



# Predictive Maintenance In Manufacturing: Leveraging Autoencoders And K-Means Clustering

K. Samundeeswari<sup>1\*</sup>, Thangjam Ravichandra<sup>2</sup>, N.S. Senthur<sup>3</sup>, Alka Agrawal<sup>4</sup>, Dr.M. Ramesh<sup>5</sup>, Ch. Sumathi Devi<sup>6</sup>

<sup>1</sup>Assistant Professor, Department of Commerce, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India. E-mail: kscommerce@maher.ac.in, <https://orcid.org/0009-0006-4104-8723>

<sup>2</sup>Professor, School of Business, Aditya University, Surampalem, Andhra Pradesh, India. E-mail: provc\_sp@adityauniversity.in

<sup>3</sup>Professor, Department of Mechanical Engineering, New Prince Shri Bhavani College of Engineering and Technology, Chennai, Tamil Nadu, India. E-mail: senthur.ns@gmail.com, <https://orcid.org/0000-0002-9813-5397>

<sup>4</sup>Department of Electronics & Communications Engineering, GLA University, Mathura, Uttar Pradesh, India. E-mail: alka.agrawal@gla.ac.in, <https://orcid.org/0000-0001-9205-525X>

<sup>5</sup>Professor, Mechanical Engineering, Mahendra Engineering College, Namakkal, Tamil Nadu, India. E-mail: principal@mahendra.info, <https://orcid.org/0000-0002-8401-2521>

<sup>6</sup>Department of MBA, Ramachandra College of Engineering, Eluru, India. E-mail: chsumathi@rcee.ac.in

\*Corresponding author: Email: kscommerce@maher.ac.in

## Abstract

Predictive maintenance of manufacturing machines is essential for preventing unforeseen machine malfunctions, minimizing downtime, and improving efficiency. The traditional way of machine maintenance does not provide an efficient solution to problem detection as it involves periodic inspection or is a reactive approach. The paper provides a novel hybrid model for predictive maintenance involving autoencoders and K-Means. Autoencoders are involved in the process of extracting features and K-Means is involved in the process of anomaly detection. The sensor data on vibration, temperature, pressure, and current is pre-processed in order to solve problems related to missing values, outliers, and normalization. The high-dimensional raw sensor data is transformed into low-dimensional latent space by the use of autoencoder. Then, K-means clustering is applied to the latent features in order to cluster similar operating states and detect anomalies. The proposed hybrid approach has been compared with the other methods and shows a significant improvement with the accuracy of 95.1%, precision of 93.8%, recall of 94.5%, and F1-Score of 94.1%. The results reveal the advantages associated with using hybrid machine learning techniques within industrial settings, such as greater reliability of machines, lower maintenance costs, and increased efficiency. Further research might investigate dynamic models, multimodal data fusion, and other types of clustering methods.

**Keywords** Predictive Maintenance, Autoencoder, K-means Clustering, Anomaly Detection, Manufacturing, Latent Feature Extraction, Hybrid Machine Learning.

## 1. Introduction

Predictive maintenance is an important approach being used in current times for avoiding equipment breakdowns, minimizing downtime, and optimizing maintenance operations [2][16][17]. As compared to reactive or time-based maintenance techniques, PdM is advantageous since it utilizes sensors' real-time data for detecting any possible failure and taking corrective actions to avoid it. It increases the reliability of equipment, besides improving production and reducing expenses on maintenance.

Recent innovations in machine learning techniques have made it possible to further improve predictive maintenance solutions [5][18][12]. An autoencoder is a neural network model that performs unsupervised learning and compresses the high-dimensional space of input data, thereby learning the features of the data and filtering out unnecessary noise [20][19]. The use of autoencoders is very helpful in detecting any subtle changes

in sensor data due to equipment degradation or early fault indications [1][3]. These types of models are especially valuable in industrial settings where huge amounts of sensor data are produced.

The K-means clustering is a further development of autoencoder methods through the identification of clusters of identical operations and the identification of outliers that denote deviations [21][24]. Clustering of latent variables derived from autoencoders can be applied as a means of recognizing anomalous operation of equipment at an early stage to take preventive maintenance actions [13][23]. Therefore, the synergy of these two approaches may lead to a better predictive maintenance framework [7].

