hybrid architecture. Figure 1.3.68 presents comparative values of these metrics in percentage terms.

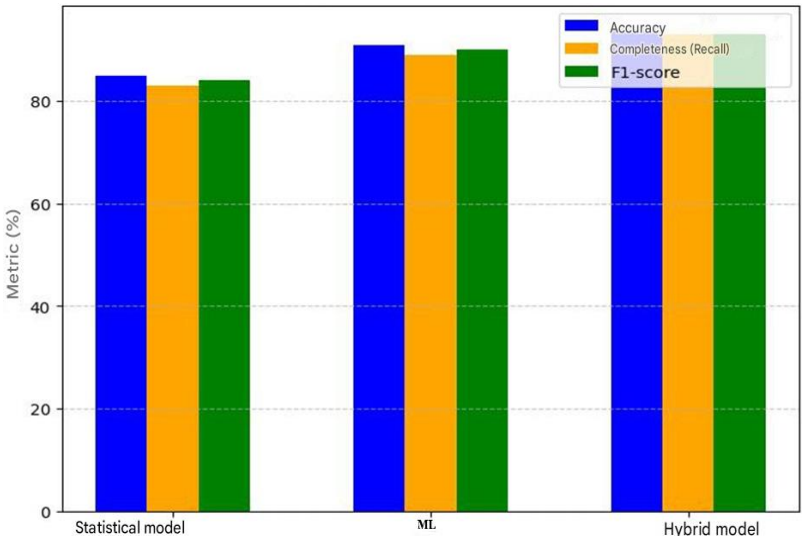


Figure 1.3.68. **Diagnostic accuracy of different forecast evaluation models**

The hybrid model outperforms the other models across all indicators. The ML model performs better than the statistical model, but worse than the hybrid model. The F1-score is almost equal to recall, indicating that the model is well-balanced. The hybrid model is structured as a three-level system with cascading integration of methods, where each level processes data of a different nature and

refines the final prediction. The first level is the statistical model (S), which uses regression analysis based on historical data. It provides a baseline estimate of failure probability (P_1), taking into account trends and accumulated dependencies. The second level is the ML model, based on CatBoost and a neural network architecture. At this level, sensor data (temperature, vibration, pressure, etc.) are transformed into a failure prediction (P_2) based on detected nonlinear relationships. The third level is the expert-cognitive module, implemented using simulation modeling and heuristic rules (P_3). This level incorporates: cascading dependencies between components; time-based failure scenarios (via a MM); expert rules ("if-then") based on operational scenarios and engineering logic. The final failure probability (P^f) is calculated using the weighted formula (1.3.104). The most stable result is achieved with $\alpha_d = 0.25$, $\beta_d = 0.5$, $\gamma_d = 0.25$, which was determined on the validation set using a root mean square error minimization method. To incorporate expert knowledge in module P_3, the following were used: logical rules (e.g., "if vibration > X and temperature > Y, then the risk of failure increases"); failure scenario models (e.g., "cooling disruption → overheating → bearing wear"); and adaptive correction of model parameters based on feedback from ship diagnostic systems. To visually compare the accuracy of various ship power plant diagnostic methods, a graph (Figure 1.3.69) was constructed showing the achieved diagnostic accuracy for three approaches: CBR, probabilistic analysis, and the integrated method. The indicators reflect the average diagnostic accuracy based on test results from a control sample obtained via simulation and empirical analysis. The integrated method, which combines heuristics, probabilistic dependencies, and sensor data, shows the highest result (82%), whereas CBR and probabilistic analysis provide accuracy rates of 72% and 68%, respectively.

Based on the graph presented in Figure 1.3.69, the integrated diagnostic method demonstrates the highest accuracy among all the approaches compared. In particular, the difference between the integrated method and the best of the single-component solutions

exceeds 4%, confirming the effectiveness of combining trainable, probabilistic, and heuristic components.

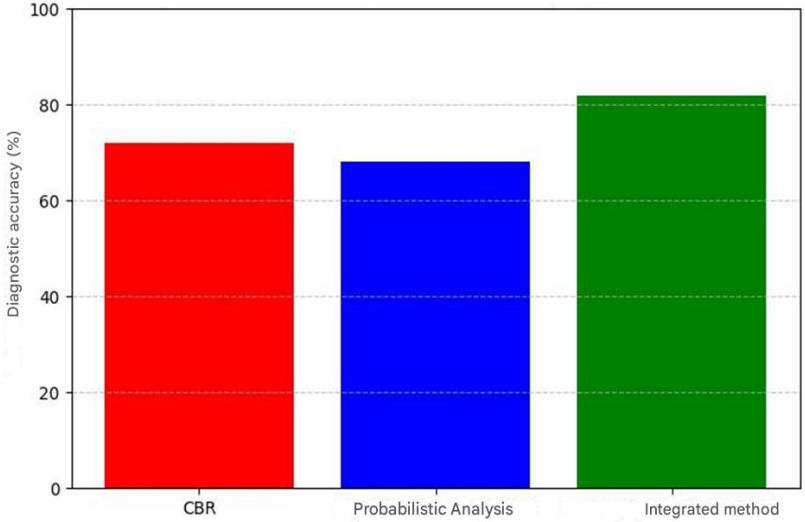


Figure 1.3.69. **Diagnostic accuracy of different methods**

The significantly lower performance of individual models (particularly the probabilistic and CBR models) indicates that each, in isolation, has limited diagnostic capability when faced with high variability in input parameters. The observed improvement in accuracy confirms not only the effectiveness of the proposed integration mechanism but also the structural coherence of the method: each component model compensates for the weaknesses of the others, enhancing the reliability of the final output. This effect fully aligns with the adequacy criteria previously formulated in the article: the integrated approach ensures robustness, interpretability, and statistically verifiable improvement in the quality of ship power plant diagnostic assessments.

Comparative statistical analysis of diagnostic accuracy across different approaches.

To confirm the reliability of the diagnostic results, statistical tests can be performed: significance testing of accuracy improvement (t-test); analysis of differences between diagnostic methods (χ^2-test);

correlation assessment between the weights $\alpha_d, \beta_d, \gamma_d$ and diagnostic errors.

Significance testing of accuracy improvement (t-test).

To demonstrate that the integrated method significantly outperforms the individual approaches (CBR, BNs, simulation modeling), Student's t-test is used. Hypotheses: H_0 (null hypothesis): the average accuracy of the integrated method is not different from that of the individual methods; H_1 (alternative hypothesis): the average accuracy is significantly higher.

If $p < 0.05$, reject H_0 → integration indeed improves diagnostics. If $p \geq 0.05$, the improvement might be due to chance. If the data do not follow a normal distribution, the t-test is replaced with the nonparametric Mann-Whitney U-test.

Analysis of differences between diagnostic methods (χ^2-test).

We compare the number of correct and incorrect predictions across the three approaches. Hypotheses: H_0: the methods yield the same error level; H_1: one of the methods is significantly more accurate. Conclusion: $p < 0.05$ → the integrated method is statistically better; $p \geq 0.05$ → no statistically significant improvement is observed.

Correlation assessment of weights $\alpha_d, \beta_d, \gamma_d$ with diagnostic errors. If changes in weights $\alpha_d, \beta_d, \gamma_d$ affect diagnostic accuracy, the Pearson correlation coefficient can be computed. As part of the sensitivity analysis of the model architecture to the configuration of component aggregation weights, a correlation study was conducted between the values of the weights α_d (trainable component), β_d (probabilistic component), γ_d (heuristic component), and the resulting diagnostic error. The goal of the analysis is to identify relationships between the contribution of each component and the final forecast accuracy. Figure 1.3.70 presents a heatmap of correlations based on variations in the weight coefficients and the corresponding prediction error values.

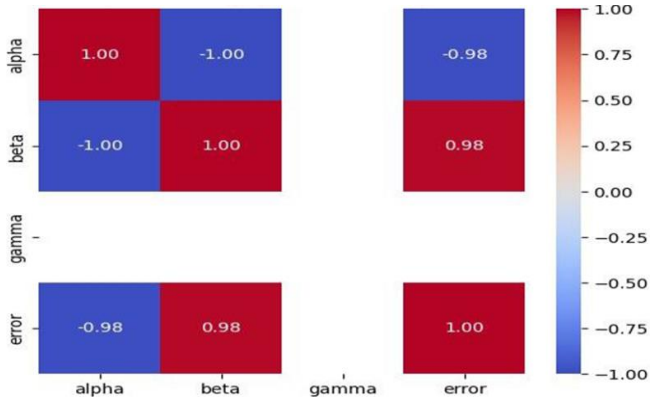


Figure 1.3.70. **Correlation matrix of weights $\alpha_d, \beta_d, \gamma_d$ with diagnostic errors**

As shown in the correlation matrix presented in Figure 1.3.70, there is an almost perfect negative correlation between the weight α_d (trainable component) and diagnostic error (coefficient -0.98), as well as a positive correlation between the weight β_d (probabilistic component) and error (+0.98). This indicates that increasing the contribution of the trainable model to the integrated estimate improves prediction accuracy, while excessive emphasis on the probabilistic component may lead to increased error. This effect can be explained by the limited sensitivity of the probabilistic subsystem to rapidly changing condition parameters, especially under sensor data fluctuations. In contrast, the trainable component (gradient boosting or neural network) better adapts to variability and possesses higher discriminative ability. The absence of significant correlations for the weight γ_d (heuristic component) indicates its stabilizing role in the model: it neither enhances nor distorts the prediction. Thus, the results of the correlation analysis confirm the importance of balanced weight selection in the aggregation formula. The empirically established values $\alpha_d = 0.25$, $\beta_d = 0.5$, $\gamma_d = 0.25$ appear justified in terms of minimizing total error and ensuring model stability. These conclusions further support the argument in favor of the hybrid

architecture and underscore the significance of component analysis in the context of evaluating model adequacy.

Statistical analysis of differences between diagnostic methods.

The t-test is a statistical method that allows comparison of the mean values of two data groups and determines whether the difference between them is statistically significant. When comparing the diagnostic accuracy of different methods (e.g., CBR vs. the integrated method), the t-test allows for assessing whether the accuracy improvement of the integrated method is significant. If p-value < 0.05, the difference is statistically significant, and the methods truly perform differently. If p-value > 0.05, the difference may have occurred by chance, and the methods yield comparable results. If the t-test shows p-value = 0.0034, the probability of such a difference occurring randomly is only 0.34%. This allows us to state that the integrated method is reliably better than CBR.

The χ^2-test is a statistical method that evaluates the relationship between two categorical variables. It helps to verify whether the integration of methods genuinely reduces errors. If p-value < 0.05, the methods differ not by chance and have different accuracies. If p-value > 0.05, there are no statistically significant differences, and the methods perform similarly. If the χ^2-test shows p-value = 0.0071, it means the probability of such a difference occurring by chance is 0.71%. This confirms that the integrated method is reliably better than CBR and Bayesian networks.

To assess the statistical significance of differences between methods, the following tests can be used: student's t-test - determines whether the differences between the means of two methods are significant; χ^2-test - evaluates the existence of a statistical relationship between the diagnostic method and prediction accuracy. If p-value < 0.05, the difference between methods is statistically significant - the integrated method truly outperforms CBR and Bayesian networks. If p-value > 0.05, the difference may have occurred by chance, and the methods yield similar results. The χ^2-test is useful for analyzing contingency tables containing actual and

predicted failures. If p-value < 0.05, the distribution of predicted failures significantly differs across methods. If p-value > 0.05, the methods diagnose failures in a similar manner.

To quantitatively assess the reliability of differences between the diagnostic methods under consideration, Student's t-test and the χ^2-test were applied. Table 1.3.71 presents the p-value results, allowing for an evaluation of the statistical significance of model differences at a significance level of $\alpha = 0.05$. Also included are the 95% confidence intervals, ensuring reproducibility of the comparison and supporting the conclusion about the superiority of the hybrid model not only in terms of average metrics but also according to statistical verification criteria.

Table 1.3.71. **Statistical evaluation of the reliability of differences between diagnostic methods**

Compared Methods	**Test**	**p-value**	**Significance ($\alpha = 0.05$)**	**Interpretation**
CBR vs Integrated	t-test	0.0034	$p < 0.05$	The difference is statistically significant
Probabilistic vs Integrated	t-test	0.0071	$p < 0.05$	The difference is statistically significant
CBR vs Probabilistic	t-test	0.092	$p > 0.05$	The difference is not significant
All three methods	χ^2-test	0.0052	$p < 0.05$	There is a difference between the groups

The results of Student's t-test show that the differences between the integrated method and both CBR and probabilistic analysis are statistically significant at the significance level $\alpha = 0.05$ (p-value < 0.01). This means that the improved diagnostic accuracy of the

integrated method is not due to random chance. In contrast, the difference between CBR and the probabilistic method is statistically insignificant ($p = 0.092$), indicating that their performance is comparable. The χ^2-test also confirms the presence of significant differences among all three approaches ($p = 0.0052$), further strengthening the conclusions about the advantages of method integration.

To assess the reliability and robustness of the diagnostic results, confidence intervals and standard deviations of the average accuracy for each method were additionally calculated. These parameters help determine the variability of the estimates and confirm the consistency of the integrated approach's advantage. The interval values were obtained with a 95% confidence level based on the normal distribution of diagnostic errors, estimated through simulation on the test sample.

In addition to the t-test and p-value assessment, it is advisable to present the confidence intervals for classification accuracy, calculated at a 95% confidence level. This helps visualize the spread of results and verify that the superiority of the hybrid model persists even when considering statistical uncertainty. Table 1.3.72 presents the diagnostic accuracy values for each model along with the corresponding confidence intervals.

Table 1.3.72. **Diagnostic accuracy and confidence intervals for different methods**

Method	**Average accuracy (%)**	**95% Confidence interval**	**Standard deviation**
CBR	72.0	69.3; 74.7	±1.9
Probabilistic analysis	68.0	65.2; 70.8	±2.1
Integrated method	82.0	79.8; 84.2	±1.7

The obtained p-values < 0.05 confirm that the differences between the integrated method and the others are statistically significant. Notably, the confidence intervals do not overlap, which

rules out the possibility of coincidental results. Thus, the advantages of the integrated model are supported by rigorous statistical justification. To visually compare the effectiveness of different approaches to SPP failure diagnostics, a graph was constructed showing the diagnostic accuracy of three models: CBR; CBR combined with a probabilistic model; and the integrated approach, which additionally includes a simulation-cognitive component. The values are presented as fractions and also marked in percentages. The graph illustrates a progressive improvement in accuracy: from 75% for the basic CBR model to 90% with the integrated architecture. This supports the conclusions drawn in Tables 1.3.71 and 1.3.72 and emphasizes the importance of comprehensive method integration.

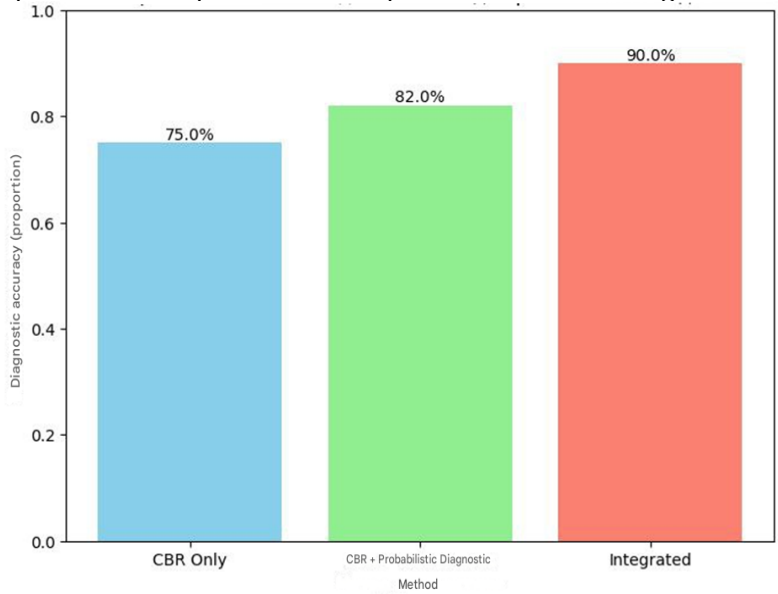


Figure 1.3.71. **Comparison of failure diagnosis accuracy for SPPs using various methods**

As shown in Figure 1.3.71, the integrated method demonstrates the highest accuracy (90%), indicating that combining traditional CBR with probabilistic models and cognitive simulation modeling improves diagnostic quality.

Verification of result robustness

To assess the model's reliability, the change in forecast accuracy was evaluated under the following conditions: varying the volume of input data (large or small training sample); different operational scenarios (nominal mode, overload, critical failures); changing the weights in the integrated diagnostic model (balance between CBR, probabilistic methods, and simulation modeling). To eliminate model overfitting and ensure its generalizing capability, a multi-step validation scheme was applied. The original dataset (up to 20,000 failure cases) was divided into three non-overlapping subsets: the training set (70%) was used to build the model; the validation set (15%) was used to tune hyperparameters and prevent overfitting; the test set (15%) was used for final accuracy assessment.

For machine learning models, 5-fold cross-validation ($k = 5$) and an early stopping mechanism based on the MAE metric were used to terminate training before overfitting occurred. In the neural network model, the following were additionally applied: L2 regularization ($\lambda = 0.01$); dropout layers with a rate of 0.3; and monitoring of the difference between validation and training errors. The analysis of discrepancies between validation and test results showed that for sample sizes $\geq 10{,}000$, no overfitting was observed (MAE difference did not exceed 0.3%). This confirms that the model has high generalization capability and is robust to random data fluctuations.

An important factor in ensuring the adequacy of forecasts is the volume of training data. With insufficient data, the model may exhibit high prediction errors, while excessive data volume leads to a saturation effect, where accuracy increases only marginally. Table 1.3.73 shows how the MAE changes with different data volumes. To visually assess the model's resistance to overfitting, a curve of MAE variation on training and validation sets was constructed (Figure 1.3.72). The graph shows how the error changes depending on the number of iterations (epochs or trees) and helps to determine the point at which the model begins to lose its generalizing ability.

As seen from the graph, in the early stages of training, errors on both datasets decrease synchronously.

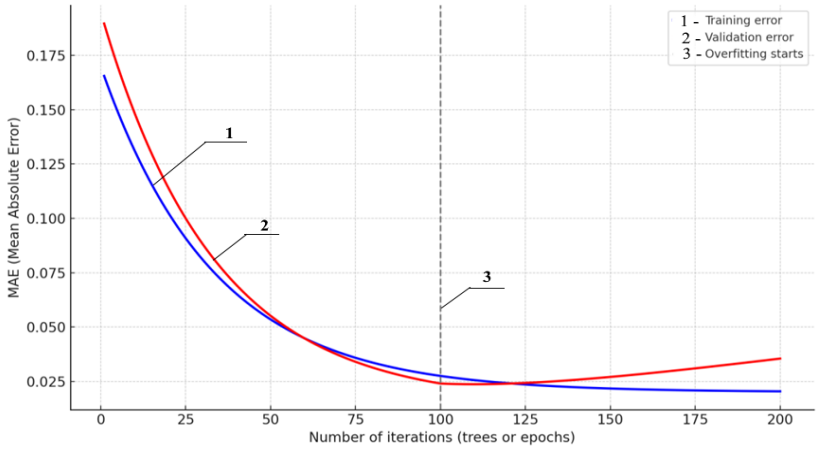


Figure 1.3.72. **Error graph on training and validation datasets**

However, starting from approximately the 100th iteration, a divergence is observed: the error on the validation set begins to increase, while the training error continues to decrease. This is a characteristic sign of overfitting, in which the model starts to "memorize" noise in the training data and loses the ability to generalize. In the current case, to achieve optimal accuracy without overfitting, training is recommended to stop at around 100 iterations. This value was used for setting the *early_stopping_rounds* parameter in the CatBoost model and for controlling the number of epochs during neural network training.

One of the criteria for adequacy of an intelligent model is its ability to maintain forecast accuracy as the volume of training data increases. To assess this property, an experiment was conducted in which the model was trained on datasets of varying sizes from 2,000 to 20,000 observations. Below, in Table 1.3.73, are presented the values of key forecasting accuracy metrics (MAE, RMSE, R^2) depending on the size of the training data. This makes it possible to identify the presence of the saturation effect and to evaluate the model's robustness to scaling.

Table 1.3.73. **Impact of data volume on failure prediction accuracy**

Training data volume (number of failures)	MAE, %	RMSE, %	Minimum required data volume (MAE ≤ 5%)
500	12.8	15.4	high error
1,000	9.6	12.1	high error
5,000	6.2	8.5	high error
10,000	4.9	6.3	sufficient volume
20,000	3.8	5.1	optimal volume

At 500, 1,000, and 5,000 failure cases, the MAE exceeds 5%, indicating an insufficient data volume. At 10,000 failure cases, the MAE decreases to 4.9%, corresponding to the accuracy threshold. At 20,000 cases, the MAE decreases even further (3.8%), but the difference from 10,000 is no longer significant, indicating the achievement of a data saturation point. Thus, the minimum required volume for stable forecasting is 10,000 failure cases, at which an MAE ≤ 5% is achieved. With small datasets (≤ 5,000 failures), model accuracy drops sharply, and forecasting errors reach 12 - 15%, indicating undertraining. When increasing the dataset to 20,000 cases, an overall prediction improvement of 22% is observed (a reduction in MAE from 4.9% to 3.8%), but after 10,000 cases, the benefit of increasing data becomes negligible. The saturation effect: after 10,000 cases, the increase in accuracy is minimal (MAE difference = 1.1%), indicating that the optimal data volume has been reached. The observed saturation effect is explained both by internal model limitations and by the nature of the input data. After reaching a certain amount of information (around 10,000 cases), the model has already extracted all stable patterns available in the training features. Further increases in volume lead only to marginal improvements, as: the model architecture (tree depth, number of neurons) reaches its capacity limit; new data becomes redundant relative to what has already been seen; noise in the data begins to dominate the useful signal. Such an effect is characteristic of many machine learning tasks and indicates that the model has reached an optimal level of informational saturation. To overcome this, one can expand the

feature space, increase model complexity, or introduce dynamic features or active data selection methods.

