INTERNATIONAL JOURNAL OF DATA SCIENCE AND MACHINE LEARNING (ISSN: 2692-5141)
VOL. 06 ISSUE 02 (2026)

AMERICAN ACADEMIC PUBLISHER

Research Article

A Survey on Machine Learning Applications in Storage Quality Monitoring and Fault Detection

Dr. Parth Gautam

Assistant Professor Department of Computer Sciences and Applications Mandsaur University, Mandsaur

check for updates

Received: 24 October 2026
Revised: 30 October 2026
Accepted: 07 November 2026
Published: 18 November 2026

DOI: 10.55640/ijdsml-06-02-17
Page No: 282-291

Copyright: © 2026 Authors retain the copyright of their manuscripts, and all Open Access articles are disseminated under the terms of the Creative Commons Attribution License 4.0 (CC-BY), which licenses unrestricted use, distribution, and reproduction in any medium, provided that the original work is appropriately cited.

Abstract

Fault detection and diagnosis (FDD) techniques, like hardware redundancy and threshold-based monitoring, are becoming increasingly ineffective in detecting early-stage faults and ensuring operational reliability in increasingly complex and interconnected mechanical, electrochemical, chemical, thermal, and electrical energy storage systems. As a result, ML and AI methods are becoming more popular because to their strong capacity to model nonlinear, high-dimensional, and noisy data settings. ML-based solutions, including supervised, unsupervised, and deep learning methods, have demonstrated superior fault detection accuracy and adaptability across various applications from wind turbine systems and gearboxes to magnetic sensor arrays in nuclear fusion reactors. Recent trends in distributed ML architectures enable real-time, in-sensor diagnostics, further enhancing responsiveness and system resilience. However, practical implementation faces challenges such as sensor data inconsistencies, scarcity of labelled fault data, real-world data noise, multi-source data integration, and the gap between lab research and field deployment. The paper delves into the various types of systems, diagnostic methods, and failure stages in energy storage and mechanical systems that use ML-driven FDD. It also addresses important implementation issues and potential future research paths.

Keywords: Machine Learning, Fault Detection, Predictive Maintenance, Storage Quality Monitoring, Deep Learning, Adaptive Models, Industrial Systems.

INTRODUCTION

Modern energy storage systems, mechanical, electrochemical, thermal, chemical, and electrical, require a plethora of intricate procedures to ensure storage quality and reliable operation [1]. Traditional fault detection and diagnosis (FDD) techniques, such as hardware redundancy and threshold-based limit checking, often fall short in these dynamic and data-rich environments. In the hardware redundancy approach, multiple sensors measure the same parameter, and discrepancies are identified through a voting scheme; however, this method increases cost and complexity, and is not easily applicable to all types of storage systems [2][3]. On the other hand, limit checking involves monitoring whether a parameter exceeds predefined thresholds, but it can only detect faults when they become severe enough to cross those limits, making it ineffective in identifying early-stage or incipient faults. Such limitations call for more adaptive and intelligent solutions capable of working under diverse and challenging conditions.

A robust and flexible solution to these problems, Machine Learning (ML) involves training models with both historical and real-time data in order to accurately detect outliers and forecast failures [4]. ML algorithms are particularly well-suited to handle high-dimensional, nonlinear data with missing values characteristics commonly found in complex energy systems. Modern ML techniques, such as multi-mode Principal Component Analysis (PCA) and Gaussian Mixture Models (GMM), have been shown to be significantly more effective than traditional clustering algorithms when used to sewage heat pump systems. These models are capable of identifying subtle deviations from normal operating patterns, which would be undetectable using conventional methods. ML has also shown its potential in a broad range

of environmental and engineering domains, including crop disease detection, product development, bias correction, and code optimization, underscoring its adaptability and effectiveness.

Distributed ML-based fault detection and diagnostic systems have started to gain popularity as a way to make fault detection even more efficient and rapid. In this approach, the sensor nodes themselves are equipped with computational capabilities that allow them to execute ML algorithms immediately after sensing the data, rather than transmitting it to a centralized location [5]. This real-time, in-sensor analysis reduces communication latency, enhances scalability [6], and supports faster decision-making, which is critical in applications requiring immediate action to prevent damage or downtime. Moreover, this distributed model reduces dependency on network infrastructure, making the system more robust and fault-tolerant. As energy storage systems become more interconnected and data-intensive, the integration of ML especially through distributed architectures represents a significant advancement in ensuring the long-term reliability, quality, and intelligence of storage operations.

A. Structure of the Paper

The following is the outline of the paper: Section II overviews storage systems. Section III explores machine learning applications for fault diagnosis. Section IV discusses challenges in fault detection. A literature assessment is provided in Section V, and important findings and ideas for future study are presented in Section VI.

II. OVERVIEW OF STORAGE SYSTEMS

System types for storing energy can be mechanical, thermochemical, chemical, electrical, or thermal, depending on the type of energy they store. There are a number of common ways to classify them, including by response time, function, or storage duration. Based on the type of energy, Figure 1 displays a hierarchy of energy storage systems. Energy storage systems are becoming more important as power networks integrate renewable energy sources more and more. Several studies have looked at the environmental effects of energy storage in transportation [7].

