



Survey paper on Anomaly Detection in Industrial Machine using Machine Learning Technique

¹Bikey Kumar Shah, ²Mr. Rakesh Kumar Lodhi, ³Rakesh Kumar Tiwari

M. Tech. Scholar, Dept. of CSE, Technocrats Institute of Technology and Science, Bhopal, India¹

Assistant Professor, Dept. of CSE, Technocrats Institute of Technology and Science, Bhopal, India^{2,3}

Abstract: The rapid advancement of Industry 4.0 and Industrial Internet of Things (IIoT) technologies has led to the widespread deployment of sensors and smart monitoring systems in industrial environments. These systems continuously generate large volumes of operational data, creating opportunities for intelligent condition monitoring and predictive maintenance. Anomaly detection plays a crucial role in identifying abnormal machine behavior, equipment faults, and potential failures before they result in costly downtime or safety hazards. Traditional rule-based and statistical approaches often struggle to handle complex and high-dimensional industrial data, leading to the adoption of Machine Learning (ML) techniques for more accurate and adaptive anomaly detection.

This survey paper presents a comprehensive review of anomaly detection methods in industrial machines using machine learning techniques. Various supervised, unsupervised, and semi-supervised learning approaches are analyzed, including One-Class Support Vector Machine (OC-SVM), Isolation Forest, Local Outlier Factor (LOF), K-Nearest Neighbors (KNN), Decision Trees, Random Forests, Neural Networks, and Deep Learning-based methods such as Autoencoders and Long Short-Term Memory (LSTM) networks. The study examines their working principles, advantages, limitations, and applicability to industrial fault diagnosis and predictive maintenance. Furthermore, the survey discusses commonly used industrial datasets, performance evaluation metrics, and recent research trends in anomaly detection.

The findings indicate that machine learning-based anomaly detection significantly improves fault identification accuracy, reduces maintenance costs, and enhances operational reliability. Among the reviewed techniques, unsupervised methods such as Isolation Forest and LOF demonstrate effectiveness in scenarios with limited labeled data, while deep learning models provide superior performance for complex and high-dimensional sensor datasets. The paper concludes by highlighting current challenges and future research directions, including explainable artificial intelligence, real-time edge deployment, federated learning, and hybrid anomaly detection frameworks for smart industrial systems.

Keywords: Anomaly Detection, Industrial Machines, Machine Learning, Predictive Maintenance, Fault Diagnosis, Condition Monitoring



I. INTRODUCTION

The increasing adoption of smart manufacturing and Industry 4.0 technologies has transformed modern industrial systems by integrating advanced sensors, automation, and data analytics into machine operations. Industrial machines continuously generate large volumes of data through sensors that monitor parameters such as temperature, vibration, pressure, current, speed, and acoustic signals. Analyzing this data is essential for ensuring reliable operation, reducing downtime, and improving production efficiency. One of the most critical tasks in industrial monitoring is anomaly detection, which involves identifying unusual patterns or behaviors that deviate from normal operating conditions and may indicate equipment faults, process abnormalities, or impending failures [1, 2].

Traditional maintenance strategies, such as corrective maintenance and preventive maintenance, often result in unexpected machine failures or unnecessary maintenance costs. Corrective maintenance addresses issues only after a failure occurs, leading to production losses and increased repair expenses. Preventive maintenance schedules maintenance activities at fixed intervals, regardless of the actual machine condition, which may cause inefficient resource utilization. To overcome these limitations, predictive maintenance has emerged as a data-driven approach that uses machine condition monitoring and intelligent analytics to predict failures before they occur. Anomaly detection serves as a fundamental component of predictive maintenance systems by providing early warnings of abnormal machine behavior. Machine Learning (ML) techniques have gained significant attention for anomaly detection due to their ability to learn complex patterns from large-scale industrial data. Unlike traditional statistical methods, ML algorithms can automatically discover hidden relationships, adapt to changing operating conditions, and handle high-dimensional datasets. Various machine learning approaches have been applied to industrial anomaly detection, including supervised, unsupervised, and semi-supervised learning methods. Popular techniques such as One-Class Support Vector Machine (OC-SVM), Isolation Forest, Local Outlier Factor (LOF), Random Forest, Artificial Neural Networks, Autoencoders, and Long Short-Term Memory (LSTM) networks have demonstrated promising results in detecting machine faults and operational anomalies [3, 4].