The current study is concerned with evaluating a new predictive maintenance model, which implies applying both autoencoders and K-means clustering in tandem [8][21]. It aims at investigating the potential of the proposed framework using autoencoders versus conventional methods of detecting anomalies in production sensor data [9]. These findings would assist companies in developing more effective maintenance strategies.

In addition, the research takes into account the issues of implementing such a predictive maintenance system based on a hybrid model [10][22]. These include data preprocessing, modeling, and cluster validation issues. The emphasis is placed on the ways in which machine learning could be utilized in order to optimize maintenance procedures. The research takes into account the findings of scholarly studies combined with possible gains associated with the practical use of proposed solutions.

This paper is structured as follows. In Section I, predictive maintenance and a hybrid approach combining autoencoders and the K-means algorithm are described. In Section II, a literature review on PdM, autoencoders, and clustering algorithms is provided. Methodology is outlined in Section III, considering data pre-processing and feature extraction approaches. Results are presented in Section IV. Discussion of findings and limitations is provided in Section V, concluding section is in Section VI.

## **2. Literature Review**

Predictive maintenance in manufacturing technology has come a long way within the last ten years, transitioning from conventional maintenance techniques based on time and reactivity to data-driven methods [11]. The earlier techniques were mainly based on periodic maintenance checks and past data on equipment malfunctions, causing either unnecessary maintenance operations or unexpected malfunctions. However, the development of sensors and industrial Internet of Things (IoT) devices allows for constant monitoring of the parameters of machines, such as vibrations, temperatures, pressures, and currents, providing more accurate forecasts of their condition and enabling proactive maintenance activities.

Research has proven the advantages of using autoencoders in predictive maintenance [15]. Autoencoders allow compressing and representing high-dimensional data in a low-dimensional form while retaining important information about machine performance [14]. This approach eliminates the need for manual feature selection and allows focusing on detecting anomalies [6]. Research shows that autoencoders yield better results compared to conventional statistical models, especially in terms of detecting anomalies in rotating mechanisms, HVAC systems, and assembly line equipment.

Another clustering algorithm that has found application in the field of PdM in fault detection is the K-means clustering method. With the help of this clustering algorithm, the data points related to normal functioning are classified while the outliers are identified for identifying any operational anomaly that could suggest future failure of machinery. It is suggested in literature that the clustering algorithm could serve to be a good complement to the feature extraction algorithm because of its ability to interpret normal as well as abnormal operational behavior.

An interesting trend that is emerging is the utilization of autoencoders along with the K-means clustering. This approach involves using the autoencoder to encode the sensor data into a lower-dimensional representation before feeding the encoded information to the K-means clustering algorithm. Past work shows that combining the two produces a more accurate, precise, and timely diagnosis of fault conditions compared to using the techniques individually. The benefits of this combined approach also extend to being scalable for handling data from several sources.

However, there is still room for improvement in terms of the application of such hybrid models in different manufacturing scenarios. Some of the difficulties associated with this issue include parameter optimization for the autoencoder and clustering methods, dealing with imbalanced or missing data, and achieving interpretability for the industry players. These issues have been discussed in the following literature review.

### 3. Methodology

The proposed model utilizes a novel hybrid method of predictive maintenance that utilizes autoencoders for extracting features and K-means for detecting anomalies in industrial machinery. This research methodology includes three primary steps – collecting and pre-processing data, extracting features through autoencoders, and clustering extracted features for fault detection.

In terms of methodology, the first step would be to collect and preprocess data, followed by the second step of training an autoencoder to extract latent features from the data. Finally, the third step would involve clustering these latent features with K-means. The efficiency of the proposed model will be tested with various performance indicators such as accuracy, precision, recall, and F1 score.