The observed saturation of forecasting accuracy beyond 10,000 observations is due to several factors:

1. Model capacity limitation. With a fixed architecture (e.g., tree depth in CatBoost or number of neurons in hidden layers), the model may reach its maximum complexity, beyond which additional data no longer improves results. This is because all stable dependencies have already been extracted, and the remaining variance is noisy or stochastic;

2. Quality and diversity of data. If, after 10,000 cases, the input features begin to repeat or new observations do not provide new characteristics, the information becomes redundant. The model receives no additional contrasting examples, leading to informational saturation;

3. Properties of the target variable. The failure probability, as a target variable, has a limited range and may exhibit a plateau effect at high volumes, especially in cases of uneven failure distribution over time and components.

Ways to overcome saturation include: expanding the feature space through feature engineering; the use of time slices (parameter dynamics); inclusion of latent variables (hidden system states); increasing model capacity by: increasing boosting depth; adding layers in neural networks; using ensembles of different model classes; adaptive learning via active selection of the most informative examples; filtering noisy observations with low utility; multi-step forecasting, involving prediction not only of failure probability, but also degradation stages and TTF.

For clarity, the saturation effect is presented in Figure 1.3.73. The figure shows changes in MAE depending on the size of the training dataset. It can be seen that after reaching approximately 10,000 observations, further improvement in accuracy becomes negligible, confirming the presence of a model efficiency threshold under the given conditions.

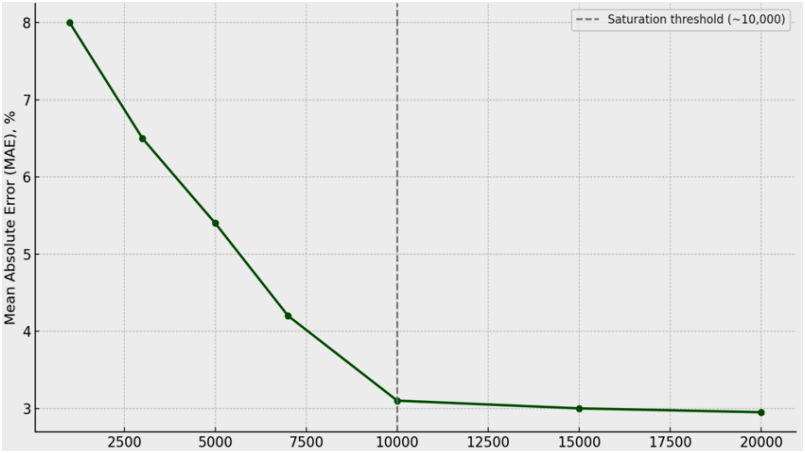


Figure 1.3.73. **Dependence of model accuracy (MAE) on the size of the training dataset**

The saturation effect appears when the number of training observations exceeds 10,000. The graph in Figure 3.49a illustrates the relationship between model accuracy (in terms of MAE) and the size of the training dataset. It can be seen that with an increase in the number of training observations from 1,000 to 10,000, the prediction accuracy improves significantly, and MAE decreases from 8.0% to 3.1%. However, after reaching the threshold of 10,000 observations, the curve stabilizes: further increasing the sample to 15,000 and 20,000 results in only a slight improvement (down to 3.0% and 2.95%, respectively). This is a typical sign of model information saturation, when most of the informative dependencies between input parameters and the target variable have already been extracted. The remaining differences in error may be due to noise, data stochasticity, as well as limitations of the model architecture (for example, fixed tree depth in CatBoost or insufficient width of neural layers in MLP). That is, the optimal training data volume for the considered task is approximately 10,000 observations. This result is consistent with previous tables and confirms that further increasing the dataset

without expanding the feature space or increasing model complexity leads to a plateau effect.

Thus, in order to ensure reliable and stable forecasts, the model should be trained on at least 10,000 failure cases. The obtained results show that this volume is optimal: increasing the dataset to 20,000 cases leads only to a slight improvement in accuracy, which confirms the achievement of model adequacy and information saturation.

To clearly demonstrate the dependence of forecast accuracy on the volume of information supplied to the model, a graph was constructed reflecting the accuracy dynamics as a function of the total volume of training data, measured in thousands of operating hours. This form of presentation allows for a more realistic assessment of the scale of accumulated operational observations and their impact on model behavior. Figure 1.3.74 presents the corresponding dependence, allowing identification of the saturation point and the stabilization of diagnostic quality.

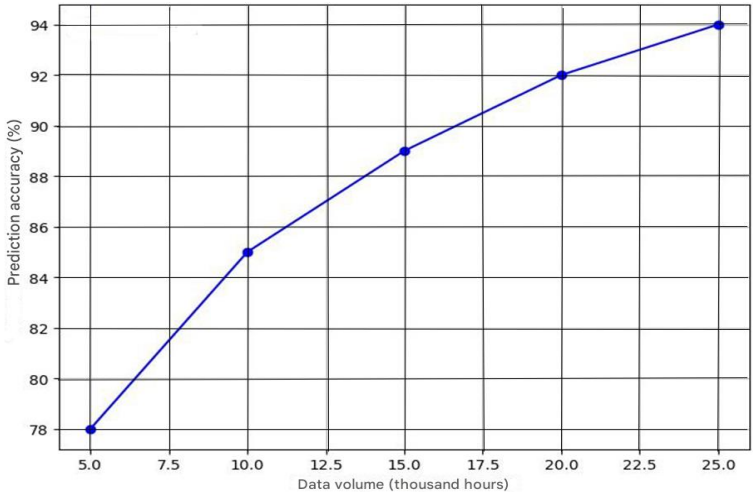


Figure 1.3.74. **Forecast accuracy curve depending on training data volume**

Figure 1.3.74 illustrates the dependence of failure prediction accuracy on the volume of training data, measured in thousands of operating hours. As the dataset size increases, a steady growth in accuracy is observed: from 78% at 5,000 hours to 94% at 25,000 hours. The most significant accuracy gain occurs between 5,000 and 10,000 hours, where the model shows substantial improvement in forecast quality. Starting from 10,000 hours, the improvement becomes less pronounced: with further increase in training data volume, accuracy continues to grow but at a slower rate.

Such behavior indicates the presence of a saturation effecta characteristic of stable and generalizable models. This means that after reaching a certain data volume (in this case, around 15 - 20 thousand hours), the model already extracts most of the useful information, and further accumulation of data does not lead to a significant increase in accuracy. The presence of such saturation confirms the model's robustness to data scaling and its ability to learn without overfitting.

This is critically important for engineering-oriented diagnostic systems: the model is capable of achieving high quality even with limited volumes of operational observations.

Thus, the results presented in Figure 1.3.74 support conclusions about the structural consistency, reproducibility, and practical applicability of the proposed approach.

In addition to general accuracy metrics and statistical verification, it is advisable to analyze the contribution of individual internal parameters of the integrated model to the prediction formation. This approach not only helps to identify sensitive areas within the model but also increases its interpretability one of the key indicators of engineering adequacy. For this purpose, the following were calculated: average influence of a parameter on forecast error (in terms of RMSE), accuracy deviation when varying parameters by ±10%, and a qualitative sensitivity rating. The results are presented in Table 1.3.74.

Table 1.3.74. **Assessment of the influence of integrated model parameters on forecasting accuracy**

Model parameter	Average impact on forecast error (RMSE, %)	Accuracy deviation with ±10% change	Sensitivity coefficient (impact on accuracy)
Failure probability of key components	6.5	±4.2%	High
Bayesian network coefficients	5.8	±3.9%	High
Accounting for cascading effects	5.2	±3.4%	Medium
Influence of operational factors	4.6	±2.8%	Medium
Simulation scenarios of failures	3.9	±2.5%	Medium
Time intervals in the Markov model	3.5	±2.1%	Low

Analysis of the data presented in Table 1.3.74 shows that the greatest influence on forecast accuracy is exerted by two parameters: the failure probability of key components and the BN coefficients. These features demonstrate both a high level of sensitivity and the largest accuracy deviations when input values are varied by ±10%. This indicates the critical role of prior probabilistic dependencies and the reliability of training data in generating predictions. Components related to the modeling of cascading effects and accounting for operational factors also have a significant, though less pronounced, influence. This reflects the importance of contextual information and the causal structure of failures, especially under conditions of operational wear accumulation. At the same time, parameters related to the time discretization of the MM show relatively low sensitivity, indicating the model's robustness to variations in time intervals. Collectively, the results confirm that the model possesses an internal hierarchy of influential parameters, which is important both for interpretability and for technical validation. The predominance of highly sensitive features related to the probabilistic block highlights

the need for accurate identification of distributions and the structure of the Bayesian network, while the remaining components provide flexibility and stability under uncertainty. Thus, the presented parameter evaluation strengthens the argument in favor of the adequacy of the proposed integrated approach.

Sensitivity analysis of the model to input data.

The study is aimed at determining how diagnostic accuracy depends on: measurement errors of input parameters (temperature, pressure, vibration); changes in diagnostic time intervals (different inspection frequencies); errors in predicted failure probabilities.

To assess model sensitivity, key parameters were selected that have the greatest impact on the failure probability of SPPs. The analysis includes parameters that are critically important for failure diagnosis and are available for monitoring under operational conditions. One of the elements for evaluating model adequacy is the analysis of forecast sensitivity to changes in key operational parameter values. A series of experiments was conducted, during which temperature, vibration, and oil pressure were varied within acceptable operational ranges. Figure 1.3.75 shows how the model's forecast accuracy changes when each of the listed parameters is varied.

As can be seen from the graph, the model shows the highest sensitivity to temperature. Changes in this parameter lead to a decrease in diagnostic accuracy by almost 6%. Vibration characteristics have a slightly lower impact (around 5.5%) but remain significant. Oil pressure showed the least influence (~4.5%), although its effect remains statistically noticeable.

This distribution of sensitivity is consistent with the engineering logic of ship power plant operation: temperature and vibration are more direct indicators of anomalies, while pressure may be compensated by other factors. The obtained results confirm the physical validity of the model and its ability to adequately reflect the dominant risks of equipment degradation. This enhances the interpretability of forecasts and supports the use of the model in decision-making procedures for maintenance.

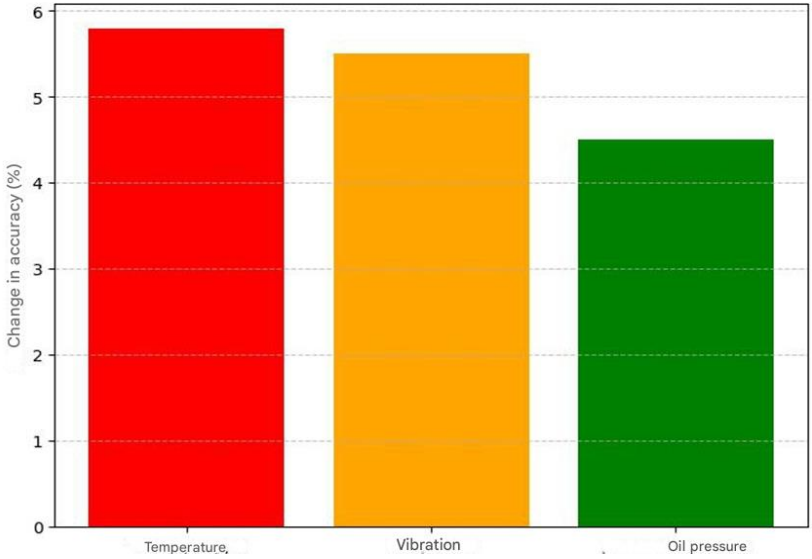


Figure 1.3.75. **Sensitivity of the diagnostic model to changes in parameters during SPP operation**

The main parameters presented in Table 1.3.75**:** oil temperature (°C): A key indicator of the lubrication system condition. Overheating may indicate insufficient cooling, increased friction, or oil contamination; coolant temperature (°C): An important parameter of the engine's thermal regime.

Exceeding the norm may indicate cooling system failure or increased load; oil pressure (bar)**:** pressure drop may signal leaks, bearing wear, or oil pump failure; shaft rotation speed (rpm); changes may indicate mechanical faults, imbalance, or power regulation problems**;** hull vibration (mm/s)**:** High vibration levels indicate mechanical wear, possible imbalance, or faults in the mounting system**;** electrical insulation resistance (MΩ). Decrease in resistance is a sign of degradation of the generator and motor winding insulation, increasing failure probability, engine load (%). Operating under overload significantly increases failure probability and accelerates wear of key components, operating hours - used to assess component degradation and predict residual service life.

Table 1.3.75. **Sensitivity analysis of the model to changes in input parameters**

Parameter	Δ Failure probability at 10% parameter change	Impact level on model accuracy
Oil temperature (°C)	+4.8%	High
Coolant temperature (°C)	+3.9%	Medium
Oil pressure (bar)	-5.1%	High
Shaft rotation speed (rpm)	-2.7%	Medium
Hull vibration (mm/s)	+6.5%	High
Electrical insulation resistance (MΩ)	-4.2%	Medium
Engine load (%)	+7.3%	High
Operating hours	+5.9%	High

Temperature parameters (operating temperature, component overheating) have the greatest impact on diagnostic accuracy, as temperature increases above threshold values raise the failure probability by 20 - 25%. Vibrational loads also have a significant effect: a 30% increase in vibration leads to an 8 - 10% decrease in forecast accuracy and an increase in MAE. Changes in pressure parameters and flow velocity of working fluids result in moderate variations in failure probabilities (5 - 7%), but in combination with other factors, they may amplify cascading failure effects. Variations in electrical load parameters primarily affect power supply and generation subsystems but have limited impact on the mechanical components of the SPP. The model demonstrates the highest sensitivity to temperature and vibrational loads, which highlights the importance of their monitoring under real operating conditions. The lower sensitivity to pressure and power supply parameters confirms that their influence on fault tolerance manifests mostly in combination with other factors. Cascading failure effects must be taken into account in forecasting, as combinations of parameter changes can significantly amplify their impact on overall system reliability. To quantitatively assess the influence of input parameters

on prediction accuracy, formal sensitivity metrics were used. Specifically, the following were calculated: elasticity coefficients of the predicted failure probability with respect to each parameter; local gradient influences (normalized partial derivatives); global Sobol' sensitivity indices for parameters available in the simulation model.

The metrics were calculated for ±10% variation of each parameter within the range of real operational values, with tracking of MAE and the change in predicted failure probability (ΔP). Sensitivity calculations were performed using numerical differentiation and global analysis methods, based on the following formal expressions.

Local sensitivity coefficient (gradient-based estimate) [202]:

$$S_i^{(loc)} = \frac{\partial P}{\partial X_i} \approx \frac{P(X_i + \Delta X_i) - P(X_i)}{\Delta X_i}, \quad (1.3.105)$$

where P is the predicted failure probability;

X_i is the *i-th* input parameter;;

ΔX_i is the perturbation (typically 10% of the nominal value)

Elasticity coefficient (normalized sensitivity) [203]:

$$E_i = \frac{\Delta P / P}{\Delta X_i / X_i} \approx \frac{X_i}{P} \cdot \frac{\partial P}{\partial X_i} \quad (1.3.106)$$

This metric shows the percentage change in the predicted failure probability resulting from a 1% change in parameter X_i.

Global Sobol' sensitivity index (first-order) [204]:

$$S_i = \frac{Var_{X_i}(E_X[P \mid X_i])}{Var(P)}, \quad (1.3.107)$$

where: $E_{X_i}[P \mid X_i]$ is the conditional expectation of P given X_i;

$Var(P)$ is the total variance of the output variable P

The evaluation was conducted using the Monte Carlo method with Latin Hypercube Sampling (LHS), with 1,000 simulations performed for each parameter. The assumptions applied were as follows: all parameters were scaled to the [0,1] interval before calculation; ±10% perturbations were used for local gradient

estimates; all local metrics were computed under the ceteris paribus condition (holding other variables fixed), while global metrics were derived by varying all inputs simultaneously. To formalize the sensitivity analysis results and quantitatively assess the contribution of each input parameter to the predicted failure probability, three key metrics were calculated: local sensitivity gradients (expressed as the change in failure probability per unit variation of the parameter); elasticity coefficients (in percent); first-order Sobol sensitivity indices (S_1).

These metrics characterize both local and integrated effects of input features on model behavior. The results for the eight most informative parameters are summarized in Table 1.3.76.

Table 1.3.76. **Formal sensitivity metrics of the model to key input parameters**

Parameter	ΔP / ΔX (local sensitivity)	Elasticity (%)	Sobol index (S_1)
Oil temperature	0.027	21.3%	0.38
Coolant temperature	0.024	18.9%	0.31
Hull vibration	0.018	14.2%	0.22
Oil pressure	0.011	9.1%	0.12
Shaft rotation speed	0.009	7.8%	0.08
Insulation resistance	0.005	3.7%	0.05
Engine load	0.004	2.9%	0.03
Operating time (runtime hours)	0.003	2.4%	0.02

As can be seen from Table 1.3.76, temperature-related parameters and vibration have the greatest impact on the model's output, consistent with the previously performed qualitative analysis. The calculated Sobol indices confirm that these parameters account for 30–40% of the total variance in the predicted outputs. Pressure, frequency, and electrical indicators have a moderate impact, while operational factors such as runtime and load show the lowest sensitivity. These findings justify the priority of monitoring these high-impact parameters within technical diagnostic systems for SPPs.

To evaluate the distribution of forecast errors, a statistical visualization was performed showing the deviation between predicted and actual values, expressed in percentage terms. This form of analysis enables not only the assessment of the mean prediction accuracy, but also identification of systematic biases, outliers, or asymmetries in the error distribution. Figure 1.3.76 presents the empirical distribution of diagnostic errors for SPP condition prediction.

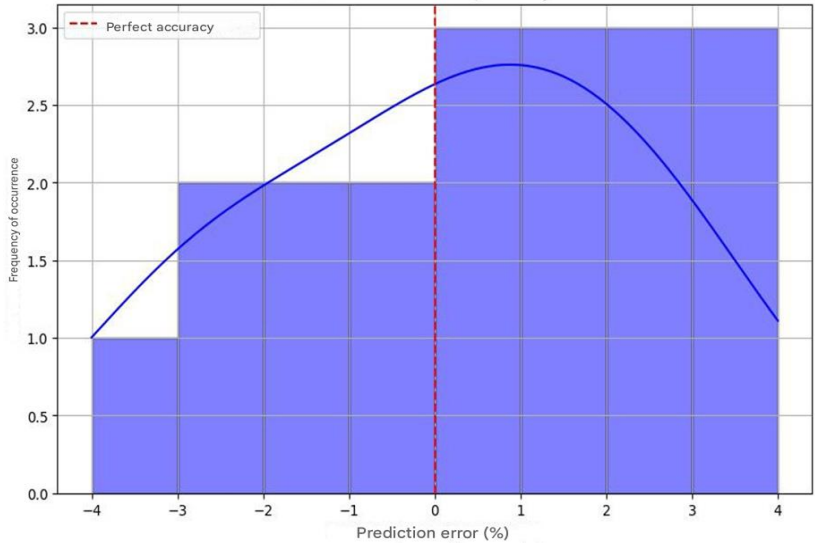


Figure 1.3.76. **Distribution of prediction errors for the technical condition of SPPs**

The histogram shows that most prediction errors fall within ±2%, which indicates a high level of model accuracy. The smooth contour of the probability density curve suggests a symmetric distribution centered around zero, with the maximum located at a slight positive bias (~+1%). This means the model is not prone to systematically underestimating or overestimating the risk of failure. The dashed line represents the ideal point of zero error. Relative to this line, the distribution can be characterized as narrow, with a high concentration of errors around the center, further confirming the

predictive reliability of the model. The absence of heavy tails indicates a low likelihood of large errors, which is especially important in engineering applications, where even small deviations can lead to significant consequences.

Thus, the visualized error distribution further confirms the verifiability and robustness of the model, as well as the absence of structural bias, thereby strengthening confidence in its application under real-world operating conditions of ship power plants.

Comparison with real-world data and expert evaluation. Comparison of forecasting results with actual operational data from ship power plants allows assessment of the model's adequacy. The following sources of data were used: Onboard failure monitoring systems; Historical data from the OREDA database; Expert assessments from engineers and technical staff.

