

**INTEGRATED APPROACH TO CREATING A CASE-BASED DATABASE FOR
DIAGNOSING FAILURES IN SHIP POWER PLANTS**

V.V. Vychuzhanin, A.V. Vychuzhanin

National Odesa Polytechnic University
1, Shevchenko Ave., Odesa, 65044, Ukraine
Email: v.v.vychuzhanin@op.edu.ua

Modern ship power plants (SPPs) represent highly complex technical systems operating under variable, high-load, and often harsh maritime conditions. These systems are characterized by a high degree of component interdependence, making the identification and prediction of failures a challenging task. Traditional deterministic diagnostic approaches often fail to accurately reflect the stochastic nature of technical failures and cascading effects. This article presents a comprehensive integrated approach to failure diagnostics based on a case-based reasoning (CBR) system. The core of the methodology is a structured case base that unites historical failure data, probabilistic models (including Bayesian networks and Markov processes), and discrete-event simulation tools. The structure of each case is standardized and includes failure descriptions, associated risk levels, diagnostic method references, operational context, and interconnection effects. The database currently includes over 5,000 unique failure scenarios derived from operational logs, maintenance records, and simulated data. To increase diagnostic accuracy and system adaptability, the authors implement dynamic updates and automated optimization of parameter weights using the L-BFGS-B algorithm. Simulation modeling is applied to reproduce rare and cascading failure conditions and to improve the generalization capacity of the system. Numerical experiments demonstrate that this integrated approach achieves a diagnostic accuracy of up to 95%, reduces false positives by 30%, and increases system flexibility in real-time operational contexts. The fusion of CBR with probabilistic and simulation models enables the system not only to diagnose known patterns but also to predict new and atypical failures, accounting for system degradation over time. The result is a knowledge-driven support tool for decision-making in ship power plant operations, significantly enhancing operational safety and maintenance planning. This work has implications for the development of intelligent diagnostic platforms in complex marine and industrial systems.

Keywords: fault diagnostics; failure prediction; stochastic processes; intelligent systems; operational risks; data analytics

Introduction. Modern ship power plants (SPPs) are complex technical systems (CTS) operating under harsh operational conditions [1]. The reliability and efficiency of SPP operation largely depend on the timely diagnosis and prediction of equipment failures [2]. However, traditional diagnostic methods based on deterministic models or expert assessments often lack flexibility and fail to account for the complexity of multicomponent systems [3, 4].

One of the promising approaches for diagnosing failures in CTS equipment is Case-Based Reasoning (CBR), which enables the use of accumulated experience to identify similar failures and predict their possible consequences [5, 6]. However, existing CBR systems face several challenges, such as optimizing the case database, accounting for probabilistic dependencies between parameters, and reducing computational load when processing large volumes of data.

An analysis of existing CBR optimization approaches shows that significant efforts have been directed toward improving case retrieval methods and database management. For instance, the study [7] proposes an enhanced case retrieval method in CBR systems for diagnosing aircraft engines. This approach considers attribute interactions, improving the accuracy of equipment condition diagnostics. However, its high computational complexity and limited applicability (aviation engines) make it less universal. The study [8] introduces a

dynamic case base maintenance method that optimizes database size without loss of accuracy. However, this method does not address case interaction effects or performance efficiency with large data volumes. Researchers in [9] presented a case base and feature dictionary update method based on the theory of belief functions. This method's advantage is the automatic adjustment of the knowledge base size, though its implementation complexity and potential increase in computational load remain open issues. Ayed et al. developed a method for removing duplicate cases in the case base, but it requires manual tuning [10].

Thus, despite extensive research in CBR optimization, the integration of dynamic updating and uncertainty processing remains a relevant challenge. For effective CBR application, a structured case base must be developed that includes not only historical failure data but also probabilistic estimates, information on cascading effects, operational parameters, and simulation-based failure modeling results for CTS.

The aim of this study is to develop and justify an integrated approach to diagnosing failures in marine power plants by creating a case-based reasoning (CBR) database that combines probabilistic models and simulation modeling. This approach enables consideration of the stochastic nature of failures, cascading effects, and dynamic changes in operational conditions, thereby improving the accuracy of diagnostics and failure prediction.

