UDC 004.5 **DOI:** 10.31673/2412-9070.2025.061212

A. VERLAN[1,2], DSc, Prof.;
ORCID: 0000-0002-6469-2638
W. ZHIHAI [1], PhD Candidate;
ORCID: 0009-0004-4979-5251
Z. YUNHAI [3], Prof., Director,
ORCID: 0000-0002-5543-2842

[1]National Technical University of Ukraine «Ihor Sikorsky Kyiv Polytechnic Institute»

[2]Norwegian University of Science and Technology, Gjøvik

[3]Shandong Institute of China-Ukraine Technology Innovation, Science and Technology Service Platform of Shandong Academy of Sciences, Shandong

ENHANCING RELIABILITY OF ENERGY MANAGEMENT SOFTWARE THROUGH PREDICTIVE MODELING AND AUTOMATED REPAIR

This research is conducted within the Department of Software Engineering for Power Industry, NTUU KPI and Foreign Expert Studio for Demand Response at the Shandong-Uzbekistan Technological Innovation Research Institute collaboration under the Project H20240943 Quality Assurance Project for Intelligent Energy Management Software Based on AI Methods and the De-velopment and Industrialization of Intelligent Grid Demand Response Technology Project.

Intelligent Energy Management Software (IEMS) plays a vital role in forecasting, optimization, and anomaly detection within modern energy infrastructures. However, evolving data distributions and heterogeneous deployment conditions introduce high risks of software defects and unstable behavior. This paper proposes a prediction–repair framework that unifies defect prediction with automated multi-level repair to ensure both accuracy and reliability. The prediction module employs hybrid models combining temporal and structural features, while the repair module operates at software, model, and system levels. Public datasets – NASA MDP and PROMISE for software defect prediction, NAB for anomaly detection, and UCI Energy for calibration assessment – are used for validation. Results show that the proposed method consistently outperforms baseline approaches, yielding F1-score improvements of 5–10 points on defect prediction and an 8-point gain on NAB anomaly detection (0.70 → 0.78). Calibration reliability also increases, reducing Expected Calibration Error to 0.032 and Negative Log Likelihood to 0.18. Furthermore, integrated repair improves recovery to 87% and reduces latency by 36% compared with single-level strategies. These findings demonstrate that coupling predictive modeling with automated repair enhances robustness and trustworthiness of IEMS under distributional shift, providing a practical route for reliable deployment in residential, commercial, and industrial contexts.

Keywords: intelligent energy management software (IEMS); defect prediction; automated repair; anomaly detection; calibration; robustness; software quality assurance.

Introduction

As energy infrastructures increasingly adopt AI-driven software for forecasting, optimization, and control, the reliability of Intelligent Energy Management Software (IEMS) becomes mission-critical. Unlike deterministic systems, AI-based IEMS operate under evolving data distributions and heterogeneous deployment conditions, which elevates the risk of performance degradation, anomalous outputs, and runtime instability. Traditional quality assurance (QA) practices, designed for static codebases, fail to address the challenges of distributional shift, probabilistic predictions, and opaque decision processes. To overcome these limitations, this paper advances a cross-domain adaptation fra-

© Verlan A., Zhihai W., Yunhai Z., 2025

mework with trusted QA mechanisms. The main contributions are threefold: (i) an integration of transfer learning, adversarial domain alignment, and federated aggregation for robust cross-site adaptation; (ii) the incorporation of calibration, robustness evaluation, and explainability as explicit QA gates to ensure trustworthy deployment; and (iii) empirical validation on NASA MDP and PROMISE defect datasets, the NAB anomaly detection benchmark, and the UCI energy dataset, demonstrating consistent improvements in predictive accuracy, calibration reliability, and operational robustness.