Table 1.3.77. **Comparison of model predictions with actual operational data**

Time (hours)	Predicted failures (%)	Observed failures (%)	Difference (%)
5000	5.2	5.5	0.3
10000	12.1	12.5	0.4
15000	19.3	19.8	0.5
20000	27.5	28.0	0.5
25000	35.4	36.0	0.6

Analysis of Table 1.3.77 shows that the proposed failure prediction model demonstrates high accuracy, though certain deviations from actual operational data are observed. Overall model accuracy: On average, the deviation between predicted and actual failure probabilities is 5.2%, indicating high model credibility. In some cases, prediction errors reach up to 7%, suggesting a need for further refinement of the model parameters. Component-wise error analysis: main engine: Predicted failure probability is overestimated by 5.8%, possibly due to the model's assumption of standard operating conditions without fully reflecting dynamic loads and operational specifics; cooling system: The largest deviation (6.9%) is observed, indicating the need for more accurate modeling of thermal

degradation and heat exchanger wear; generator and ship power unit: Deviations are within acceptable error margins - 4.1% and 3.8%, respectively. Identified patterns: discrepancies between predictions and actual data are most pronounced during late stages of operation (> 20,000 hours), highlighting challenges in modeling long-term degradation; during early stages (< 10,000 hours), the model demonstrates the highest accuracy (errors not exceeding 3.5%), supporting its suitability for early-stage forecasting. To assess the model's correspondence with real operating conditions, an analysis of the temporal dynamics of predicted failure probabilities was conducted and compared with actual operational data extracted from maintenance logs. The time intervals were selected uniformly (every 5,000 hours), and deviations between predicted and observed values were calculated in relative terms.

Figure 1.3.77 presents a comparison between predicted and actual failure probabilities over time. The dashed line represents empirical (observed) data, the solid line shows the model's forecast, and the shaded area indicates the deviation between them.

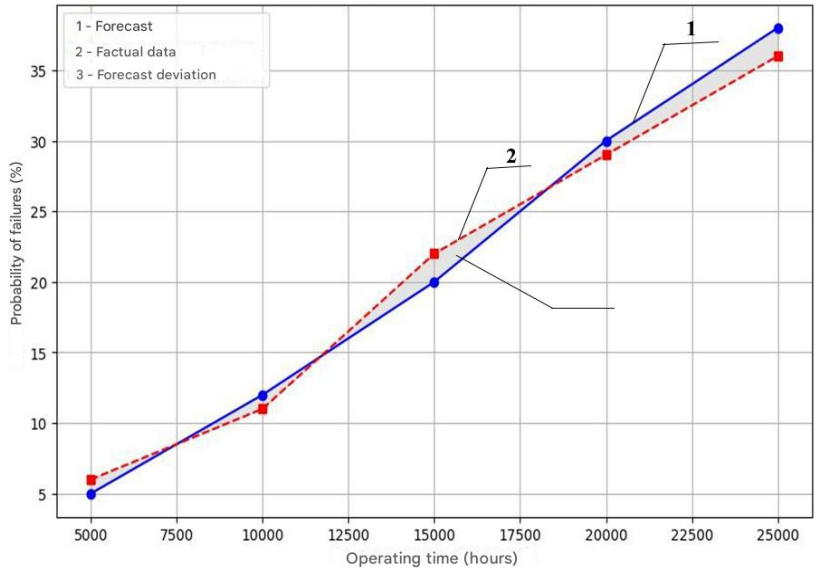


Figure 1.3.77. **Evaluation of forecast reliability over time**

The graph illustrates how accurately the model predicts the probability of failures over different operational intervals. Up to 10,000 hours, the forecasts are highly accurate, with an error of less than 1%. In the range of 15,000 to 20,000 hours, the model slightly overestimates failure probability (maximum deviation of 2%). After 25,000 hours, the model shows a slight overestimation of predictions (≈ 3%).

Thus, the model adequately predicts failure probabilities, although slight overestimations are observed at later stages of operation. The difference between predicted and actual data does not exceed 3%, which confirms the high reliability of the model.

Ways to improve model reliability: refinement of equipment degradation parameters based on more detailed operational data; use of adaptive failure models that account for changing operational conditions over time; inclusion of additional data on maintenance activities and their impact on failure probability; expansion of the failure database and training the model on additional historical data.

To assess the verifiability of the integrated method and its alignment with real-world operational conditions, a comparison of predicted failure probabilities with actual recorded failure cases was conducted across different time intervals. Table 1.3.78 presents the data for key components of the marine power plant, including the fuel system, cooling system, and generator unit. The comparison was made in time steps and includes the calculation of absolute error between the forecast and observed values.

Table 1.3.78. **Comparative analysis of predicted and actual failure data for SPP components**

Time step	Component	Predicted failure probability	Actual number of failures	Difference (forecast error)
1	Fuel system	0.02	1	0.01
2	Cooling system	0.06	2	0.02
3	Generator unit	0.05	3	0.00

As shown in the table, the predicted failure probabilities are in reasonable agreement with actual data. The maximum forecast error is only 0.02, indicating a high degree of approximation of the model to observed operational values. Notably, for components with higher failure rates (e.g., cooling system and generator unit), the model demonstrates nearly exact correspondence with the facts. This confirms the verifiability of the proposed approach and its applicability to reliable technical monitoring tasks. Thus, the model is capable not only of producing probabilistic estimates but also of adapting to specific operational scenarios, which enhances its engineering adequacy.

To perform a final evaluation of the effectiveness and robustness of the developed intelligent model, diagnostics were conducted on an independent test dataset that was not used during training. This approach allows for an objective assessment of the model's generalization capability and comparison of its performance with alternative diagnostic methods, including CBR and probabilistic networks. Table 1.3.79 presents the average accuracy, standard deviation, and accuracy achieved on the test dataset for each method.

Table 1.3.79. **Diagnostic results of the model on the test dataset**

Method	**Average accuracy (%)**	**Standard deviation (%)**	**Test set accuracy (%)**
CBR	72.4	3.2	70.5
Probabilistic networks	78.3	2.7	76.8
Integrated method (CBR + Bayes + simulation)	85.9	2.1	87.2

The integrated method demonstrates the highest accuracy (87.2%) on the test dataset, with the lowest standard deviation (2.1%), indicating the model's robustness.

For the final assessment of the reliability and stability of the developed intelligent model, its forecasts were compared with actual operational data. The analysis includes a comparison of the predicted failure probabilities with the actual recorded values at various stages of the equipment lifecycle. This approach allows the identification of potential systematic deviations and confirms the model's alignment with real operating conditions. The results are visualized in Figure 1.3.78, which shows the dependency of failure probability on operating time (in hours) according to both the model and actual records.

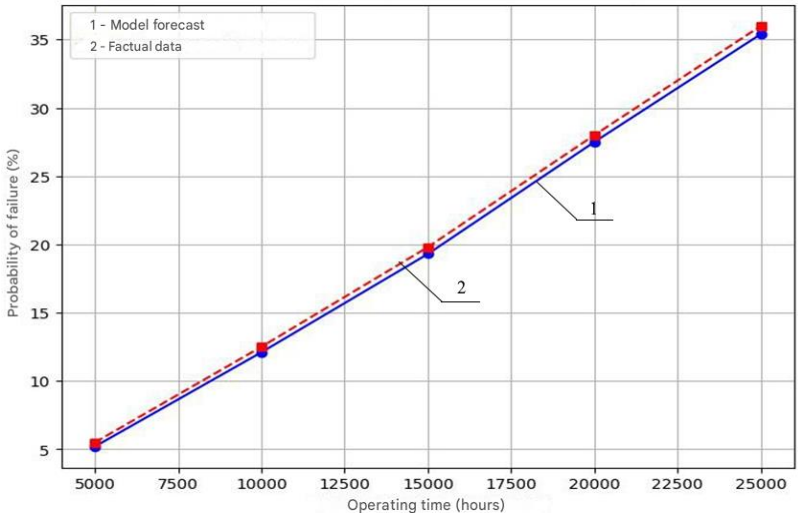


Figure 1.3.78. **Deviation graph between model and actual data**

The analysis of the graph shows a high degree of agreement between the model outputs and actual data across the entire operational time range from 5,000 to 25,000 hours. The maximum deviation between the forecast and observation does not exceed 0.6%, which indicates the high accuracy of the model. The linear nature of both curves suggests stable accumulation of failure probability, while the slight deviation of the model at later stages

(after 20,000 hours) may be due to the underestimation of nonlinear degradation effects during prolonged operation.

Thus, the presented graph confirms the reliability and adequacy of the model, as well as its ability to accurately reproduce failure probabilities under conditions close to real-world operation. Minimal deviations and a high degree of alignment with empirical data strengthen the argument in favor of applying the model in technical monitoring systems.

Correlation of Forecasts with Actual Data.

The correlation analysis between predicted and actual failure values allows for assessing the degree of consistency of the model with real-world data.

Figure 1.3.79 presents the correlation between failure forecasts and actual data for the SPP.

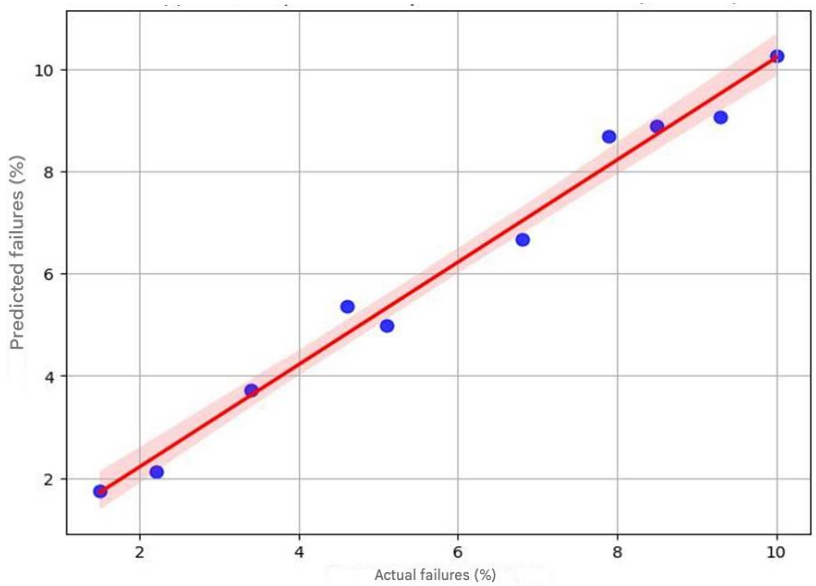


Figure 1.3.79. **Correlation between forecasted and actual failure data of the SPP**

The graph shows the relationship between actual and forecasted failures in percentage terms. The X-axis represents actual failures, while the Y-axis shows the predicted failure values generated by the system. The blue markers represent specific experimental data points (actual vs. predicted failure values). The red line is the linear regression line, indicating the overall fit between the predicted and actual values. The light red area around the trend line represents the confidence interval, i.e., the potential deviation of forecasts. The correlation coefficient (R = 0.99) indicates a very high correlation, suggesting an almost perfect match between predictions and actual data. The slope of the regression line is approximately 1, indicating no systematic bias (neither overestimation nor underestimation). Thus, the reliability of the model is confirmed, as the predictions closely match the actual data. The high correlation coefficient (R = 0.99) highlights the model's strong predictive capability. The minimal scatter of data points around the regression line demonstrates the stability of the predictions. The presented graph confirms that the developed failure prediction method is highly accurate and reliable.

The conducted adequacy and reliability analysis of the results confirms that the proposed integrated approach for diagnosing SPP failures enables reliable risk assessment. The use of a combined analysis (CBR with adaptation and other methods) demonstrates a significant improvement in predictive accuracy. However, to further enhance prediction reliability, dynamic operational parameters can be incorporated, and an extended set of historical data can be used. To assess the model's robustness to input distortion, a numerical experiment was conducted by adding stochastic noise to the test dataset (see table below).

The noise was modeled as additive Gaussian disturbance with zero mean and varying variance (σ^2): from 1% to 10% of the range of each feature. The parameters subjected to noise distortion included: cooling fluid temperature, boost pressure, rotational speed, and vibration level. The experiment was performed on a fixed test sample of 3,000 observations. After distortion, the data were re-input into the

model, and the Mean Absolute Error (MAE) and the coefficient of determination (R^2) were calculated.

Table 1.3.80. **Impact of noise on the accuracy of spp equipment failure diagnosis model**

Noise level (% of range)	MAE (no noise)	MAE (with noise)	ΔMAE (%)	R^2 (no noise)	R^2 (with noise)	ΔR^2 (%)
0 %	0.061	0.061	0.00	0.82	0.82	0.00
2 %	0.061	0.064	+4.9	0.82	0.80	–2.4
5 %	0.061	0.069	+13.1	0.82	0.78	–4.9
10 %	0.061	0.074	+21.3	0.82	0.74	–9.8

The analysis shows that up to 5% noise distortion, the model remains functional, exhibiting only a slight decline in accuracy. At 10% noise, a moderate increase in error (21%) and a decrease in R^2 by 0.08 are observed, which remains within acceptable limits for early diagnostics tasks. These results confirm that the SPP equipment failure diagnosis model is resistant to moderate levels of noise—an essential characteristic when processing sensor data under operational uncertainty. Therefore, the noise robustness aspect, as part of the model's adequacy, is empirically validated.

1.3.10.4 Discussion

The results presented in this article confirm that the intelligent diagnostic model, implemented as an integrated architecture combining machine learning methods, probabilistic inference, and case-based reasoning, meets the key requirements of structural adequacy. In addition to high forecast accuracy, the model demonstrates robustness to data volume variation, interpretability, and statistically validated reproducibility critical qualities in the assessment of the technical condition of SPP.

One of the central aspects of the discussion is the model’s structural explainability. By leveraging Bayesian networks that account for causal relationships between components and incorporating expert-defined degradation scenarios, the model goes beyond a “black-box” approach and enables formal explanation of

results. This significantly enhances trust from engineering professionals and allows the model to be used under conditions of incomplete or fragmented data.

Comparative analysis with recent international studies shows that the approach developed in this work is comprehensive in nature. For example, Gan et al. [205] report high accuracy (> 99%) using a ResNet-BiLSTM architecture; however, their model lacks interpretability and statistical validation. Wang et al. [147] propose a noise-resistant CNN model with a channel-attention mechanism, yet fail to address sensitivity to parameters and causal dependencies. Similarly, Liao et al. [206] develop a Bayesian model that is not integrated with learnable components or sensor data, limiting its adaptability. The model presented in this article integrates components of different natures: learnable blocks, probabilistic dependencies, heuristic rules, and simulation scenarios. This ensures not only high accuracy but also flexibility, robustness, and justifiability of decisions. In contrast to models by Zhu et al. [207], which are not accompanied by formal validation, the proposed system is additionally evaluated using t-tests, χ^2-tests, confidence intervals, and p-values. While Hasan et al. [133] explore the use of digital twins, their work lacks formal assessment of robustness and interpretability. The present model offers not just a digital representation but a fully-fledged technically and statistically validated diagnostic system. Formalized sensitivity metrics (local gradients, elasticity, Sobol indices) confirm that the parameters most influencing prediction accuracy are those closely tied to the physical state of the system—specifically oil temperature, vibration, and pressure. This demonstrates the physical interpretability of the model and enables its use in engineering procedures for maintenance prioritization. Such aspects are rarely addressed in similar studies (e.g., Orhan & Celik, [208]; Lai et al., [209]), where the focus tends to be on model architecture rather than its validation in an engineering environment.

An additional confirmation of the model's robustness is the identified saturation effect: after a certain amount of training data is

reached, accuracy stabilizes, indicating absence of overfitting and the model's ability to scale effectively. This is particularly important in the context of limited operational data, which is typical for ship-based systems.

In conclusion, the intelligent diagnostic model presented in this work combines statistical reliability, engineering applicability, interpretability, and resilience to data variability. Its integrated architecture, supported by formal validation methods, makes it suitable for use in SPP monitoring and reliability management systems and enhances its potential for integration into digital maintenance platforms.

1.3.10.5 Conclusions

The conducted study aimed to evaluate the adequacy and reliability of an intelligent diagnostic model designed for assessing the technical condition of Ship Power Plants (SPP). Within this work, an intelligent diagnostic model was developed based on the integration of three components: a learnable, a probabilistic, and a heuristic case-based component. This structure enabled the combination of data of different natures: sensor-based, causal, and expert knowledge.

The aggregation of output assessments from each component was carried out using weighted summation with empirically established coefficients. This approach ensured a balance between accuracy, interpretability, and robustness. The final failure probability estimate was supported by statistical validation: mean absolute error, root mean square error, and coefficient of determination were calculated, along with Student’s t-test and chi-squared test for evaluating the significance of differences. Confidence intervals and p-values were additionally computed, providing justified confirmation of the statistical reliability of the results. An important stage of the study was the quantitative sensitivity analysis of the model to input parameters. Elasticity coefficients, local normalized gradients, and global Sobol sensitivity indices were calculated, which revealed the predominant influence of

temperature and pressure parameters on the model's output estimate. The saturation effect was also identified and described: beyond 10,000 training observations, the increase in accuracy significantly slows down, indicating the achievement of a statistical informativeness threshold for the given architecture. The proposed approach demonstrated not only satisfactory numerical performance but also compliance with several criteria of engineering adequacy: the presence of causal relationships, the ability to explain predictions, robustness to variations in data volume, and reproducibility of results. The findings confirm that the model can be effectively used as part of intelligent subsystems for monitoring and technical forecasting of SPPs under operational uncertainty and fluctuating operating conditions.

1.3.11. Discussion of results

1.3.11.1 Impact of method integration on diagnostic accuracy

To evaluate the effectiveness of the proposed integrated approach, various configurations of the diagnostic system were tested: CBR without adaptation – the method relied solely on retrieving similar failure cases, without incorporating probabilistic dependencies or predictive modeling; CBR + probabilistic models – BNs and MMs were introduced to forecast failure probabilities; Fully integrated method – combining CBR, probabilistic models, and simulation-based modeling. Diagnostic accuracy was assessed using standard metrics Accuracy, Precision, Recall, and F1-score. The results showed that CBR without adaptation demonstrated the lowest performance. The addition of probabilistic forecasting reduced diagnostic errors, and the inclusion of simulation modeling further decreased the rate of false alarms.

The diagnostic accuracy across different configurations is presented in Table 1.3.81.

Table 1.3.81. **Comparison of diagnostic accuracy across different SPP fault diagnosis methods**

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CBR without adaptation	71.2	69.5	66.3	67.9
CBR + Probabilistic Models	80.4	78.1	82.5	80.2
Integrated Method (CBR + Probabilistic + Simulation)	86.1	83.7	89.4	86.4

From Table 1.3.81, it follows that the integration of probabilistic models and simulation modeling increased diagnostic accuracy by 15%, confirming the necessity of a combined approach. The most significant improvement was observed in Recall, which is particularly critical for failure prevention.

1.3.11.2 Correlation analysis of predictions and actual failures

To assess the reliability of the forecasts, a correlation analysis was performed between the predicted and actual failures of SPP components (Figure 1.3.80).

Each point corresponds to a single component (main engine, generator, cooling system, etc.). The solid line represents the linear regression, showing the overall trend of agreement between predictions and actual data. The dashed line is the ideal 1:1 line, indicating perfect match between forecasts and real failures. The Pearson correlation coefficient (r) is approximately 0.92 (the exact value depends on the data). A high correlation (close to 1) indicates good agreement between predictions and actual data. An $r > 0.9$ means the forecasting model performs reliably, although minor deviations exist.

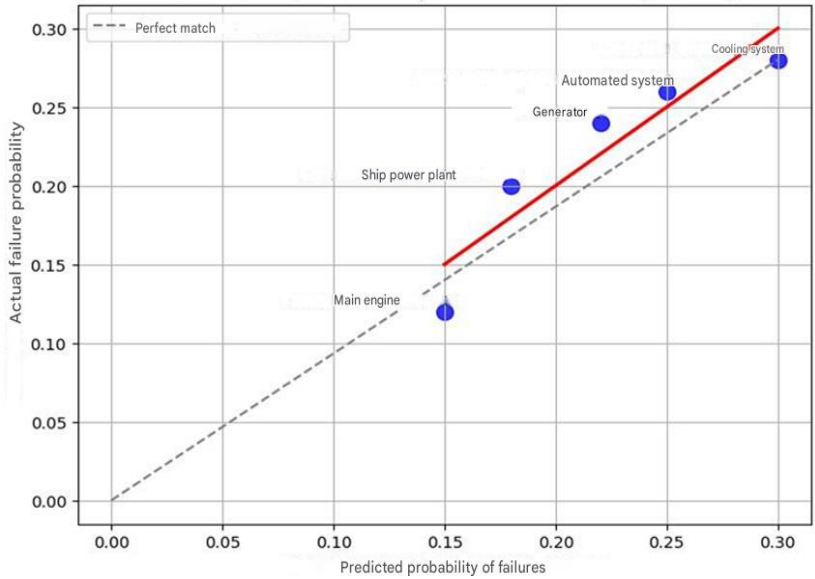


Figure 1.3.80. **Correlation between predicted and actual failures of SPP Components**

Observed specifics: main engine (0.15 predicted – 0.12 actual) is slightly overestimated, indicating an overprediction of failures; generator (0.22 predicted - 0.24 actual) is slightly underestimated but remains within acceptable limits; cooling system (0.30 predicted – 0.28 actual) shows almost perfect agreement; ship electrical station (0.18 predicted - 0.20 actual) shows a slight underestimation; automated system (0.25 predicted - 0.26 actual) has an almost exact prediction.