Industrial anomaly detection presents several challenges, including data imbalance, scarcity of labeled fault data, noisy sensor measurements, changing operating environments, and real-time processing requirements. Many industrial systems operate under normal conditions for extended periods, making anomalous events rare and difficult to capture. Consequently, unsupervised and semi-supervised machine learning techniques have become particularly important because they can identify abnormal behavior without requiring extensive labeled datasets [5].

This survey paper provides a comprehensive review of machine learning techniques used for anomaly detection in industrial machines. The study examines various ML-based approaches, their methodologies, advantages, limitations, and industrial applications. Furthermore, it discusses commonly used datasets, evaluation metrics, and recent advancements in the field.



The survey aims to provide researchers and practitioners with a clear understanding of current developments and future research directions for intelligent anomaly detection systems in industrial environments [6, 7].

II. LITERATURE REVIEW

Mastriani et al. (2025) presented a comprehensive study on improving anomaly detection in industrial time-series data through the integration of segmentation techniques and heterogeneous ensemble learning models. The authors highlighted that industrial systems generate large volumes of sequential sensor data characterized by complex temporal dependencies, noise, and varying operational conditions, making anomaly detection a challenging task. To address these challenges, the proposed methodology first segments the time-series data into meaningful intervals, enabling the detection model to focus on local patterns and temporal variations. Subsequently, multiple heterogeneous machine learning models are combined in an ensemble framework to improve detection robustness and reduce false alarm rates. Experimental evaluations demonstrated that segmentation significantly enhances the ability of models to capture anomalous behaviors, while the ensemble strategy improves generalization across diverse industrial environments. The study reported higher detection accuracy, improved precision, and better adaptability compared to traditional single-model approaches. The authors concluded that combining segmentation with heterogeneous ensembles offers an effective and scalable solution for industrial anomaly detection systems, particularly in manufacturing and process automation applications.

Yang and Xie (2025) proposed an attention-based autoencoder framework for graph anomaly detection, focusing on industrial environments where data are naturally represented as graph structures. Traditional anomaly detection techniques often struggle to capture the complex relationships and dependencies among interconnected nodes in industrial networks. To overcome this limitation, the authors incorporated an attention mechanism within the autoencoder architecture, allowing the model to selectively focus on important nodes and edges while reconstructing graph representations. The attention layer effectively captures both local and global structural information, improving the identification of abnormal patterns that may indicate equipment faults, network intrusions, or operational irregularities. Extensive experiments on benchmark graph datasets demonstrated that the proposed method achieved superior anomaly detection performance compared to conventional graph-based and deep learning approaches. The model showed improved reconstruction accuracy, enhanced sensitivity to subtle anomalies, and better scalability for large industrial networks. The study emphasized the potential of attention mechanisms in enhancing graph representation learning and enabling more accurate anomaly detection in complex industrial systems.

Habib et al. (2025) developed a stacked autoencoder integrated with machine learning classifiers for anomaly detection in Industrial Internet of Things (IIoT) environments. The rapid adoption of IIoT technologies has increased the volume and complexity of sensor-generated data, creating significant challenges for identifying abnormal events and system failures. The proposed framework utilizes a stacked autoencoder to perform deep feature extraction and



dimensionality reduction from high-dimensional IIoT data. The extracted features are then fed into machine learning classifiers to distinguish normal and anomalous operating conditions. The study demonstrated that deep feature learning significantly enhances the representation of complex industrial data patterns, resulting in improved classification performance. Experimental results showed that the hybrid framework outperformed traditional machine learning methods in terms of detection accuracy, precision, recall, and F1-score. Furthermore, the model exhibited strong robustness against noisy and incomplete data, which are common challenges in real-world IIoT applications. The authors concluded that combining deep learning-based feature extraction with machine learning classification provides a powerful solution for intelligent anomaly detection and predictive maintenance in industrial environments.

M. B. A et al. (2025) introduced an intelligent failure detection framework for industrial machines by combining hybrid machine learning algorithms with anomaly score analysis. The research addressed the growing need for reliable fault detection mechanisms capable of preventing unexpected machine breakdowns and minimizing production losses. The proposed methodology employs multiple machine learning techniques to analyze operational parameters collected from industrial machinery and generates anomaly scores that quantify deviations from normal behavior. These anomaly scores are subsequently used to identify early signs of machine degradation and potential failures. The hybrid framework effectively integrates supervised and unsupervised learning strategies, enabling the detection of both known and previously unseen fault conditions. Experimental validation demonstrated that the proposed system achieved high detection accuracy while maintaining low false positive rates. Additionally, the model provided interpretable anomaly scores that assist maintenance engineers in understanding machine health conditions and prioritizing maintenance activities. The study highlighted the importance of combining anomaly detection with machine learning techniques to support predictive maintenance and improve industrial operational reliability.