Main part. Case Structure. To standardize the representation of failures, a case structure has been developed, including: failure description (type, causes, consequences); failure risk assessment (Harrington’s desirability function [11], probabilistic failure assessment); component characteristics (unit condition, remaining resource, failure intensity); diagnostic methods (CBR, Bayesian networks [12, 13, 14], simulation modeling [15]); data sources (maintenance logs, OREDA databases [16], expert assessments).

As an example, the structure of a single case can be presented in Table 1.

Table 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 (0-1)	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	Markov models
Diagnostic methods	Applied algorithms and similarity measures	CBR. Bayesian networks, simulation modeling

The formalized structure of a failure precedent can be represented 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"]
}
}
}
}

```

The presented code reveals: structuredness – data connections are interrelated; flexibility – easily expandable with new parameters; processing readiness – applicable in diagnostic and analysis systems.

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

Case Database Formation.

Updating the knowledge base includes recalculating failure probabilities, filtering data, and dynamically adapting the diagnostic system. The main data categories are presented in Table 2.

Table 2.

Main Data Categories in the Knowledge Base

Data Category	Description	Data Source
Case 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, Bayesian networks, expert assessments
Simulation Scenarios	Simulated failure situations, including rare and cascading events	Discrete-event modeling, Monte Carlo method, 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 database includes over 5,000 records of ship power plant equipment failures collected from operational reports, technical inspections, and emergency situations. The database structure provides 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%), and anomalies in electrical system operation (12%).

The automatic knowledge base update process consists of several key stages: adjusting failure probabilities, including recalculations based on Bayesian analysis; filtering data to eliminate unreliable information; and adapting diagnostic models, including parameter weight adjustments in CBR.

Simulation Model for Predicting the Technical Condition of Ship Power Plant Equipment.

To improve the accuracy of predicting the technical condition of SPP equipment, a simulation model is used, which includes: modeling of component failures (Weibull and exponential distributions); consideration of cascading failure effects; analysis of operational modes and maintenance procedures.

The relationships between the model components are shown in Figure 1. The comprehensive SPP schematic (Fig. 1) [17] includes not only the main power units (e.g., the main engine, power plant) but also auxiliary systems that ensure their operation and the safety of the vessel.

The following relationships can be established between the components and subsystems in Figure 1: main engine corresponds to the main engine (8); cooling system is not explicitly specified in the structure but may be part of the remote-automated control subsystem (9) or included in the main engine support system; fuel system is not directly highlighted but is logically linked to the boiler plant (5) and the main engine (8); generator may be part of the ship's power plant (6); pumping system is likely connected to the ballast and bilge subsystem (10) and other SPP support systems; power supply corresponds to the ship's power plant (6); automation system may relate to the remote-automated control subsystem (9).

