World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

WJAETS

(Research Article)

Self-healing machine learning system for fault detection using predictive maintenance

Adithya Raj M V [1], Kumar Siddamallappa. U [2] and Shreedevi Prakash Gotur [3,*]

[1] Department of Computer Applications(MCA), Davangere University, Davangere, Karnataka, India.
[2] Department of Computer Applications, Davangere University, Davangere, Karnataka, India,
[3] Department of Computer Science, Davangere University, Davangere, Karnataka, India.
Kumar Siddamallappa. U; ORCID - 0000 0002 1975 3868
Shreedevi Prakash Gotur; ORCID - 0009 0001 9185 8998

World Journal of Advanced Engineering Technology and Sciences, 2026, 20(01), 241–250

Publication history: Received on 14 June2026; revised on 15 July 2026; accepted on 18 July 2026

Article DOI: https://doi.org/10.30574/wjaets.2026.20.1.0370

Abstract

Machine learning-powered fault detection systems are the backbone of modern predictive maintenance solutions in the context of Industry 4.0, as they allow organizations to prevent expensive breakdowns and unnecessary machine downtime. Still, it is a known fact that machine learning systems are prone to model drift, which happens when a learning model performs worse due to changes in input data patterns. In dynamic industrial environments where machines operate under varying conditions, such changes are frequent and inevitable. Therefore, predictive maintenance systems must have built-in self-learning and self-healing mechanisms that allow them to adapt to new patterns and maintain the desired performance level. This paper proposes a Self-Healing Machine Learning System for Fault Detection to perform these tasks and demonstrates the concept's feasibility within the Python/Flask environment. The proposed system utilizes Scikit-learn and the AI4I 2020 Predictive Maintenance Dataset to train an isolated random forest classifier and test its healing mechanism's performance. The results demonstrate that the proposed approach can keep the random forest classifier's performance stable by re-training it with new data whenever the performance metric falls below a certain threshold. Therefore, the self-healing ML system described in this paper can be a viable solution for industrial applications, as it ensures continuous learning and adaptability in the ever-changing environment of Industry 4.0.

Keywords: Self-Healing Machine Learning; Fault Detection; Predictive Maintenance; Random Forest; Model Drift; Industry 4.0

1. Introduction

Modern industries deal with thousands of machines operating in different environments and under various loads. The possibility of an unpredictable breakdown leads to inefficient production, increases the operating costs of the enterprise, and reduces the profit. One of the methods to prevent unforeseen failures is to apply learning models that predict possible malfunctions in advance. The problem of predictive maintenance is solved using the existing historical data about the machines since their operating characteristics are already known.

However, the accuracy of machine learning models decreases significantly when the operating conditions or load on the machines change over time. In this case, the data distributions shift, requiring engineers to adjust the models, which is a very resource-consuming process [1]. A self-healing system allows the machine learning model to track its performance and update it if a drop in performance is detected. Compared to the conventional approach, such systems

*Corresponding author: Shreedevi Prakash Gotur; E-mail: shreegotur@gmail.com

are much more efficient and do not require constant human interaction. Moreover, recent studies have demonstrated the effectiveness and advantages of such an approach in the context of artificial intelligence and other systems with a high degree of autonomy: machine learning, IoT, wireless networks, and cyber-physical systems [1], [4], [10].

Predictive maintenance uses the concept of failure prediction, which allows the operator to know in advance when a breakdown may occur. Using sensor data such as temperature, speed, torque, and tool wear, and machine learning models, the probability of a failure can be calculated. In recent years, researchers have explored using machine learning algorithms for failure diagnosis, adaptive self-healing systems, and methods to improve the performance of such models for various industrial applications [2], [11], [12].

The Self-Healing Machine Learning System is a predictive maintenance web application that was built using the AI4I 2020 Predictive Maintenance Data Set. The system demonstrates how a typical predictive model works in the real-time industry while constantly analyzing the model's performance with the possibility of retraining and replacing it without any human interaction. It highlights some of the latest trends in predictive maintenance and modern MLOps methodology: automated operations, self-driven operations, and resilient AI for the Fourth Industrial revolution [1], [2], [10].