Velesaca et al. (2025) conducted a case study on anomaly detection in tinplate lid manufacturing, comparing the effectiveness of supervised and unsupervised learning approaches. Quality inspection in industrial production environments often requires the detection of subtle defects that may affect product performance and customer satisfaction. The authors evaluated various anomaly detection models using real-world manufacturing data collected from tinplate lid production processes. Supervised methods were trained using labeled defect samples, while unsupervised techniques learned normal production patterns without requiring extensive labeling efforts. The comparative analysis revealed that supervised models generally achieved higher detection accuracy when sufficient labeled data were available. However, unsupervised approaches demonstrated greater flexibility and adaptability in scenarios with limited or unavailable defect labels. The study also examined computational efficiency, robustness, and deployment feasibility in industrial settings. Results indicated that both approaches possess distinct advantages depending on data availability and operational requirements. The authors concluded that selecting an appropriate anomaly detection strategy



the learning process. By dynamically adjusting learning priorities, the framework effectively utilizes both labeled and unlabeled data to improve anomaly detection capabilities. Experimental results demonstrated that the proposed method achieved superior detection accuracy, recall, and robustness compared to traditional machine learning and deep learning approaches. The system was particularly effective in identifying rare and previously unseen anomalies that could compromise device security and operational reliability. The authors concluded that prioritized deep reinforcement learning provides a promising direction for handling partially labeled datasets and enhancing anomaly detection performance in modern intelligent electronic systems.

III. ANOMALY DETECTION USING MACHINE LEARNING

Anomaly Detection using Machine Learning is the process of automatically identifying unusual patterns, abnormal behaviors, or rare events in data that deviate significantly from normal operating conditions. In industrial environments, machines generate large amounts of sensor data such as temperature, vibration, pressure, current, speed, and acoustic signals. These data are collected continuously over time and form time-series datasets. Machine learning algorithms analyze these datasets to learn the characteristics of normal machine behavior and detect anomalies that may indicate equipment faults, system failures, process deviations, or maintenance requirements [8, 9].

Traditional anomaly detection methods often rely on predefined thresholds and manual inspections, which may not effectively capture complex relationships and hidden patterns in industrial data. Machine learning techniques overcome these limitations by automatically learning patterns from historical data and adapting to changing operating conditions. These methods can identify subtle anomalies that are difficult to detect through conventional approaches. Depending on the availability of labeled data, anomaly detection can be performed using supervised, unsupervised, or semi-supervised learning approaches. Supervised methods require labeled examples of normal and abnormal conditions, whereas unsupervised methods identify anomalies without prior labeling by recognizing observations that significantly differ from the majority of data. Semi-supervised methods learn primarily from normal data and detect deviations as anomalies [10].

The anomaly detection process typically involves data collection, preprocessing, feature extraction, model training, anomaly identification, and performance evaluation. During preprocessing, noise removal, normalization, and handling of missing values improve data quality. Feature extraction techniques identify relevant characteristics from raw sensor measurements. Machine learning models such as Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), K-Nearest Neighbor (KNN), Isolation Forest, Autoencoders, and Long Short-Term Memory (LSTM) networks are then trained to distinguish between normal and abnormal patterns. Once trained, the models continuously monitor incoming data and generate alerts whenever anomalous behavior is detected [11, 12].