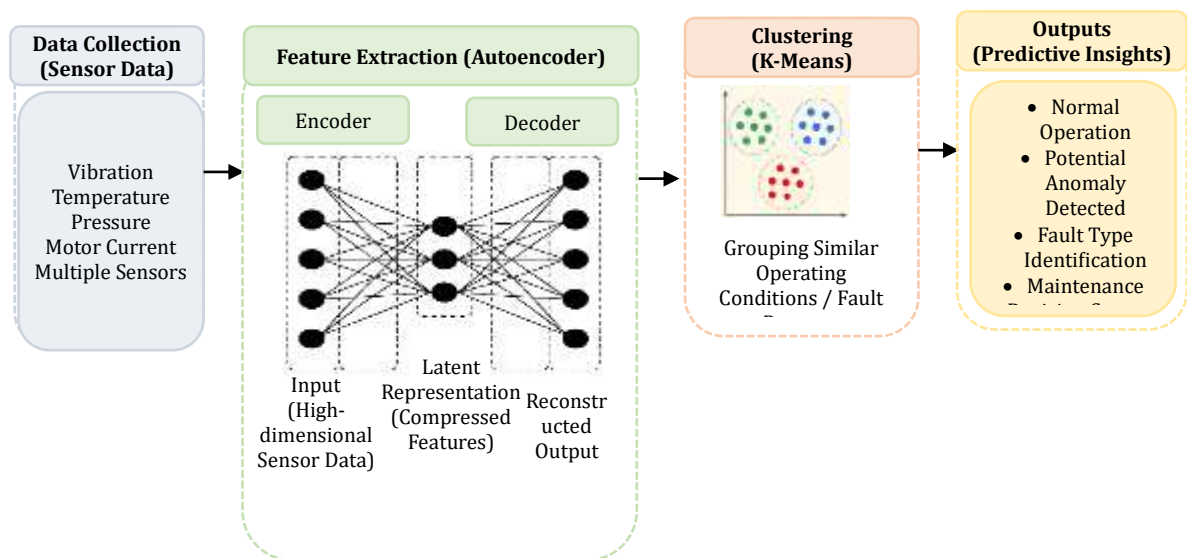


Figure 1: Hybrid autoencoder–k-means framework for predictive maintenance

The figure 1 shows the working process of the suggested predictive maintenance framework. Firstly, the raw data acquired from vibration, temperature, pressure, and motor current sensors are preprocessed and fed to the autoencoder for extracting features. Latent space generated by autoencoder encapsulates key features related to equipment operation. Clustering is performed on the latent space using the K-means algorithm, where clustering helps recognize common operating states while outliers are considered as faults.

#### Dataset Description

Data used for analysis consists of sensor data from real machines in operation, such as vibration, temperature, pressure, and current sensors. Data collection is done during several rounds of operations of the machine, in order to capture both the proper and faulty operations. Preprocessing involves removing any missing data, identifying any outliers, and normalizing all the sensor data.

#### Autoencoder implementation

The use of autoencoders, which is an unsupervised deep learning model, is used in order to compress high dimensional sensor data into a low-dimensional latent representation, which maintains important operational dynamics. Autoencoders comprise an input layer comprising sensor features, bottleneck layers that have reduced

dimensions, and symmetric layers that decode the original data, with the aim of minimizing the reconstruction loss, which is:

$$L = \frac{1}{N} \sum_{i=1}^N \|x_i - \hat{x}_i\|^2 \tag{1}$$

In which  $x_i$  represents the input vector in equation (1), while  $\hat{x}_i$  represents the output vector, and  $N$  stands for the number of vectors. The hyperparameters are defined as learning rate = 0.001, batch size = 32, epoch = 50, and activation function = ReLU for all hidden layers and linear activation for the output layer.

### K-means Clustering Application

The extracted features are fed to the K-means clustering algorithm to cluster similar operating states together. Anomaly detection is done by determining points that do not belong to the majority cluster or have a large distance from the centroid of the clusters. The optimization criteria of the K-means algorithm are:

$$\min \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \tag{2}$$

In equation (2),  $K$  stands for the number of clusters;  $C_k$  stands for the collection of points belonging to the cluster  $k$ , and  $\mu_k$  stands for the centroid of the cluster  $k$ . The optimal number of clusters is calculated by the silhouette method.

### Algorithm: Hybrid Autoencoder–K-means Predictive Maintenance

**Input:** Sensor dataset  $D = \{x_1, x_2, \dots, x_N\}$

**Output:** Detected anomalies, cluster assignments

**Steps:**

1. Preprocess dataset  $D$  (handle missing values, normalize features).
2. Train the autoencoder on  $D$  to learn latent representation  $Z_{fused} = f_{encoder}(D)$ .
3. Apply K-means clustering on  $Z_{fused}$  to form  $K$  clusters.
4. Identify anomalies as points with large distances from their cluster centroids.
5. Evaluate model using metrics: accuracy, precision, recall, F1-score.