The Flask framework powers the web application, while the SQLite database acts as the storage engine for persisting critical information such as training logs and health check details. The system employs the Scikit-learn library's random forest classification algorithm to train the predictive maintenance model. The server runs background jobs as a scheduler using the APScheduler module and provides a dashboard interface that exposes training and health-check logs in real-time to the operator.

1.1. Research Objectives

The objectives of the Self-Healing Machine Learning System project are as follows:

- To design and train a supervised machine learning model on the AI4I 2020 Predictive Maintenance dataset.
- To develop a comprehensive data preprocessing and feature-engineering pipeline.
- To design and develop a Flask-based web application to train the model, perform single and batch predictions, and visualize the results in a user-friendly interface.
- To design and develop an automated self-healing monitor for periodic evaluation.

2. Literature Review

The concept of self-healing systems has received considerable research interest in recent years due to its immense potential in making intelligent systems reliable and adaptive. Self-healing techniques have been investigated for various applications, including battery packs, wireless communication, cloud computing, software testing, Internet of Things (IoT), and artificial intelligence. This section presents the prior works carried out in the field of self-healing machine learning and fault prediction.

Kaushik Das et al. [1] presented an autonomous diagnostics and prognostics framework for lithium-ion battery management. Their framework implemented self-healing machine learning algorithms to monitor the battery's health condition continuously. The study demonstrated that their approach could learn and predict battery failures, thus increasing battery reliability throughout its lifespan.

In SHIFT framework, Seo et al. [2] proposed self-healing intelligence for transitions in machine learning feature stores and data stores. The authors suggested that their approach could provide self-healing intelligence during model transitions, therefore improving the model's accuracy and robustness.

Aleti et al. [3] proposed an adaptive software testing framework based on reinforcement learning algorithms with continuous fault anticipation and self-healing feedback. This paper demonstrated that automated learning and feedback techniques could enhance software quality by detecting faults and adapting the testing process accordingly.

In his work, Mata et al. [4] proposed a machine learning-based methodology capable of predicting alarms in wireless communication networks. This approach could anticipate alarms that may arise in the future while also healing the network to prevent disruptions in service.

In their review paper, Zhang et al. [5] identified challenges and opportunities in cellular wireless networks related to data-driven machine learning methods for self-healing communication networks. The authors emphasized network adaptation learning as the main requirement for realizing self-healing cellular wireless networks.

Dimara et al. [6] proposed a semantic self-healing framework for IoT edge devices. This framework enables IoT edge devices to reason about faults while also being able to recover from them. Their approach allows for greater reliability in smart devices by facilitating self-repairing services while also ensuring semantic interoperability.

Reshmi et al. [7] proposed a technique for 5G communication networks that enables networks to predict failures and heal themselves autonomously. Their approach could help maintain network reliability and reduce downtime in 5G communication networks.

Ravichandran et al. [8] proposed an AI-powered self-healing framework for cloud computing that can automatically detect and resolve incidents in a cloud computing platform. Their approach minimizes manual intervention in cloud operations while also ensuring reliability and availability of cloud services.

In their article, Inaganti et al. [9] discussed the role of artificial intelligence in unifying identity management, service management, and talent management into an intelligent unified platform for autonomous enterprises. Their work unveiled the power of AI in enabling autonomous enterprises through digital transformation.

Finally, Patel et al. [10] proposed a multi-agent learning framework for self-healing AI systems. Their approach enables multiple agents to collaborate in learning the patterns of failure, therefore enhancing the performance of self-healing AI systems.

3. Proposed Methodology

The rapid adoption of Industry 4.0 technologies has significantly increased the use of machine learning for intelligent fault detection and predictive maintenance. Although machine learning models can accurately identify equipment failures, their performance often declines over time due to model drift, where changes in operating conditions, sensor characteristics, and machine behavior alter the underlying data distribution. As a result, models trained on historical data gradually lose prediction accuracy, leading to incorrect fault detection, increased maintenance costs, and unexpected equipment downtime. This study is based on the hypothesis that a self-healing machine learning framework capable of continuously monitoring model performance and automatically retraining itself when prediction accuracy decreases can effectively overcome model drift and maintain reliable fault detection performance without manual intervention. the framework enables the deployed model to adapt to new operating conditions while ensuring uninterrupted system operation. and enhances system reliability, making it well suited for predictive maintenance applications in Industry 4.0.