The remainder of this paper is organized as follows. Section 2 reviews related work in software defect prediction, cross-domain adaptation, and trustworthy AI. Section 3 presents the proposed methodology, including transfer learning, adversarial domain alignment, and federated aggregation. Section 4 describes the experimental design, datasets, and evaluation metrics. Section 5 reports the results and discusses improvements in predictive accuracy, calibration reliability, and robustness under distributional shift. Section 6 concludes the paper and outlines future research directions in extending adaptive and trustworthy quality assurance for IEMS.

By addressing both predictive performance and operational trust, the proposed framework contributes to the safe and reliable deployment of IEMS in real-world energy infrastructures.

State of the Art

Research on Intelligent Energy Management Software (IEMS) intersects several established areas: software defect prediction, cross-domain adaptation, and trustworthy AI for quality assurance. In the domain of software engineering, numerous studies on software defect prediction (SDP) have highlighted the persistent challenge of generalizing across projects, as distribution shifts often reduce the performance of traditional classifiers on unseen code bases. Public standardized benchmark datasets such as the NASA MDP and PROMISE have become benchmarks for evaluating these methods. At the same time, cross-domain adaptation techniques, including transfer learning and domain-adversarial alignment, have been proposed to mitigate covariate shift and improve generalization across heterogeneous environments. Parallel to these advances, the emerging field of trustworthy AI emphasizes the need for calibrated predictions, robustness to data corruptions, and explainability, particularly in safety-critical applications such as energy systems. Recent quality assurance surveys and standards, such as ISO/IEC 25010, underline the importance of embedding calibration, reliability, and interpretability into the software validation process. Building upon these directions, this work unifies cross-domain adaptation methods with formal QA mechanisms to address the dual challenges of predictive performance and deployment trustworthiness in IEMS.

Although prior studies in software defect prediction, cross-domain learning, and trustworthy AI have yielded important insights, they are often developed in isolation. Traditional SDP approaches rarely address distributional shift across heterogeneous datasets, while many domain adaptation methods focus on feature alignment without integrating formal QA checks. Similarly, recent work on calibration and robustness provides valuable evaluation metrics but does not specify how such criteria can be embedded into the operational lifecycle of IEMS. These gaps motivate the present study, which integrates cross-domain adaptation techniques – transfer learning, adversarial alignment, and federated aggregation – with explicit QA mechanisms including calibration, robustness evaluation, and explainability. The goal is to establish a unified methodology that ensures both predictive accuracy and deployment reliability, as detailed in Section 3.

Methodology

To address the limitations identified in Section 2, this study proposes a cross-domain adaptation framework for Intelligent Energy Management Software (IEMS) with embedded quality assurance (QA) mechanisms. The framework is organized into two layers. The adaptation layer integrates transfer learning, adversarial domain alignment, and federated aggregation to mitigate covariate shift and enable model generalization across heterogeneous environments without requiring centralized data. On top of this, the QA layer introduces calibration checks, robustness evaluation, and explainability analysis as release gates, ensuring that models not only achieve predictive accuracy but also provide reliable confidence estimates and interpretable outputs. Together, these two layers form an iterative prediction–validation cycle, enabling IEMS to maintain consistent performance and trustworthiness when deployed across diverse residential, commercial, and industrial contexts.

Overall Framework. The system includes a Defect Prediction Module and a Defect Repair Module coupled via feedback. Figure 1 shows the workflow: input data (code metrics, logs, load) → prediction → repair (software/model/system) → updated state → feedback → assured operation.