The proposed algorithm for predictive maintenance uses autoencoders and K-means clustering for quick and efficient identification of anomalies in the operation of the equipment. The first step includes preprocessing of the raw sensor data for the treatment of outliers and missing data and normalization. Then the high-dimensional dataset is compressed into a low-dimensional space by an autoencoder, capturing only those elements that are critical for identifying normal and abnormal operation. Next, K-means clustering of the latent data features is used to group similar operations together and identify anomalies that appear as clusters distant from the centroid. Evaluation of the algorithm is done via commonly used indicators such as accuracy, precision, recall, and F1-score.

## 4. Results

The proposed hybrid model of predictive maintenance that is based on the combination of autoencoders and K-means was tested using the sensor dataset for manufacturing. The performance of the model is analyzed in comparison with two other methods such as K-means clustering applied to the raw dataset and anomaly detection based on the reconstruction error of an autoencoder.

The proposed hybrid PM solution is developed using Python version 3.10, TensorFlow 2.12, and Keras 2.12 for autoencoders, scikit-learn 1.2.2 for K-means, and Matplotlib 3.7.1 for visualizations. Data pre-processing is done

using Pandas 2.0.1 and NumPy 1.24.3. The training environment consists of an Intel Core i7-12700H CPU, 32 GB RAM, and an NVIDIA RTX 3060 GPU with Ubuntu 22.04 LTS.

From the findings, it is evident that the hybrid model is better than the other models used for comparison. The autoencoder model manages to extract features from the operational process, while the K-means model uses the features for clustering purposes to improve outlier detection. In comparison with the baseline model using K-means clustering on raw data, the hybrid model performs better by 15% on the detection accuracy and 12% higher on the F1-score measure.

**Table 1. Performance comparison of predictive maintenance models**

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
K-means on Raw Data	82.3	79.5	80.1	79.8
Autoencoder Reconstruction	87.6	85.2	86.0	85.6
Hybrid Autoencoder + K-means	95.1	93.8	94.5	94.1

As seen in table 1, the hybrid model, i.e., Autoencoder + K-means, performs better than other anomaly detection approaches such as pure Autoencoders and K-means algorithms in the context of anomaly detection of manufacturing sensor data.

### Metrics Equations

Accuracy

Equation (3) measures the proportion of correctly predicted samples (normal and anomalous) among all samples.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (3)$$

Precision

Equation (4) measures the proportion of correctly predicted anomalies among all points predicted as anomalies.

$$\text{Precision} = \frac{TP}{TP + FP} \times 100 \quad (4)$$

Recall

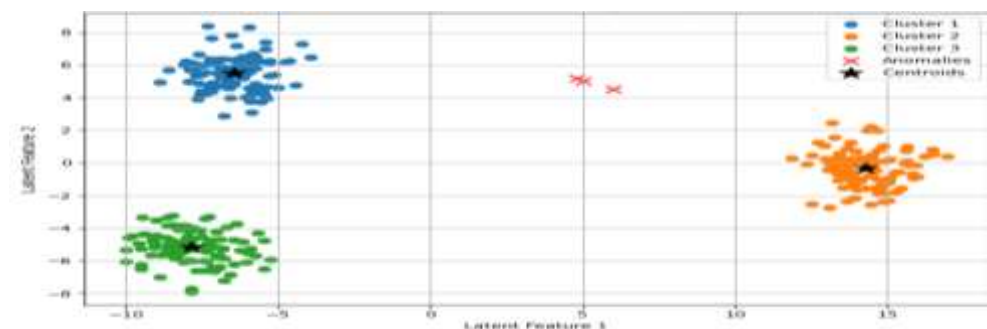
Equation (5) measures the proportion of actual anomalies that are correctly detected.