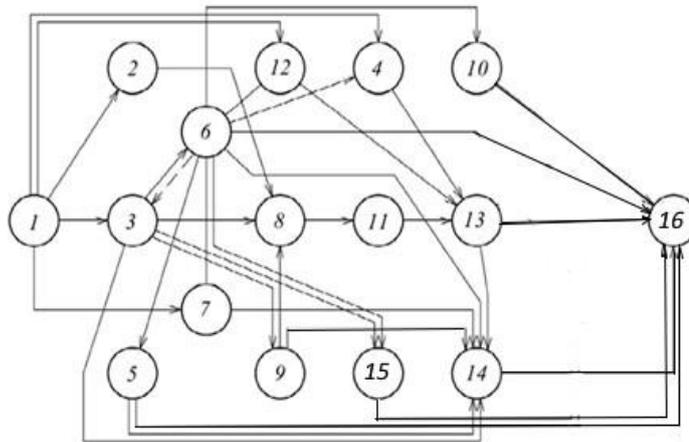


Fig.1. Structure of the SPP: input component 1; manual control of the main engine 2; compressed air subsystem 3; propulsion and steering control subsystem (PSC) 4; boiler plant 5; ship power plant 6; fire protection system 7; main engine 8; remote-automated control subsystem 9; ballast and bilge subsystem 10; power transmission from the main engine to the propeller 11; emergency PSC drive 12; PSC 13; measuring instruments subsystem 14; sanitary water treatment subsystem 15; output component 16.

The key parameters of the simulation scenarios for SPP equipment operation are presented in Table 3. Table 3 contains the parameters of failure simulation scenarios, allowing for an analysis of system behavior under various operating conditions.

Table 3.

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, SEU failure

The analysis of accumulated data has made it possible to establish the dependence of the probability of SEU equipment failures on operating time, as shown in Figure 2.

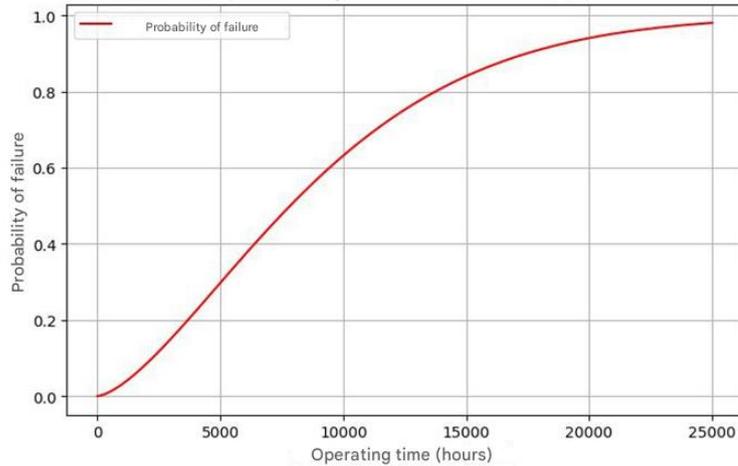


Fig. 2. Probability of Failure with Increasing Operating Time of SPP Equipment

The graph illustrates the increase in failure probability with extended equipment operation. In the initial phase (up to 5000 hours), the probability of failure is low, corresponding to the period of normal operation. After 15,000 hours, the probability of failure rises more sharply, indicating the need for more thorough maintenance. By 25,000 hours, the probability approaches 1, meaning the equipment is almost certain to fail. The curve demonstrates the characteristic growth in failure probability with increased operating time, confirming the necessity of using probabilistic models and simulation modeling in developing the case base. The obtained data confirm that accounting for the stochastic nature of failures and integrating the CBR method with probabilistic models not only enables failure prediction but also enhances diagnostic accuracy at various stages of the equipment life cycle.

Integration of Diagnostic Methods

The simulation model incorporates Bayesian networks for analyzing failure interdependencies, Markov processes [18] for modeling state transitions of the equipment, and cognitive modeling that accounts for expert knowledge and operational conditions. To improve diagnostic accuracy, CBR is used to identify similar failures, Bayesian networks to assess probabilistic dependencies, Markov processes to predict degradation, and simulation modeling to generate rare failures.

Table 4 presents the distribution of diagnostic methods and their functional purposes.

Table 4.

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 3 presents a comparison of equipment failure probabilities when using traditional diagnostic methods versus employing a case-based database integrated with the CBR method. It is evident that utilizing the case-based database reduces the probability of diagnostic errors, as accumulated failure experience is used to refine predictions. This demonstrates that the proposed approach not only enhances diagnostic accuracy but also improves the system's adaptation to changing operating conditions.

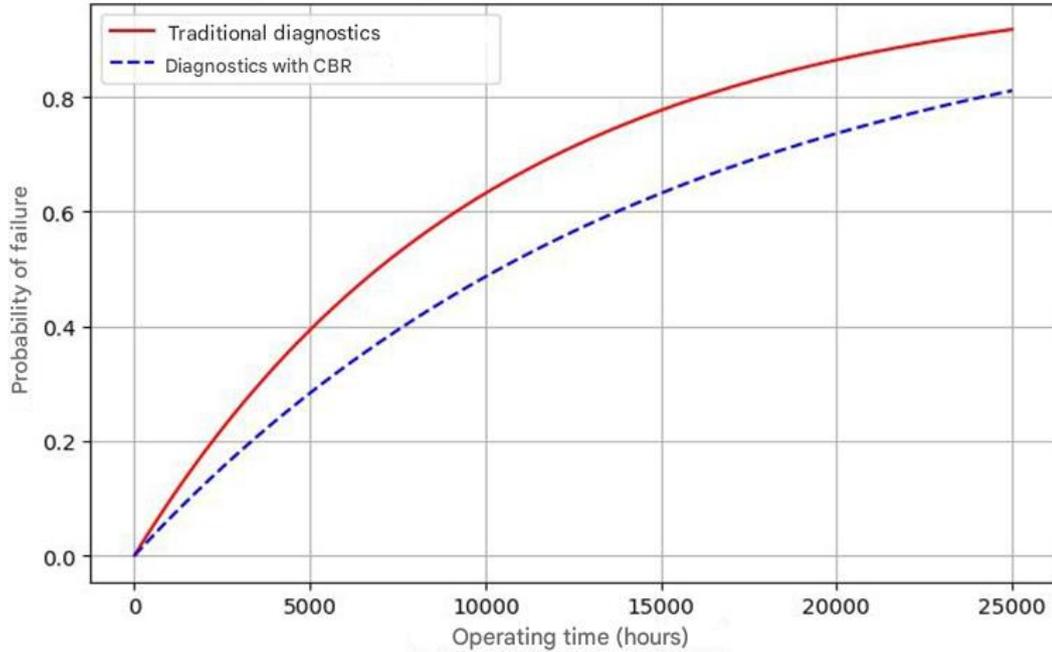


Fig. 3. Probability of Equipment Failures in SPP with and without CBR

The graphs illustrate the effect of applying CBR, which manifests in a reduction in diagnostic errors and an increase in prediction accuracy. The connection to the case-based database reflects the stochastic nature of failures: the graphs show how the probability of equipment failure increases over time. This confirms the importance of considering stochastic factors in diagnostics, which is achieved through the integration of probabilistic models in the case-based database.

