DOI: https://doi.org/10.33216/1998-7927-2025-291-5-5-17

UDC 004.03

INTEGRATED APPROACH TO DIAGNOSING COMPLEX TECHNICAL SYSTEMS: EXPERIMENTAL VALIDATION AND MULTIDIMENSIONAL EFFICIENCY ASSESSMENT

Vychuzhanin V.V., Vychuzhanin O.V.

ІНТЕГРОВАНИЙ ПІДХІД ДО ДІАГНОСТИКИ СКЛАДНИХ ТЕХНІЧНИХ СИСТЕМ: ЕКСПЕРИМЕНТАЛЬНА ВАЛІДАЦІЯ ТА БАГАТОВИМІРНА ОЦІНКА ЕФЕКТИВНОСТІ

Вичужанін В.В., Вичужанін О.В.

This paper presents a comprehensive experimental validation of an integrated approach to the diagnosis of the technical condition (TC) of complex technical systems (CTS), using ship power plants (SPPs) as an example. The proposed methodology combines precedent-based logic (Case Based Reasoning – CBR), probabilistic forecasting using Bayesian networks and Markov chains, and simulation modeling of degradation scenarios and cascading failures. Testing was conducted under three scenarios: normal operating mode, high-load mode, and a scenario with limited data availability, which enabled a thorough assessment of the algorithms' adaptability and resilience to changing operational factors. Classical binary classification metrics (Accuracy, Precision, Recall, and F1 score) were used for quantitative evaluation of diagnostic quality, along with newly introduced extended indicators: weighted accuracy (WAcc), F1 score accounting for the criticality of component failures (F1W), recall weighted by failure risk (RecallR), cost-adjusted precision for false alarms (PrecisionC), and the Diagnostic Stability Index (DSI). The results of the multi-scenario experiment showed a consistent improvement in all major indicators: Accuracy increased from 78.5% to 85.3%, Precision from 75.2% to 83.1%, Recall from 80.1% to 87.6%, F1 score from 77.5% to 85.3%, RecallR reached 91.0%, and DSI was 0.983. Five-fold cross-validation yielded a standard deviation of F1 score at 2.2%, confirming the reproducibility and reliability of the proposed method for experimental testing of the integrated diagnostic approach for CTS. The implementation of a cyclic procedure "simulation, probabilities, CBR adaptation" significantly reduced the number of false alarms and missed critical failures in SPP equipment. The practical significance of the approach lies in its potential integration into SCADA/PMS systems of marine CTS and ground power stations, facilitating a shift to intelligent predictive maintenance, thereby reducing unplanned downtime, lowering costs, and enhancing equipment reliability. Future research prospects include increasing the adaptability of the approach, expanding the precedent base, and developing tools for automated processing of heterogeneous data.

Keywords: *predictive diagnostics, Bayesian networks, CBR adaptation, failure simulation modeling, risk-oriented metrics, diagnostic stability, intelligent maintenance*

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 [1, 2]. 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 (Bayesian networks and Markov chains), 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 [3, 4].

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 [5]. Researchers Nikpour & Aamodt [6] 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. Yang et al. [7] integrated CBR and Bayesian Networks for diagnosis and prognosis of the TC of complex systems, with validation based on real sensor data. However, their integration of CBR and Bayesian Networks 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. [8] developed a combined HMM (Hidden Markov Model 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 Bayesian Network 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 [9] proposed Simulation Supported Bayesian Networks (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 TS 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 review, Zhong et al. [10] 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. [11]. 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. [12] used a Bayesian Network 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. [13] 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.

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 (Bayesian networks and Markov chains) – for estimating current and future probabilities of component failures; Remaining Useful Life (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 (Bayesian networks, Markov processes) 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 [14]); simulated degradation scenarios of components, mimicking different operational modes; Bayesian networks accounting for probabilistic interrelations between component failures; Markov processes 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 – the proportion 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: Mean Absolute Error (MAE) – average absolute error in RUL prediction; Root Mean Square Error (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.

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 –Table 1), 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

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). Bayesian networks 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 2.

The data presented in Table 2 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.

Table 2

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)

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; 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.

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

$$Precision = \frac{TP}{TP + FP};$$

$$Recall = \frac{TP}{TP + FN};$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

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 Bayesian networks and Markov models, enabling the consideration of cascading failure risks and the remaining useful life of components.

Table 3

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 3 confirms that adapting the CBR method using probabilistic models (Bayesian networks 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 illustrates how diagnostic metrics improve with the addition of probabilistic methods and simulation modeling.