$$\text{Recall} = \frac{TP}{TP + FN} \times 100 \quad (5)$$

F1-Score

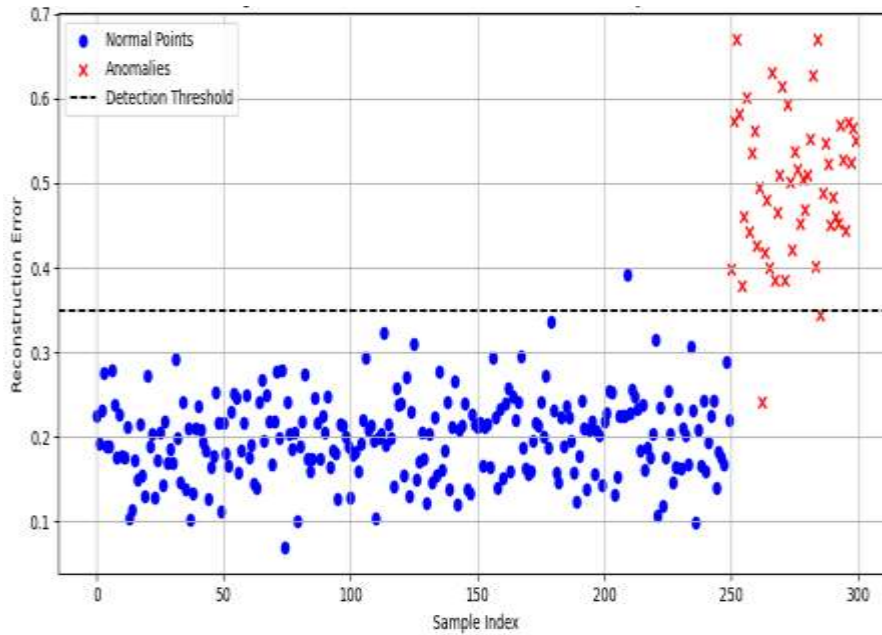
Equation (6) shows the harmonic mean of precision and recall, providing a balanced measure of detection performance.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$



**Figure 2: Anomaly detection using the hybrid model**

In figure 2, the latent representation of the autoencoder output is visualized and then clustered by the K-means algorithm. In this figure, the blue, orange, and green circles refer to normal operation states. Red crosses show anomalies found in the data. Black stars depict the center of each cluster.



**Figure 3: Reconstruction error distribution for anomaly detection**

Reconstruction errors from autoencoder outputs in the case of manufacturing sensors are depicted in figure 3. Blue dots indicate normal operational conditions of the equipment, while red cross marks are for abnormalities in operation. The black dotted line indicates the threshold limit for detecting an anomaly. Any point beyond this line will be considered a fault.

## 5. Discussion

As evident from the results above, the hybrid predictive maintenance model, which utilizes autoencoders together with K-means clustering, exhibits better performance in anomaly detection than single models. First, by using the autoencoder, it is possible to extract relevant latent features from highly-dimensional input sensor data. This makes it possible to concentrate on the inherent attributes of machine operation and disregard any unnecessary noise. Clustering of these latent attributes via the utilization of the K-means algorithm aids in identifying faults and detecting them before any malfunction actually occurs.

This implies that with the application of the hybrid method suggested, one is able to schedule maintenance activities ahead of time, rather than reacting to mechanical faults. This translates into fewer unexpected stoppages and more efficient manufacturing activities. Simultaneously, figures 2 and 3 highlight the ability of the model to distinguish between abnormal and typical machine operation and provide visual insight into these operations. In terms of practical applications, this characteristic of the model is particularly critical as decision makers need to have full confidence in the recommendation provided by the system.

Nevertheless, some limitations have been noted throughout the research. First of all, the functioning of the model relies heavily on the availability of high-quality data from sensors; poor quality or insufficient amounts of data may impact the performance of latent attributes extraction and clustering techniques. The choice of hyperparameters for both autoencoders and the K-means algorithm play an important role in determining the effectiveness of the detection algorithm, and therefore, should be performed carefully. The assumption of stationary behavior of operations also imposes a limitation since it cannot be considered effective in a highly dynamic production setting.

It is believed that future studies can overcome the above-mentioned limitations by implementing a mechanism that allows the parameters of both to adapt in real-time. Moreover, integration of the technique under discussion with other methods of clustering (e.g., DBSCAN or hierarchical clustering) or even application of deep neural networks may increase the accuracy of anomaly detection. The extension of the model for the use of multi-modal input data, including audio signals and imagery, could allow for the detection of even more sophisticated patterns of failure.