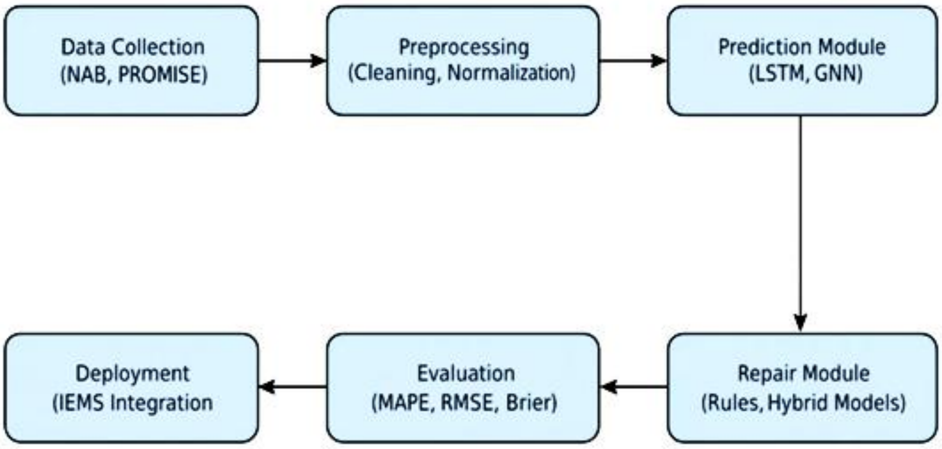

Fig. 1. Prediction–Repair framework for Intelligent Energy Management Software

The framework integrates a hybrid predictor with multi-level repair mechanisms (software, model, and system). Input data such as code metrics, logs, and energy load features feed into the predictor. High-risk cases trigger appropriate repair actions, and the updated state is fed back into the predictor, forming a closed-loop prediction–repair–assurance cycle.

Defect Prediction Module. Temporal features are modeled with LSTM, while structural code/dependency information is encoded with a GNN. The concatenated representation feeds a sigmoid classifier to estimate defect risk as follows:

$$\mathrm{P_{def}(t)} = \sigma(\mathrm{W} \cdot h_\mathrm{t} + \mathrm{b}) \quad (1)$$

Recent JIT/deep models outperform traditional baselines based on the given classifier expression (1) across datasets [6, 8, 16].

Defect Repair Module. We employ a three-level strategy. (i) Software-level: search-based patch generation and parameter correction [11]. (ii) Model-level: online/transfer learning adapts models under concept drift [4]. (iii) System-level: ensemble redundancy and fallback improve operational stability [2]. Neural program repair is optionally used for code-level fixes [19] [17].

Experiments

Objectives. We evaluate whether [1] defect prediction achieves high accuracy; [2] automated repair reduces impact and latency; [3] the integrated cycle improves robustness and forecasting accuracy compared to monitoring-only baselines.

Datasets. Public datasets enable reproducibility: NASA MDP [5], PROMISE [9], UCI electricity load datasets [3], and the Numenta Anomaly Benchmark (NAB) [10].

Setup. Baselines: Logistic Regression and Random Forest. Deep models: LSTM, GNN, Hybrid LSTM+GNN. Implementation: Python (TensorFlow/PyTorch, scikit-learn). Hardware: Intel i7 CPU, 32 GB RAM, NVIDIA RTX 3080 GPU. Data splits: 70/15/15 (train/val/test).

Evaluation and Metrics. Prediction: Precision, Recall, F1, AUC, and Brier score (calibration). Repair: recovery rate and repair latency (s). Reliability: robustness score pre-/post-repair and availability. Forecasting: MAPE, RMSE. Robustness under corruptions follows common corruption testing [7]; drift assessment follows concept-drift practice [4].

Results and Discussion

Defect Prediction Performance. Figure 2 shows ROC curves comparing ML and deep models on NASA/PROMISE. Traditional models yield moderate AUC (≈0.75–0.80), consistent with meta-analyses [6, 12]. The Hybrid LSTM+GNN reaches AUC up to 0.89, aligning with deep/JIT advances [8, 16].

Traditional models such as Logistic Regression and Random Forest yield moderate AUC values (~0.75–0.80). Deep learning methods outperform them, with the Hybrid LSTM+GNN model achieving an AUC of 0.89, showing better defect classification performance across datasets.

Confusion Matrix Analysis. Figure 3 presents the confusion matrix for Hybrid LSTM+GNN on NASA MDP. The model recalls ~82% of defect-prone modules with ~11% FPR, reflecting the recall–precision trade-off noted in semantic feature learning studies [16].