A. Classification of Energy Storage

The most appropriate way to classify energy storage devices is into five categories: Hydrogen, various forms of sensible and latent heat, pumped hydro and flywheels, flow batteries and batteries in general, mechanical, electrochemical, chemical, and electrical phenomena are all examples. The chemical or physical processes that each kind employs to store and transform energy is what differentiates them.

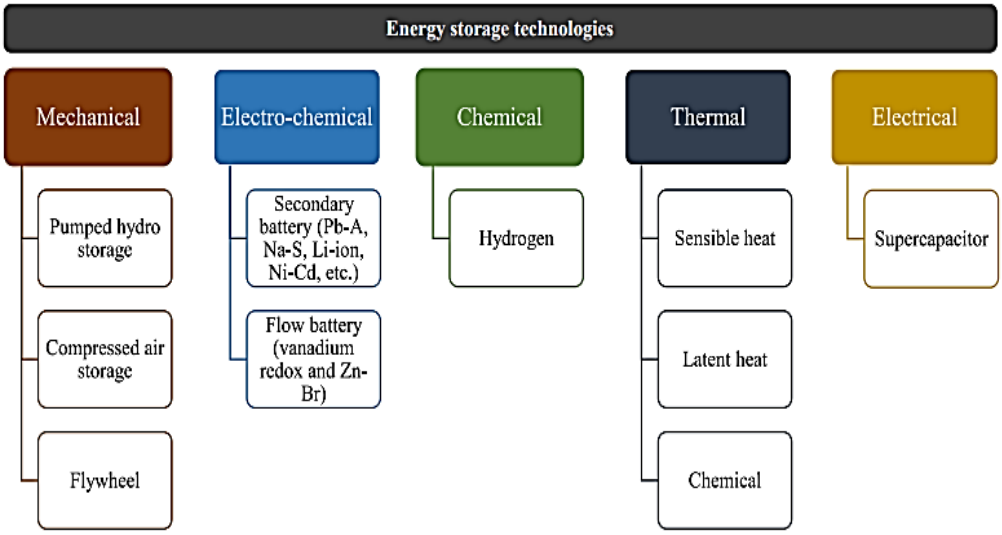


Fig. 1. Classification of Energy Storage Technology

Mechanical, electrochemical, chemical, thermal, and electrical energy storage are the five primary forms. Some mechanical storage systems that use the aforementioned technique to transform electrical energy into mechanical energy are flywheels, compressed air storage, and pumped hydro storage. Figure 1 shows that one form of electrochemical storage technology takes advantage of reversible chemical reactions to store energy: rechargeable batteries. Sodium-sulfur, lead acid, lithium ion, nickel cadmium, and zinc bromine are some of the most prevalent types of flow batteries. Hydrogen and other gases with good combustion characteristics and high energy densities are part of the chemical storage category. Chemical thermal storage, sensible heat, and latent heat are the three main methods for storing thermal energy. Sensible heat is stored whenever a material's temperature changes. Latent heat can be released through phase shifts. Supercapacitors are the last form of electrical storage, it can store energy as an electric field, have a high-power density, and can be charged and discharged quickly.

1) Mechanical storage:

Modern mechanical storage systems have gone a long way, yet 94% of the world's energy is still stored in pumped

hydro storage (PHS) systems. Worldwide, 322 PHS projects with a total capacity of 164.63 GW were operating in 2017. Power generating stations (PHS) with indicated capacity between 100 MW and 2000 MW. Its long cycle life, high efficiency, and duration of discharge make it ideal for use in applications requiring large amounts of energy.

2) Electrochemical Storage:

At 2.03 GW, electro-chemical battery storage systems rank third in terms of installed capacity. Lead acid, sodium sulphate, lithium ion, nickel cadmium, and flow batteries make up the bulk of utility-scale electrochemical batteries. It can see the market shares of several battery technologies here: Capacity estimates for sodium-based batteries are 204.32 MW, Li-ion batteries 1.66 GW, and flow batteries 71.94 MW.

3) Chemical Energy Storage (CES) System:

Chemical energy storage systems using CES technology are the best option for long-term storage. Atomic and molecular bond energy is analogous to the electricity discharged from a battery in a chemical reaction. The formation of new chemical bonds and the breaking of old ones alter the chemical makeup of materials during energy release. Chemical fuels have surpassed all others as the go-to for transportation and power generation on a global scale. Chemical fuels comprise a variety of substances, including coal, diesel, gasoline, propane, butane, ethanol, and hydrogen [8]. Chemical reactions convert molecular potential into mechanical energy, which is a precursor to electrical energy and the first step in producing electricity.

4) Hydrogen and thermal storage:

The United States Department of Energy reports that one of Germany's power plants has the highest production of any power plant in the world, at 6 MW. Deploying proton exchange membrane (PEM) electrolysis, the ground-breaking multi-MW Energeia Mainz project had a discharge length of four hours and twenty minutes. One nineteen million dollars went into the project's capital. Problems with conversion efficiency and a large initial investment prevent this technology from being used widely in commercial-scale grid applications. Various energy storage systems can have vastly different self-discharge, cycle time, specific energy, and efficiency. The degree to which a storage device meets certain demands is influenced by these features.