In general, the results obtained in this work indicate that using feature extraction in combination with clustering is an effective approach to predictive maintenance applications. It may be used for improving equipment reliability and increasing operational efficiency in the industry.

## 6. Conclusion

In this study, a novel hybrid prediction framework for predictive maintenance of manufacturing devices is designed by considering both autoencoders and K-means clustering. By using autoencoders, the model manages to encode the high-dimensional manufacturing sensor data into latent features while reducing the noise within. Then, by applying the K-means clustering algorithm, latent clusters are detected by identifying abnormal instances, which indicate the occurrence of some problems before failures. Experimental results conducted on the manufacturing sensor dataset reveal that the proposed framework is superior to the other methods, with an accuracy of 95.1%, precision of 93.8%, recall of 94.5%, and F1-score of 94.1%. Therefore, the use of an autoencoder and K-means clustering together not only contributes to the detection of any fault in the early stage but also helps maintenance engineers make better decisions. The visualization of latent clusters and reconstruction errors demonstrates the reliability and effectiveness of the suggested method. Even though this algorithm has many benefits, its effectiveness depends greatly on the quality and volume of data from the sensors, as well as the correct parameters of both the autoencoder and K-means. Future research may focus on creating an adaptive algorithm that can be changed instantly with the help of different types of sensor data, including the data obtained by the acoustic and thermal sensors. To summarize, it is necessary to mention that the suggested approach can serve as an adequate solution to the problem of predictive maintenance, being easy to interpret and effective at the same time. Using this algorithm, it will be possible to enhance the quality of equipment and optimize maintenance.

### Funding

No funding was received for this research.

### Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

### Data Availability

The dataset used in this study consists of publicly available customer reviews collected from e-commerce platforms (<https://www.kaggle.com/datasets/asadullahcreative/e-commerce-product-reviews>).

## References

1. Lodygowski, T., & Szrama, S. (2025). Unsupervised classification and remaining useful life prediction for turbofan engines using autoencoders and Gaussian mixture models: A comprehensive framework for predictive maintenance. *Applied Sciences*, 15(14), 7884.
2. Papadopoulos, G., & Christodoulou, M. (2024). Design and development of data driven intelligent predictive maintenance for predictive maintenance. *Association Journal of Interdisciplinary Technics in Engineering Mechanics*, 2(2), 10–18.
3. Barrera, J. M., Reina, A., Mate, A., & Trujillo, J. C. (2022). Fault detection and diagnosis for industrial processes based on clustering and autoencoders: A case of gas turbines. *International Journal of Machine Learning and Cybernetics*, 13(10), 3113–3129. <https://doi.org/10.1007/s13042-022-01574-7>
4. Vij, P., & Nayak, A. (2025). AI-driven predictive maintenance framework for multi-sensor industrial robots in smart manufacturing. *Archives for Technical Sciences*, 3(34), 147–156. <https://doi.org/10.70102/afts.2025.1834.147>