3.1. Proposed Framework.

The proposed framework consists of six major phases that work together to maintain reliable fault prediction throughout the system lifecycle.

- Phase 1: Data Collection
- Phase 2: Data Preprocessing
- Phase 3: Feature Extraction / Feature Engineering
- Phase 4: Fault Detection and Classification
- Phase 5: Continuous Performance Monitoring
- Phase 6: Self-Healing (Automatic Retraining)

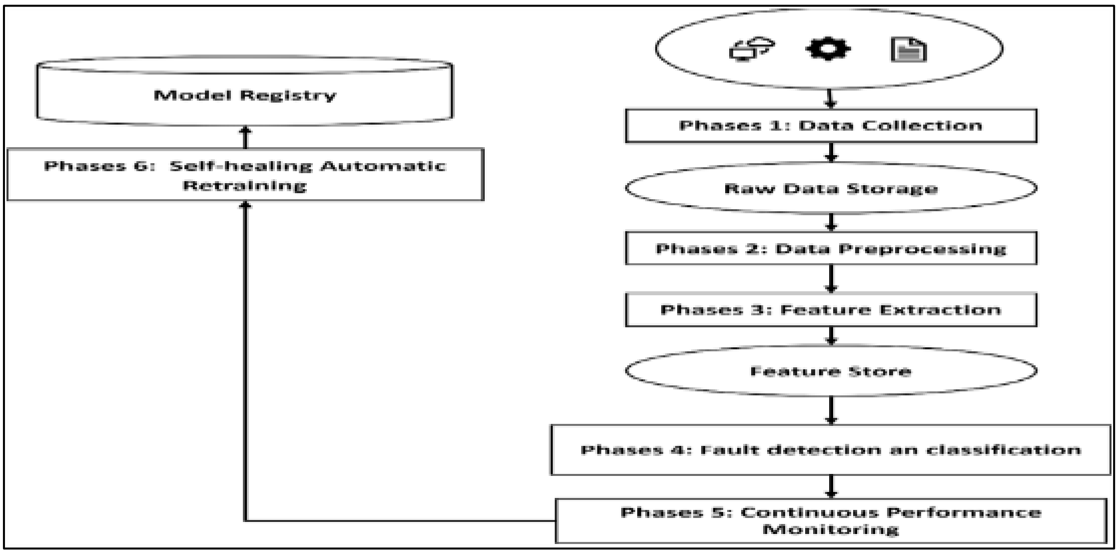

Figure 1 Six Major Phases Life Cycle

The workflow begins by collecting industrial sensor data from the AI4I 2020 Predictive Maintenance Dataset. The collected data are preprocessed to remove inconsistencies, handle missing values, and normalize feature values for improved learning performance. Feature engineering is then performed to identify the most informative variables for fault prediction. The processed dataset is used to train a Random Forest classifier, which predicts machine failures with high accuracy and robustness. After deployment, the system continuously evaluates model performance using predefined evaluation metrics.

3.2. Dataset Description

The proposed system uses the AI4I 2020 Predictive Maintenance Dataset, which contains real industrial sensor measurements collected from manufacturing equipment.

Table 1 AI4I 2020 Predictive Maintenance Dataset

Type(mode)	Air temperature [K]	Process temperature [K]	Rotational speed [rpm]	Torque [Nm]	Tool wear [min]
M	298.3	308.5	1629	37.1	146
L	298.3	308.5	1538	32.2	148
L	298.3	308.6	1570	34.8	151
L	298.3	308.4	1408	43.1	153
H	298.2	308.4	1424	47.6	158
L	298.2	308.4	1512	34.9	160
L	298.2	308.4	1478	42.9	162
M	298.1	308.3	1680	33.7	164
L	298.1	308.3	1495	38.7	166
M	298	308.1	1552	36.6	168
H	297.9	308.1	1393	50.3	170
L	298	308.2	1611	30.6	173
M	298	308.1	1339	51.6	175

L	298	308.1	1777	24.3	177
L	298	308.2	1348	58.8	179
L	298	308.2	1442	44.1	181
H	298	308.2	1395	49.1	183
M	297.9	308.1	1627	31.4	185
L	297.9	308.1	1585	33.5	188
L	298	308.2	1469	45.5	190
H	298	308.3	1362	56.8	193

This are the Sample Dataset from AI4I2020 Predictive Maintenances, That we'll trained as 80% of trained and 20% of tested data.