The advantage of CBR over traditional diagnostic and failure prediction methods for SEU equipment is evident in the differences in failure forecasts: traditional approaches do not utilize accumulated experience, leading to a higher probability of diagnostic errors. In contrast, CBR analyzes historical data and adapts diagnostic decisions, reducing the likelihood of errors. This confirms that the case-based database refines predictions by incorporating previously recorded failures.

Simulation modeling enhances forecasting accuracy by accounting for not only frequent but also rare, cascading failures. Utilizing accumulated data improves the prediction of potential malfunctions, reducing the number of false alarms and increasing diagnostic accuracy.

The integration of these methods ensures not only precise fault diagnostics but also system adaptability to new operating conditions, increasing prediction reliability and minimizing the likelihood of missed failures.

Numerical experiments have shown that applying the case-based database in combination with CBR adaptation increases diagnostic accuracy from 85% (when using only traditional probabilistic methods) to 92–95%. The integration of Bayesian networks and simulation modeling reduced the number of false-positive diagnostic decisions by 30%, while

using optimization algorithms for parameter weighting reduced the mean absolute error in failure prediction by 12% (Table 5).

Table 5.
Impact of the Case-Based Database 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%	18
Integrated Approach	95	8%	5%	1,2

Optimization of the Case Base

To enhance diagnostic accuracy and system adaptability, the case base is regularly updated and optimized. One of the key mechanisms is the adjustment of 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:

$$\omega^* = \operatorname{argmin}_{\omega} \sum_i (S_{true}(i) - S_{pred}(i, \omega))^2$$

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

$S_{pred}(i, \omega)$ is the system-predicted similarity value, dependent on parameter weights ω , ω represents the diagnostic parameter weights, defining the contribution of various factors to failure similarity assessment

The L-BFGS-B (Limited-memory Broyden-Fletcher-Goldfarb-Shanno with Box constraints) method is a modification of the quasi-Newton optimization method BFGS that: allows minimization of a function dependent on a large number of parameters; uses an approximate inverse Hessian matrix representation to save memory; supports constraints on parameters, which is essential when optimizing weights (e.g., ensuring weights remain positive and sum to 1).

The main steps in updating the case base include: dynamic adjustment of parameter weights based on new operational data; filtering out outdated and irrelevant data to maintain database relevance; adapting the case base structure considering the results of simulation modeling.

Applying this approach ensures more accurate matching of failure cases and improves the diagnostic system's adaptability to changing operating conditions.

Conclusions. The presented approach to case base development justifies the necessity of structured failure data storage and the creation of a unified knowledge base that enables efficient failure diagnostics and prediction. Within the study, a precedent structure has been developed, incorporating key failure parameters, their probabilistic assessments, and links to operational factors. A knowledge base concept has been formulated, integrating the case base, probabilistic models, and failure simulation scenarios. Methods for automatic data updates have been defined, including CBR adaptation, failure probability recalculations, and data verification, ensuring system relevance. The integration of probabilistic and simulation