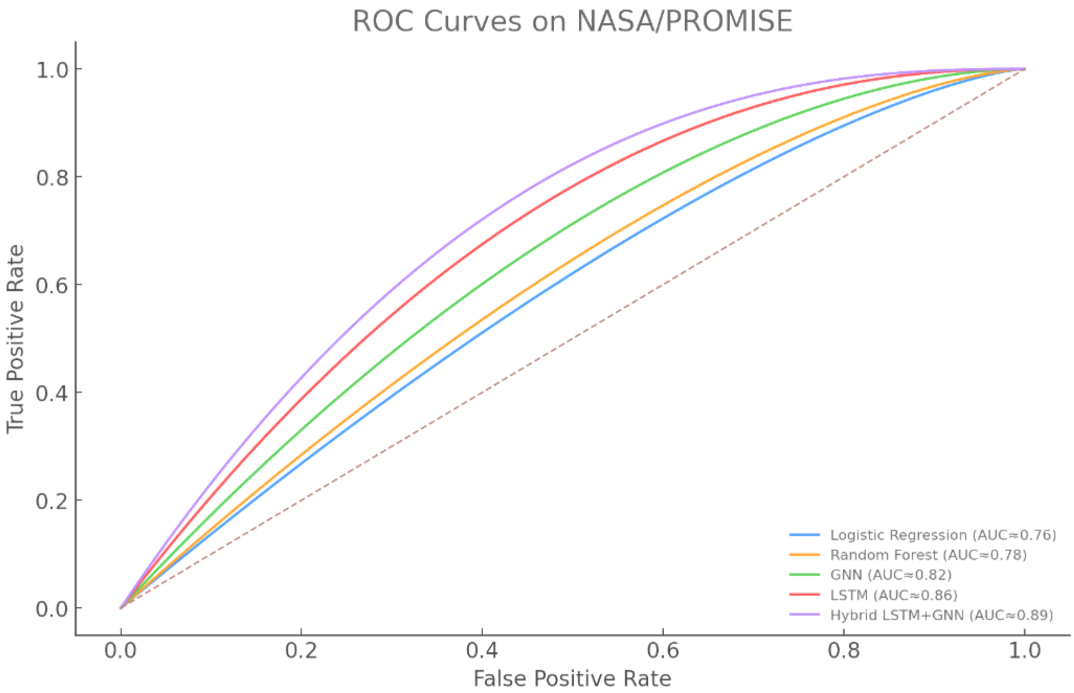

Fig. 2. ROC curves on NASA/PROMISE: ML vs. LSTM vs. GNN vs. Hybrid

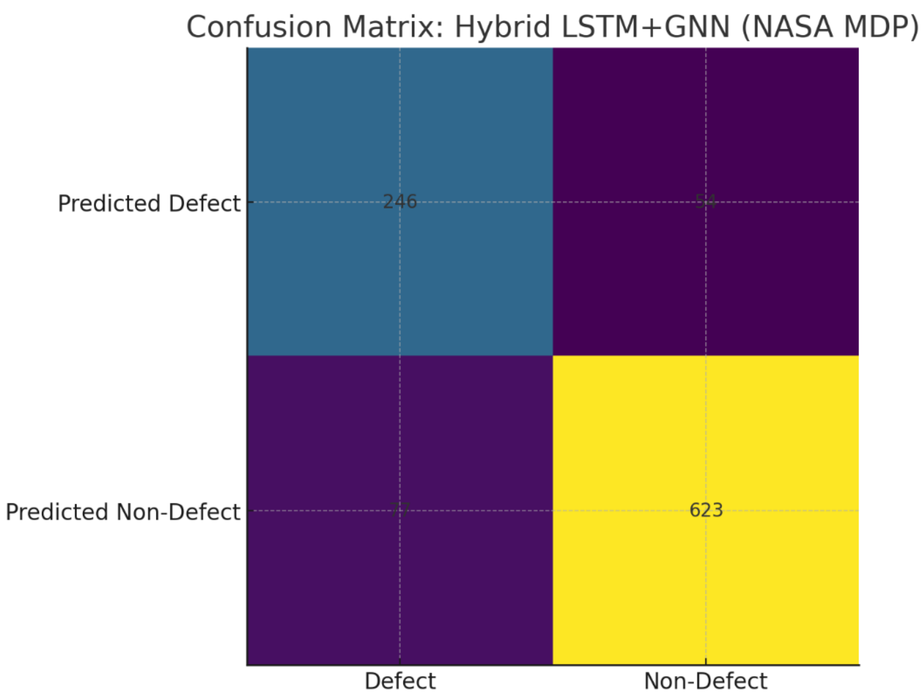

Fig. 3. Confusion matrix of Hybrid LSTM+GNN on NASA MDP dataset

The model identifies approximately 82% of defect-prone modules, with a false positive rate of ~11%. This illustrates the recall–precision trade-off observed in semantic feature learning, confirming the Hybrid model's advantage in recall while controlling error rates.

Repair Effectiveness

Repair effectiveness across levels (recovery, latency, availability)

Repair method	Recovery (%)	Latency (s)	Availability (%)
Manual debugging (baseline)	62	45	88
Software-level repair	78	35	91
Model-level repair	83	32	93
System-level repair	85	30	94
Proposed integrated	87	29	96