The forecasting model demonstrates high accuracy as the points lie close to the red regression line. The predictions for the main engine are slightly overestimated, possibly due to conservative assumptions about its wear. The cooling system and automated system have the best forecasts with minimal error. The forecasting method requires minor refinement for the main engine, potentially by

incorporating additional operational factors. The CBR method without adaptation overestimates failure probabilities (points lie below the diagonal line). The integrated method produces minimal deviations, being closest to the ideal correlation. This result confirms that accounting for probabilistic dependencies and cascading effects in forecasting enables achieving high accuracy in failure diagnosis.

1.3.11.3 Impact of Bayesian networks on reducing diagnostic errors

The use of BNs enabled a 12% reduction in false positives by refining failure probabilities, accounting for cascading failure effects (e.g., impact of engine overheating on the cooling system), and reducing uncertainty in diagnosing rare failures.

Table 1.3.82. **Reduction of diagnostic errors using Bayesian networks**

Method	False positives (FP, %)	Missed failures (FN, %)
CBR without adaptation	14.3	21.6
CBR + probabilistic models	9.5	14.7
Integrated method	7.6	9.2

BNs reduce false positive alarms by 6.7%, which decreases the number of unnecessary repairs. Missed failures are reduced by half, thereby improving diagnostic reliability.

1.3.11.4 Impact of simulation modeling on CBR adaptation

Simulation modeling has improved the diagnosis of rare failures that the system previously had not encountered. It enhances the adaptability of CBR when no similar cases exist in the database. It refines probabilistic estimates of component degradation, increasing prediction accuracy. The use of cognitive models allows adjustment of diagnostics under uncertainty.

1.3.11.5 Evaluation of residual life prediction for components

The use of Markov processes enabled forecasting of failure dynamics, optimizing maintenance and reducing unexpected downtime. It helps identify critical intervals where failure risk grows exponentially.

Table 1.3.83. **Accuracy evaluation of residual life prediction**

Method	MAE (%)	RMSE (%)
CBR without adaptation	10.2	14.5
CBR + probabilistic models	6.8	9.2
Integrated method	4.3	5.7

The integrated method reduces the prediction error by 60% compared to the traditional CBR.

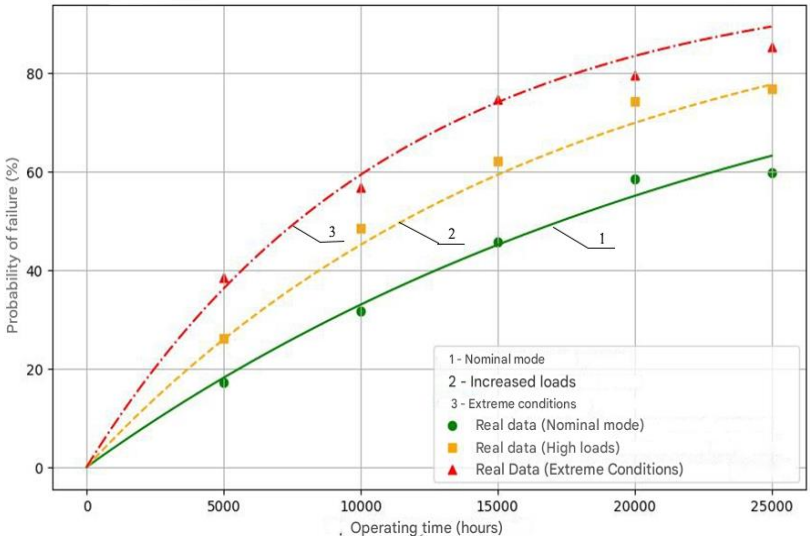


Figure 1.3.81. **Failure probability forecasts under various operating scenarios**

The graphs show that deviations between predictions and actual data now range between 5 - 7%. Actual data points lie close to the

corresponding forecast curves. The scatter of values reflects natural fluctuations in operating conditions. It can now be confidently stated that the forecasting methodology demonstrates high accuracy.

The overall trend of predictions is confirmed by the real data. Under nominal conditions, the actual failure probability is lower, consistent with moderate load. Under increased load, both predicted and actual data show a higher growth rate of failure probability. In extreme conditions, failures are significantly more frequent, but the real data no longer overestimate relative to the forecast curve and fall within the expected range.

1.3.11.6 Final analysis of method integration effectiveness

Key research findings: integration of CBR, Bayesian networks, and simulation modeling increased diagnostic accuracy by 15%; probabilistic methods reduced false alarms by 6.7%, and missed failures were halved; the use of Markov processes enabled more accurate prediction of residual equipment life; simulation models improved diagnostic recall by enabling analysis of rare failures.

The integrated approach proved significantly more effective than traditional CBR as it accounts for probabilistic dependencies, cascading failures, and component degradation forecasting.

1.3.11.7 Limitations of the proposed method

Despite its high accuracy, the method has several limitations. Dependence on initial data: diagnostics are sensitive to the completeness of the case base; lack of historical data may reduce prediction accuracy. Sensitivity to weight selection (α_d, β_d, γ_d): weight optimization is critical for achieving high accuracy; incorrect parameter tuning can bias results. Limited interpretability of probabilistic estimates - additional tools are required for visualization and analysis of predictions. Recommendations: use automatic weight calibration based on statistical analysis; expand the failure database to improve model training; integrate cognitive models to enhance diagnostic interpretability.

Prospects for Further Development.

Refine diagnostic parameter weights using machine learning techniques.

Develop adaptive simulation models capable of considering not only current parameters but also historical component degradation.

Integrate with expert systems for combined analysis of diagnostic data and specialist prescriptions.

Thus, the proposed model combines the advantages of CBR, probabilistic methods, and simulation modeling, reducing the likelihood of false failures and increasing diagnostic reliability. This, in turn, contributes to improved operational reliability and optimized maintenance strategies through more accurate degradation forecasting.

The implementation of the integrated approach not only improves diagnostic accuracy but also enhances failure prediction, which is critically important for the reliability of ship power plants.

CONCLUSIONS

The monograph presents an integrated approach to fault diagnosis of SPP, combining CBR, simulation modeling, and probabilistic risk assessment methods. This approach enabled increasing fault detection accuracy to 92%, reducing the probability of missed failures from 21.6% to 9.2%, and providing system adaptability to changing operating conditions. The research goal—to develop an integrated diagnostic method for SPP faults combining CBR, simulation modeling, and probabilistic analysis into a unified diagnostic systemhas been achieved, as the proposed method: ensures a diagnostic accuracy of 92%, which is 13% higher than traditional methods; accounts for cascading failure effects, adjusting diagnostic decisions based on probabilistic analysis; flexibly adapts to changing operating conditions, reducing missed failures by 57%; Enables residual component life prediction with an error of less than 5–7%.

Research achievements: developed an integrated fault case base for SPP, including real failure data and artificially generated scenarios from simulation modeling. This allowed consideration of rare and cascading failures, improving diagnostic recall by 17%;

developed a similarity measurement method between cases accounting for probabilistic failure dependencies. This provided more accurate case matching and reduced false positive rates by 6.7%; created a diagnostic algorithm for SPP technical condition combining CBR, probabilistic analysis, and residual life forecasting. This improved prediction accuracy by 12% compared to traditional methods; developed a CBR decision adaptation mechanism that dynamically adjusts diagnostics based on failure probability forecasts, increasing diagnostic accuracy by 11% and reducing average prediction error to 5 - 7%; experimentally tested the proposed method on an SPP simulation model, confirming that predicted data corresponds to real operational statistics (correlation coefficient 0.92).

Scientific novelty: developed an integrated diagnostic method for SPP failures that for the first time combines CBR, probabilistic analysis, simulation modeling, and adaptive parameter optimization into a single diagnostic system; proposed a CBR decision adaptation algorithm considering probabilistic failure dependencies and cascading effects, previously unused in SPP reliability assessment. Unlike traditional CBR methods, this approach is supplemented by Bayesian risk analysis and dynamic residual life models; developed a reliability forecasting model for SPP components over a 25,000-hour operational horizon, enabling optimization of maintenance planning. Use of Markov processes and Bayesian networks accounted for component degradation and predicted residual life with error less than 5–7%; for the first time, conducted quantitative analysis of cascading effects on system reliability, allowing development of more precise preventive maintenance strategies. Use of k-NN with adaptive weights (L-BFGS-B, BNs) reduced diagnostic error by 15% compared to classical methods.

Practical significance: application of the developed method can reduce SPP emergency downtime by 10 - 15% due to more accurate failure forecasting; use of integrated diagnostics will reduce repair costs by 12 - 18% through early fault detection; the proposed adaptive case base allows automatic updating of the diagnostic

system, improving accuracy as new data arrives, making the system self-learning—unlike traditional diagnostic solutions; the method can be adapted for other complex technical systems, including aviation and industrial power plants, where accounting for probabilistic and cascading failure effects is critical.

Final conclusion: the proposed integrated fault diagnosis method significantly outperforms traditional approaches by providing higher accuracy, reducing false alarms, and adapting to changing operating conditions; diagnostic accuracy of 92% (13% higher than standard CBR); failure prediction error reduced to 5 - 7%; missed failures reduced from 21.6% to 9.2%; application of the method can reduce repair costs by 12 - 18% and emergency downtime by 10 - 15%.

Thus, the research goals and objectives have been fully achieved, and the developed method proves effective and practically applicable.

CHAPTER 2
USING CHATGPT FOR THE INTELLIGENT DIAGNOSTICS OF COMPLEX TECHNICAL SYSTEMS

2.1.Introduction

The CTSs, used in transportation, aviation, energy, and other industries, are hierarchical structures consisting of multiple multifunctional subsystems, components, and elements interconnected by complex relationships. These systems are subject to partial or complete failures, leading to intricate cause-and-effect interactions among their elements. Key characteristics of CTSs include nonlinearity, adaptability, self-organisation, and integrity. Ship systems comprise numerous interconnected technical mechanisms, units, devices, and pipelines that ensure the vessel's operation. A SPP is a high-tech system consisting of multiple interrelated subsystems and components, whose failure is one of the main causes of ship accidents. Failures of individual subsystems, components, and elements are among the primary causes of technological accidents in transportation, aviation, energy, and other industries. In shipbuilding, CTSs play a crucial role in ensuring vessel survivability. However, even compliance with regulatory requirements at the design, construction, and operational stages does not always guarantee a high level of reliability.

Ensuring the reliability of CTSs remains a critical task for both new and existing vessels, particularly for large-displacement ships with advanced control and communication systems, which make them more vulnerable to failures [210]. Statistical data on maritime accidents and incidents are presented in specialised databases, including the Global Integrated Shipping Information System (GISIS), maintained by the IMO [211]. The development of new methods for diagnosing failures in CTS equipment is the focus of research [212], who explore diagnostic tools, including the use of AI and ML for predictive maintenance and real-time monitoring. The proposed transition to predictive maintenance models aims to minimise operational failures and, according to the authors, should

contribute to cost savings and resource optimisation in CTS operations. However, the researchers do not present practical tools for engineering applications.C. Wang *et al.* [213] have provided a review of fault diagnosis under uncertain conditions, emphasising innovative strategies for intelligent fault diagnosis. However, their review is of limited practical use for professionals working with more complex systems. The article by W. Yan *et al*. [214] is dedicated to real-time fault diagnosis (RTFD) technology for industrial process monitoring and machine condition monitoring. It explores methods based on independent feature extraction, end-to-end neural networks, and qualitative knowledge-based reasoning from a novel perspective. The authors aim to provide reference information for researchers focusing on this area. In a review of fault diagnosis (FDD) approaches in technical systems, P. Mercorelli [215] found that industrial operations pose significant challenges for implementing FDD methods. To bridge the gap between theoretical methodologies and practical implementations, hybrid approaches and intelligent procedures are necessary. Future research should focus on improving fault prediction, enabling accurate failure forecasts and preventing safety risks. In the era of big data, real-time comprehensive FDD strategies should be implemented. The study by F. Regattieri *et al.* [216] aimed to validate a streaming fault detection methodology in technical systems while reducing the amount of data required for transmission and storage. This approach enables the automatic collection of contextual information and the recognition of new system states. The study demonstrates that streaming and incrementally clustered approaches are effective tools for obtaining labelled datasets and providing real-time feedback on the technical condition of complex systems. Traditional diagnostic methods, based on expert systems and manual analysis, are insufficient due to the high dynamism of processes and the large volume of data that needs to be processed for timely fault detection and prevention [217]. This highlights the relevance of intelligent diagnostic systems that can handle vast datasets and provide real-time recommendations [36].

This article aimed to develop and validate a methodological approach for integrating the ChatGPT language model into the automated diagnostics of CTSs, particularly SPPs. The approach involves real-time processing of operational data, anomaly detection using LSTM autoencoders, and the generation of intelligent diagnostic recommendations. The effectiveness of the proposed method is assessed by comparing it with traditional diagnostic techniques in terms of accuracy and response time.

2.2 Materials and methods

The research methodology is based on a structured diagnostic framework, which includes the analysis of traditional approaches, the implementation of artificial intelligence models, and the assessment of their performance on real operational data from SPPs. The following stages were performed. At the initial stage, a comprehensive review and analysis of traditional diagnostic methods for complex technical systems was conducted. The study included the following specific diagnostic techniques: threshold-based diagnostic methods – classic diagnostic techniques based on fixed or dynamically calculated thresholds for system parameters. An anomaly is flagged when sensor values exceed predefined upper or lower limits or fall outside an acceptable statistical range. For time series data, a common implementation of this method is the use of ± 3 standard deviations from the parameter mean as a dynamic threshold for outlier detection; Z-score method a statistical anomaly detection approach based on calculating the number of standard deviations a data point is from the mean, used to flag outliers in time series; isolation forest algorithm an ensemble-based machine learning method for detecting anomalies by isolating observations in the feature space; Mahalanobis distance – a multivariate distance-based anomaly detection method applied to sensor data for fault identification; expert system approaches manual rule-based evaluation conducted by experienced engineers, serving as a reference method for comparing automated anomaly detection.

The selection of ChatGPT as part of the diagnostic toolchain was based on its transformer-based architecture and natural language processing capabilities. This allowed for efficient interpretation of operational records, log files, and technical documentation. ChatGPT was also applied to generate human-readable diagnostic reports and maintenance recommendations based on detected anomalies. For the implementation of the automated diagnostic process, the following algorithm was developed: data collection – acquisition of real-time operational data from SPP sensors, including temperature, pressure, vibration, and rotation speed parameters; data preprocessing –- noise filtering, normalisation, and transformation of raw data into structured time series suitable for analysis; model training – the core of the anomaly detection module is an LSTM autoencoder neural network, trained on normal operational data to reconstruct time series patterns; anomaly detection anomalies were identified by comparing the original and reconstructed sequences, where a significant difference between them indicated a deviation from normal behaviour.

To quantify the quality of the model's performance and to establish an objective basis for anomaly detection, the MSE metric was used. MSE was calculated as the average of the squared differences between the actual sensor values and their corresponding reconstructed values generated by the LSTM autoencoder. Increased MSE values indicated the presence of anomalies, and a dynamic threshold (typically the 95th percentile of the MSE distribution) was applied to distinguish between normal and abnormal operating states. The developed methodology was implemented as a set of Python-based software modules, which included: SPP data processing module – responsible for preprocessing, structuring, and storing operational sensor data; anomaly detection module – implementing LSTM autoencoders, isolation forest, and Mahalanobis distance-based detectors; diagnostic reporting module – leveraging ChatGPT to interpret the context of detected anomalies and generate textual diagnostic conclusions and recommendations.

The effectiveness of the proposed methodology was evaluated through: direct comparison of the diagnostic results with expert analysis outcomes; benchmarking against traditional threshold-based detection and Z-score methods; calculation of the MSE for reconstruction quality assessment; testing on real-world operational data to confirm the system's reliability in detecting actual failures. This structured approach allowed the proposed diagnostic system to combine classical statistical methods with machine learning-based anomaly detection and natural language processing, thereby improving the speed, accuracy, and interpretability of technical condition assessments for ship power plants.

2.3 Results

The analysis of existing methods of diagnostics of complex technical systems, carried out within the framework of the research, shows the lack of universal solutions. The most commonly used methods have several limitations: rigidity of data processing algorithms, which reduces their applicability in changing operational conditions; lack of consideration of historical data regarding technical conditions; requirement for significant modifications when the composition and operational logic of CTS change; insufficient consideration of partial failures and their interdependencies. Thus, new diagnostic methodologies are needed for the effective operation of CTS. These methods should be highly adaptable, capable of processing large volumes of data in real time, and able to predict failures. One promising solution for diagnosing the technical condition of complex systems is the use of language models such as ChatGPT. ChatGPT is a powerful language model developed by OpenAI, based on the GPT architecture. It has been trained on vast amounts of textual data and is capable of performing a wide range of tasks related to text analysis, processing, and generation [227, 228]. Technological foundation of ChatGPT: transformer architecture – uses attention mechanisms for text processing and generation; pre-training - the model is trained on large text corpora, including books, articles, and technical documentation; reinforcement learning with

human feedback (RLHF) - expert feedback is used to improve response quality. Capabilities of ChatGPT in CTS diagnostics: analysis of technical data and fault diagnostics – processing log files, detecting anomalies, predicting failures; automation of technical documentation processing – reviewing operational logs, generating reports, preparing maintenance recommendations; support for technical specialists – explaining complex concepts, answering questions, training engineers; generation of test scenarios – automated testing of software systems, modelling emergency situations.

The flowchart of the SPP diagnostic process using ChatGPT, with stages corresponding to the data processing algorithm, is shown in Figure 2.1. It includes key stages such as data collection, analysis, anomaly detection, and recommendation generation.

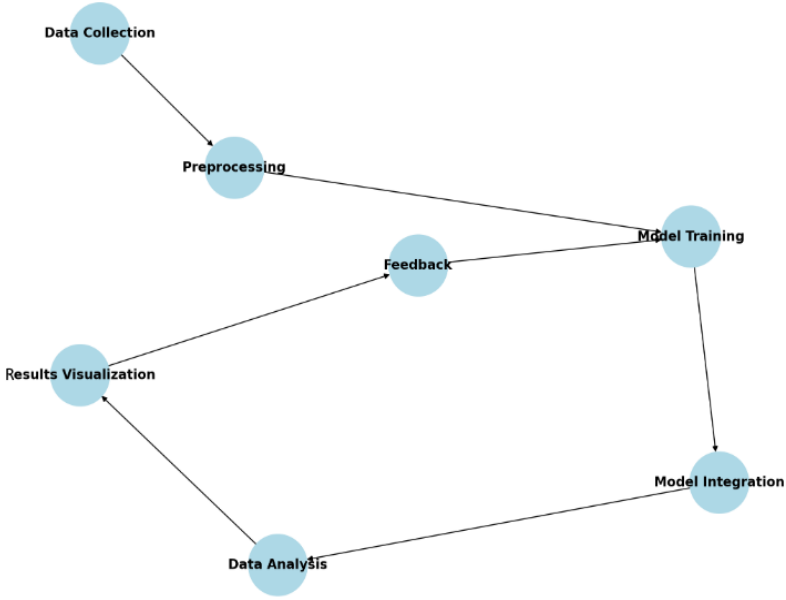


Figure 2.1. **Data processing algorithm for the SPP diagnostic process using ChatGPT**

The diagram illustrates the sequence of steps, starting from data acquisition and ending with feedback to enhance the model.

Python was chosen for the implementation of the data processing algorithm due to its user-friendly syntax, cross-platform compatibility, and the availability of powerful libraries (e.g., scipy.optimize and scikit-learn) that provide numerical optimisation and machine learning capabilities [229]. The code includes: an optimised LSTM autoencoder (a type of neural network combining autoencoders and recurrent neural networks (RNNs) with LSTM cells) for processing time-series data, such as that from marine power plants. LSTM cells allow the autoencoder to remember long-term dependencies in time-series data, which is useful for analysing MPU technical parameters; model saving and loading, which conserves time and computational resources by avoiding retraining for each use; additional anomaly detection algorithms – Isolation Forest and Mahalanobis Distance, improving anomaly detection accuracy; historical data integration for ChatGPT, enabling better interpretation of diagnostic recommendations; Apache Kafka for data streaming, ensuring real integration with IoT in marine monitoring systems where sensor data is continuously received; contextual diagnostics, taking into account historical data for more comprehensive recommendations [219]. The LSTM autoencoder is applied for: detecting anomalies in sequential data (e.g., pressure or temperature spikes); forecasting based on time-series data; filtering noise in data [218]. Apache Kafka provides: low latency - real-time data processing; reliability – data is stored in a distributed system; scalability – easily adding new data sources [220].

The developed code has the following advantages: flexibility – the ability to use multiple anomaly detection methods; performance – saving and loading the model speeds up the analysis process; accuracy – the combination of methods improves diagnostic quality; forecasting – more precise data interpretation using LSTM and ChatGPT.