5) Electrical Energy Storage (EES) System:

Energy conversion is not a part of electric energy storage (EES) devices; instead, it uses electric fields to store electrical energy [9]. There are two primary types of EES, and these are magnetic and electrostatic [10]. Electrostatic energy storage devices come in two varieties: capacitors and supercapacitors. A SMES, which stands for superconducting magnetic energy storage.

B. Fault Detection in Machine Learning

Machines' increasingly autonomous and self-optimized operations, along with the deluge of real-time data gathered, make issue identification and diagnosis an essential but challenging task. At present, functional tests conducted using diagnostic software use real-world manufacturing data, which leads to inaccurate diagnoses and lengthy repair times (days or weeks) because technicians need to physically probe and debug the system multiple times to get to the bottom of problems:

1) Supervised Learning-Based Solutions:

One supervised learning project aimed to automate the process of diagnosing industrial faults by building a circuit-board diagnostic tool that would use ML methods to reduce human error and increase diagnostic accuracy [11]. An intelligent detection and diagnosis system used three ML classification methods: weighted-majority voting (WMV), artificial neural networks (ANNs), and SVM. The best features of ANNs and SVMs were combined by WMV. In order to train the algorithms, use repair history data. To teach the classification models, use fault situations obtained from failure records and repair operations that followed. Unbalanced data and concept drifts are challenges in using supervised ML for fault detection tests. Concept drifts refer to changes in fault patterns caused by aged machinery or repairs/replacements:

2) Unsupervised Learning-Based Solutions:

Problem identification solutions using unsupervised learning algorithms have been presented in a number of publications [12]. By creating an intelligent fault detection technique, the Reference writers hoped to reduce the workload of manually extracting problem information using prior knowledge and diagnostic abilities. The outcome was a two-stage NN-based unsupervised learning method. In the first stage, features were extracted from the mechanical vibration signals by sparse filtering. In the subsequent phase, softmax regression was used to automatically categorize the machines' health statuses. The suggested method was validated in two scenarios: motor bearing fault detection and locomotive bearing fault detection, and it was found that the accuracy performance improved with increasing volumes of unlabelled data.

3) Deep Learning-Based Solutions:

DL-based solutions have been developed to enhance fault detection performance and overcome the limitations of standard ML algorithms. DL uses raw data and several layers, with each layer being simplistic and non-linear, to automatically generate new data representations and deliver precise detection and classification. LiftingNet, a DL solution, was introduced in the reference. Without the need for pre-defined features, LiftingNet can be trained adaptively using mechanical data sets and carry out classification tasks [13]. Early fault identification and how to deal with signal feature changes were among the topics covered.

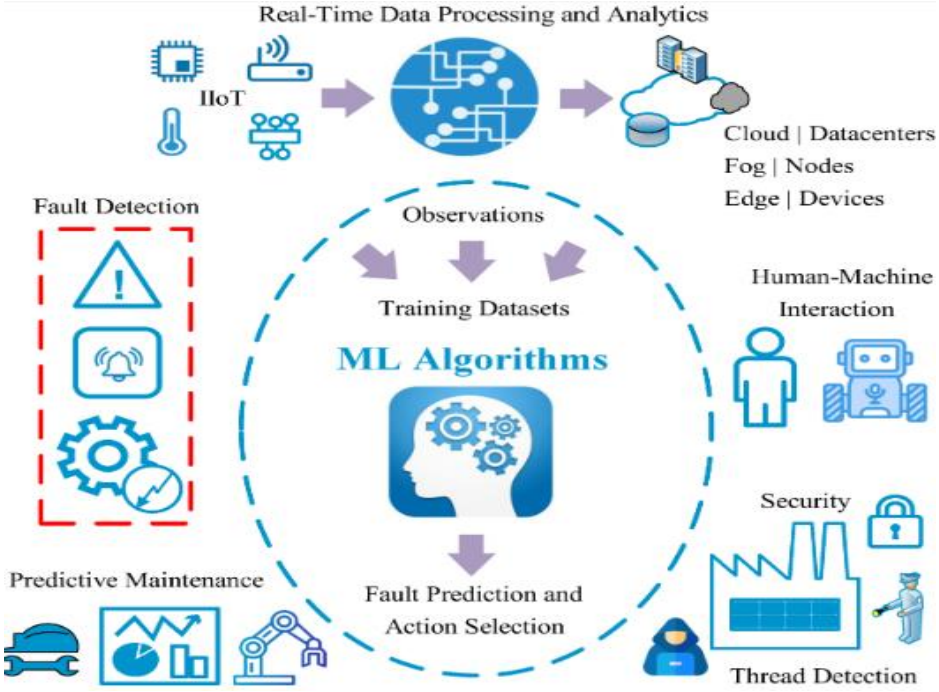


Fig. 2. AI/ML Fault Detection Solutions

Figure 2 displays a number of examples of AI/ML technologies that are used for defect detection, prediction, and prevention [14]. All of these topics are covered in depth in the sections that follow, with an emphasis on the effects of ML-based solutions.