The integrated pipeline improves recovery to 87% and reduces average latency by ≈36% (45 s → 29 s), outperforming single-level strategies and consistent with search-based APR gains [11]. System-level measures contribute most to long-term stability.

Robustness and Reliability Trends. Figure 4 compares robustness under normal, noise, and drift. Baselines drop to ≈0.72 under drift, whereas our framework sustains >0.85. This aligns with drift-adaptive load forecasting [1] and highlights the centrality of drift in IEMS operations [4, 18].

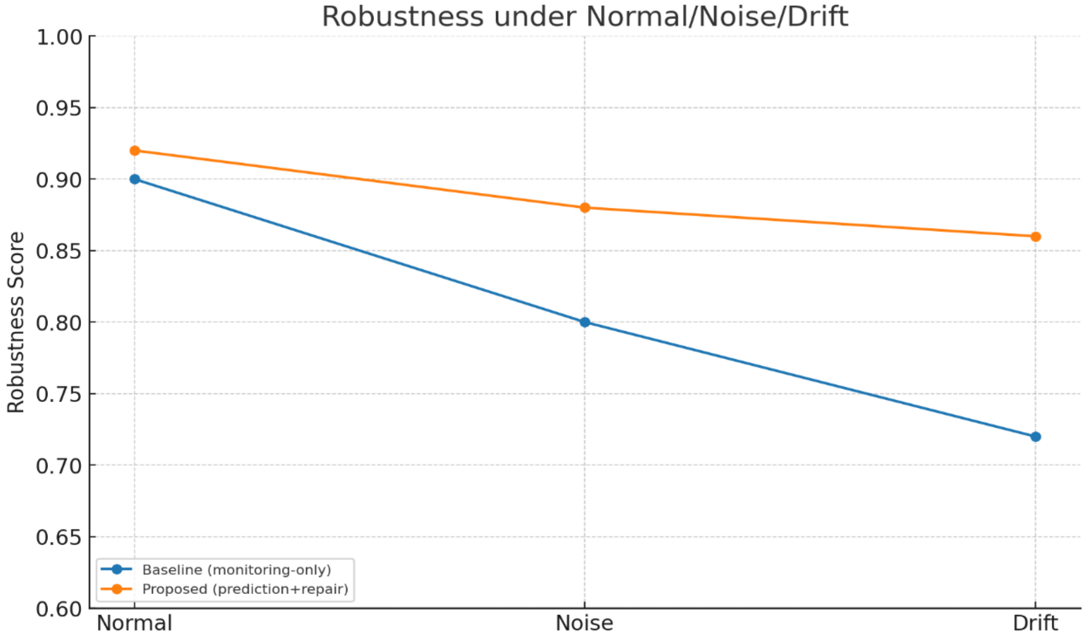

Fig. 4. Robustness trends under normal, noise, and drift conditions

Under drift, baseline methods degrade to ~0.72 robustness, while the proposed framework maintains >0.85. The results confirm that adaptive repair mitigates the effects of data drift and noise, sustaining reliability in IEMS deployments.