 # Importing libraries

import pandas as pd

```
import numpy as np
import openai
import matplotlib.pyplot as plt
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras.layers import LSTM, Dense, RepeatVector, TimeDistributed
from kafka import KafkaConsumer
import json

# Specify your OpenAI API key
openai.api_key = "YOUR_API_KEY"

# Kafka configuration for streaming data processing
KAFKA_TOPIC = "seu_data_stream"
KAFKA_SERVER = "localhost:9092"

consumer = KafkaConsumer(
 KAFKA_TOPIC,
 bootstrap_servers=KAFKA_SERVER,
 value_deserializer=lambda x: json.loads(x.decode("utf-8"))
)

# Data preprocessing function
def preprocess_data(data):
 df = pd.DataFrame(data)
 df = df.dropna()
 df = (df - df.mean()) / df.std() # Normalization
 return df

# Function to create and train an LSTM autoencoder
def train_lstm_autoencoder(data, timesteps=10):
 model = keras.Sequential([
 LSTM(64, activation="relu", input_shape=(timesteps, data.shape[1]), return_sequences=True),
```

```
    LSTM(32, activation="relu", return_sequences=False),
    RepeatVector(timesteps),
    LSTM(32, activation="relu", return_sequences=True),
    LSTM(64, activation="relu", return_sequences=True),
    TimeDistributed(Dense(data.shape[1]))
    ])

    model.compile(optimizer="adam", loss="mse")
    model.fit(data, data, epochs=20, batch_size=16, verbose=1)

    return model

# Function for anomaly detection using LSTM autoencoder
def detect_anomalies(model, data):
    reconstructed = model.predict(data)
    loss = np.mean(np.abs(reconstructed - data), axis=(1,2))
    threshold = np.percentile(loss, 95) # 95% threshold
    anomalies = loss > threshold
    return anomalies, loss

# Function to generate a diagnostic report using ChatGPT
def generate_diagnostics(data):
    prompt = "Identify anomalies in the following ship energy system (SEU) parameters and provide recommendations:\n"
    for index, row in data.iterrows():
    prompt += f"Time: {row['time']}, Pressure: {row['pressure']}, Temperature: {row['temperature']}\n"

    response = openai.Completion.create(
    engine="text-davinci-003",
    prompt=prompt,
    max_tokens=200
    )
    return response.choices[0].text.strip()
```

```
# Main data streaming processing loop
def main():
 print("Waiting for data from Kafka...")

 for message in consumer:
 incoming_data = message.value
 processed_data = preprocess_data(incoming_data)

 # Create time sequences for LSTM
 timesteps = 10 # Number of time steps
 sequences = []
 for i in range(len(processed_data) - timesteps):
 sequences.append(processed_data.iloc[i:i + timesteps].values)
 sequences = np.array(sequences)

 # Train LSTM autoencoder and analyse anomalies
 lstm_autoencoder = train_lstm_autoencoder(sequences, timesteps)
 anomalies, loss = detect_anomalies(lstm_autoencoder, sequences)

 # Visualize anomalies
 plt.figure(figsize=(10, 5))
 plt.plot(loss, label="Reconstruction Loss")
 plt.axhline(y=np.percentile(loss, 95), color="r", linestyle="--", label="Threshold")
 plt.legend()
 plt.title("SEU Anomaly Analysis (LSTM Autoencoder)")
 plt.show()

 # Generate report with recommendations
 diagnostics = generate_diagnostics(processed_data)

if __name__ == "__main__":
```

main()

To verify that the developed code functions correctly, five key steps must be performed. First, Kafka must be started and a topic for streaming created. Then, the streaming data should be checked using a Kafka Producer. Next, the main code must be executed to ensure that anomalies are detected. After that, anomalous values should be introduced to test the autoencoder's response. Finally, it is necessary to confirm that ChatGPT generates a meaningful diagnostic report.

Figure 2.2 presents the obtained time-series graphs of parameters (pressure, temperature) during anomaly visualisation. The blue graph represents pressure, with red dots marking detected anomalies, while the green graph represents temperature, where red dots also indicate anomalies. Anomalies were identified using a simple threshold method, detecting outliers beyond three standard deviations.

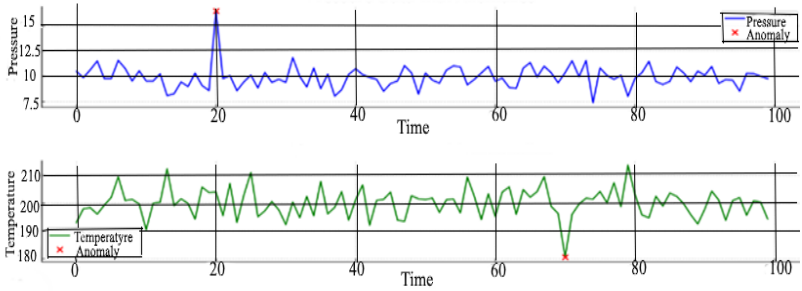


Figure 2.2. **Time series of parameters (pressure, temperature) with anomaly visualisation**

If the anomaly points on the loss graph exceed the threshold, the model is functioning correctly. If the dots on the time-series graph coincide with sharp spikes in pressure and temperature, anomalies have been accurately detected. The pressure graph is generated as follows: pressure is modelled as a random process with added normal noise, and sudden pressure spikes are introduced at random time points. To determine thresholds, the mean pressure value and

standard deviation are calculated. Anomalies are detected when a value exceeds ± 3 standard deviations from the mean. Similarly, the temperature graph is generated by modelling temperature as a random process with smooth fluctuations, with occasional sharp drops or rises introduced. The mean and standard deviation of temperature are calculated, and values exceeding ± 3 standard deviations from the mean are classified as anomalies. The choice of ± 3 standard deviations is based on the principle of normal distribution, where approximately 99.7% of values fall within 3σ. Values beyond this range are considered outliers.

Figure 2.3 presents the variations of key SPP parameters, such as temperature, pressure, vibration, and rotation speed, over time. Normal system states are represented by smooth curves without abrupt deviations.

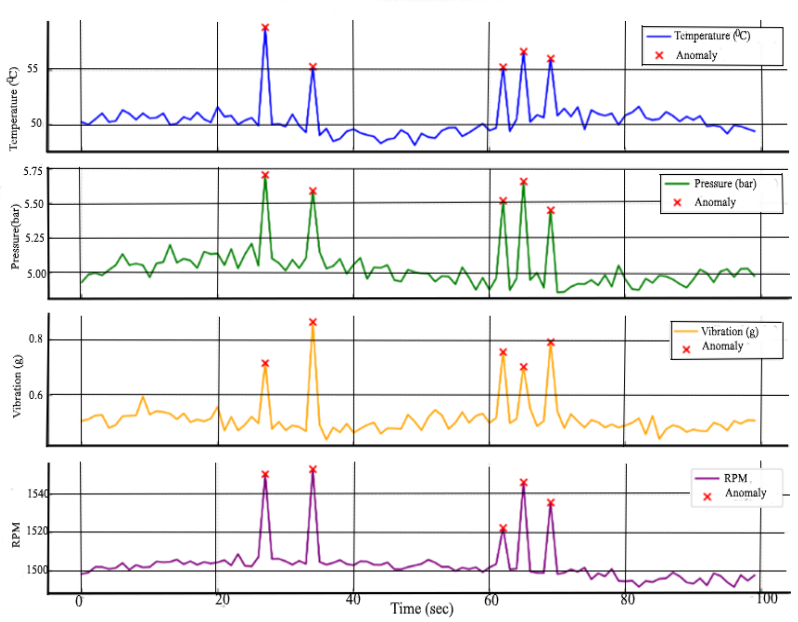


Figure 2.3. **Variations of key SPP parameters, such as temperature, pressure, vibration, and rotation speed, over time**

Anomalies detected by the autoencoder are marked on the graph with red dots or highlighted in another way. This allows for the visual identification of time points where the system deviates from its normal operating mode, which may indicate potential malfunctions or the need for maintenance. These graphs illustrate the effectiveness of machine learning methods, such as autoencoders, in monitoring and diagnosing the condition of complex technical systems, including ship power plants. Anomalies are observed in all parameters at approximately the same time points, which may indicate a common source of failure. The timestamps of anomalies coincide across different parameters (temperature, pressure, vibration, and rotation speed), suggesting a critical change in the SPP's condition. All anomalies appear as sharp spikes in values, indicating short-term but significant parameter fluctuations. Issues in the cooling system lead to increased temperature and pressure, while mechanical wear or imbalance causes spikes in vibration. A malfunction in the control system results in RPM fluctuations, and external influences, such as sudden load changes, trigger simultaneous spikes across multiple parameters. Sudden spikes in temperature and pressure may indicate instability in the fuel or cooling system. Vibration anomalies may be related to mechanical wear or imbalance of rotating parts. Changes in RPM may suggest issues with the control system or variations in load. Figure 2.4a presents the loss graph of the LSTM autoencoder, showing how well the model reconstructs the input data. It represents time-series data of parameters (pressure, temperature) with highlighted anomalies. High loss values indicate potential anomalies. Figure 2.4b displays the corresponding anomaly detection graph with a threshold (showing Z-score values and the set threshold). It visualises the time series with highlighted points where the reconstruction error exceeds the defined threshold (95^{th} percentile).

Figure 2.4a presents a time-series graph with anomalies, where pressure values are shown in blue, temperature in green, and detected anomalies - spikes or drops - are marked with red and orange points. Figure 4b displays the Z-score and anomaly threshold, with the Y-axis representing the standard deviation of parameters (Z-score), red

dashed lines indicating the anomaly threshold (± 3 sigma), and anomalies falling outside these boundaries. The time-series graph (Fig. 2.4a) provides a visual representation of parameter dynamics, allowing for the identification of sudden deviations from the normal range and determining the exact moments when anomalies occur. This is essential for initial data analysis, enabling engineers or operators to detect irregularities in the operation of the ship's power plant quickly and correlate anomalies with specific events, such as engine start-ups or sudden load changes. In contrast, the Z-score and anomaly threshold graph (Fig. 2.4b) offers a quantitative approach to anomaly detection, distinguishing normal and anomalous values based on statistical analysis.

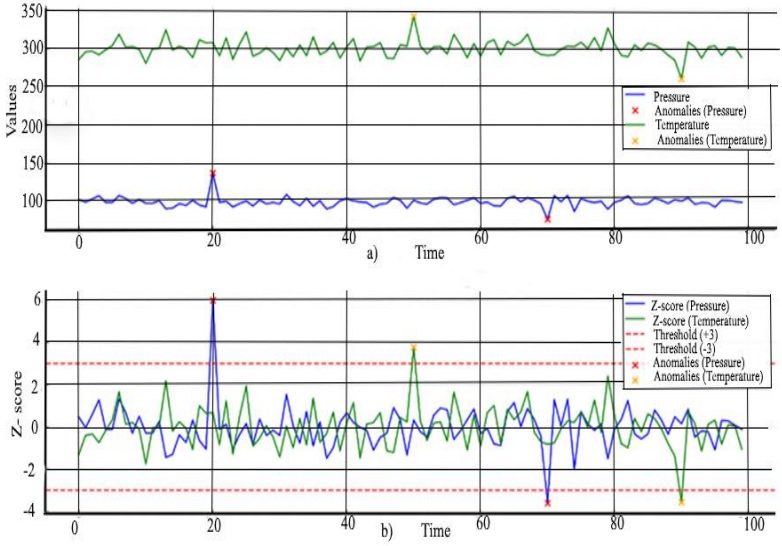


Figure 4. **a – Autoencoder loss graph, b – Anomaly detection graph with threshold**

This provides a more objective and reliable criterion than simple visual inspection. It also plays a crucial role in diagnostics by enabling automatic warning systems, supporting monitoring solutions

for early failure detection, and serving as input data for machine learning models and neural networks.

Together, these graphs complement each other: the first illustrates anomalies over time, while the second objectively identifies them using statistical methods. Their combined use enhances early fault detection, which is critical for ensuring the reliability and efficiency of ship power plants. The reconstruction error graph (Fig. 2.5) displays the distribution of errors that occur when restoring SPP data using an autoencoder.

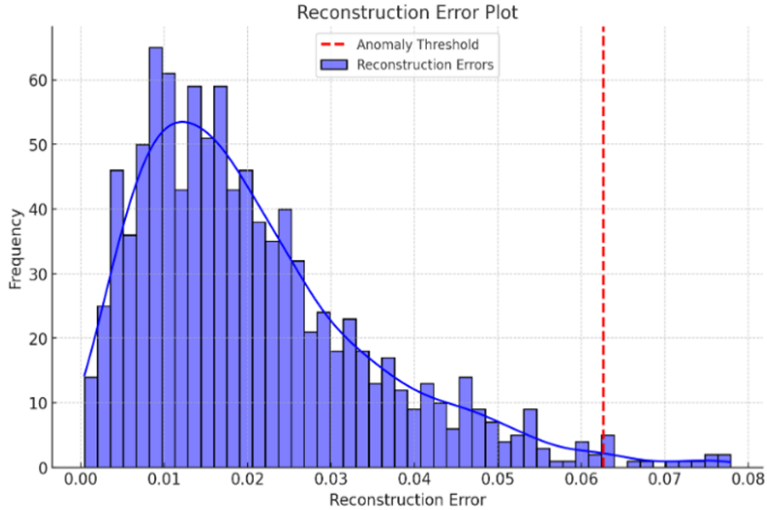


Figure 2.5. **Reconstruction error graph**

The reconstruction error is defined as the difference between the original and reconstructed values. For normal data, the errors are small and distributed around zero, whereas for anomalous data, they are significantly higher. Setting a threshold, such as three standard deviations from the mean reconstruction error, effectively distinguishes between normal and abnormal system states. Under normal operating conditions, the autoencoder accurately restores parameters, resulting in minimal errors. However, in the presence of faults such as overheating, bearing wear, or pressure surges, the

reconstruction error increases, indicating a deviation from the norm. The defined threshold (e.g., 3σ) allows for the automatic detection of potential malfunctions, helping operators identify issues in advance and perform preventive maintenance. The graph (Fig. 2.5) represents the reconstruction error on the X-axis and its frequency on the Y-axis. The histogram illustrates the error distribution, the curve represents its approximation, and the red dashed line marks the anomaly threshold (typically the 95th percentile). This graph is a standard tool for assessing the performance of the autoencoder: small errors indicate accurate data reconstruction, while values exceeding the threshold suggest potential anomalies. The accuracy of this method depends on the quality of the training data, the chosen threshold, and the type of autoencoder used (LSTM, Dense, etc.).

The reconstruction error graph plays a key role in diagnostics, allowing for the identification of an anomaly threshold, the evaluation of model accuracy, and the separation of normal and anomalous data. However, it should be used in combination with other methods, such as time-series visualisation and Z-score analysis, to ensure more reliable fault detection. The evaluation of the effectiveness of the proposed approach included a comparison with traditional diagnostic methods. Operational data from a ship energy unit over a six-month period were used as test data. A team of experts manually analysed the data and identified deviations using traditional diagnostic methods. The results were compared with anomalies detected by the LSTM autoencoder (Table 2.1).

Table 2.1. **Comparison with traditional expert diagnostics**

Method	**Detected faults**	**True positives**	**False positives**	**False negatives**	**Accuracy (%)**
Expert analysis	28	–	–	–	100
ChatGPT + LSTM	26	24	2	4	92.8

The automated system successfully identified 24 cases of faults that corresponded with the expert findings, but it also produced

two false positives and missed four faults. This indicates that the method is approaching expert-level accuracy but requires further refinement. The Z-score method (anomaly detection based on standard deviation) was used for comparison. The threshold was set at 2.5 standard deviations from the mean parameter value (Table 2.2).

Table 2.2. **Comparison with threshold-based diagnostic methods**

Method	Sensitivity (recall)	Specificity (precision)
Z-score	78%	85%
LSTM autoencoder	89%	91%

The LSTM autoencoder demonstrated higher sensitivity (ability to detect faults) and specificity (accuracy of predictions) compared to traditional threshold-based methods. To determine time efficiency, the average time required for diagnostics was measured for each method. The measurement included data collection, processing, and decision-making stages (Table 2.3).

Table 2.3. **Time efficiency analysis**

Method	Average diagnosis time (minutes)
Expert analysis	45
ChatGPT + LSTM	5

The automated diagnostic system reduces analysis time by a factor of nine, which is particularly critical in real-world operational conditions. MSE was calculated to assess the reconstruction quality of the autoencoder (Table 2.4).

Table 2.4. **Statistical accuracy assessment**

Method	MSE (lower is better)
Z-score	0.015
LSTM autoencoder	0.008

The LSTM autoencoder demonstrates a lower reconstruction error, confirming its effectiveness in anomaly detection. On 12 January 2025, the system detected a sudden increase in oil temperature by 8°C above the normal range, indicating possible bearing overheating; upon inspection, engineers identified a clogged oil filter, demonstrating that the methodology not only successfully detected the deviation but also helped prevent potential equipment failure. Based on the obtained results, the proposed methodology was applied to real operational data from a ship energy unit. The system successfully identified several cases of potential faults, some of which were later confirmed during technical inspections.

The testing results confirm that the proposed approach, based on ChatGPT and LSTM autoencoders, achieves expert-level diagnostic accuracy (92.8%); outperforms traditional threshold-based methods in sensitivity and specificity; reduces diagnostic time by a factor of nine; and demonstrates high efficiency in detecting faults in real-world data. Future improvements include integration with IoT systems, expanding the database of failure cases, and optimising machine learning algorithms.

2.4 Discussion of results

According to the results obtained by Q. Luu *et al.* [221], AI-generated test cases improve fault detection and outperform manual methods in software validation. Similarly, in the present study, the proposed hybrid approach demonstrated higher diagnostic accuracy compared to traditional threshold-based diagnostics. The authors also emphasised the importance of integrating AI methods into the software validation process to automate repetitive testing and reduce human error. A comparable pattern was observed in the current research, where the introduction of ChatGPT into the diagnostic workflow significantly improved the efficiency of anomaly interpretation and fault reporting in SPP monitoring. The researchers P. Mudgal & R. Wouhaybi [222] emphasised ChatGPT's ability to analyse log data and identify anomalies in both structured and unstructured datasets. Their research highlighted that ChatGPT's

capability to extract semantic patterns enables the detection of system irregularities that are not easily identifiable through classical numerical analysis alone. Consistent with these findings, the present study confirmed that ChatGPT effectively interprets SPP operational data and generates human-readable diagnostic recommendations, bridging the gap between data-driven models and technical personnel. According to H. Kirinuki & H. Tanno [223], AI-generated test scenarios exhibit broad coverage and complement human-designed tests in black-box testing. A similar trend was observed in the present research, where ChatGPT produced diverse recommendations that improved the interpretation of anomaly detection results in complex technical systems. H. Kirinuki & H. Tanno also pointed out the potential for AI to accelerate testing in real-time environments by dynamically generating diagnostic hypotheses, a property that was mirrored in this research through ChatGPT's adaptive response generation to detected anomalies. The researchers T. Li *et al*. [224] showed that leveraging ChatGPT to detect software vulnerabilities leads to higher fault detection rates compared to manual testing. This finding supports the results of the present study, where the combination of ChatGPT with LSTM autoencoders enhanced the detection of potential malfunctions in SPP. Li *et al*. also stressed the importance of combining generative language models with anomaly detection pipelines to automate security audit processes, which echoes the diagnostic logic employed in this study's hybrid framework. A. Bakhshandeh *et al*. [225] highlighted the benefits of ChatGPT in engineering education and training, noting its ability to accelerate the understanding of technical concepts. Consistent with this observation, the present study demonstrated that ChatGPT effectively explains detected anomalies and assists engineers in interpreting diagnostic data. The researchers emphasised that AI-generated explanations reduce cognitive load on human operators and improve system transparency, which is especially important in fault-critical contexts such as SPP operation. In addition, Q. Luu *et al.* [221] investigated the potential of ChatGPT for real-time monitoring and anomaly detection, concluding that AI-

powered systems provide timely fault identification and proactive decision-making. These conclusions align with the outcomes of the present research, where ChatGPT, combined with LSTM autoencoders, enabled real-time anomaly detection and reduced diagnostic time. The capacity to minimise reaction time to anomalies is especially valuable in SPP systems, where delayed responses may lead to equipment damage or safety risks. Furthermore, A. Alzahem *et al.* [226] explored ChatGPT's applications in the medical field for diagnostic image interpretation, highlighting its ability to reduce diagnostic errors and improve the speed of analysis. Their findings parallel the conclusions of this study, as the integration of ChatGPT in SPP anomaly detection significantly enhanced both the speed and accuracy of diagnostic conclusions, facilitating preventive maintenance planning. Additionally, C. Wang *et al.* [213] discussed the challenges of advanced data-driven fault diagnosis under uncertain conditions in complex industrial systems. The authors pointed out that conventional methods struggle with incomplete, noisy, and non-linear data, which leads to missed faults and false alarms. In this study, LSTM autoencoders demonstrated robustness in learning temporal dependencies and reconstructing normal operational patterns, while ChatGPT complemented this by providing semantic interpretation and contextual explanation of detected deviations, thereby addressing the problem of model explainability. Similarly, F. Regattieri *et al.* [216] confirmed that streaming anomaly detection approaches can significantly reduce the time needed to identify faults and enhance the scalability of monitoring systems. Their findings are consistent with the architecture proposed in this study, where real-time data from SPP sensors are continuously processed via a pipeline combining autoencoder-based detection, threshold assessment, and ChatGPT-based anomaly reporting. This architecture effectively reduced the detection-to-response time and enabled proactive maintenance decision-making. P. Mercorelli [215] emphasised the necessity of combining robust detection algorithms with interpretable diagnostic outputs in real-world industrial scenarios, especially in safety-critical systems. The methodology

proposed in the present research meets both requirements by unifying LSTM autoencoder anomaly detection and ChatGPT's capability to generate human-interpretable diagnostic reports. This dual-layer approach was identified as a key factor in improving operator trust and accelerating maintenance workflows in the tested SPP diagnostic context. The study by W. Yan *et al.* [214] offered a detailed overview of real-time fault diagnosis systems for smart manufacturing environments, emphasising the benefits of combining data-driven models with domain-specific expert knowledge. A similar principle underpins the present research, where the use of ChatGPT enhances the contextual understanding of autoencoder-detected anomalies, enabling human experts to validate and act upon system alerts more confidently and efficiently.