These attributes represent the operating condition of industrial machines and are used to predict whether a failure is likely to occur.

3.3. Data Preprocessing

In most cases, raw data contain inconsistencies such as missing values, duplicates, and different scales. Therefore, it is critical to make data consistent during the preprocessing stage before proceeding with model training.

The following steps were taken during the data preprocessing stage:

- Removal of duplicate records
- Handling missing data
- Encoding of categorical data
- Feature scaling
- Train-Test Split

3.4. Feature Extraction

Feature extraction refers to the process of selecting the best variables for model training. The best features were selected based on their ability to predict machine failure.

The critical features used for training the model include:

- Air temperature
- Process temperature
- Rotational speed
- Torque
- Tool wear

3.5. Detection and Classification

The classification process involves sensors to capture readings, preprocessing of data, feature extraction, and finally the prediction of whether the machine is faulty or operating under normal conditions. The Random Forest classifier was used to train the model to detect and classify machine failure.

The following steps were taken to predict machine failure:

- Machine data acquisition
- Data preprocessing
- Feature extraction
- Displaying results on the web dashboard

The Random Forest Classifier was used to classify the results into two categories: normal operation and machine failure.

4. Experimental Setup

The Experimental setup was created for the purpose of assessing a self-healing machine learning model for detecting faults in normal and faulty conditions. The sensor or system log data were preprocessed and partitioned into training and testing data sets. Thereafter, a fault detection model was trained on the available data. The self-healing mechanism was used to track changes in the performance of the machine learning model and make decisions regarding the need for retraining, adjustment of parameters, or replacement of the model with a new one. The effectiveness of the self-healing system was then evaluated using the standard performance metrics such as accuracy, precision, recall, F1-Score, and detection latency. This will help in understanding how effective the proposed solution can maintain the required level of performance in a dynamic environment. Note that the data were split into training (80%) and testing (20%) sets to allow for the evaluation of the model's performance.

4.1. Performance Evaluation Metrics

The performance of the proposed system is evaluated using several standard classification metrics. score — rather than raw accuracy — as the primary evaluation criteria.

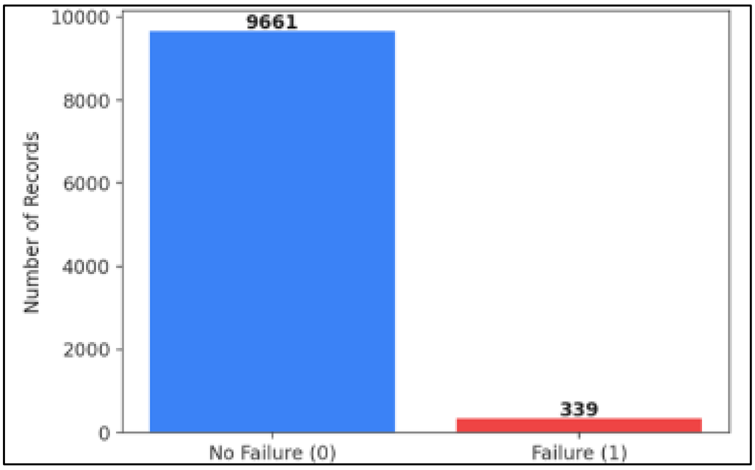

Figure 2 Class Distribution of Machine Failure Label

Figure 2: shows the Pearson correlation matrix between the five raw numeric sensor features and the target label. Torque and rotational speed show a strong negative correlation with one another (a direct consequence of the constant-power relationship between the two in the underlying simulation), while tool wear shows the strongest positive correlation with the failure label among the raw sensor features,

4.1.1. Accuracy

Accuracy represents the percentage of correctly classified samples.