methods allows consideration of cascading failure effects and the prediction of rare malfunctions, significantly improving diagnostic accuracy.

The developed case base integrates the CBR method with probabilistic models and simulation modeling, enhancing the accuracy of ship power plant failure diagnostics. The analysis of the proposed approach has shown that using the case base increased diagnostic accuracy to 95% due to the accumulation and structuring of failure experience, while probabilistic methods ensure the consideration of the stochastic nature of failures, which is particularly critical for complex technical systems. Simulation modeling enabled the consideration of rare and cascading failures, expanding prediction capabilities. The developed case base covers the main types of failures in ship power plants, including fuel system malfunctions, gas turbine engine overheating, and electrical system anomalies. The process of accumulating and analyzing cases contributes to identifying failure patterns and their early prediction, reducing the number of false positives by 30%.

Thus, the developed case base serves as a key element of an intelligent ship power plant failure diagnostics system. It facilitates diagnostic data accumulation, decision support, enhanced failure prediction accuracy, as well as consideration of the stochastic nature of failures and cascading effects. The integrated approach, combining CBR, probabilistic models, and simulation modeling, significantly improves diagnostic efficiency and the reliability of ship power plants.

References

1. Vychuzhanin V., Vychuzhanin A. Stochastic Models and Methods for Diagnostics, Assessment, and Prediction of the Technical Condition of Complex Critical Systems: monograph. Lviv-Torun: Liha-Pres, 2025. 176 c. DOI: 10.36059/978-966-397-457-6.
2. Zhang P., GaoCao Z., Dong L. Marine Systems and Equipment Prognostics and Health Management. Systematic Review from Health Condition Monitoring to Maintenance Strategy. *Machines*. 2022. № 10. C. 72. DOI: 10.3390/machines10020072.
3. Vychuzhanin V., Rudnichenko N. Assessment of risks structurally and functionally complex technical systems. *Eastern-European Journal of Enterprise Technologies*. 2014. № 1(2). C. 18–22. DOI: 10.15587/1729-4061.2014.19846.
4. Vychuzhanin V., Rudnichenko N. Complex Technical System Condition Diagnostics and Prediction Computerization. *CMIS-2020 Computer Modeling and Intelligent Systems*. Ceur-ws.org. 2020. № 2608. C. 1–15. DOI: 10.32782/cmis/2608-4.
5. Glukhikh D., Shchinnikov I., Glukhikh I. Using hybrid-CBR for intelligence monitoring and decision-making systems on SMART grid. *Intelligent Decision Technologies*. 2022. № 16. C. 1–8. DOI: 10.3233/IDT-210239.
6. Aamodt A., Plaza E. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communications*. IOS Press. 1994. T. 7, № 1. C. 39–59. DOI: 10.3233/AIC-1994-7104.
7. Chen M., Xia J., Huang R., Fang W. Case-Based Reasoning System for Aeroengine Fault Diagnosis Enhanced with Attitudinal Choquet Integral. *Applied Sciences*. 2022. № 12. C. 5696. DOI: 10.3390/app12115696.
8. Salamó M., Golobardes E. Dynamic Case Base Maintenance for a Case-Based Reasoning System. *Lecture Notes in Computer Science*. 2004. DOI: 10.1007/978-3-540-30498-2_10.
9. Ben Ayed S., Elouedi Z., Lefevre E. An evidential integrated method for maintaining case base and vocabulary containers within CBR systems. *Information Sciences*. 2020. T. 529. C. 214–229. DOI: 10.1016/j.ins.2019.11.009.

10. Ayed S.B., Elouedi Z., Lefevre E. CIMMEP: constrained integrated method for CBR maintenance based on evidential policies. *Applied Intelligence*. 2022. DOI: 10.1007/s10489-020-02159-4.