C. Machine-Learning Based Fault Detection

This section provides the general overview of fault detection discussed below:

1) Generating the Fault Signature Database:

The ML models cannot be trained without a defect signature database. Data from the Tokamak Aditya's sensors were used in its design. A complete plasma discharge, which includes the starting, ramping up, flat-top, ramping down, and ending phases, requires 3,500 samples (readings) [15]. thought the sensor values were correct and consistent with predictions. Following this, the initial database was enhanced with labels that indicated that every one of the 3,500 samples had been adjusted to account for stuck-at-zero problem.

The following are the four types of sensor data that are generated when Tokamak's magnetic position sensors experience faults:

- No Faults: This is the raw sensor data that matches the predicted values for the measurements.
- Stuck-at-Zero fault: In order to create this kind of error, all the sensor values are set to zero.
- Offset Fault: All of the readings from the sensors are multiplied by a constant.
- Noise fault: The sensor values are supplemented with a zero-mean Gaussian noise that is created within the range of -π to +ó. The bin sizes for the Gaussian noise were determined using the number of samples of the measurement signal length.

2) ML-based Fault Classifier:

A model learning defect classifier is depicted in Figure 3 as a block diagram. First, compile a list of all known fault signs; second, train a model to identify faults using machine learning. This will form the basis of the fault detection system. It is possible to categorize sensor errors in real-time after the ML model has been trained and implemented in hardware.

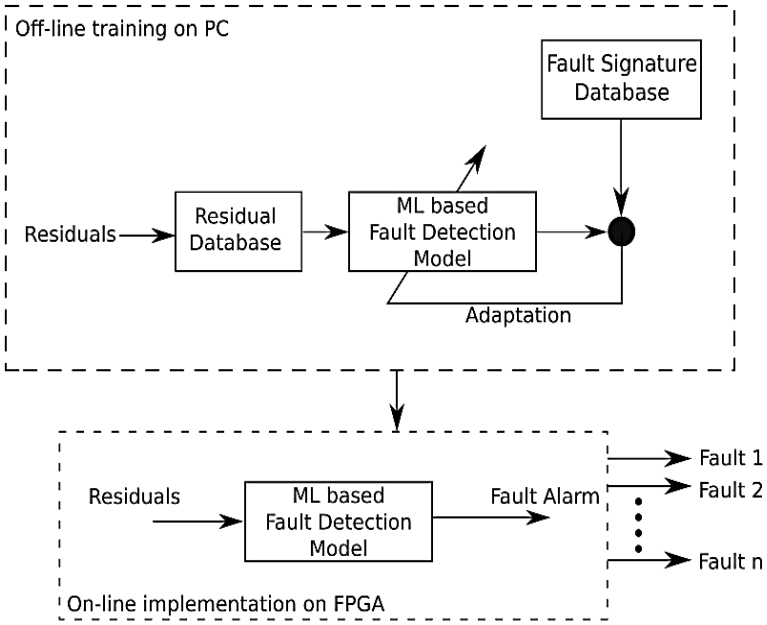


Fig. 3. Fault Identification and Classification Scheme

- The resulting residuals for each sensor are input into a Residual database, which takes into account the Tokamak's physics model and the sensor configurations.
- The ML model is trained to classify the residuals using a limited number of offline data representing expected fault categories.
- Online fault classification using the trained ML model is possible using residual data.

3) Simulation Tools:

Using three separate Python libraries, Pybrain, Theano [15], and Scikit-Learn the defect detection model has been developed. Scikit-Learn provides the greatest leeway to the user among these three ML libraries.

III. MACHINE LEARNING APPLICATIONS FOR FAULT DIAGNOSIS

Faults and malfunctions are inevitable parts of any physical system's lifespan and can disrupt its regular performance. Because of this, in order to keep production running smoothly, a plant needs a system to detect defects [16]. A fault diagnosis system is able to detect potential malfunctions inside the monitored system.

A. Applications for fault detection.

This section only covers a subset of common errors because the WT system is so complicated.

1) Fault Diagnosis Applied to the Gearbox:

A SVM incorporating a clustering binary tree and a diagonal spectrum was suggested for application in defect diagnosis. The technique described in this paper outperformed generic SVMs and radial basis function (RBF) neural networks in terms of diagnostic accuracy while using a smaller sample size of defects. A crude neighborhood set was used to decrease the sample attributes [17]. To prevent the loss of crucial traits, the rough neighborhood set used a forward search strategy to add relevant attributes to the reduced attribute set. The proposed method aimed to improve defect identification accuracy while decreasing computational time in comparison to SVM without rough neighborhood set.

2) Fault Diagnosis Applied to WT Bearing.

A regression SVM model was used to establish the WT output power [18]. The generator's output power drops down steadily as the fault level rises in a WT failure, and the difference between the two sets of numbers eventually above the predetermined threshold. As a result, a persistent shift in the residual's trend can reveal impending failure. Diagnosing faults in low-speed rotating bearings became much easier with the use of an alpha stable distribution (ASD).

3) Fault Diagnosis Applied to the Rotational Imbalance

A method for diagnosing direct-drive wind turbine malfunctions using SVM and feature selection was presented. The results of the experiments demonstrated the effectiveness of the proposed method for issue detection. Utilized principal component analysis (PCA) to select representative features for training the proximal support vector machine (PSVM).