Figure 1 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. The addition of probabilistic models (Bayesian networks, Markov processes) 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 2) further confirms that the proposed integrated approach to diagnosing SPP significantly improves accuracy compared to standalone methods.

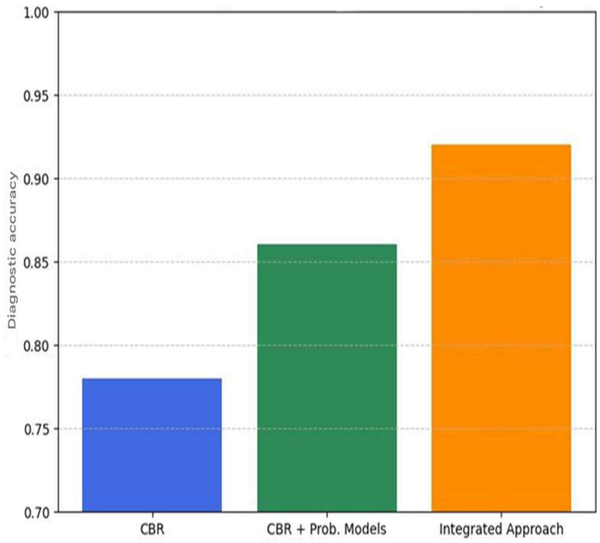

Fig. 1. Comparison of fault diagnosis accuracy metrics across different scenarios

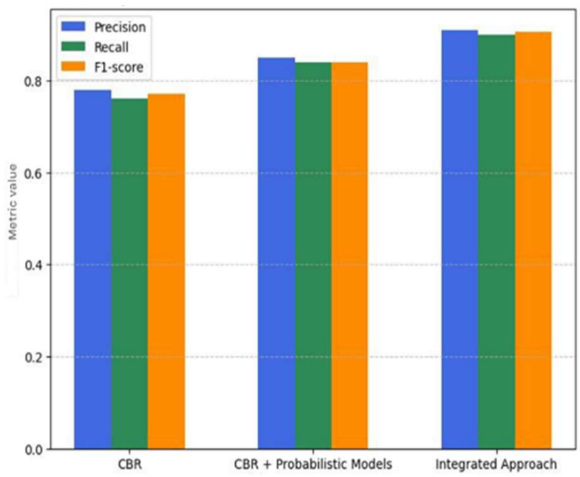

Fig. 2. Diagnostic Accuracy Metrics Chart

Based on Figure 2, 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. 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 and 2) illustrate how well the model identifies both faults and healthy states.

Figure 3 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.

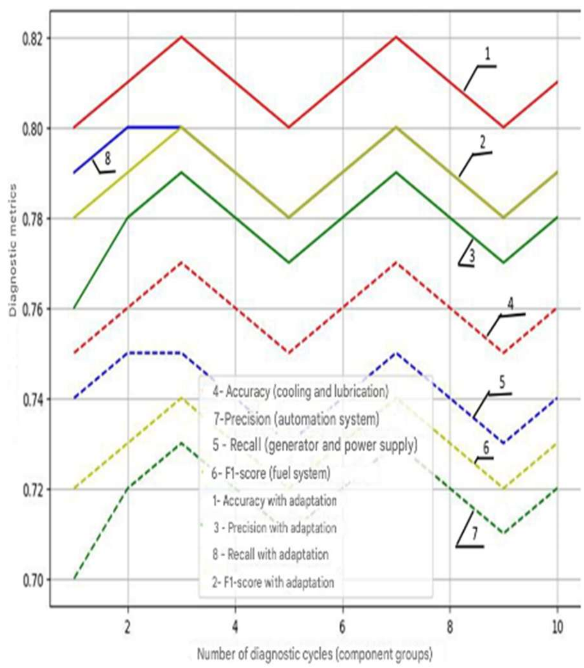

Fig. 3. Diagnostic Accuracy Dynamics with CBR Adaptation

Figure 3 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 4

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 4 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 ($\sigma = 3.0\%$). 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 4 presents the cross-validation results for various fault diagnosis methods.