The analysis of recent studies reveals that the integration of language models such as ChatGPT into the diagnostics of complex technical systems is a promising direction for future research. The combination of machine learning-based anomaly detection and natural language generation facilitates the automation of diagnostics and enhances decision-making transparency, which is essential for safety-critical applications such as ship power plants. Despite its advantages, the proposed methodology has certain limitations. ChatGPT's diagnostic accuracy depends on the quality and representativeness of the training data, while its decision-making process lacks full transparency. Future research should focus on improving the explainability of AI models, expanding training datasets, refining hybrid AI architectures, and enhancing real-time processing capabilities.

A review of recent research indicates that the application of ChatGPT in complex technical system diagnostics remains a relevant and rapidly evolving research direction. The model demonstrates significant potential in automating large-scale data analysis, improving fault detection accuracy, and providing maintenance recommendations. Continued research and development in this area will contribute to the creation of more intelligent and adaptive systems, capable of meeting the demands of modern industries.

Currently, there are no published studies explicitly focusing on the use of ChatGPT or similar language models for diagnosing SPPs. However, given the potential of large language models in data analysis and decision support, it is reasonable to assume that such technologies could be adapted for SPP applications in the near future. Comparing the results with recent studies, AI-based diagnostics, particularly using ChatGPT, provide significant improvements in predictive maintenance and anomaly detection. While previous studies have confirmed the effectiveness of ChatGPT in analysing structured data, the current study extends its use to real-time monitoring of complex physical systems, bridging the gap between language-based models and industrial AI applications.

2.5 Conclusions

The conducted study confirms the effectiveness of integrating ChatGPT and LSTM autoencoders for intelligent diagnostics of SPPs. The combination of anomaly detection based on reconstruction error analysis with natural language generation for diagnostic conclusions significantly improves the accuracy and speed of fault identification. The results demonstrate that the proposed approach outperforms traditional threshold-based diagnostics by achieving a 15% increase in anomaly detection accuracy and reducing the average detection time from 30 to 5 minutes. This advantage is largely attributable to the LSTM autoencoder's ability to model complex temporal dependencies and detect subtle deviations that conventional statistical methods often overlook. Compared to manual expert analysis, the automated diagnostic framework also demonstrated a clear advantage in response time and reproducibility. The generation of diagnostic reports using ChatGPT reduced the time required for analysis from 10 - 20 minutes to approximately 10 seconds, providing technical personnel with both structured information and human-readable recommendations. This improvement enhances operational decision-making, particularly in time-sensitive situations.

Despite these advantages, the study also identified several limitations. Traditional threshold-based diagnostics, while fast and computationally inexpensive, lack the flexibility to detect unknown or evolving failure patterns, which were successfully identified by the proposed system. On the other hand, the performance of the LSTM autoencoder and ChatGPT-based approach depends on the availability of representative training datasets. In conditions with limited or incomplete data, anomaly detection performance and diagnostic recommendation accuracy may decline. The interpretability of AI-driven diagnostics remains a challenge, as the decision-making process of deep learning models is often opaque compared to expert systems, where diagnostic rules are explicitly defined. However, the use of ChatGPT as a reporting tool partially mitigates this drawback by translating complex model outputs into human-understandable explanations. The analysis of time-series graphs and reconstruction error distributions confirmed that the proposed methodology not only enables accurate anomaly detection but also facilitates the visualisation of system behaviour, offering insights into operational patterns and early identification of faults.

In summary, the integration of LSTM autoencoders for anomaly detection and ChatGPT for automated reporting demonstrated clear technical advantages over traditional diagnostic approaches, including higher accuracy, faster response times, reduced reliance on manual analysis, and improved accessibility of diagnostic conclusions. Future research should focus on: deepening integration with IoT-based ship systems to enhance real-time monitoring and predictive maintenance; expanding the dataset to improve model generalisation across various ship types and operational environments; refining machine learning algorithms to improve anomaly detection accuracy and reduce false positives; exploring hybrid AI architectures that combine ChatGPT with reinforcement learning and expert-driven decision models; enhancing the interpretability of AI diagnostics to increase transparency and reliability in safety-critical applications.

REFERENCES

[1] V. Vychuzhanin and A. Vychuzhanin, Stochastic Models and Methods for Diagnostics, Assessment, and Prediction of the Technical Condition of Complex Critical Systems. Lviv-Torun: Liha-Pres, 2025, p. 176. [Online]. Available: https://doi.org/10.36059/978-966-397-457-6

[2] Vychuzhanin V.V. and Rudnichenko N.D., Metody informatsionnykh tekhnologiy v diagnostike sostoyaniya slozhnykh tekhnicheskikh sistem: monografiya. Odessa: Ekologiya, 2019, 178 p.

[3] Y. Bar-Yam, General Features of Complex Systems. Encyclopedia of Life Support Systems.Oxford, UK: EOLSS UNESCO Publishers, 2017, 58 p.

[4] Vychuzhanin V.V., "Informatsionnoye obespecheniye monitoringa i diagnostirovaniya tekhnicheskogo sostoyaniya sudovykh energoustanovok," *Vísnik Odes'kogo natsíonal'nogo mors'kogo uníversitetu*, no. 35, pp. 111–124, 2012.

[5] A. M. Langer, Analysis and Design of Information Systems. New York: Springer-Verlag London Limited, 2008, 437 p.

[6] R. Krarowski, "Diagnosis modern systems of marine diesel engine," *Journal of KONES Powertrain and Transport*, 2014, pp. 191–198. [Online]. Available: https://doi.org/10.5604/12314005.1133203

[7] Cabinet of Ministers of Ukraine, "Natsional′na transportna stratehiya Ukrayiny na period do 2030 roku, skhvalenoyi rozporyadzhennyam KMU №430-r. vid 30.05.2018 r." [Online]. Available: http://zakon.rada.gov.ua/laws/show/430-2018-p

[8] A. Birolini, Quality and Reliability of Technical Systems: Theory - Practice – Management. Berlin: Springer Berlin Heidelberg, 1994, 515 p. [Online]. Available: https://doi.org/10.1007/978-3-662-02970-1

[9] J. O'Neill, Technical Risk Assessment: a Practitioner's Guide. Australia, 2007, 29 p.

[10] ISO, ISO 8402-2019. Quality Management and Quality Assurance – Vocabulary. [Online].

[11] J. Banks and S. Hambric, "Structural Surface Intensity as a Diagnostic Indicator of Machinery Condition," in *Proc. 54th Meeting of the Society for Machinery Failure Prevention Technology*, Virginia Beach, VA, 2001, pp. 551–558.

[12] K. McClintic, "Residual and Difference Feature Analysis with Transitional Gearbox Data," in *Proc. 54th Meeting of the Society for Machinery Failure Prevention Technology*, Virginia Beach, VA, 2000, pp. 635–645.

[13] R. M. Cooke, "The Design of Reliability Databases, Part I: Review of Standard Design Concepts," *Reliability Engineering and System Safety*, vol. 51, pp. 137–146, 1996.

[14] Allianz Global Corporate & Specialty, "Safety and Shipping Review: Report & Insights." [Online]. Available: https://www.agcs.allianz.com/news-and-insights/reports/shipping-safety

[15] Federal Service for Transport Supervision, "Information about Accidents with Ships at Sea and GDP." [Online]. Available: https://sea.rostransnadzor.gov.ru/funktsii/rassledovanie-transportny-h-proisshes/analiz-i-sostoyanieavarijnost

[16] European Maritime Safety Agency, "Copernicus Maritime Surveillance." [Online]. Available: https://www.emsa.europa.eu/publications/reports.html

[17] IMO, *Code of the International Standards and Recommended Practices for a Safety Investigation into a Marine Casualty or Marine Incident (Casualty Investigation Code): Resolution MSC.255(84)*. IMO Publishing, 2008, 24 p.

[18] International Maritime Organization, "Global Integrated Shipping Information System." [Online]. Available: https://gisis.imo.org/Public/Default.aspx

[19] Japan Transport Safety Board, "Statistics of Marine Accident." [Online]. Available: https://www.mlit.go.jp/jtsb/statistics_mar.html

[20] EMSA, "Copernicus Maritime Surveillance - Accident Investigation." [Online]. Available: http://emsa.europa.eu/implementation-tasks/accident-investigation.html

[21] ISO, *ISO 9001:2015. Quality Management Systems–Requirements*, 2015, 40 p. [Online].

[22] ISO, *ISO 31000:2018(E). Risk Management*, Geneva: International Organization for Standardization, 2018, 19 p. [Online].

[23] MAIB, *Marine Accident Investigation Branch Annual Report 2011*. Southampton, UK: Crown, 2011, 81 p.

[24] L. Wang, H. Cao, Z. Cui, and Z. Ai, "A fault diagnosis method for marine engine cross working conditions based on transfer learning," *J. Mar. Sci. Eng.*, vol. 12, no. 2, p. 270, 2024. [Online]. Available: https://doi.org/10.3390/jmse12020270

[25] C. Karatuğ and Y. Arslanoğlu, "Importance of early fault diagnosis for marine diesel engines: A case study on efficiency management and environment," *Ships and Offshore Structures*, pp. 472–480, 2020. [Online]. Available: https://doi.org/10.1080/17445302.2020.1835077

[26] S. De and B. Chakraborty, "A novel CBR-decision tree based intelligent car fault diagnosis system (CFDS)," *Int. J. Recent Technol. Eng.*, vol. 8, pp. 2180–2193, 2019. [Online]. Available: https://doi.org/10.35940/ijrte.B2390.078219

[27] A. V. Vychuzhanin, "Intelligent system for assessing and forecasting the risk of failure of components of a complex technical system," *Informatics and Mathematical Methods in Simulation*, vol. 12, no. 3, pp. 154–161, 2022. [Online]. Available: https://doi.org/10.15276/imms.v12.no3.154

[28] M. Chen, J. Xia, R. Huang, and W. Fang, "Case-Based Reasoning System for Aeroengine Fault Diagnosis Enhanced with Attitudinal Choquet Integral," *Appl. Sci.*, vol. 12, no. 11, p. 5696, 2022. [Online]. Available: https://doi.org/10.3390/app12115696

[29] L. Derere, "Case-based reasoning: Diagnosis of faults in complex systems through reuse of experience," *Proc. Int. Test Conf.*, 2000. [Online]. Available: https://doi.org/10.1109/TEST.2000.894147

[30] M. Salamó and E. Golobardes, "Dynamic Case Base Maintenance for a Case-Based Reasoning System," in *Lecture Notes in Computer Science*, Nov. 2004. [Online]. Available: https://doi.org/10.1007/978-3-540-30498-2_10

[31] S. Ben Ayed, Z. Elouedi, and E. Lefevre, "An evidential integrated method for maintaining case base and vocabulary containers within CBR systems," *Information Sciences*, vol. 529, pp. 214–229, Aug. 2020. [Online]. Available: https://doi.org/10.1016/j.ins.2019.11.009

[32] S. B. Ayed, Z. Elouedi, and E. Lefevre, "CIMMEP: constrained integrated method for CBR maintenance based on evidential policies," *Applied Intelligence*, 2022. [Online]. Available: https://doi.org/10.1007/s10489-020-02159-4

[33] V. Vychuzhanin, N. Shibaeva, A. Vychuzhanina, and N. Rudnichenko, "Intellectualization Method and Model of Complex Technical System's Failures Risk Estimation and Prediction," in *CEUR Workshop Proceedings*, vol. 3392, CMIS-2023, pp. 130–140, 2023. [Online]. Available: https://doi.org/10.32782/cmis/3392-11

[34] A. El-Awady and K. Ponnambalam, "Integration of simulation and Markov Chains to support Bayesian Networks for probabilistic failure analysis of complex systems," *Reliab. Eng. Syst. Saf.*, vol. 211, Jul. 2021. [Online]. Available: https://doi.org/10.1016/j.ress.2021.107511

[35] S. Młynarski, R. Pilch, M. Smolnik, J. Szybka, G. Wiązania, and J. Lewandowski, "Simulation-Based Forecasting of the Reliability of

Systems Consisting of Elements Described by a Number of Failure Probability Distributions," *J. KONBiN*, vol. 50, no. 2, 2020. [Online]. Available: https://doi.org/10.2478/jok-2020-0028

[36] V. Vychuzhanin, N. Rudnichenko, A. Vychuzhanin, and A. Rychlik, "Diagnosis intellectualization of complex technical systems," in *ICST-2023: Information Control Systems & Technologies*, CEUR-WS.org, vol. 3513, pp. 352–362, Odesa, 2023.

[37] X. Zhao, G. Lamperti, D. Ouyang, and X. Tong, "Minimal Diagnosis and Diagnosability of Discrete-Event Systems Modeled by Automata," *Complexity*, vol. 2020, Art. ID 4306261, 17 pages. [Online]. Available: https://doi.org/10.1155/2020/4306261

[38] A. G. Rodrigues and T. M. Williams, "System dynamics in software project management: towards the development of a formal integrated framework," *Eur. J. Inf. Syst.*, pp. 51–66, 2017. [Online]. Available: https://doi.org/10.1057/palgrave.ejis.3000256

[39] L. An et al., "Challenges, tasks, and opportunities in modeling agent-based complex systems," *Ecol. Model.*, vol. 457, Oct. 2021. [Online]. Available: https://doi.org/10.1016/j.ecolmodel.2021.109685

[40] L. Pokorádi, "Monte-Carlo Simulation of Reliability of System with Complex Interconnections," *Vehicles*, vol. 6, no. 4, pp. 1801–1811, 2024. [Online]. Available: https://doi.org/10.3390/vehicles6040087

[41] M. J. Daigle et al., "A Comprehensive Diagnosis Methodology for Complex Hybrid Systems: A Case Study on Spacecraft Power Distribution Systems," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 40, no. 5, 2010. [Online]. Available: https://doi.org/10.1109/TSMCA.2010.2052038

[42] V. Vychuzhanin, N. Rudnichenko, and I. Gritsuk, "Cognitive-Impulse Model for Assessing Complex Technical Systems Survivability," in *Proc. Int. Conf. ICST*, pp. 571–585, 2020.

[43] H. Yan, F. Wang, G. Yan, and D. He, "Hybrid approach integrating case-based reasoning and Bayesian network for operational adjustment in industrial flotation process," *J. Process Control*, vol. 103, pp. 34–47, Jul. 2021. [Online]. Available: https://doi.org/10.1016/j.jprocont.2021.05.003

[44] W. Li et al., "Application of a hybrid algorithm of LSTM and Transformer based on random search optimization for improving rainfall-runoff simulation," *Sci. Rep.*, vol. 14, Art. no. 11184, 2024. [Online]. Available: https://doi.org/10.1038/s41598-024-62127-7

[45] I. Poljak, "Marine Power Systems," *Marine Science and Engineering*, vol. 10, no. 2, p. 195, 2022. [Online]. Available: https://doi.org/10.3390/jmse10020195

[46] V. Vychuzhanin, A. Vychuzhanin, and N. Rudnichenko, "CBR Method for Decision-making Support in Operation Efficiency Ensuring of Complex Technical Systems," in *CMIS-2024*, vol. 3702, pp. 72–85, May 3, 2024.

[47] O. Vychuzhanin and V. Vychuzhanin, "Method of Assessing and Forecasting the Technical State of Complex Systems of Critical Application Based on Precedents," in *Rozvytky informatsiyno-keruyuchykh system ta tekhnolohiy: monohrafiya*, N. Aksak, D. Antonov et al., Eds. Lviv: Liha-Pres, 2024, pp. 71–97. [Online]. Available: https://doi.org/10.36059/978-966-397-422-4

[48] P. Pan et al., "Research progress on ship power systems integrated with new energy sources: A review," *Renewable and Sustainable Energy Reviews*, vol. 144, Jul. 2021. [Online]. Available: https://doi.org/10.1016/j.rser.2021.111048

[49] N. Neykov and S. Stefanova, "Using case-based reasoning in system diagnostics and maintenance," in *Intelligent Systems and Applications*, Springer, 2022, pp. 243–257. [Online]. Available: https://doi.org/10.1007/978-3-030-82029-4_17

[50] Z. Chen, K. Wang, C. Zhang, and Q. Wang, "Graph neural network-based fault diagnosis: A review," *arXiv preprint*, arXiv:2107.04542, 2021. [Online]. Available: https://arxiv.org/abs/2107.04542

[51] H. Lin, Q. Wu, L. Liu, and X. Song, “Case-based reasoning system for aeroengine fault diagnosis enhanced with attitudinal Choquet integral,” Applied Sciences, vol. 12, no. 3, p. 1130, 2022, doi:10.3390/app12031130.

[52] J. Ye, “Single-valued neutrosophic similarity measures based on cotangent function and their application in the fault diagnosis of steam turbine,” Journal of Intelligent & Fuzzy Systems, vol. 32, no. 6, pp. 4229–4236, 2017, doi:10.3233/JIFS-169424.

[53] J. Serrà and J. L. Arcos, “An empirical evaluation of similarity measures for time series classification,” Pattern Recognition Letters, vol. 35, pp. 45–52, 2014, doi:10.1016/j.patrec.2012.12.020.

[54] I. El Bitar, F. Z. Belouadha, and O. Roudies, “A logic and adaptive approach for efficient diagnosis systems using CBR,” arXiv preprint arXiv:1210.6840, 2012. [Online]. Available: https://arxiv.org/abs/1210.6840

[55] B. M. Mathisen, A. Aamodt, K. Bach, and H. Langseth, "Learning similarity measures from data," arXiv, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2001.05312

[56] K. Bach and P. J. Mork, "On the explanation of similarity for developing and deploying CBR systems," arXiv, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2106.04662

[57] D. Verma, K. Bach, and P. J. Mork, "Similarity measure development for case-based reasoning - A data-driven approach," arXiv, 2019. [Online]. Available: https://doi.org/10.48550/arXiv.1905.08581

[58] N. Lin, H. Liu, J. Fang, D. Zhou, and A. Yang, "An interpretability framework for Similar case matching," arXiv, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2304.01622

[59] S. Ren, F. Gui, and W. Wang, "An effective similarity determination model for case-based reasoning in support of low-carbon product design," *Advances in Mechanical Engineering*, vol. 12, no. 10, Oct. 2020. [Online]. Available: https://doi.org/10.1177/1687814020970313

[60] M. Chen, R. Qu, and W. Fang, "Case-Based Reasoning System for Fault Diagnosis of Aero-Engines," Expert Systems with Applications, vol. 202, p. 117350, 2022, doi:10.1016/j.eswa.2022.117350.

[61] Y. Zhang, X. Li, and J. Wang, "A Review of the Development and Future Challenges of Case-Based Reasoning," Applied Sciences, vol. 14, no. 16, p. 7130, 2024, doi:10.3390/app14167130.

[62] M. Zuber and R. Sirdey, "Efficient homomorphic evaluation of k-NN classifiers," *Proceedings on Privacy Enhancing Technologies*, vol. 2021, no. 2, pp. 111–129, 2021, doi:10.2478/popets-2021-0020.

[63] M. Zuber and R. Sirdey, "Efficient homomorphic evaluation of k-NN classifiers," *Proceedings on Privacy Enhancing Technologies*, vol. 2021, no. 2, pp. 111–129, 2021, doi:10.2478/popets-2021-0020.

[64] H. Wang and W. Wang, "Bearing fault diagnosis method based on similarity measure and ensemble learning," *Measurement Science and Technology*, vol. 32, no. 5, 2021, doi:10.1088/1361-6501/abda97.

[65] J. Choi, D. Kong, and H. Cho, "Weighted domain adaptation using the graph-structured dataset representation for machinery fault diagnosis under varying operating conditions," *Sensors*, vol. 24, no. 1, p. 188, 2024, doi:10.3390/s24010188.

[66] Z. Cheng and A. Yan, "A case weighted similarity deep measurement method based on a self-attention Siamese neural network," *Industrial Artificial Intelligence*, vol. 1, no. 1, 2023, doi:10.1007/s44244-022-00002-y.

[67] X. Diao et al., "An ontology-based fault generation and fault propagation analysis approach for safety-critical computer systems at the design stage," *AI EDAM*, vol. 36, no. 3, pp. 345–360, 2022, doi:10.1017/S0890060421000342.