4) Fault Diagnosis Applied in the Misalignment

The feature vector was derived from the signal's demodulation and reconstruction through the use of DTCWT. The problem of misalignment in WT was tackled by implementing a technique that fuses diverse information. Using t-distributed stochastic neighbor embedding (t-SNE), dimensionality reduction processing was performed after creating

the feature vectors by integrating the fault characteristics of the multi-source signals.

B. Fault Detection and Diagnosis Stages.

Audio emissions (AE), vibration, pressure, oil analysis, and temperature analysis are some of the numerous monitoring tools used in the four stages of fault detection and diagnosis. problems identification [19], fault identification, fault severity evaluation, fault progression prediction, and remaining useful life estimation.

1) Fault Detection (Stage 1):

The process of examining machine components for problems is known as fault detection [20]. The majority of fault detection investigations, according to recent studies, aim to predict future failures in order to go to the FDD phase. Early defect detection is essential for preventing catastrophic failure. Learn the ins and outs of identifying potential issues with moving equipment components in this section.

2) Fault Identification (Stage 2):

The aim of fault identification is to locate and assess mechanical component issues. In order to diagnose bearing components, it is common practice to inspect the outer race, inner race, and rolling element for signs of wear [21]. Gear components undergo defect type identification, which includes types like chipped gear teeth. The majority of DL models have found significant use in stage 2.

3) Fault Severity Assessment (Stage 3):

The process of determining the seriousness of a mechanical component's failure is called fault severity evaluation [22]. They have put out a possible approach to determining the extent of bearing component faults by combining signal processing with an ML model [23].

4) Fault Growth and Remaining Useful Life Prediction (Stage 4):

One method for estimating how long a piece of machinery last is the remaining usable life (RUL) methodology. Simultaneously, the fault's size following a certain cycle is utilized to foretell its future growth. The relevance of a Stage 4 analysis in an FDD system can be determined by considering the output RUL forecast in relation to the amount and time of fault progression. examination of the possible consequences through the use of a DL model.

IV. CHALLENGES OF FAULT DETECTION

A massive quantity of data is necessary to implement sophisticated learning techniques for MFD. Unfortunately, the performance of the models used is limited by the constraints of the datasets that are currently accessible. Using machinery fault datasets presents researchers with the following challenges:

A. Smart Sensors:

Energy efficiency, network traffic, and privacy and security are some of the problems with sensor networks. Protecting data both internally and while stored in the cloud is of the highest significance from a security and privacy standpoint [24][25]. Congestion and data loss can occur when numerous sensors transmit data at once, which adds to the difficulty of managing network traffic [26].

B. Multi-Sensor Data Fusion:

Integrating data from heterogeneous and homogeneous sensors, effectively analyzing data from varied sources, and dealing with uncertainty and imprecision are all challenges in multi-sensor data fusion. To keep real-time performance in dispersed systems, data synchronization is crucial.

C. Challenges with Real-World Data

Fault analysis is made more difficult by the fact that data collected from real-world machinery is frequently noisy, inconsistent, non-stationary, and partial, and it comes from a variety of different sensors. This could happen if there are problems with the sensors, the data transmission, or any interruptions in the process of data collecting [27]. The data quality difference between controlled laboratory conditions and real-world scenarios poses significant challenges for real-world applications [28][29].

D. Faulty Data Scarcity

The lack of real-world data on broken machinery makes it difficult to train sophisticated ML models [30]. The regular operating conditions of machinery make it difficult to collect enough defective data for fault identification.

E. Insufficient Labeled Data:

Labelled data is required for supervised learning, which is the main method used in most works. There is a dearth of labelled datasets in part because accurately identifying problem kinds and severity levels is difficult. Due to the time and effort required for labelling and the fact that it is not always diverse enough to guarantee effective generalization to invisible defects, the scarcity of labelled data poses a serious problem in practical industrial settings.

F. Data Incompatibility

Machine failure detection becomes more challenging when dealing with data from various sources, each of which has its own objectives, level of complexity, and data treatment needs. Problems with modelling machine failure detection are also caused by differences in the depth of data storage within databases.

G. Transferring Knowledge from Research to Practice

Datasets collected in controlled laboratory settings and made publicly available are the backbone of the majority of research. One eventual goal is to use this information to forecast actual machinery failures using data gathered in a lab, as well as to detect invisible machinery flaws. But several technological hurdles must be surmounted first.

V. LITERATURE REVIEW

This literature review explores into various ML and IoT applications in predictive maintenance, fault detection, health diagnostics, water and food quality monitoring, and other industrial and healthcare domains. Improved system reliability, data-driven decision-making, and real-time monitoring are the main points.

Leija et al. (2025) study set out to evaluate various machine learning approaches to AMS defect diagnosis and draw comparisons between them. Discovering the optimal model for classifying equipment failures based on historical data is the primary objective. ANN, DT, boosting, RF, KNN, stacking, and SVM were among the many methods that were selected. In the first step of the investigation, data was collected using an Arduino-based system that tracked the operational status of the equipment. This system had sensors that measured electrical current, temperature, vibration, differential pressure, and sound. The failures of the motor, electrical valve, and pneumatic cylinder were intentionally brought about [31].