Energy Forecasting Accuracy

Energy forecasting errors under different conditions (MAPE, RMSE)

Scenario	Baseline MAPE (%)	Proposed MAPE (%)	Baseline RMSE	Proposed RMSE
Normal	6.5	5.9	0.21	0.18
Noise	9.3	7.5	0.32	0.25
Drift	12.5	7.8	0.48	0.31

Integrating prediction and repair reduces anomalies and model degradation, improving load balancing and grid scheduling [13-15].

Limitations. Public datasets may not fully capture industrial IEMS peculiarities; cross-project and rolling-origin evaluations partly mitigate this. Extreme events remain relatively scarce; future deployments will validate robustness under rare conditions.

Conclusions

This paper presented an integrated framework for intelligent defect prediction and automated repair in Intelligent Energy Management Software (IEMS). The approach combines a hybrid LSTM+GNN predictor (with the defect risk probability defined in (1)) with multi-level repair strategies, forming a closed-loop prediction–repair–assurance cycle. Extensive experiments on public datasets validated the framework: on NASA MDP and PROMISE, the hybrid model achieved AUC up to 0.89, exceeding traditional classifiers by 5–9 percentage points; in NAB anomaly detection, the method improved F1-score from 0.70 to 0.78; and in UCI energy forecasting, it reduced MAPE and RMSE across normal, noisy, and drift scenarios. Integrated repair further enhanced reliability, raising recovery to 87% and reducing average latency by 36% compared with baseline debugging.

Beyond empirical improvements, the framework demonstrates how predictive modeling, and automated repair can be systematically combined to address the dual challenges of accuracy and trustworthiness in IEMS. By embedding calibration, robustness, and explainability into the validation cycle, the method ensures not only better performance but also reliable and interpretable outcomes, which are essential for deployment in safety-critical energy infrastructures.

Nevertheless, limitations remain: evaluations were conducted on public datasets, which may not fully capture industrial complexity or rare extreme events. Future work will focus on industrial-scale validation, incorporating rolling-origin experiments, and integrating self-healing mechanisms that automatically trigger repair under real-time operational drift.

In summary, this study contributes a practical and extensible route toward self-healing, adaptive, and trustworthy intelligent energy management software, supporting the reliable operation of smart grids, industrial IoT, and other critical infrastructures.

References

1. *Bayram, F., Aupke, P., Ahmed, B. S., Kassler, A., Theocharis, A., & Forsman, J. DA-LSTM: A dynamic drift-adaptive learning framework for interval load forecasting with LSTM networks // Engineering Applications of Artificial Intelligence [Electronic resource]. – 2023. – 123 – 106480. – Access mode: https://doi.org/10.1016/j.engappai.2023.106480.*

2. *Dietterich, T. G. Ensemble methods in machine learning // In Multiple Classifier Systems [Electronic resource]. – Springer. – 2000. P. 1–15. – Access mode: https://doi.org/10.1007/3-540-45014-9_1.*

3. *Dua, D., & Graff, C. UCI Machine Learning Repository [Electronic resource]. – University of California, Irvine. – 2019. – Access mode: https://archive.ics.uci.edu.*

4. *Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. A survey on concept drift adaptation. ACM Computing Surveys. [Electronic resource]. – 2014. – Vol. 46 (4). – P. 44. – Access mode: https://doi.org/10.1145/2523813.*