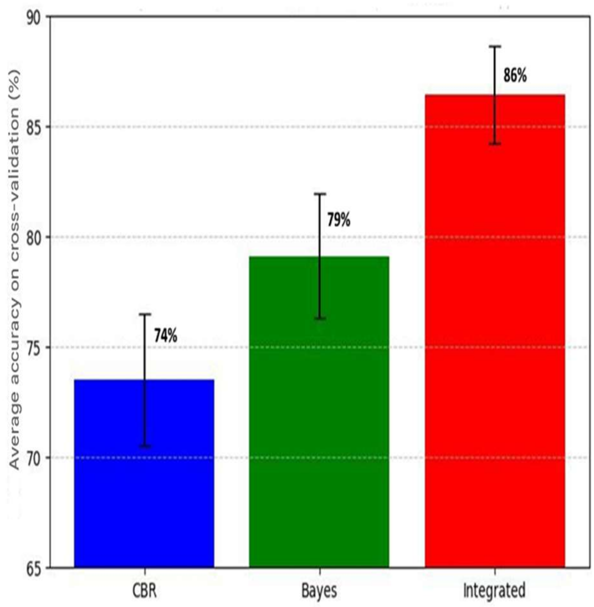

Fig. 4. Cross-validation results for different diagnostic methods

Comparison of methods based on Figure 4: 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:

$$WA_{cc} = \frac{\omega_{TP} \cdot TP + \omega_{TN} \cdot TN}{\omega_{TP} \cdot TP + \omega_{TN} \cdot TN + \omega_{FP} \cdot FP + \omega_{FN} \cdot FN},$$

where $\omega_{TP}, \omega_{TN}, \omega_{FP}, \omega_{FN}$ are weights reflecting the relative importance of each classification outcome type.

For example, $\omega_{FN} > \omega_{FP}$, 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):

$$F1_W = 2 \cdot \frac{P \cdot R}{P + R} \cdot$$

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):

$$Recall_R = \frac{\sum_i r_i \cdot TP_i}{\sum_i r_i \cdot (TP_i + FN_i)},$$

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:

$$Precision_C = \frac{TP}{TP + c_{FP} \cdot FP},$$

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:

$$DSI = 1 - \frac{\sigma_{F_1}}{\overline{F_1}}$$

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 5 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%).

Table 5

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

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 5 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 5 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.

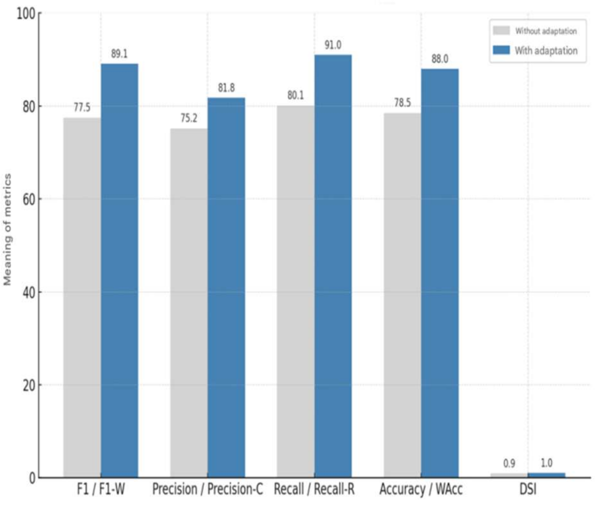

Fig. 5. Comparison of Modified Diagnostic Metrics in the Integrated System With and Without Adaptation

This study presents extensive experimental testing of an integrated approach to diagnosing complex technical systems, combining CBR, probabilistic methods (Bayesian networks and Markov chains), 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 (σ 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 [15] is limited to a review of digital twins without CBR or classification metrics; Jovanović [16] combines FTA and BN without simulations or case-based mechanisms; Velasco Gallego et al. [17] assess only RMSE/MAE without considering recall and precision; Schultheis [18] applies a hybrid CBR without probabilistic or simulation components; Daya & Lazakis [19] use DFTA, FMECA, and BBN without multi-scenario testing; Neupane et al. [20] review ML approaches without a hybrid implementation; Lv et al. [21] study FDD models without comprehensive integration or F1 evaluation; and Yan Li et al. [22] 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.

Conclusions

The effectiveness of the experimentally validated integrated approach to diagnosing complex technical systems -combining CBR, probabilistic methods (Bayesian networks and Markov chains), 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.

References

1. Vychuzhanin V., Vychuzhanin A. Stochastic Models and Methods for Diagnostics, Assessment, and Prediction of the Technical Condition of Complex Critical Systems / V. Vychuzhanin, A. Vychuzhanin. Kyiv : Liha Pres, 2025. 360 p. https://doi.org/10.36059/978-966-397-457-6
2. Vychuzhanin V. V., Rudnichenko N. D. Metody informatsionnykh tekhnologiy v diagnostike sostoyaniya slozhnykh tekhnicheskikh sistem: monografiya / V. V. Vychuzhanin, N. D. Rudnichenko. Odessa : Ekologiya, 2019. 178 p.
3. Yan H., Wang F., Yan G., He D. Hybrid approach integrating case-based reasoning and Bayesian network for operational adjustment in industrial flotation process / H. Yan, F. Wang, G. Yan, D. He // Journal of Process Control. 2021. Vol. 103. Pp. 34–47. https://doi.org/10.1016/j.jprocont.2021.05.003
4. Xu X., Lin Y., Ye C. Fault diagnosis of marine machinery via an intelligent data-driven framework / X. Xu, Y. Lin,