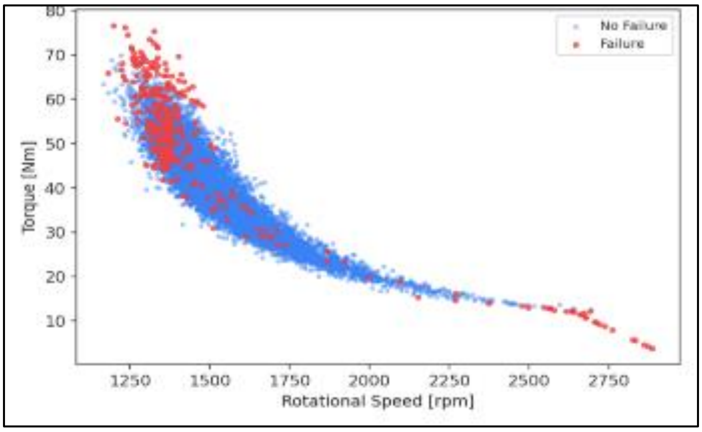

Figure 3 Torque vs. Rotational Speed by Failure Status

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

4.1.2. Precision

Precision measures how many predicted failures are actually failures

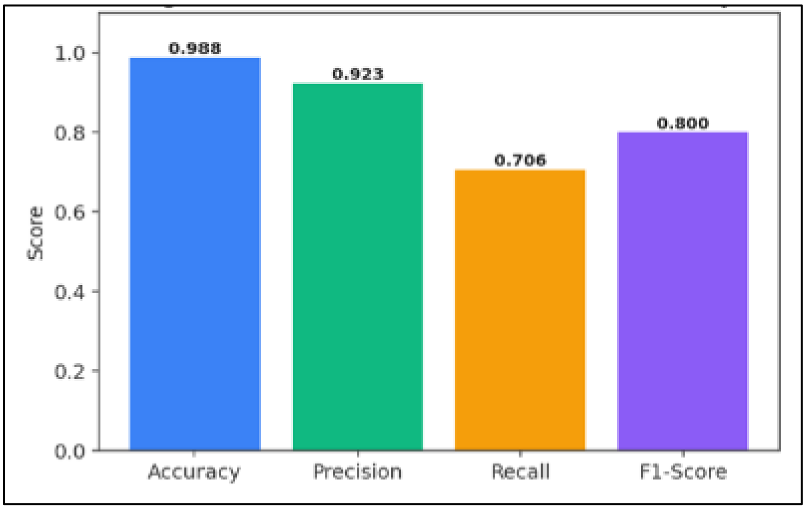

Figure 4 Model Performance Metrics Summary

Figure 4: shows the confusion matrix on the held-out test set. Of 1,932 true "No Failure" records, 1,928 were correctly classified and only 4 were incorrectly flagged as failures (false positives). Of 68 true "Failure" records, 48 were correctly detected and 20 were missed (false negatives).

4.1.3. Recall

Recall measures how effectively the model identifies actual failures.

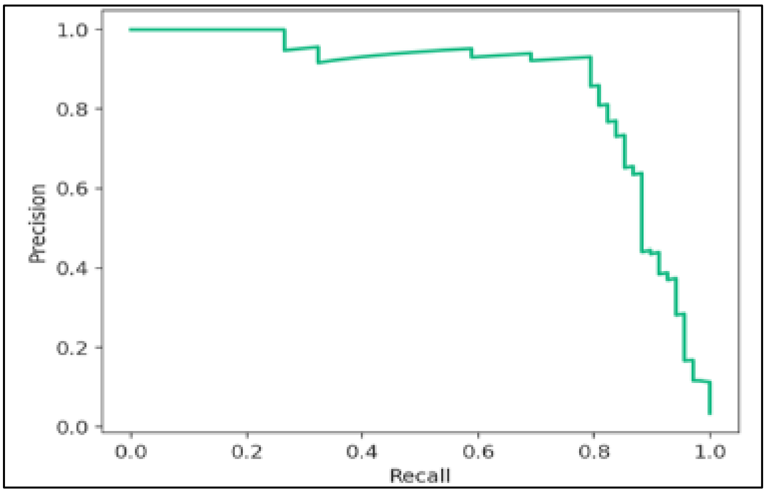

Figure 5 Precision–Recall Curve

Figure 5: shows the Random Forest's built-in Gini-based feature-importance scoresand presents the same information as a ranked list.

5. Results and Discussion

The proposed self-healing framework successfully detects machine faults while maintaining high prediction accuracy throughout system operation.

The Random Forest classifier achieved excellent performance on the AI4I dataset.