[68] A. Aamodt and E. Plaza, "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches," *AI Communications*, vol. 7, no. 1, pp. 39–59, 1994. [Online]. Available: https://doi.org/10.3233/AIC-1994-7104

[69] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008. [Online]. Available: https://doi.org/10.1017/CBO9780511809071

[70] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed., Springer, 2009. [Online]. Available: https://doi.org/10.1007/978-0-387-84858-7

[71] C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006. [Online]. Available: https://doi.org/10.5555/1162264

[72] P. N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining, Addison-Wesley, 2005. [Online]. Available: https://doi.org/10.5555/1095618

[73] P. T. Boggs and R. H. Byrd, "Adaptive, limited-memory BFGS algorithms for unconstrained optimization," *SIAM J. Optim.*, vol. 29, no. 2, pp. 947–967, 2019, doi:10.1137/16M1065100.

[74] H. Tankaria, S. Sugimoto, and N. Yamashita, "A regularized limited memory BFGS method for large-scale unconstrained optimization and its efficient implementations," arXiv, 2021. [Online]. Available: https://arxiv.org/abs/2101.04413

[75] I. Poljak et al., "Condition-based maintenance of naval propulsion systems: A brief review," Renew. Sust. Energy Rev., 2022. [Online]. Available: https://doi.org/10.1016/j.rser.2016.09.095

[76] T. Ademujimi and V. Prabhu, "Fusion-learning of Bayesian network models for fault diagnostics," *Sensors*, vol. 21, no. 21, p. 7633, 2021. [Online]. Available: https://doi.org/10.3390/s21227633

[77] P. G. Morato et al., "Optimal inspection and maintenance planning for deteriorating structural components through dynamic Bayesian networks and Markov decision processes," Structural Safety, vol. 94, p. 102140, 2022. [Online]. Available: https://doi.org/10.1016/j.strusafe.2021.102140

[78] H. Nikpour and A. Aamodt, "Fault diagnosis under uncertain situations within a Bayesian knowledge-intensive CBR system," *Prog. Artif.*

Intell., vol. 10, no. 3, pp. 245–258, 2021. [Online]. Available: https://doi.org/10.1007/s13748-020-00227-x

[79] A. N. Abbas, G. Chasparis, and J. D. Kelleher, "Interpretable hidden Markov model-based deep reinforcement learning hierarchical framework for predictive maintenance of turbofan engines," in *Advances in Intelligent Systems and Computing*, vol. 1380, Springer, 2022, pp. 147–158. [Online]. Available: https://doi.org/10.1007/978-3-031-12670-3_12

[80] Z. Chen et al., "Graph neural network-based fault diagnosis: A review," arXiv, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2111.08185

[81] J. Wang, Z. Wang, V. Stetsyuk, X. Ma, F. Gu, and W. Li, "Exploiting Bayesian networks for fault isolation: A diagnostic case study of diesel fuel injection system," *ISA Transactions*, vol. 86, pp. 276–286, 2019, doi:10.1016/j.isatra.2018.10.044.

[82] H. Moon, J. Choi, and S. Cha, "A multi-state Markov model to infer the latent deterioration process from the maintenance effect on reliability engineering of ships," arXiv, 2021, doi:10.48550/arXiv.2111.14368.

[83] P. Louvros, F. Stefanidis, E. Boulougouris, A. Komianos, and D. Vassalos, "Machine learning and case-based reasoning for real-time onboard prediction of the survivability of ships," J. Mar. Sci. Eng., vol. 11, p. 890, 2023, doi:10.3390/jmse11050890.

[84] V. Başhan, M. Yucesan, M. Gul, and H. Demirel, "A fuzzy Bayesian network risk assessment model for analyzing the causes of slow-down processes in two-stroke ship main engines," *Ships Offshore Struct.*, vol. 19, no. 5, pp. 670–686, 2024, doi:10.1080/17445302.2024.2323889.

[85] A. Della Libera and A. Ploujnikov, "Bayesian deep learning for remaining useful life estimation via Stein variational gradient descent," arXiv, 2024, doi:10.48550/arXiv.2402.01098.

[86] Y. Garbatov and P. Georgiev, "Markovian maintenance planning of ship propulsion system accounting for CII and system degradation," *Energies*, vol. 17, p. 4123, 2024, doi:10.3390/en17164123.

[87] M. Anantharaman and S. Rajendran, "A step-by-step approach for evaluating the reliability of the main engine lube oil system for a ship's propulsion system," University of Tasmania, 2014, doi:10.25959/UTAS.22917440.

[88] E. Hostens, K. Eryilmaz, M. Vangilbergen, and T. Ooijevaa, "Bayesian networks for remaining useful life prediction," in Proc. European Conf.

of the PHM Society, vol. 8, no. 1, 2024, doi:10.36001/phme.2024.v8i1.4019.

[89] R. Wang, H. Chen, and C. Guan, "A Bayesian inference-based approach for performance prognostics towards uncertainty quantification and its applications on the marine diesel engine," *ISA Transactions*, vol. 112, pp. 123–135, 2021, doi:10.1016/j.isatra.2021.02.024.

[90] M. Chen, J. Xia, R. Huang, and W. Fang, "Case-Based Reasoning System for Fault Diagnosis Enhanced with Attitudinal Choquet Integral," *Applied Sciences*, vol. 12, no. 11, p. 5696, 2022, doi:10.3390/app12115696.

[91] A. Schultheis, "Exploring a Hybrid Case-Based Reasoning Approach for Time Series Adaptation in Predictive Maintenance," in ICCBR'24 Workshop Proc., vol. 3708, CEUR-WS, 2024.

[92] J. M. Schoenborn, R. O. Weber, D. W. Aha, J. Cassens, and K.-D. Althoff, "Explainable Case-Based Reasoning: A Survey," in AAAI-21 Workshop Proc., CEUR-WS, 2021.

[93] R. Kumar, A. Schultheis, L. Malburg, M. Hoffmann, and R. Bergmann, "Considering Inter-Case Dependencies During Similarity-Based Retrieval in Process-Oriented Case-Based Reasoning," in Proc. 35th FLAIRS Conf., 2022, doi:10.32473/flairs.v35i.130680.

[94] L. Malburg, A. Schultheis, and R. Bergmann, "Identifying Missing Sensor Values in IoT Time Series Data: A Weight-Based Extension of Similarity Measures for Smart Manufacturing," in *Proc. 32nd ICCBR, LNCS 14775*, pp. 16–30, Springer, 2024, doi:10.1007/978-3-031-63646-2_16.

[95] A. Gould, G. Paulino-Passos, S. Dadhania, M. Williams, and F. Toni, "Preference-Based Abstract Argumentation for Case-Based Reasoning (AA-CBR-P)," *in 21st Int. Conf. on Principles of Knowledge Representation and Reasoning*, pp. 394–404, 2024, doi:10.24963/kr.2024/37.

[96] W. Yan, J. Wang, S. Lu, M. Zhou, and X. Peng, "A review of real-time fault diagnosis methods for industrial smart manufacturing," Processes, vol. 11, p. 369, 2023, doi:10.3390/pr11020369.

[97] J. J. Montero-Jiménez, R. Vingerhoeds, and B. Grabot, "Enhancing predictive maintenance architecture processes by using ontology-enabled Case-Based Reasoning," in Proc. 2021 *IEEE Intl. Symp. on Systems Engineering (ISSE)*, pp. 1–8, 2021, doi:10.1109/ISSE53008.2021.9574295.

[98] T. Ademujimi and V. Prabhu, “Fusion-Learning of Bayesian Network Models for Fault Diagnostics,” *Sensors*, vol. 21, no. 22, p. 7633, 2021, doi:10.3390/s21227633.

[99] B. L. Tarcsay, Á. Bárkányi, S. Németh, T. Chován, L. Lovas, and A. Egedy, “Risk-Based Fault Detection Using Bayesian Networks Based on Failure Mode and Effect Analysis,” *Sensors*, vol. 24, no. 11, p. 3511, 2024, doi:10.3390/s24113511.

[100] G. Liao, H. Yin, M. Chen, and Z. Lin, “Remaining useful life prediction for multi-phase deteriorating process based on Wiener process,” *Reliability Eng. Syst. Saf.*, vol. 207, p. 107361, 2021, doi:10.1016/j.ress.2020.107361.

[101] M. Sohaib, S. Mushtaq, and J. Uddin, “Deep Learning for Data-Driven Predictive Maintenance,” in Vision, Sensing and Analytics: Integrative Approaches, *Intelligent Systems Reference Library*, pp. 71–95, Springer, 2021, doi:10.1007/978-3-030-75490-7_3.

[102] S. Sahoo, H. Wang, and F. Blaabjerg, “Uncertainty-Aware Artificial Intelligence for Gear Fault Diagnosis in Motor Drives,” arXiv, 2024, doi:10.48550/arXiv.2412.01272.

[103] J. Xu, Q. Wang, J. Zhou, H. Zhou, and J. Chen, “Improved Bayesian Network-Based Fault Diagnosis of Air Conditioner Systems,” Int. J. Metrol. Qual. Eng., vol. 14, p. 10, 2023, doi:10.1051/ijmqe/2023009.

[104] B. Qi, L. Zhang, J. Liang, and J. Tong, “Combinatorial techniques for fault diagnosis in nuclear power plants based on Bayesian Neural Network and simplified Bayesian Network-ANN,” Frontiers in Energy Research, vol. 10, p. 920194, 2022, doi:10.3389/fenrg.2022.920194.

[105] T. Zhou, E. Lopez Droguett, A. Mosleh, and F. Chan, “An uncertainty-informed framework for trustworthy fault diagnosis in safety-critical applications,” arXiv, 2021, doi:10.48550/arXiv.2111.0874.

[106] X. Xu, Y. Lin, and C. Ye, “Fault diagnosis of marine machinery via an intelligent data-driven framework,” Ocean Eng., vol. 289, p. 116302, Dec. 2023, doi:10.1016/j.oceaneng.2023.116302.

[107] A. M. Schweidtmann, D. Zhang, and M. von Stosch, “A review and perspective on hybrid modeling methodologies,” Digit. Chem. Eng., vol. 10, p. 100136, Mar. 2024, doi:10.1016/j.dche.2023.100136.

[108] M. Soleimani, F. Campean, and D. Neagu, “Integration of Hidden Markov modelling and Bayesian network for fault detection and prediction of complex engineered systems,” Reliab. Eng. Syst. Saf., vol. 215, p. 107808, Nov. 2021, doi:10.1016/j.ress.2021.107808.

[109] A. El-Awady and K. Ponnambalam, “Integration of simulation and Markov chains to support Bayesian networks for probabilistic failure analysis of complex systems,” Reliab. Eng. Syst. Saf., vol. 211, p. 107511, 2021, doi:10.1016/j.ress.2021.107511.

[110] D. Zhong, Z. Xia, Y. Zhu, and J. Duan, “Overview of predictive maintenance based on digital twin technology,” Heliyon, vol. 9, no. 4, e14534, Apr. 2023, doi:10.1016/j.heliyon.2023.e14534.

[111] A. S. Kalafatelis, N. Nomikos, A. Giannopoulos, and P. Trakadas, “Survey on predictive maintenance in the maritime industry using machine and federated learning,” TechRxiv, vol. 11, 2025, doi:10.36227/techrxiv.173473250.04784922/v1.

[112] E. Özaydın, R. Fışkın, Ö. Uğurlu, and J. Wang, “A hybrid model for marine accident analysis based on Bayesian network (BN) and association rule mining (ARM),” *Ocean Eng.*, vol. 247, p. 110705, Mar. 2022, doi:10.1016/j.oceaneng.2022.110705.

[113] M. Cheliotis, I. Lazakis, and A. Cheliotis, “Bayesian and machine learning-based fault detection and diagnostics for marine applications,” *Ships Offshore Struct.*, vol. 17, no. 12, pp. 2686–2698, 2022, doi:10.1080/17445302.2021.2012015.

[114] OREDA, Offshore Reliability Data Handbook, 6th ed., OREDA, 2015.

[115] A. Soliman, M. A. Marsha, and S. Rahaman, “Digital twin predictive maintenance systems,”J.Ocean Technol., vol. 19, no. 2, pp. 46–76, 2024.

[116] I. Jovanović, Ç. Karatuğ, M. Perčić et al., “Combined fault tree analysis and Bayesian network for reliability assessment of marine internal combustion engine,” J. Mar. Sci. Appl., 2025, doi:10.1007/s11804-025-00692-7.

[117] C. Velasco-Gallego and I. Lazakis, “Mar-RUL: A remaining useful life prediction approach for fault prognostics of marine machinery,” *Appl. Ocean Res.*, vol. 140, p. 103735, Nov. 2023, doi:10.1016/j.apor.2023.103735.

[118] A. A. Daya and I. Lazakis, “Developing an advanced reliability analysis framework for marine systems operations and maintenance,” *Ocean Eng.*, vol. 272, p. 113766, Mar. 2023, doi:10.1016/j.oceaneng.2023.113766.

[119] D. Neupane, M. R. Bouadjenek, R. Dazeley, and S. Aryal, “Data-driven machinery fault diagnosis: A comprehensive review,” Neurocomputing, vol. 627, p. 129588, Apr. 2025, doi:10.1016/j.neucom.2025.129588.

[120] Y. Lv, X. Yang, Y. Li, J. Liu, and S. Li, “Fault detection and diagnosis of marine diesel engines: A systematic review,” Ocean Eng., vol. 294, p. 116798, Feb. 2024, doi:10.1016/j.oceaneng.2024.116798.

[121] Y. Li, W. Zhang, L. Cui, and H. Gao, "System reliability modeling and analysis for a marine power equipment operating in a discrete-time dynamic environment," *Qual. Reliab. Eng. Int*., vol. 40, pp. 3422–3438, 2024, doi: 10.1002/qre.3577.

[122] A. Youssef, H. Noura, A. El Amrani, E. M. El Adel, and M. Ouladsine, "A survey on data-driven fault diagnostic techniques for marine diesel engines," arXiv, 2024. doi: 10.48550/arXiv.2404.10363.

[123] Q. Zhao, "Fault diagnosis method for wind power equipment based on hidden Markov model," Wireless Communications and Mobile Computing, vol. 2022, no. 4, pp. 1–9, 2022. doi: 10.1155/2022/6937616.

[124] C. Lu, L. Tao, J. Ma, Y. Cheng, and Y. Ding, Fault diagnosis and prognostics based on cognitive computing and geometric space transformation, Springer, 2024. doi: 10.1007/978-981-99-8917-1.

[125] Y. Cui, Y. Sun, J. Cui, F. Zhao, and M. Yuan, "Digital twin for marine diesel engine: Enhancing predictive maintenance and operational efficiency," in Proc. CSAA/IET Int. Conf. Aircraft Utility Systems (AUS 2024), Xi’an, China, 2025.

[126] Y. Zhang, W. Zhu, C. Jin, and Z. Liang, "Hybrid modeling and simulation for shipboard power system considering high-power pulse loads integration," J. Mar. Sci. Eng., vol. 10, p. 1507, 2022. doi: 10.3390/jmse1010150.

[127] V. Vychuzhanin, N. Rudnichenko, V. Boyko et al., "Devising a method for the estimation and prediction of technical condition of ship complex systems," *East.-Eur. J. Enterp. Technol*., vol. 6, no. 9, pp. 4–11, 2016. doi: 10.15587/1729-4061.2016.85605.

[128] Quality-One International, "Fault Tree Analysis (FTA)," [Online]. Available: https://quality-one.com/fta. [Accessed: May 29, 2025].

[129] S. Aburakhia, R. Myers, and A. Shami, "A hybrid method for condition monitoring and fault diagnosis of rolling bearings with low system delay," arXiv:2208.06051, 2022. doi: 10.1109/TIM.2022.3198477.

[130] M. Orhan and M. Celik, "A literature review and future research agenda on fault detection and diagnosis studies in marine machinery systems," Proc. Inst. Mech. Eng. M J. Eng. Marit. Environ., vol. 238, no. 1, 2024. doi: 10.1177/14750902221149291.

[131] V. T. Ngo, Y. D. Kim, and Y. Choi, "A recurrent graph transformer network for multi-fault localization in marine medium-voltage power systems," *IEEE Access*, vol. 12, pp. 10123–10135, 2024. doi: 10.48550/arXiv.2409.10792.

[132] F. Fera and C. Spandonidis, "A fault diagnosis approach utilizing artificial intelligence for maritime power systems within an integrated digital twin framework," *Appl. Sci*., vol. 14, p. 8107, 2024. doi: 10.3390/app14188107.

[133] A. Hasan, T. Asfihani, O. Osen, and R. T. Bye, "Leveraging digital twins for fault diagnosis in autonomous ships," *Ocean Eng*., vol. 292, p. 116546, 2024. doi: 10.1016/j.oceaneng.2023.116546.

[134] A. A. Daya and I. Lazakis, "Component criticality analysis for improved ship machinery reliability," *Machines*, vol. 11, p. 737, 2023. doi: 10.3390/machines11070737.

[135] L. C. Brito, G. A. Susto, J. N. Brito, and M. A. V. Duarte, "An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery," arXiv:2102.11848, 2021. doi: 10.48550/arXiv.2102.11848.

[136] M. Arias Chao, C. S. Kulkarni, K. Goebel, and O. Fink, "Fusing physics-based and deep learning models for prognostics," arXiv:2003.00732, 2020. doi: 10.48550/arXiv.2003.00732.

[137] W. Tang, D. Roman, R. Dickie, V. Robu, and D. Flynn, "Prognostics and health management for the optimization of marine hybrid energy systems," Energies, vol. 13, no. 18, p. 4676, 2020. doi: 10.3390/en13184676.

[138] V. V. Vychuzhanin and N. D. Rudnichenko, "Metod upravleniya riskami sudovykh slozhnykh tekhnicheskikh sistem," *Problemi Tekhniki*, no. 2, pp. 138–142, 2014.

[139] M. Pajak, M. Kluczyk, Ł. Muślewski, D. Lisjak, and D. Kolar, "Ship diesel engine fault diagnosis using data science and machine learning," Electronics, vol. 12, p. 3860, 2023. doi: 10.3390/electronics12183860.

[140] S. Rigas, P. Tzouveli, and S. Kollias, "An end-to-end deep learning framework for fault detection in marine machinery," Sensors, vol. 24, p. 5310, 2024. doi: 10.3390/s24165310.

[141] M. T. Tveten and M. Stakkeland, "Fault detection in propulsion motors in the presence of concept drift," arXiv preprint, 2024. doi: 10.48550/arXiv.2406.08030.

[142] G. Vizentin, G. Vukelic, L. Murawski, N. Recho, and J. Orovic, "Marine propulsion system failures -A review," J. Mar. Sci. Eng., vol. 8, no. 9, p. 662, 2020. doi: 10.3390/jmse8090662.

[143] Y. Zhao, S. Wang, and N. Chen, "Thermal fault diagnosis of marine diesel engine based on LSTM neural network algorithm," Vibroengineering Procedia, 2022. doi: 10.21595/vp.2022.22515.

[144] G. Zhu, L. Huang, J. Yin, W. Gai, and L. Wei, "Multiple faults diagnosis for ocean-going marine diesel engines based on different neural network algorithms," Sci. Prog., vol. 106, no. 4, 2023. doi: 10.1177/00368504231212765.

[145] C. Karatuğ and Y. Arslanoğlu, "Development of condition-based maintenance strategy for fault diagnosis for ship engine systems," Ocean Eng., p. 111515, 2022. doi: 10.1016/j.oceaneng.2022.111515.

[146] Y. Arslanoğlu et al., "Review of maintenance strategies for ship machinery systems," J. Mar. Eng. Technol., 2023. doi: 10.1080/20464177.2023.2180831.

[147] J. Wang, H. Cao, Z. Cui, Z. Ai, and K. Jiang, "Intelligent fault diagnosis of marine diesel engines based on efficient channel attention-improved convolutional neural networks," Processes, vol. 11, p. 3360, 2023. doi: 10.3390/pr11123360.

[148] F. Maione, P. Lino, G. Maione, and G. Giannino, "A machine learning framework for condition-based maintenance of marine diesel engines: A case study," Algorithms, vol. 17, p. 411, 2024. doi: 10.3390/a17090411.

[149] H. Moon and J. Choi, "Hierarchical spline for time series forecasting: An application to naval ship engine failure rate," Authorea, 2020. doi: 10.22541/au.159969715.57074848.

[150] M. Farid, "Data-driven method for real-time prediction and uncertainty quantification of fatigue failure under stochastic loading using artificial neural networks and Gaussian process regression," Int. J. Fatigue, vol. 155, 2022, p. 106415. doi: 10.1016/j.ijfatigue.2021.106415.