Akyaz and Engın (2024), A DL method for predictive modelling and four additional ML techniques were examined in this research. By analyzing past data, the algorithms were able to learn to correlate certain operational characteristics with certain failure occurrences. Data collection from industrial equipment, preprocessing, feature selection, training of models, and deployment are all steps in the process. Extensive experimental and case study validation in an industrial context confirms the efficacy of the suggested system [32].

Rathore et al. (2024) method not only makes maintenance procedures more efficient, but it also provides substantial savings and safety improvements. The research addresses practical challenges in deploying ML solutions, including data acquisition, model accuracy, & integration into existing maintenance systems. The efficacy of methods like decision trees, neural networks, and clustering algorithms in identifying outliers, foretelling failures, and suggesting maintenance regimens is investigated [33]

Sathish et al. (2023) propose a method for monitoring water quality that utilizes the SVM algorithm and the Raspberry Pi. Several sensors are used by the system to measure water quality indicators, including temperature, pH, total dissolved solids, and turbidity. The sensor data is processed using the SVM algorithm, which then sorts the water quality into 3 levels: drinkable, mildly contaminated, and highly contaminated. Water quality can be tracked in real-time by the suggested system, which notify the user if it drops below the safe threshold. The system is simple to set up, doesn't break the bank, and can be utilized even in outlying locations where finding reliable water quality monitors is a challenge. The outcomes prove that the suggested approach is effective in keeping track on water quality measures and giving out live notifications when problems arise [34].

Tutul et al. (2023) results from testing the system with a variety of foodstuffs demonstrated that it accurately anticipated when food fresh. The food sector might greatly benefit from the suggested system's ability to cut down on food wastage while simultaneously increasing customer access to nutritious, freshly prepared foods. Home and kitchen use cases benefit greatly from the system's real-time monitoring and notifications. Lastly, the proposed smart food monitoring system utilizing the internet of things and ML offers a dependable and cost-effective method of tracking the freshness of consumed food [35].

Parbat et al. (2022), Vital organs in system are constantly watched over, data is sent to doctors in the cloud and patients are warned of possible risks. An Internet of Things (IoT) device has been created that links several sensors to a microcomputer and sends the data to a cloud server for the Modified Stochastic Gradient Descent (SGD) Algorithm, which uses DL. This suggested method should work because a set of rules has been made to deal with and, to some extent, solve the problems listed above [36].

Table I summarizes recent studies applying ML and IoT for predictive maintenance, health, food, and water quality monitoring. These approaches improve real-time decision-making but face challenges in scalability, integration, sensor reliability, and broader applicability.

TABLE I. SUMMARY OF LITERATURE REVIEW BASED ON STORAGE QUALITY MONITORING AND FAULT DETECTION.

Reference	Focus On	Approach	Key Findings	Challenges	Limitations/Gaps
Leija et al. (2025),	Fault diagnosis in AMS	ML models (ANN, SVM, DT, RF, KNN, stacking) + Arduino sensors	ML models classify equipment failures effectively	Data acquisition & induced faults	Limited to specific industrial machines
Akyaz and Engin (2024)	Predictive maintenance in industrial machines	ML (4 algorithms) + DL on Thing Speak IoT; data-driven framework	Real-time health monitoring, improved maintenance, validated via case studies	Integration into existing industrial setups	Generalizability to diverse machine types
Rathore et al. (2024)	Fault prediction and maintenance optimization	ML with decision trees, neural networks, clustering on sensor & historical data	Enables preemptive maintenance, enhances safety and cost efficiency	Model accuracy, system integration	Long-term validation under varying operational conditions
Sathish et al. (2023)	Water quality monitoring	SVM + Raspberry Pi with multi-sensors (pH, turbidity, etc.)	Real-time water classification and alerts, suitable for remote use	Sensor precision, environmental variations	Limited scalability and environmental robustness
Tutul et al. (2023)	Smart food freshness monitoring	IoT + ML system with sensors for freshness tracking	High accuracy, reduces food waste, enhances home/industrial use	Real-time updates and freshness categorization	Sensor reliability and lifespan in varied storage conditions
Parbat et al. (2022)	IoT-based health monitoring with cloud support	IoT + Modified SGD + DL; cloud doctor alerts	Real-time health alerts, proactive healthcare via cloud feedback	Secure transmission, cloud reliability	Requires further medical testing and standardization

VI. CONCLUSION AND FUTURE

The role of AI and ML in fault detection, predictive maintenance, and real-time monitoring across diverse domains, including industrial machinery, healthcare, water quality, and energy systems. Techniques such as supervised learning, unsupervised learning, and DL have demonstrated effectiveness in identifying faults, predicting failures, and enhancing operational efficiency. Studies have shown that integrating ML models with IoT platforms enables continuous monitoring, data-driven decision-making, and early fault diagnosis, reducing downtime and improving system reliability. Additionally, ML methods outperform traditional fault detection systems by handling high-dimensional, nonlinear, and incomplete datasets. Despite promising results, challenges persist in real-world deployment, such as handling imbalanced data, managing concept drift, integrating with legacy systems, and ensuring model interpretability. Future work should focus on developing adaptive ML models that can learn from dynamic environments and unlabeled data, ensuring scalability and robustness. Moreover, federated and edge learning approaches may help address privacy and latency concerns in distributed systems. Further research is also needed to explore hybrid models combining statistical methods with AI techniques for improved accuracy and resilience. Standardizing evaluation benchmarks and datasets can enhance comparability and accelerate the adoption of AI/ML-based fault detection solutions in critical infrastructure systems.