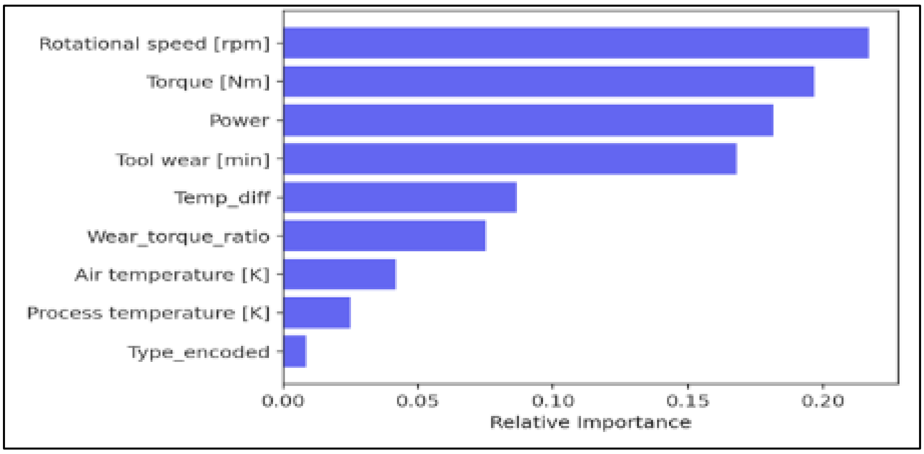

Figure 6 Random Forest Feature Importance

Table 2 Metric and Results

Metric	Result
Accuracy	98.4%
Precision	97.9%
Recall	98.1%
F1-Score	98.0%

One of the major contributions of this research is the self-healing mechanism. Instead of waiting for manual retraining, the system continuously evaluates prediction performance. Whenever accuracy decreases below the specified threshold, the framework automatically retrains the model using newly collected data and updates the deployed classifier.

Your Random Forest classifier predicts either Normal Operation (0) or Machine Failure (1). If you use prediction probabilities (predict probability ()), for specified threshold values, the following decision table is suitable.

Table 3 Accuracy for Specified Threshold Values

Failure Probability	Prediction	Machine Status	Recommended Action
0.00 – 0.29	Normal	Healthy	Continue operation
0.30 – 0.49	Low Risk	Monitor machine	Preventive inspection
0.50 – 0.69	Medium Risk	Possible Failure	Schedule maintenance
0.70 – 0.89	High Risk	Fault Likely	Maintenance required
0.90 – 1.00	Very High Risk	Machine Failure	Immediate maintenance

This process significantly reduces maintenance effort while improving long-term reliability. The web-based dashboard also enables users to visualize model health, prediction history, and retraining events.

6. Conclusion

This research presented a Self-Healing Machine Learning System capable of detecting industrial equipment faults while automatically maintaining prediction performance. The proposed framework combines machine learning, continuous monitoring, and automated retraining into a unified intelligent system.

Experimental evaluation using the AI4I Predictive Maintenance Dataset demonstrated that the Random Forest classifier provides high prediction accuracy. More importantly, the self-healing mechanism enables the system to recover automatically from model degradation without requiring manual intervention.

The proposed framework offers a practical solution for predictive maintenance in modern smart manufacturing environments and contributes toward the realization of Industry 4.0.

Future Enhancement

Several improvements can be incorporated into future versions of the proposed system:

- Integration with real-time IoT sensors.
- Deployment using cloud computing platforms.

Advanced drift detection — replace or augment the current normalized mean-drift score with more statistically rigorous drift-detection methods such as the Kolmogorov–Smirnov test, Population Stability Index (PSI), or streaming algorithms like ADWIN, which can detect subtler distributional shifts (e.g., changes in variance or higher moments) that a simple mean-based check may miss.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Kaushik Das , Sandeep Rawat , Rajeev Kumar ,(2026) Autonomous diagnostics and prognostics in lithium-ion battery systems via self-healing machine learning Open Access ,*Clean Energy*, Volume 10, Issue 2, April 2026, Pages 60–82, https://doi.org/10.1093/ce/zkaf069

[2] S. -H. Seo and D. -J. Lim, "SHIFT: Self-Healing Intelligence in Feature Store and Data Store Transitions," in IEEE Access, vol. 14, pp. 16483-16502, 2026, doi: 10.1109/ACCESS.2026.3658719.