C. Ye // Ocean Engineering. 2023. Vol. 289, No. 1. Article 116302. https://doi.org/10.1016/j.oceaneng.2023.116302

5. Schweidtmann A. M., Zhang D., von Stosch M. A review and perspective on hybrid modeling methodologies / A. M. Schweidtmann, D. Zhang, M. von Stosch // Digital Chemical Engineering. 2024. Vol. 10. Article 100136. https://doi.org/10.1016/j.dche.2023.100136
6. Nikpour H., Aamodt A. Fault diagnosis under uncertain situations within a Bayesian knowledge-intensive CBR system / H. Nikpour, A. Aamodt // Progress in Artificial Intelligence. 2021. Vol. 10, No. 1. Pp. 245–258. https://doi.org/10.1007/s13748-020-00227-x
7. Chen M., Xia J., Huang R., Fang W. Case-based reasoning system for aeroengine fault diagnosis enhanced with attitudinal Choquet integral / M. Chen, J. Xia, R. Huang, W. Fang // Applied Sciences. 2022. Vol. 12, No. 11. Article 5696. https://doi.org/10.3390/app12115696
8. Soleimani M., Campean F., Neagu D. Integration of Hidden Markov modelling and Bayesian network for fault detection and prediction of complex engineered systems / M. Soleimani, F. Campean, D. Neagu // Reliability Engineering & System Safety. 2021. Vol. 215. Article 107808. https://doi.org/10.1016/j.ress.2021.107808
9. El-Awady A., Ponnambalam K. Integration of simulation and Markov chains to support Bayesian networks for probabilistic failure analysis of complex systems / A. El-Awady, K. Ponnambalam // Reliability Engineering & System Safety. 2021. Vol. 211. Article 107511. https://doi.org/10.1016/j.ress.2021.107511
10. Zhong D., Xia Z., Zhu Y., Duan J. Overview of predictive maintenance based on digital twin technology / D. Zhong, Z. Xia, Y. Zhu, J. Duan // Heliyon. 2023. Vol. 9, No. 4. Article e14534. https://doi.org/10.1016/j.heliyon.2023.e14534
11. Kalafatelis A. S., Nomikos N., Giannopoulos A., Trakadas P. Survey on predictive maintenance in the maritime industry using machine and federated learning / A. S. Kalafatelis, N. Nomikos, A. Giannopoulos, P. Trakadas // TechRxiv. 2025. Vol. 11. https://doi.org/10.36227/techrxiv.173473250.04784922/v1
12. Özaydın E., Fışkın R., Uğurlu Ö., Wang J. A hybrid model for marine accident analysis based on Bayesian network (BN) and association rule mining (ARM) / E. Özaydın, R. Fışkın, Ö. Uğurlu, J. Wang // Ocean Engineering. 2022. Vol. 247. Article 110705. https://doi.org/10.1016/j.oceaneng.2022.110705
13. Cheliotis M., Lazakis I., Cheliotis A. Bayesian and machine learning-based fault detection and diagnostics for marine applications / M. Cheliotis, I. Lazakis, A. Cheliotis // Ships and Offshore Structures. 2022. Vol. 17, No. 12. Pp. 2686–2698. https://doi.org/10.1080/17445302.2021.2012015
14. OREDA. Offshore Reliability Data Handbook. 6th ed. OREDA, 2015.
15. Soliman A., Marsha M. A., Rahaman S. Digital twin predictive maintenance systems / A. Soliman, M. A. Marsha, S. Rahaman // Journal of Ocean Technology. 2024. Vol. 19, No. 2. Pp. 46–76.
16. Jovanović I., Karatuğ Ç., Perčić M., et al. Combined fault tree analysis and Bayesian network for reliability assessment of marine internal combustion engine / I. Jovanović, Ç. Karatuğ, M. Perčić, et al. // Journal of Marine Science and Application. 2025. https://doi.org/10.1007/s11804-025-00692-7
17. Velasco-Gallego C., Lazakis I. Mar-RUL: A remaining useful life prediction approach for fault prognostics of marine machinery / C. Velasco-Gallego, I. Lazakis // Applied Ocean Research. 2023. Vol. 140. Article 103735. https://doi.org/10.1016/j.apor.2023.103735
18. Schultheis A. Exploring a hybrid case-based reasoning approach for time series adaptation in predictive maintenance / A. Schultheis // ICCBR ’24 Workshop Proceedings. CEUR WS, Vol. 3708, 2024.
19. Daya A. A., Lazakis I. Developing an advanced reliability analysis framework for marine systems operations and maintenance / A. A. Daya, I. Lazakis // Ocean Engineering. 2023. Vol. 272. Article 113766. https://doi.org/10.1016/j.oceaneng.2023.113766
20. Neupane D., Bouadjenek M. R., Dazeley R., Aryal S. Data-driven machinery fault diagnosis: A comprehensive review / D. Neupane, M. R. Bouadjenek, R. Dazeley, S. Aryal // Neurocomputing. 2025. Vol. 627. Article 129588. https://doi.org/10.1016/j.neucom.2025.129588
21. Lv Y., Yang X., Li Y., Liu J., Li S. Fault detection and diagnosis of marine diesel engines: A systematic review / Y. Lv, X. Yang, Y. Li, J. Liu, S. Li // Ocean Engineering. 2024. Vol. 294. Article 116798. https://doi.org/10.1016/j.oceaneng.2024.116798
22. Li Y., Zhang W., Cui L., Gao H. System reliability modeling and analysis for a marine power equipment operating in a discrete-time dynamic environment / Y. Li, W. Zhang, L. Cui, H. Gao // Quality and Reliability Engineering International. 2024. Vol. 40. Pp. 3422–3438. https://doi.org/10.1002/qre.3577