REFERENCES

1. G. Bode, S. Thul, M. Baranski, and D. Müller, “Real-world application of machine-learning-based fault detection trained with experimental data,” Energy, vol. 198, p. 117323, 2020, doi: https://doi.org/10.1016/j.energy.2020.117323.
2. S. U. Jan, Y. D. Lee, and I. S. Koo, “A distributed sensor-fault detection and diagnosis framework using machine learning,” Inf. Sci. (Ny)., vol. 547, pp. 777–796, 2021, doi: https://doi.org/10.1016/j.ins.2020.08.068.
3. P. Jinka, “DATA WAREHOUSE OPTIMIZATION TECHNIQUES: A SURVEY OF QUERY PROCESSING, STORAGE, AND PERFORMANCE ENHANCEMENT METHODS,” World J. Adv. Eng. Technol. Sci., vol. 20, no. 2, pp. 216–226, Aug. 2026, doi: 10.30574/wjaets.2026.20.2.0408.
4. H. Ahmad, “Machine Learning Applications in Oceanography,” Aquat. Res., vol. 2, no. 3, pp. 161–169, 2019, doi: 10.3153/ar19014.
5. G. Hu, T. Zhou, and Q. Liu, “Data-Driven Machine Learning for Fault Detection and Diagnosis in Nuclear Power Plants: A Review,” 2021. doi: 10.3389/fenrg.2021.663296.
6. H. S. Chandu, “Enhancing Manufacturing Efficiency : Predictive Maintenance Models Utilizing IoT Sensor Data,” IJSART, vol. 10, no. 9, 2024.
7. J. F. Peters, M. Baumann, B. Zimmermann, J. Braun, and M. Weil, “The environmental impact of Li-Ion batteries and the role of key parameters--A review,” Renew. Sustain. Energy Rev., vol. 67, pp. 491–506, 2017.
8. K. P. Kampouris, V. Drosou, C. Karytsas, and M. Karagiorgas, “Energy storage systems review and case study in the residential sector,” IOP Conf. Ser. Earth Environ. Sci., vol. 410, no. 1, Jan. 2020, doi: 10.1088/1755-1315/410/1/012033.
9. J. Mitali, S. Dhinakaran, and A. A. Mohamad, “Energy storage systems: a review,” Energy Storage Sav., vol. 1, no. 3,

pp. 166–216, Sep. 2022, doi: 10.1016/j.enss.2022.07.002.

10. S. Garg, "AI-Driven Innovations in Storage Quality Assurance and Manufacturing Optimization," Int. J. Multidiscip. Res. Growth Eval., vol. 1, no. 1, pp. 143–147, 2020, doi: 10.54660/.IJMRGE.2020.1.1.143-147.

11. C.-C. Lin, D.-J. Deng, C.-H. Kuo, and L. Chen, "Concept Drift Detection and Adaption in Big Imbalance Industrial IoT Data Using an Ensemble Learning Method of Offline Classifiers," IEEE Access, vol. 7, pp. 56198–56207, 2019, doi: 10.1109/ACCESS.2019.2912631.

12. V. Rajavel, "Optimizing Semiconductor Testing: Leveraging Stuck-At Fault Models for Efficient Fault Coverage," Int. J. Latest Eng. Manag. Res., vol. 10, pp. 69–76, 2025, doi: 10.56581/IJLEMR.10.02.69-76.

13. B. Luo, H. Wang, H. Liu, B. Li, and F. Peng, "Early Fault Detection of Machine Tools Based on Deep Learning and Dynamic Identification," IEEE Trans. Ind. Electron., vol. 66, no. 1, Jan. 2019, doi: 10.1109/TIE.2018.2807414.

14. V. Prajapati, "Improving Fault Detection Accuracy in Semiconductor Manufacturing with Machine Learning Approaches," J. Glob. Res. Electron. Commun., vol. 1, no. 1, pp. 20–25, Jan, 2025, doi: https://doi.org/10.5281/zenodo.14935091.

15. D. Mohapatra, B. Subudhi, and R. Daniel, "Real-Time Sensor Fault Detection in Tokamak Using Different Machine Learning Algorithms," Fusion Eng. Des., vol. 151, Feb. 2020, doi: 10.1016/j.fusengdes.2019.111401.

16. G. Ciaburro, "Machine fault detection methods based on machine learning algorithms: A review," aimspress.com, vol. Volume 19, no. Issue 11: 11453-11490, 2022, doi: doi: 10.3934/mbe.2022534.

17. M. Heidari, H. Homaei, H. Golestanian, and A. Heidari, "Fault diagnosis of gearboxes using wavelet support vector machine, least square support vector machine and wavelet packet transform," J. Vibroengineering, vol. 18, no. 2, pp. 860–875, 2016.