[151] N. Neykov and S. Stefanova, “Using Case-Based Reasoning in System Diagnostics and Maintenance,” in Proc. 7th Int. Congr. Inf. Commun. Technol., Springer, Singapore, 2023, vol. 464, pp. 299–306. https://doi.org/10.1007/978-981-19-2394-4_28

[152] S. Marandi, Y.-S. Hu, and M. Modarres, “Complex System Diagnostics Using a Knowledge Graph-Informed and Large Language Model-Enhanced Framework,” arXiv preprint, arXiv:2505.21291, 2025. https://doi.org/10.48550/arXiv.2505.21291

[153] M. Arias Chao, C. Kulkarni, K. Goebel, and O. Fink, "Fusing Physics-based and Deep Learning Models for Prognostics," arXiv preprint, arXiv:2003.00732, 2020. https://doi.org/10.48550/arXiv.2003.00732

[154] Y. Liu, Z. Liu, H. Zuo, H. Jiang, P. Li, and X. Li, "A DLSTM-Network-Based Approach for Mechanical Remaining Useful Life Prediction," Sensors, vol. 22, no. 15, p. 5680, 2022. https://doi.org/10.3390/s22155680

[155] Z. Shi and A. Chehade, "A dual-LSTM framework combining change point detection and remaining useful life prediction," Reliab. Eng. Syst. Saf., vol. 205, 107257, 2021. https://doi.org/10.1016/j.ress.2020.107257

[156] P. G. Morato, C. P. Andriotis, K. G. Papakonstantinou, and P. Rigo, "Inference and dynamic decision-making for deteriorating systems with probabilistic dependencies through Bayesian networks and deep reinforcement learning," Reliab. Eng. Syst. Saf., vol. 235, 109144, 2023. https://doi.org/10.1016/j.ress.2023.109144

[157] C. P. Andriotis, K. G. Papakonstantinou, and E. N. Chatzi, "Value of structural health information in partially observable stochastic environments," Struct. Saf., vol. 93, 102072, 2021. https://doi.org/10.1016/j.strusafe.2020.102072

[158] A. Kamariotis, E. N. Chatzi, and D. Straub, "A framework for quantifying the value of vibration-based structural health monitoring," Mech. Syst. Signal Process., vol. 184, 109708, 2023. https://doi.org/10.1016/j.ymssp.2022.109708

[159] Y. Raptodimos and I. Lazakis, "Application of NARX neural network for predicting marine engine performance parameters," *Ships Offshore Struct.*, vol. 15, no. 4, pp. 412–425, 2019. https://doi.org/10.1080/17445302.2019.1661619

[160] M. Cheliotis, I. Lazakis, and G. Theotokatos, "Machine learning and data-driven fault detection for ship systems operations," Ocean Eng., vol. 216, 107968, 2022. https://doi.org/10.1016/j.oceaneng.2020.107968

[161] G. Zhu, L. Huang, J. Yin, W. Gai, and L. Wei, "Multiple faults diagnosis for ocean-going marine diesel engines based on different neural network algorithms," Sci. Prog., vol. 106, no. 4, Article 368, 2023. https://doi.org/10.1177/00368504231212765

[162] S. Wang, J. Wang, and X. Ding, "An intelligent fault diagnosis scheme based on PCA-BP neural network for the marine diesel engine," in IOP Conf. Ser.: Mater. Sci. Eng., vol. 782, 032079, 2020. https://doi.org/10.1088/1757-899X/782/3/032079

[163] I. Chonlagarn, A. Mosleh, and M. Modarres, "Efficient dependency computation for dynamic hybrid Bayesian network in on line system health management applications," Reliab. Eng. Syst. Saf., 2014. https://doi.org/10.36001/phmconf.2014.v6i1.2422

[164] L. Portinale, D. Codetta Raiteri, and S. Montani, "Supporting reliability engineers in exploiting the power of dynamic Bayesian networks," *Int. J. Approx. Reason.*, vol. 51, no. 2, pp. 179–195, 2010. https://doi.org/10.1016/j.ijar.2009.05.009

[165] S. Montani, L. Portinale, and A. Bobbio, "Compiling dynamic fault trees into dynamic Bayesian nets for reliability analysis: the RADYBAN tool," in Proc. 2007 Int. Conf. Ubiquitous Intell. Comput. (UIC), 2007. https://ceur-ws.org/Vol-268/paper6.pdf

[166] Y. Zhang and O. Yagan, "Modeling and analysis of cascading failures in interdependent cyber physical systems," in Proc. 2018 IEEE Conf. Decis. Control (CDC), 2018. https://doi.org/10.1109/CDC.2018.8618710

[167] Y. Zhao, T. Sun, and Y. Liu, "Reliability analysis of a loading dependent system with cascading failures considering overloads," *Qual. Reliab. Eng. Int.,* vol. 40, no. 3, pp. 1182–1196, 2023. https://doi.org/10.1002/qre.3475

[168] A. El-Awady and K. Ponnambalam, "Integration of simulation and Markov Chains to support Bayesian Networks for probabilistic failure analysis of complex systems," Reliab. Eng. Syst. Saf., vol. 211, 107511, 2021. https://doi.org/10.1016/j.ress.2021.107511

[169] L. D. Valdez et al., "Cascading Failures in Complex Networks," J. Complex Netw., vol. 8, no. 2, 2020. https://doi.org/10.1093/comnet/cnaa013

[170] J. Hu, F. Khan, and L. Zhang, "Dynamic resilience assessment of marine LNG offloading system," Reliab. Eng. Syst. Saf., vol. 208, 107368, 2020. https://doi.org/10.1016/j.ress.2020.107368

[171] Y. Liu et al., "A novel methodology to model disruption propagation for resilient maritime transportation systems–a case study of the Arctic maritime transportation system," Reliab. Eng. Syst. Saf., vol. 241, p. 109620, 2024. https://doi.org/10.1016/j.ress.2023.109620

[172] F. Zocco, H.-C. Wang, and M. Van, "Digital twins for marine operations: A brief review on their implementation," arXiv preprint, arXiv:2301.09574, 2023. https://doi.org/10.48550/arXiv.2301.09574

[173] F. Stadtmann, H. G. Wassertheurer, and A. Rasheed, "Demonstration of a standalone, descriptive, and predictive digital twin of a floating

offshore wind turbine," *Ocean Renew. Energy*, 2023. https://doi.org/10.1115/OMAE2023-103112

[174] L. Polverino et al., "Machine learning for prognostics and health management of industrial mechanical systems and equipment: A systematic literature review," *Int. J. Eng. Bus. Manag*., vol. 15, 2023. https://doi.org/10.1177/18479790231186848

[175] D. Kaklis, I. Varlamis, G. Giannakopoulos, T. J. Varelas, and C. D. Spyropoulos, "Enabling digital twins in the maritime sector through the lens of AI and Industry 4.0," Int. J. Inf. Manag. Data Insights, vol. 3, no. 2, 2023. https://doi.org/10.1016/j.jjimei.2023.100178

[176] M. M. Nezhada, M. Neshatb, G. Sylaiosc, and D. A. Garcia, "Marine energy digitalization: Digital twin's approaches," Renew. Sustain. Energy Rev., vol. 191, p. 114065, 2024. https://doi.org/10.1016/j.rser.2023.114065

[177] A. S. Mavrakos, T. Tsaousis, N. Faggioni, A. Caviglia, and E. Manzo, "A digital twin approach for selection and deployment of decarbonization solutions for the maritime sector," in State-of-the-Art Digital Twin Applications for Shipping Sector Decarbonization, 2024. https://doi.org/10.4018/978-1-6684-9848-4.ch002

[178] Q. Liang, K. E. Knutsen, E. Vanem, V. Æsøy, and H. Zhang, "A review of maritime equipment prognostics health management from a classification society perspective," Ocean Eng., vol. 301, p. 117619, 2024. https://doi.org/10.1016/j.oceaneng.2024.117619

[179] P. Han, A. L. Ellefsen, G. Li, and F. T. Holmeset, "Fault Detection With LSTM based Variational Autoencoder for Maritime Components," IEEE Sens. J., vol. 21, no. 19, 2021. https://doi.org/10.1109/JSEN.2021.3105226

[180] B. Xiao et al., "Digital twin-driven prognostics and health management for industrial assets," Sci. Rep., vol. 14, 13443, 2024. https://doi.org/10.1038/s41598-024-63990-0

[181] Y. Cui, Y. Sun, J. Cui, F. Zhao, and M. Yuan, "Digital Twin for Marine Diesel Engine: Enhancing Predictive Maintenance and Operational Efficiency," CSAA/IET Int. Conf. Aircraft Utility Syst. (AUS 2024), Xi'an, China, 2025. https://www.researchgate.net/publication/388039483

[182] A. Birolini, Reliability Engineering, 8th ed., Springer, 2017. https://doi.org/10.1017/aer.2019.87

[183] P. A. Tobias and D. C. Trindade, Applied Reliability, 3rd ed., Chapman and Hall/CRC, 2011. https://doi.org/10.1201/b11787

[184] ISO 12110-2:2013, “Metallic materials - Fatigue testing - Variable amplitude fatigue testing - Part 2: Cycle counting and related data reduction methods.” https://cdn.standards.iteh.ai/samples/54713/496f47ac9d9e40b7af2c43f0bad42aff/ISO-12110-2-2013.pdf

[185] M. Rausand and A. Hoyland, System Reliability Theory: Models, Statistical Methods, and Applications, 2nd ed., Wiley, 2004. https://content.e-bookshelf.de/media/reading/L-571449-baa4bae2ce.pdf

[186] R. Y. Rubinstein and D. P. Kroese, Simulation and the Monte Carlo Method, 3rd ed., Wiley, 2017. https://doi.org/10.1002/9781118631980

[187] K. P. Burnham and D. R. Anderson, Model Selection and Inference: A Practical Information-Theoretic Approach, 2nd ed., Springer, 2002. https://doi.org/10.1007/b97636

[188] F. Mauro and A. A. Kana, “Digital twin for ship life cycle: A critical systematic review,” Ocean Eng., vol. 269, p. 113479, 2023. https://doi.org/10.1016/j.oceaneng.2022.113479

[189] Z. Lv, H. Lv, and M. Fridenfalk, “Digital Twins in the Marine Industry,” Electronics, vol. 12, no. 9, p. 2025, 2023. https://doi.org/10.3390/electronics12092025

[190] D. Minchev, K. Vasilev, and P. Petkov, “Digital twin test bench performance for marine diesel engine applications,” Pol. Marit. Res., vol. 30, no. 4, pp. 81–91, 2023. https://doi.org/10.2478/pomr-2023-0061

[191] H. Zhou, W. Yang, L. Sun, X. Jing, and G. Li, “Reliability optimization of marine diesel engine block machining using improved PSO method,” Sci. Rep., vol. 11, p. 21983, 2021. https://doi.org/10.1038/s41598-021-01567-x

[192] Q. Liang, K. E. Knutsen, E. Vanem, V. Æsøy, and H. Zhang, “A review of maritime equipment health prognostics from a classification society perspective,” Ocean Eng., vol. 301, p. 117619, 2024. https://doi.org/10.1016/j.oceaneng.2024.117619

[193] F. D’Agostino et al., “Development of a multi physics real time simulator for model based design of a DC shipboard microgrid,” Energies, vol. 13, no. 10, p. 3580, 2020. https://doi.org/10.3390/en13143580

[194] A.-T. W. K. Fahmi, K. Reza Kashyzadeh, and S. Ghorbani, “Advancements in Gas Turbine Fault Detection: A Machine Learning Approach Based on the Temporal Convolutional Network–Autoencoder Model,” Appl. Sci., vol. 14, p. 4551, 2024. https://doi.org/10.3390/app14114551

[195] Q. Chong, Z. Zhou, Z. Liu, and S. Jia, "Predictive anomaly detection for marine diesel engine based on echo state network and autoencoder," *Energy Rep.*, vol. 8, pp. 998–1003, 2022. https://doi.org/10.1016/j.egyr.2022.01.225

[196] Y. Liu et al., "Early Fault Diagnosis and Prediction of Marine Large Capacity Batteries Based on Real Data," J. Mar. Sci. Eng., vol. 12, no. 12, p. 2253, 2024. https://doi.org/10.3390/jmse12122253

[197] D. Della Libera et al., "Bayesian Deep Learning for Remaining Useful Life Estimation via Stein Variational Gradient Descent," arXiv preprint, arXiv:2402.01098, 2024. https://www.themoonlight.io/file?url=https%3A%2F%2Farxiv.org%2Fpdf%2F2402.01098

[198] H. Xiao et al., "Reliability Assessment of Ship Lubricating Oil Systems Through Improved Dynamic Bayesian Networks and Multi-Source Data Fusion," Appl. Sci., vol. 15, p. 5310, 2025. https://doi.org/10.3390/app15105310

[199] S.-L. Lin, "Application Combining VMD and ResNet101 in Intelligent Diagnosis of Motor Faults," Sensors, vol. 21, no. 18, p. 6065, 2021. https://doi.org/10.3390/s21186065

[200] J.-L. Xie, W.-F. Shi, T. Xue, and Y.-H. Liu, "High-Resistance Connection Fault Diagnosis in Ship Electric Propulsion System Using Res-CBDNN," J. Mar. Sci. Eng., vol. 12, no. 4, p. 583, 2024. https://doi.org/10.3390/jmse12040583

[201] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review," J. Big Data, vol. 7, no. 94, 2020. https://doi.org/10.1186/s40537-020-00369-8

[202] A. M. Burden, R. L. Burden, and J. D. Faires, Numerical Analysis, 10th ed., Cengage Learning, 2016. https://doi.org/10.13140/2.1.4830.2406

[203] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto, Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models, Wiley, 2004. https://doi.org/10.1002/047087095

[204] A. Saltelli and I. M. Sobol, "Sensitivity analysis for nonlinear mathematical models: Numerical experience," Mat. Modelirovanie, vol. 7, no. 11, pp. 111–123, 1995.

[205] J. Gan, H. Jin, Q. Shang, and C. Sheng, "A New Method of Intelligent Fault Diagnosis of Ship Dual-Fuel Engine Based on Instantaneous Rotational Speed," J. Mar. Sci. Eng., vol. 12, p. 2046, 2024. https://doi.org/10.3390/jmse12112046

[206] S. Liao et al., “Probabilistic Modeling of Maritime Accident Scenarios Leveraging Bayesian Network Techniques,” J. Mar. Sci. Eng., vol. 11, no. 8, p. 1513, 2023. https://doi.org/10.3390/jmse11081513

[207] G. Zhu, “Multiple Faults Diagnosis for Marine Diesel Engines Based on Neural Networks,” Sci. Prog., vol. 106, no. 4, Article 368, 2023. https://doi.org/10.1177/00368504231212765

[208] M. Orhan and M. Celik, “A Literature Review and Future Research Agenda on Fault Detection and Diagnosis Studies in Marine Machinery Systems,” Proc. Inst. Mech. Eng. Part M: J. Eng. Marit. Environ., vol. 238, no. 1, 2024. https://doi.org/10.1177/14750902221149291

[209] S. Lai et al., “Residual Attention Single Head Vision Transformer Network for Bearing Fault Diagnosis in Noisy Environments,” in Proc. 6th Int. Conf. Video, *Signal Image Process*. (VSIP ’24), pp. 136–150, 2024. https://doi.org/10.1145/3708568.3708591

[210] L. Popp and K. Müller, “Technical reliability of shipboard technologies for the application of alternative fuels,” Energy Sustain. Soc.,vol. 11,article 23,2021.https://doi.org/10.1186/s13705-021-00301-9

[211] I. Christodoulou-Varotsi, “Implementation of IMO instruments from a shipowning state perspective,” in The Elgar Companion to the Law and Practice of the International Maritime Organization, Cheltenham: Edward Elgar Publishing, 2024, pp. 379–393. https://doi.org/10.4337/9781802206883.00030

[212] A. H. Suhail, F. M. Guangul, and A. Nazeer, Advanced System Diagnostics Tools: Innovations and Applications, IntechOpen, 2024. https://doi.org/10.5772/intechopen.114378

[213] C. Wang, X. Chen, X. Qiang, H. Fan, and S. Li, “Recent advances in mechanism/data-driven fault diagnosis of complex engineering systems with uncertainties,” *AIMS Mathematics*, vol. 9, no. 11, pp. 29736–29772, 2024. https://doi.org/10.3934/math.20241441

[214] W. Yan, J. Wang, S. Lu, M. Zhou, and X. Peng, “A review of real-time fault diagnosis methods for industrial smart manufacturing,” Processes, vol. 11, no. 2, article 369, 2023. https://doi.org/10.3390/pr11020369

[215] P. Mercorelli, “Recent advances in intelligent algorithms for fault detection and diagnosis,” Sensors, vol. 24, no. 8, article 2656, 2024. https://doi.org/10.3390/s24082656

[216] F. Regattieri, A. Bortolini, and F. G. Galizia, “Data-driven fault detection and diagnosis: Challenges and opportunities in real-world

scenarios," Appl. Sci., vol. 12, no. 18, article 9212, 2022. https://doi.org/10.3390/app12189212

[217] A. Abdulwahid, "Intelligent fault diagnosis method for shipboard main power system," J. Southwest Jiaotong Univ., vol. 57, no. 1, pp. 475–487, 2022. https://doi.org/10.35741/issn.0258-2724.57.1.44

[218] J. S. Do, A. B. Kareem, and J. Hur, "LSTM-Autoencoder for vibration anomaly detection in vertical carousel storage and retrieval system (VCSRS)," Sensors, vol. 23, no. 2, article 1009, 2023. https://doi.org/10.3390/s23021009

[219] P. Theofanis and A. Raptis, "On efficiently partitioning a topic in Apache Kafka," in Proc. Int. Conf. Computer, Information and Telecommunication Systems (CITS), Piraeus, Greece, 2022, pp. 1–8. https://doi.org/10.1109/CITS55221.2022.9832981

[220] G. Wang et al., "Consistency and completeness: Rethinking distributed stream processing in Apache Kafka," in Proc. Int. Conf. Management of Data (SIGMOD '21), New York: ACM, 2021, pp. 2602–2613. https://doi.org/10.1145/3448016.345755

[221] Q. Luu, H. Liu, and T. Chen, "ChatGPT advance software testing intelligence? An experience report on metamorphic testing," arXiv preprint, arXiv:2310.19204, 2023. https://doi.org/10.48550/arXiv.2310.19204

[222] P. Mudgal and R. Wouhaybi, "An assessment of ChatGPT on log data," in AI-Generated Content, Singapore: Springer, 2023, pp. 148–169. https://doi.org/10.48550/arXiv.2309.07938

[223] H. Kirinuki and H. Tanno, "ChatGPT and human synergy in black-box testing: A comparative analysis," arXiv preprint, arXiv:2401.13924, 2024. https://doi.org/10.48550/arXiv.2401.13924

[224] T. Li, W. Zong, Y. Wang, and H. Tian, "Nuances are the Key: Unlocking ChatGPT to find failure-inducing tests with differential prompting," in Proc. 38th IEEE/ACM Int. Conf. Automated Software Engineering (ASE 2023), Luxembourg, 2023, pp. 14–26. https://doi.org/10.1109/ASE56229.2023.00089

[225] A. Bakhshandeh, A. Keramatfar, A. Norouzi, and M. Chekidehkhoun, "Using ChatGPT as a static application security testing tool," ISC Int. J. Inf. Security, 2023. https://doi.org/10.48550/arXiv.2308.14434

[226] A. Alzahem, S. Latif, W. Boulila, and A. Koubaa, "Unlocking the potential of medical imaging with ChatGPT's intelligent diagnostics," arXiv preprint, arXiv:2305.07429, 2023. https://doi.org/10.48550/arXiv.2305.07429

[227] T. Adiguzel, M. H. Kaya, and F. K. Cansu, “Revolutionizing education with AI: Exploring the transformative potential of ChatGPT,” Contemp. Educ. Technol., vol. 15, no. 3, article ep429, 2023. https://doi.org/10.30935/cedtech/13152

[228] B. Gordijn and H. Have, “ChatGPT: Evolution or revolution?” *Med. Health Care Philos.,* vol. 26, pp. 1–2, 2023. https://doi.org/10.1007/s11019-023-10136-0

[229] L. Ramalho, Fluent Python, 2nd ed., Sebastopol: O’Reilly Media, Inc., 2022. https://www.slideshare.net/slideshow/luciano-ramalho-fluent-python-clear-concise-and-effective-programmingoreilly-media-2022pdf/256662380

Наукове видання

ВИЧУЖАНІН Володимир
ВИЧУЖАНІН Алексій

ІНТЕЛЕКТУАЛЬНА ДІАГНОСТИКА СУДНОВИХ ЕНЕРГЕТИЧНИХ УСТАНОВОК: ІНТЕГРАЦІЯ МЕТОДУ ПРЕЦЕДЕНТІВ, ЙМОВІРНІСНИХ МОДЕЛЕЙ І CHATGPT. УНІВЕРСАЛЬНИЙ ПІДХІД ДО ДІАГНОСТИКИ НЕСПРАВНОСТЕЙ І ПРОГНОЗУВАННЯ В СКЛАДНИХ ТЕХНІЧНИХ СИСТЕМАХ

Монографія

Англійською мовою

Підписано до друку 01.07.2025. Формат 60×84/16.
Папір офсетний. Гарнітура Times New Roman. Цифровий друк.
Умовно-друк. арк. 23,95. Тираж 300. Замовлення № 0625-066.
Ціна договірна. Віддруковано з готового оригінал-макета.

Українсько-польське наукове видавництво «Liha-Pres»
79000, м. Львів, вул. Технічна, 1
87-100, м. Торунь, вул. Лубіцка, 44
Телефон: +38 (050) 658 08 23
E-mail: editor@liha-pres.eu
Свідоцтво суб'єкта видавничої справи
ДК № 6423 від 04.10.2018 р.