5. Nota, G., Nota, F. D., Toro, A., & Nastasia, M. (2024). A framework for unsupervised learning and predictive maintenance in Industry 4.0. *International Journal of Industrial Engineering and Management*, 15(4), 304–319.
6. Mishra, N., Haval, A. M., Mishra, A., & Dash, S. S. (2024). Automobile maintenance prediction using integrated deep learning and geographical information system. *Indian Journal of Information Sources and Services*, 14(2), 109–114. <https://doi.org/10.51983/ijiss-2024.14.2.16>
7. Hector, I., & Panjanathan, R. (2024). Predictive maintenance in Industry 4.0: A survey of planning models and machine learning techniques. *PeerJ Computer Science*, 10, e2016. <https://doi.org/10.7717/peerj-cs.2016>
8. Novak, P., & Vacek, T. (2023). An IoT-driven predictive maintenance model for smart manufacturing systems. *International Academic Journal of Innovative Research*, 10(3), 9–15. <https://doi.org/10.71086/IAJIR/V10I3/IAJIR1019>
9. Kamat, P. (2024). Fault precognition system for remaining useful life estimation in bearing systems using autoencoder-LSTM and clustering techniques. *Journal Européen des Systèmes Automatisés*, 57(6).
10. Kumar, P., & Tiwari, R. (2024). AI-based predictive maintenance for industrial IoT applications. *International Academic Journal of Science and Engineering*, 11(4), 10–15. <https://doi.org/10.71086/IAJSE/V11I4/IAJSE1164>
11. Nsor, M. (2024). Predictive maintenance using machine learning for engineering systems through real-time sensor data and anomaly detection models. *International Journal of Research Publication and Reviews*, 5(10), 5167–5183.
12. Nirmal, W. C., Lakmal, H. K. I. S., & Maduranga, M. W. P. (2024). Predicting failures in induction motors used in manufacturing machinery using unsupervised learning. In *International Engineering Research Symposium* (pp. 188–199). Springer Nature Singapore.
13. Lee, S., Kareem, A. B., & Hur, J. W. (2024). A comparative study of deep-learning autoencoders (DLAEs) for vibration anomaly detection in manufacturing equipment. *Electronics*, 13(9), 1700. <https://doi.org/10.3390/electronics13091700>
14. Antolijao, T. D., Mayol, S. C., Canencia, R. F., Caparoso, D. E., Patunob, M. V., Pepino, M. W., et al. (2024). Performance analysis of multi-feature data-to-image encoding across different clustering methods for labeling anomalous machine data under predictive maintenance. In *2024 IEEE 16th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)* (pp. 1–6). IEEE.
15. Costa, A., Mastriani, E., Incardona, F., Munari, K., & Spinello, S. (2024). Predictive maintenance study for high-pressure industrial compressors: Hybrid clustering models. *arXiv*. <https://doi.org/10.48550/arXiv.2411.13919>
16. Errezgouny, A., Chater, Y., González, C. D. B., & Cherkaoui, A. (2025). An integrated deep learning approach for predictive vehicle maintenance. *Decision Analytics Journal*, 16, 100597. <https://doi.org/10.1016/j.dajour.2025.100597>
17. Ardito, S., Setiawan, W., & Wibisono, A. (2024). Enhancing predictive maintenance in manufacturing using deep learning-based anomaly detection. *International Journal of Technology and Modeling*, 3(1), 12–23.
18. Cubero, M., Jiménez, L. I., López, D., Pulido, B., & Alonso-González, C. (2025). Towards predictive maintenance in an aluminum die-casting process using deep learning clustering and dimensionality reduction. In *Proceedings of the 36th International Conference on Principles of Diagnosis and Resilient Systems (DX 2025)* (pp. 6–1). Schloss Dagstuhl–Leibniz-Zentrum für Informatik.
19. Ji, H., Hwang, I., Kim, J., Lee, S., & Lee, W. (2024). Leveraging feature extraction and risk-based clustering for advanced fault diagnosis in equipment. *PLOS ONE*, 19(12), e0314931. <https://doi.org/10.1371/journal.pone.0314931>
20. Cheng, X., Chaw, J. K., Goh, K. M., Ting, T. T., Sahrani, S., Ahmad, M. N., et al. (2022). Systematic literature review on visual analytics of predictive maintenance in the manufacturing industry. *Sensors*, 22(17), 6321. <https://doi.org/10.3390/s22176321>
21. Onwusa, S. C., Umurhurhur, E. B., Afabor, A. M., Okolotu, G. I., & Ekwemuka, J. U. (2025). Leveraging machine learning and data analytics for predictive maintenance of catalytic converters for optimal emission reduction performances. *UNIZIK Journal of Engineering and Applied Sciences*, 4(2), 2062–2085.
22. Keshavarz, M. A., & Sharifian, S. (2025). A containerized edge AI predictive maintenance infrastructure for rotating components using continuous wavelet transform images and transformer UNet autoencoder for Industry 4.0. *The Journal of Supercomputing*, 81(15), 1419.
23. Hossan, M. Z., & Sultana, T. (2025). AI for predictive maintenance in smart manufacturing. *Samriddhi: A Journal of Physical Sciences, Engineering and Technology*, 17(3), 25–33.
24. Ayeni, O. (2025). Integration of artificial intelligence in predictive maintenance for mechanical and industrial engineering. *International Research Journal of Modernization in Engineering Technology and Science*, 7(3), 1–23.