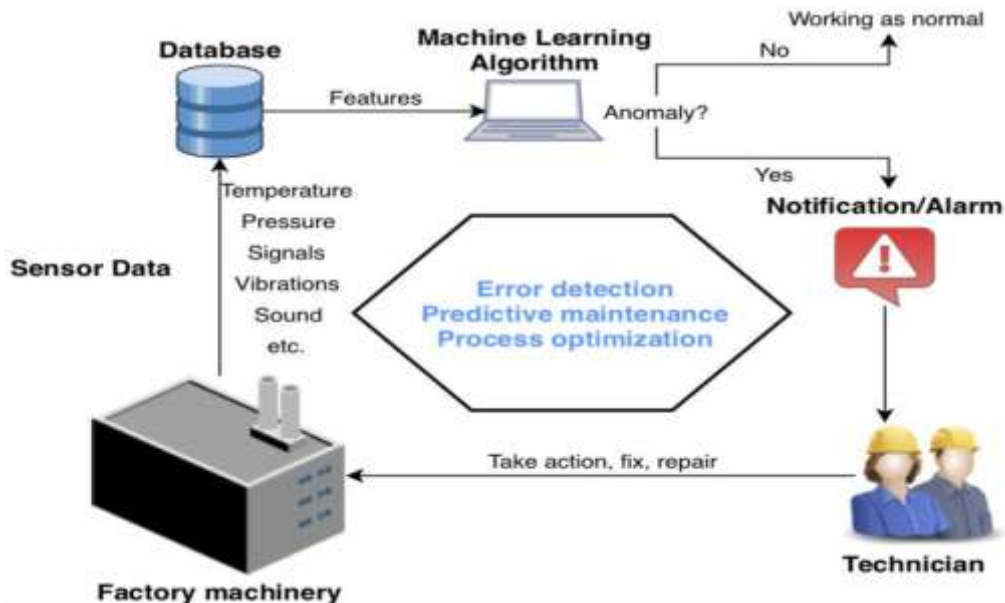


Figure 1: Anomaly Detection using Machine Learning

IV. EXPECTED OUTCOME

The proposed survey on Anomaly Detection in Industrial Machines using Machine Learning Techniques is expected to provide a comprehensive understanding of existing machine learning approaches used for detecting anomalies in industrial environments. The study will identify the strengths, limitations, and practical applicability of various techniques such as One-Class Support Vector Machine (OC-SVM), Isolation Forest, Local Outlier Factor (LOF), and statistical variance-based methods for industrial fault detection.

The survey is expected to demonstrate how machine learning algorithms improve the accuracy and reliability of anomaly detection compared to traditional monitoring approaches. It will provide insights into the effectiveness of different methods in identifying abnormal machine behavior, equipment degradation, and potential failures at an early stage. Furthermore, the study will highlight the importance of data preprocessing, feature extraction, and performance evaluation metrics in achieving accurate anomaly detection results.

Another expected outcome is the identification of current research challenges, including limited labeled datasets, data imbalance, noisy sensor measurements, and real-time implementation requirements. The survey will also explore recent advancements in predictive maintenance, Industrial Internet of Things (IIoT), and smart manufacturing systems that integrate machine learning-based anomaly detection.

Finally, the survey is expected to serve as a valuable reference for researchers, academicians, and industry professionals by providing a structured overview of existing techniques and future research directions. The findings will contribute to the development of more intelligent, robust, and efficient anomaly detection systems capable of improving machine reliability, reducing maintenance costs, minimizing downtime, and enhancing overall industrial productivity.



V. CONCLUSION

Anomaly detection has become a critical component of modern industrial systems, enabling the early identification of equipment faults, abnormal operating conditions, and potential machine failures. With the rapid growth of Industry 4.0 and Industrial Internet of Things (IIoT) technologies, industrial environments generate vast amounts of sensor data that can be effectively analyzed using Machine Learning (ML) techniques. Compared with traditional rule-based and statistical approaches, machine learning methods offer greater adaptability, improved accuracy, and the ability to handle complex and high-dimensional datasets.

This survey reviewed various machine learning techniques used for industrial anomaly detection, including supervised, unsupervised, and semi-supervised approaches. Among these techniques, unsupervised and semi-supervised methods are particularly valuable in industrial applications where labeled fault data is limited or unavailable.

The survey also highlighted the advantages, limitations, and practical applications of different anomaly detection approaches in monitoring industrial equipment such as motors, turbines, pumps, compressors, and manufacturing systems. Machine learning-based anomaly detection helps industries reduce unexpected downtime, improve equipment reliability, lower maintenance costs, enhance operational safety, and increase production efficiency.

Despite considerable progress, several challenges remain, including data imbalance, noisy sensor measurements, real-time processing requirements, model interpretability, and the scarcity of fault-labeled datasets. Addressing these challenges will be essential for the widespread deployment of intelligent anomaly detection systems. Overall, machine learning has emerged as a powerful and effective solution for industrial anomaly detection and will continue to play a vital role in the development of smart manufacturing and predictive maintenance systems in the future.