11. Harrington E.C. The desirability Function. *Industrial Quality Control*. 1965. № 21(10). C. 494–498.
12. Fam M.L., He X., Konovessis D. Dynamic Bayesian Belief Network for long-term monitoring and system barrier failure analysis: Decommissioned wells. *MethodsX*. 2022. № 9. C. 1–10. DOI: 10.1016/j.mex.2021.101600.
13. Jensen F.V. Bayesian networks basics. Tech. Rep. Denmark: Aalborg Univ., 1996. 12 c.
14. Wang Chen R., Guan C. A Bayesian inference-based approach for performance prognostics towards uncertainty quantification and its applications on the marine diesel engine. *ISA Transactions*. 2021. № 118. C. 159–173.
15. Vychuzhanin V., Rudnichenko N. Cognitive Model of the Internal Combustion Engine. *SAE Technical Paper*. 2018. № 2018-01-1738. 10 c. DOI: 10.4271/2018-01-1738.
16. OREDA-Offshore Reliability Data Handbook. 6th Edition. OREDA, 2015. 831 c.
17. Pan P., Sun Y., Yuan C., Yan X., Tang X. Research progress on ship power systems integrated with new energy sources: A review. *Renewable and Sustainable Energy Reviews*. 2021. № 144. DOI: 10.1016/j.rser.2021.111048.
18. Lisnianski A., Elmakias D., Laredo D., BennHaim H. A multistate Markov model for a short-term reliability analysis of a power generating unit. *Reliability Engineering & System Safety*. 2012. № 98. C. 1–6. DOI: 10.1016/j.ress.2011.10.008.

ІНТЕГРОВАНІЙ ПІДХІД ДО СТВОРЕННЯ БАЗИ ПРЕЦЕДЕНТІВ ДЛЯ ДІАГНОСТИКИ ВІДМОВ СУДНОВИХ ЕНЕРГЕТИЧНИХ УСТАНОВОК

В.В. Вичужанін, А.В. Вичужанін

Національний університет «Одеська політехніка»

1, Шевченка пр., м.Одеса, 65044, Україна

Email: v.v.vychuzhanin@op.edu.ua

Сучасні суднові енергетичні установки (СЕУ) є надзвичайно складними технічними системами, що функціонують в умовах високих навантажень, змінних середовищних параметрів та значної взаємозалежності між підсистемами. Ці фактори значно ускладнюють процеси своєчасної діагностики та прогнозування відмов. Традиційні детерміновані методи виявляються недостатньо ефективними, оскільки не враховують стохастичний характер збоїв і каскадну природу наслідків. У статті запропоновано інтегрований підхід до створення системи діагностики на основі методу прецедентів (CBR). У його основі лежить уніфікована база прецедентів, що включає історичні дані, ймовірнісні моделі (баєсівські мережі, марковські процеси) та результати імітаційного моделювання. Структура кожного прецеденту передбачає опис типу відмови, оцінку ризиків, умови експлуатації, джерела даних, методи діагностики та взаємозв'язки між компонентами. Сформована база містить понад 5000 випадків відмов, отриманих із експлуатаційної документації, журналів обслуговування та результатів моделювання. Для підвищення точності діагностики та адаптивності системи застосовано механізм автоматичного оновлення, оптимізації вагових коефіцієнтів (метод L-BFGS-B), а також постійне доповнення бази новими сценаріями збоїв. Імітаційне моделювання дозволяє враховувати як часті, так і рідкісні або каскадні відмови, формуючи повну картину можливих технічних ризиків. Результати чисельних експериментів підтверджують ефективність підходу: точність діагностики досягає 95%, зменшується частка хибнопозитивних результатів на 30%, а сама система виявляє гнучкість до змін умов експлуатації. Інтеграція CBR з ймовірнісними та імітаційними методами дозволяє прогнозувати розвиток відмов на різних етапах життєвого циклу обладнання, знижуючи ризики експлуатації та оптимізуючи планування обслуговування. Запропонована база знань є ефективним інструментом підтримки прийняття рішень у сфері морської енергетики й може бути адаптована для інших технічно складних систем.

Ключові слова: діагностика несправностей, прогнозування відмов, стохастичні процеси, інтелектуальні системи, експлуатаційні ризики, аналітика даних.