Вичужанін В. В., Вичужанін О. В. Інтегрований підхід до діагностики складних технічних систем: експериментальна валідація та багатовимірна оцінка ефективності

У цій статті представлено всебічну експериментальну верифікацію інтегрованого підходу до діагностики технічного стану (ТС) складних технічних систем (СТС) на прикладі суднових енергетичних установок (СЕУ). Запропонована методологія поєднує логіку прецедентів (Case Based Reasoning – CBR), ймовірнісне прогнозування за допомогою байєсівських мереж і марковських ланцюгів, а також імітаційне моделювання сценаріїв деградації та каскадних відмов. Тестування проводилось за трьома сценаріями: нормальний режим експлуатації, режим підвищеного навантаження та сценарій з обмеженою доступністю даних, що дозволило всебічно оцінити адаптивність алгоритмів і їх стійкість до змінних експлуатаційних факторів. Для кількісної оцінки якості діагностики використовувались класичні метрики бінарної класифікації (Accuracy, Precision, Recall і F1 score), а також нововведені розширені показники: зважена точність (WAcc), F1-оцінка з урахуванням критичності відмов компонентів (F1W), повнота зі зважуванням на ризик відмови (RecallR), скоригована точність з урахуванням вартості помилкових тривог (PrecisionC) та Індекс стабільності діагностики (DSI). Результати багатосценарного експерименту показали стабільне покращення всіх основних показників: Accuracy зросла з 78.5% до 85.3%, Precision – з 75.2% до 83.1%, Recall – з 80.1% до 87.6%, F1 score – з 77.5% до 85.3%, RecallR досягла 91.0%, а DSI становив 0.983. П'ятикратна крос-валідація показала стандартне відхилення F1 score на рівні 2.2%, що підтверджує відтворюваність і надійність запропонованого методу для експериментального тестування інтегрованого підходу до діагностики СТС. Реалізація циклічної процедури "імітація – ймовірності – адаптація CBR" значно зменшила кількість помилкових тривог та пропущених критичних відмов обладнання СЕУ. Практична значущість підходу полягає у можливості його інтеграції в SCADA/PMS-системи морських СТС та наземних енергетичних станцій, що сприятиме переходу до інтелектуального прогнозного обслуговування, зменшенню незапланованих простоїв, зниженню витрат і підвищенню надійності обладнання. Перспективи майбутніх досліджень включають підвищення адаптивності підходу, розширення бази прецедентів та розробку інструментів для автоматизованої обробки гетерогенних даних.

Ключові слова: *прогнозна діагностика, байєсівські мережі, адаптація CBR, моделювання відмов, ризик-орієнтовані метрики, стабільність діагностики, інтелектуальне обслуговування*

Вичужанін Володимир Вікторович – д.т.н, професор, завідувач кафедри інформаційних технологій, національний університет «Одеська політехніка», Одеса, email: v.v.vychuzhanin@op.edu.ua

Вичужанін Олексій Володимирович – доктор філософії, асистент, національний університет «Одеська політехніка», Одеса, email: v.v.vychuzhanin@op.edu.ua

Стаття подана 12.05.2025.