REFERENCES

- [1] E. Mastriani, A. Costa, F. Incardona, K. Munari, and S. Spinello, "Improving anomaly detection in industrial time series: The role of segmentation and heterogeneous ensemble," in Proc. 2025 11th Int. Conf. Control, Decision and Information Technologies (CoDIT), Split, Croatia, 2025, pp. 2623–2628.
- [2] S. Yang and C. Xie, "Attention-based auto-encoder for graph anomaly detection," in Proc. 2025 IEEE 23rd Int. Conf. Industrial Informatics (INDIN), Kunming, China, 2025.
- [3] A. A. Habib, M. K. Hasan, S. Islam, R. Hassan, T. R. Gadekallu, and X. Yu, "Stacked autoencoder with machine learning for anomaly detection in industrial Internet of Things," in Proc. 2025 5th Int. Conf. Advanced Algorithms and Neural Networks (AANN), Qingdao, China, 2025, pp. 595–599.
- [4] M. B. A, C. N. S, D. N. Marangappanavar, H. R. G, and I. S, "Intelligent failure detection in industrial machines using hybrid ML and anomaly scores," in Proc. 2025 9th Int. Conf. Inventive Systems and Control (ICISC), Coimbatore, India, 2025, pp. 1149–1156.



- [5] H. O. Velesaca, A. D. Sappa, and J. A. Holgado-Terriza, “A case study of anomaly detection in tinplate lids: Supervised vs unsupervised approaches,” in Proc. 2025 11th Int. Conf. Automation, Robotics, and Applications (ICARA), Zagreb, Croatia, 2025, pp. 370–374.
- [6] T. Akyaz and D. Engin, “Machine learning-based predictive maintenance system for artificial yarn machines,” 2024.
- [7] C. Wang, X. Li, T. Zhou, and Z. Cai, “Unsupervised time series segmentation: A survey on recent advances,” *Computers, Materials & Continua*, vol. 80, no. 2, pp. 2657–2673, Aug. 2024.
- [8] S. Qin et al., “A partially labeled anomaly data detection approach based on prioritized deep reinforcement learning for consumer electronics security,” *IEEE Transactions on Consumer Electronics*, 2024.
- [9] D. Coelho, D. Costa, E. M. Rocha, D. Almeida, and J. P. Santos, “Predictive maintenance on sensorized stamping presses by time series segmentation, anomaly detection, and classification algorithms,” *Procedia Computer Science*, vol. 200, pp. 1184–1193, 2023.
- [10] D. Velasquez, E. Perez, X. Oregui, A. Artetxe, J. Manteca, J. E. Mansilla, M. Toro, M. Maiza, and B. Sierra, “A hybrid machine-learning ensemble for anomaly detection in real-time Industry 4.0 systems,” *IEEE Access*, vol. 10, pp. 72024–72036, 2022.
- [11] C. -W. Lee, S. -H. Cho, J. -H. Lee, J. -S. Lee, S. Choi and W. -C. Seo, "Empirical Research in Anomaly Detection for Rotating Machines Diagnosis using Deep Learning," *2022 IEEE International Conference on Big Data (Big Data)*, Osaka, Japan, 2022, pp. 6758-6760.
- [12] D. Vranješ and O. Niggemann, "Anomaly detection based on time series data from industrial automatic sewing machines," *2022 IEEE 5th International Conference on Industrial Cyber-Physical Systems (ICPS)*, Coventry, United Kingdom, 2022, pp. 01-08.
- [13] A. Buabeng, A. Simons, N. K. Frempong, and Y. Z. Yevenyo, “A novel hybrid predictive maintenance model based on clustering, SMOTE and multilayer perceptron neural network optimised with grey wolf algorithm,” 2021.
- [14] J. Carrasco, I. Markova, D. Lopez, I. A. Martos, D. Garcia-Gil, M. Garcia-Barzana, M. Arias-Rodil, J. Luengo, and F. Herrera, “Anomaly detection in predictive maintenance: A new evaluation framework for temporal unsupervised anomaly detection algorithms,” 2021.
- [15] D. Imbassahy, H. Marques, G. Rocha, and A. Martinetti, “Empowering predictive maintenance: A hybrid method to diagnose abnormal situations,” *Applied Sciences*, vol. 10, no. 27, 2020.