18. Q. W. Gao, W. Y. Liu, B. P. Tang, and G. J. Li, "A novel wind turbine fault diagnosis method based on intergral extension load mean decomposition multiscale entropy and least squares support vector machine," Renew. energy, vol. 116, pp. 169–175, 2018.

19. X. Chen, S. Wang, B. Qiao, and Q. Chen, "Basic research on machinery fault diagnostics: Past, present, and future trends," Front. Mech. Eng., vol. 13, pp. 264–291, 2018.

20. S. Pandya, "Predictive Analytics in Smart Grids: Leveraging Machine Learning for Renewable Energy Sources," Int. J. Curr. Eng. Technol., vol. 11, no. 06, 2021.

21. M. Yuwono, Y. Qin, J. Zhou, Y. Guo, B. G. Celler, and S. W. Su, "Automatic bearing fault diagnosis using particle swarm clustering and Hidden Markov Model," Eng. Appl. Artif. Intell., vol. 47, pp. 88–100, 2016.

22. Kumar and R. Kumar, "Least square fitting for adaptive wavelet generation and automatic prediction of defect size in the bearing using Levenberg--Marquardt backpropagation," J. Nondestruct. Eval., vol. 36, no. 1, p. 7, 2017.

23. S. R. Thota, S. Arora, and S. Gupta, "Hybrid Machine Learning Models for Predictive Maintenance in Cloud-Based Infrastructure for SaaS Applications," 2024, pp. 1–6. doi: 10.1109/ICDSNS62112.2024.10691295.

24. S. Kabade and A. Sharma, "Intelligent Automation in Pension Service Purchases with AI and Cloud Integration for Operational Excellence," Int. J. Adv. Res. Sci. Commun. Technol., pp. 725–735, Dec. 2024, doi: 10.48175/IJARSCT-14100J.

25. S. Singamsetty, "AI-Enabled Data Stewardship Real-Time Alignment of Privacy and Storage Policies Across Global Systems using Deep CNN-RNN Techniques," in 2025 5th Asian Conference on Innovation in Technology (ASIANCON), PIMPRI, India: IEEE, Aug. 2025, pp. 1–6. doi: 10.1109/ASIANCON66527.2025.11281010.

26. R. Duggasani, "Scalable and Optimized Load Balancing in Cloud Systems: Intelligent Nature-Inspired Evolutionary Approach," Int. J. Innov. Sci. Res. Technol., vol. 10, no. 5, pp. 2153–2160, May 2025, doi: 10.38124/ijisrt/25may1290.

27. López, Á. Naranjo, S. Lu, and K. J. Moore, "Hidden Markov Model based Stochastic Resonance and its Application to Bearing Fault Diagnosis," J. Sound Vib., vol. 528, Jun. 2022, doi: 10.1016/j.jsv.2022.116890.

28. S. Singamsetty, "Efficacy of Data Governance a Cutting Edge Approach to Ensuring Data Quality in Machine Learning for Banking Industry," in 2024 2nd International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), IEEE, Dec. 2024, pp. 1–7. doi: 10.1109/SCOPES64467.2024.10991944.

29. R. Jariwala, A Guide to Construction Execution, Quality Systems, Turnover, Commissioning and Digital Project Delivery. Chennai, India: Notion Press, 2026.

30. N. Malali, "Predictive AI for Identifying Lapse Risk in Life Insurance Policies: Using Machine Learning to Foresee and Mitigate Policyholder Attrition," Int. J. Adv. Res. Sci. Commun. Technol., vol. 5, no. 5, pp. 1–11, 2025.

31. B. M. Leija, E. R. Beltrán, J. L. O. Mora, and J. O. V. Valadez, "Performance of Machine Learning Algorithms in Fault Diagnosis for Manufacturing Systems: A Comparative Analysis," Processes, vol. 13, no. 6, May 2025, doi:

10.3390/pr13061624.

32. T. Akyaz and D. Engın, "Machine Learning-Based Predictive Maintenance System for Artificial Yarn Machines," IEEE Access, vol. 12, pp. 125446–125461, 2024, doi: 10.1109/ACCESS.2024.3454548.

33. S. P. S. Rathore, A. Gupta, A. Parashar, D. Upadhyay, R. K. Deb, and A. Gupta, "Machine Learning in Predictive Maintenance for Industrial Equipment," in 2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N), IEEE, Dec. 2024, pp. 1537–1541. doi: 10.1109/ICAC2N63387.2024.10895271.

34. P. Sathish, A. S. Reddy, G. S. Teja, G. U. Kiran, and A. Kireeti, "Design of Water Quality Monitoring System using SVM Algorithm," in 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), 2023, pp. 1196–1201. doi: 10.1109/ICESC57686.2023.10193737.

35. M. J. I. Tutul, M. Alam, and M. A. H. Wadud, "Smart Food Monitoring System Based on IoT and Machine Learning," in 2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM), IEEE, Jun. 2023, pp. 1–6. doi: 10.1109/NCIM59001.2023.10212608.

36. T. Parbat, R. S. Benhal, and H. Jain, "IoT Based Health Care Data Monitoring Using Machine Learning," in 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), 2022, pp. 282–286. doi: 10.1109/ICSCDS53736.2022.9760770.