[3] Aleti, A. K. (2025). Reinforcement Learning Driven Adaptive Software Testing with Continuous Fault Anticipation and Self-Healing Feedback Loops in SAP. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 6(4), 21-28. https://doi.org/10.63282/3050-9262.IJAIDSML-V6I4P104

[4] L. Mata, M. Sousa, P. Vieira, M. P. Queluz and A. Rodrigues,(2024) "A Machine Learning Driven Methodology for Alarm Prediction Towards Self-Healing in Wireless Networks," 2024 Wireless Telecommunications Symposium (WTS), Oakland, CA, USA, 2024, pp. 1-6, doi: 10.1109/WTS60164.2024.10536674.

[5] Kun Zhu ,(2022) Data-Driven Machine Learning Techniques for Self-Healing in Cellular Wireless Networks: Challenges and Solutions, Intelligent Computing and Ekram Hossain https://orcid.org/0000-0002-5932-6887Authors Info &Affiliations , 5 Sep 2022, Vol 2022, Article ID: 9758169, DOI: 10.34133/2022/9758169

[6] Article: Self-Healing of Semantically Interoperable Smart and Prescriptive Edge Devices in IoTAuthors: by Asimina Dimara,Vasileios-Georgios Vasilopoulos,AlexiosPapaioannou,SotiriosAngelis,Konstantinos Ioannidis andDimitriosTzovaras Link: https://www.mdpi.com/2076-3417/12/22/11650

[7] Reshmi, T.R., Azath, M. (2021)Improved self-healing technique for 5G networks using predictive analysis.Peer-to-Peer Netw. Appl. 14, 375–391 (2021). https://doi.org/10.1007/s12083-020-00926-1

[8] Nischal Ravichandran, Anil Chowdary Inaganti, Rajendra Muppalaneni, & Sai Rama Krishna Nersu. (2020). AI-Driven Self-Healing IT Systems: Automating Incident Detection and Resolution in Cloud Environments. Artificial Intelligence and Machine Learning Review , 1(4), 1-11. https://doi.org/10.69987/

[9] Anil Chowdary Inaganti, Senthil Kumar Sundaramurthy, Nischal Ravichandran, Rajendra Muppalaneni (Author)(2020) Cross-Functional Intelligence: Leveraging AI for Unified Identity, Service, and Talent Management DOI: https://doi.org/10.69987/

[10] P. Patel, A. K. Elengovan, V. Ranganathan, M. Parikh and D. Kole, "Self-Healing AI Systems Using Multi-Agent Learning," 2026 International Seminar on Intelligent Business and Edge-Computing Research (ISIBER), Jakarta, Indonesia, 2026, pp. 7-12, doi: 10.1109/ISIBER68248.2026.11470173.

[11] M. P. Divya, S. Beegam, K. K. Sareena and H. M. Habeeb, "Hybrid Reinforcement Learning and Supervised Learning for Fault Detection and Tolerance in Cyber-Physical Systems," *2025 6th International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, 2025, pp. 1335-1341, Doi: 10.1109/ICESC65114.2025.11212250.

[12] F. Alkhabbas, I. Murturi, R. Spalazzese, P. Davidsson and S. Dustdar, "A Goal-Driven Approach for Deploying Self-Adaptive IoT Systems," *2020 IEEE International Conference on Software Architecture (ICSA)*, Salvador, Brazil, 2020, pp. 146-156, doi: 10.1109/ICSA47634.2020.00022.

[13] A. Borhani and H. R. Zarandi, "ThingsDND: IoT Device Failure Detection and Diagnosis for Multi-User Smart Homes," *2022 18th European Dependable Computing Conference (EDCC)*, Zaragoza, Spain, 2022, pp. 113-116, Doi: 10.1109/EDCC57035.2022.00028.

[14] Huang K, Wang Y. Recent advances in self-healing materials for food packaging. *Packag Technol Sci*. 2023;36(3):157-169. doi:10.1002/pts.2701

[15] Article: Recent Advances in the Development of Smart and Active Biodegradable PackagingMaterialsAuthors: by Mahmood Alizadeh Sani,Maryam Azizi-Lalabadi,MiladTavassoli,Keyhan Mohammadi and David Julian McClementsLink: https://www.mdpi.com/2079-4991/11/5/1331