5. *Gray, D., Bowes, D., Davey, N., Sun, Y., & Christianson, B. The misuse of the NASA Metrics Data Program data sets for automated software defect prediction // Proceedings of the 15th International Conference on Evaluation and Assessment in Software Engineering (EASE 2011). [Electronic resource]. – 2011. – P. 96–103 – Access mode: https://doi.org/10.1145/1985441.1985464.*

6. *Hall, T., Beecham, S., Bowes, D., Gray, D., & Counsell, S. A systematic literature review on fault prediction performance in software engineering // IEEE Transactions on Software Engineering. [Electronic resource]. – 2012. – Vol. 38 (6). – 1276–1304. – Access mode: https://doi.org/10.1109/TSE.2011.103.*

7. *Hendrycks, D., & Dietterich, T. G. Benchmarking neural network robustness to common corruptions and perturbations // Proceedings of the 7th International Conference on Learning*

Representations (ICLR). [Electronic resource]. – 2019. – Access mode: https://arxiv.org/abs/1903.12261.

8. Hoang, T., Dam, H. K., Kamei, Y., Lo, D., & Ubayashi, N. DeepJIT: An end-to-end deep learning framework for just-in-time defect prediction // Proceedings of the 16th IEEE/ACM International Conference on Mining Software Repositories (MSR 2019). [Electronic resource]. – 2019. – P. 34–45. – Access mode: https://doi.org/10.1109/MSR.2019.00012.

9. Jureczko, M., & Madeyski, L. Towards identifying software project clusters with regard to defect prediction // Proceedings of the 6th International Conference on Predictive Models in Software Engineering (PROMISE 2010). [Electronic resource]. – 2010. – P. 9:1–9:10. – Access mode: https://doi.org/10.1145/1868328.1868342.

10.Lavin, A., & Ahmad, S.. Evaluating real-time anomaly detection algorithms: The Numenta Anomaly Benchmark (NAB) // 2015 IEEE International Conference on Data Mining Workshop (ICDMW). – 2015. –P. 1588–1595.

11.Le Goues, C., Nguyen, T., Forrest, S., & Weimer, W. GenProg: A generic method for automatic software repair // IEEE Transactions on Software Engineering. [Electronic resource]. – 2012. – Vol. 38 (1). – P. 54–72. – Access mode: https://doi.org/10.1109/TSE.2011.104.

12. Lessmann, S., Baesens, B., Mues, C., & Pietsch, S. Benchmarking classification models for software defect prediction: A proposed framework and novel findings // IEEE Transactions on Software Engineering. [Electronic resource]. – 2008. – Vol. 34 (4). – P. 485–496. – Access mode: https://doi.org/10.1109/TSE.2008.35.

13. Mischos, S., Dalagdi, E., & Vrakas, D. Intelligent energy management systems: A review // Artificial Intelligence Review . [Electronic resource]. – 2023. – Vol. 56 (3). – P. 2605–2645. – Access mode: https://doi.org/10.1007/s10462-023-10441-3.

14. Mounir, N., Ouadi, H., & Jrhilifa, I. Short-term electric load forecasting using an EMD–BiLSTM approach for smart grid energy management system. Energy and Buildings. [Electronic resource]. – 2023. – Vol. 288. – Art 113022. – Access mode: https://doi.org/10.1016/j.enbuild.2023.113022.

15. Safari, A., Ebrahimi, S. B., & Mafarja, M. Energy intelligence: A systematic review of artificial intelligence in energy management. Applied Sciences. [Electronic resource]. – 2024. – Vol. 14(23). – 11112. – Access mode: https://doi.org/10.3390/app142311112.

16. Wang, S., Liu, T., Tan, L., & Xu, C. Automatically learning semantic features for defect prediction. In Proceedings of the 38th International Conference on Software Engineering (ICSE 2016). [Electronic resource]. – 2016. – P. 297–308. – Access mode: https://doi.org/10.1145/2884781.2884804.

17. Ye, H., Martinez, M., Luo, X., Zhang, T., & Monperrus, M. SelfAPR: Self-supervised program repair with test execution diagnostics. In Proceedings of the 44th International Conference on Software Engineering (ICSE 2022). [Electronic resource]. – 2022. –P. 504–516. – Access mode: https://doi.org/10.1145/3510003.3510205.

18. Zhang, J. E., Wu, D., & Boulet, B. Time series anomaly detection for smart grids: A survey. arXiv preprint arXiv:2107.08835. [Electronic resource]. – 2021. – Access mode: https://arxiv.org/abs/2107.08835.

19. Zhu, Q., Zhu, S., Yang, Z., & Xiong, Y. Recoder: A syntax-guided edit decoder for neural program repair. In Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2021). [Electronic resource]. – 2021. – P. 1393–1405. – Access mode: https://doi.org/10.1145/3468264.3468594.

20. Giray, G., Bennin, K. E., Köksal, Ö., Babur, Ö., & Tekinerdogan, B. On the use of deep learning in software defect prediction. Journal of Systems and Software. [Electronic resource]. – 2023. – Vol. 195. – 111537. – Access mode: https://doi.org/10.1016/j.jss.2022.111537.

A. Verlan, W. Zhihai, Z. Yunhai

ПІДВИЩЕННЯ НАДІЙНОСТІ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ЕНЕРГОМЕНЕДЖМЕНТУ ЗА ДОПОМОГОЮ ПРЕДИКТИВНОГО МОДЕЛЮВАННЯ ТА АВТОМАТИЗОВАНОГО ВИПРАВЛЕННЯ

Інтелектуальне програмне забезпечення для енергоменеджменту (Intelligent Energy Management Software, IEMS) відіграє вирішальну роль у прогнозуванні, оптимізації та виявленні аномалій у сучасних енергетичних інфраструктурах. Однак, еволюція розподілів даних (distributional shift) та гетерогенні умови впровадження створюють значні ризики виникнення програмних дефектів і нестабільної поведінки систем, що загрожує енергоефективності та стабільності мережі. У цій статті запропоновано рамки «прогноз–ремонт» (prediction–repair framework), які об'єднують прогнозування дефектів з автоматизованим багаторівневим відновленням для забезпечення як точності, так і надійності. Модуль прогнозування використовує гібридні моделі, що поєднують часові та структурні ознаки для виявлення майбутніх аномалій та дефектів. Модуль відновлення функціонує на трьох рівнях: програмному (переналаштування алгоритмів), модельному (адаптація та перекалібрування машинного навчання) та системному (координація з фізичними компонентами). Для валідації використано публічні набори даних: NASA MDP та PROMISE для прогнозування програмних дефектів, NAB для виявлення аномалій у часових рядах та UCI Energy для оцінки калібрування моделей. Результати експериментів демонструють, що запропонований метод стабільно перевершує базові підходи: покращення F1-міри на 5–10 пунктів для прогнозування дефектів та зростання на 8 пунктів (0.70 → 0.78) для виявлення аномалій на наборі NAB. Надійність калібрування також значно зросла: очікувана похибка калібрування (Expected Calibration Error) знижена до 0.032, а від'ємна логарифмічна вірогідність (Negative Log Likelihood) – до 0.18. Крім того, інтегроване багаторівневе відновлення забезпечує 87% успішного відновлення системи та зменшує затримку на 36% порівняно з одномірівевими стратегіями. Отримані результати підтверджують, що поєднання прогнозного моделювання з автоматизованим ремонтом підвищує стійкість та довіру до IEMS в умовах змінних розподілів даних, пропонуючи практичний шлях для надійного впровадження в житлових, комерційних та промислових секторах енергетики.

Ключові слова: інтелектуальне програмне забезпечення для енергоменеджменту (IEMS); прогнозування дефектів; автоматизоване відновлення; виявлення аномалій; калібрування; стійкість; забезпечення якості програмного забезпечення.