JANOLI International Journal of
Applied Engineering and Management
(A Peer Review Journal)

journal homepage: https://janolijournals.com/index.php/jijaem

Volume 2, Issue 2, 2025 ISSN: 3048-6939

Enhanced Predictive Maintenance for Industrial Machinery using Hybrid Machine Learning and IoT Sensor Fusion

Anjali Vasishtha
NIET, NIMS University, Jaipur, India,

ARTICLE INFO

Article History:
Received April 5, 2025
Revised April 17, 2025
Accepted April 24, 2025
Available online April 30, 2025

Keywords:
.Predictive Maintenance, Machine Learning, IoT, Sensor Fusion, Condition Monitoring, Anomaly Detection, Industrial Automation, Equipment Health, Remaining Useful Life (RUL)

Correspondence:
E-mail:anjali.vashistha06@gmail.com

ABSTRACT

This paper proposes a upgraded predictive maintenance (PdM) framework for industrial equipment based on a hybrid machine learning strategy combined with IoT sensor fusion. The proposed framework combines information from diverse sensor modalities (acoustic emissions, temperature, pressure, vibration) to yield a holistic evaluation of equipment condition. A new hybrid approach, where a deep learning-based autoencoder is employed for feature learning and a Random Forest classifier is used to detect anomalies and predict Remaining Useful Life (RUL), is suggested. The proposed framework is tested using a real-world industrial dataset and found to achieve considerable improvements in prediction accuracy and lower false alarm rates than conventional methods. The findings emphasize the promise of this methodology to maximize maintenance schedules, reduce downtime, and enhance overall industrial operation efficiency.

1.Introduction

In the competitive industrial environment of today, reducing downtime and maintenance schedules to a minimum are essential if productivity and profitability are to be maximized. The conventional maintenance approaches of reactive (run-to-failure) and preventive (time-based) maintenance tend to lead to either surprise equipment failure or routine maintenance actions, which in turn raise costs and compromise efficiency. Predictive Maintenance (PdM) is a more forthcoming methodology that takes advantage of data-driven methods in forecasting future equipment failure and arranging maintenance actions only when required.

The Industrial Internet of Things (IIoT) has made it possible to deploy sensors in large numbers on industrial equipment, which collect a huge amount of data in relation to equipment condition. If analyzed correctly, the data can offer useful information regarding the working condition of machines and enable the early identification of anomalies that would lead to future failures. Machine learning (ML) methods have proven to be strong tools for analyzing data and creating reliable PdM models.

Nonetheless, the construction of good PdM models is faced with various challenges. Firstly, industrial data have high dimensionality, noise, and class imbalance (failure is a rare occurrence). Secondly, complex and non-linear relationships among sensor readings and equipment health demand advanced modeling techniques. Thirdly, precise prediction of Remaining Useful Life (RUL) is a big challenge since it needs to capture the degradation process over time.

This work tries to overcome these limitations by proposing an improved PdM paradigm that brings together the power of IoT sensor fusion and hybrid machine learning methods. The proposed paradigm fuses information from various modalities of sensors to achieve a holistic evaluation of equipment condition. A new hybrid model, which integrates a deep learning-based autoencoder for feature learning and Random Forest classifier for anomaly identification and RUL estimation, is put forward.

Problem Statement: Current PdM systems tend to have limitations in feature engineering, anomaly detection precision, and RUL prediction accuracy, resulting in inefficient maintenance schedules and higher operational expenses. This work addresses the challenge of developing a more robust and accurate PdM paradigm capable of taking full advantage of the abundance of data provided by IoT sensors.

1. Create an IoT sensor fusion framework for gathering and fusing information from various sensor modalities on industrial equipment.

2. Design a hybrid machine learning model that integrates a deep learning-based autoencoder for feature extraction and a Random Forest classifier for anomaly detection and RUL estimation.

3. Test the performance of the proposed system utilizing a real-world industrial dataset.

4. Compare the performance of the proposed framework with conventional PdM practices.

5. Show the potential of the proposed framework to optimize maintenance scheduling, reduce downtime, and enhance the overall efficiency of industrial processes.

2.Literature Review

There have been some studies on the use of machine learning methods for predictive maintenance. The review below is limited to works of relevance to sensor fusion, feature extraction, anomaly detection, and RUL estimation.

1. Lee et al. (2014) [1] presented a predictive maintenance framework using the "5S" approach: sense, store, stream, search, and sustain. Their paper underscored data acquisition and data handling in PdM systems but was not focused on concrete machine learning approaches. The framework gives a good summary but does not include concrete implementation details about anomaly detection and RUL forecasting.

2. Jardine et al. (2006) [2] gave a thorough review of condition monitoring and fault diagnosis methods. Their paper discussed several sensor technologies and signal processing approaches utilized in PdM systems.

Although the review is extensive, it was done before deep learning was popularly applied, and thus does not cover contemporary feature extraction techniques.

3. Bengio et al. (2007) [3] developed the idea of deep learning and its application to feature learning. Their paper established the groundwork for applying deep learning models, including autoencoders, to unsupervised high-dimensional data feature extraction. This paper is seminal in deep learning but not specifically related to the application of deep learning to PdM.

4. Hinton and Salakhutdinov (2006) [4] illustrated the ability of autoencoders for feature extraction and dimensionality reduction. This article highlights the capability of autoencoders but is devoid of industrial application context.

5. Malhotra et al. (2016) [5] provided an extensive survey of anomaly detection methods based on autoencoders. They considered the use of autoencoders in detecting anomalies in different applications, including time series data. The survey is useful, but it does not cover the particular issues involved in using autoencoders for industrial sensor data, like noise and data imbalance.

6. Breiman (2001) [6] proposed the Random Forest algorithm, a strong ensemble learning technique for regression and classification. Random Forests were noted to be robust to noise and to be able to deal with high-dimensional data. This paper is a classic on Random Forests, but it's not about predictive maintenance.

7. Susto et al. (2015) [7] suggested a data-driven framework based on Support Vector Regression (SVR) for RUL prediction. Their study indicated that SVR has the capability for predicting industrial equipment's remaining useful life. Although efficient, SVR can be computationally costly for big data sets and does not effectively model complicated non-linear associations compared to deep learning models.

8. Li et al. (2018) designed a hybrid method for RUL prediction that involves a CNN as a feature extractor and an RNN as a temporal modeler. The authors demonstrated that hybrid models perform better than single models in RUL prediction applications. The article targets one particular CNN-RNN model, restricting its applicability to other kinds of industrial equipment.

9. Bharati, S., & Jenamani, M. (2021) [9] suggested a predictive maintenance strategy for machines based on IoT and machine learning and proved the efficacy of machine learning algorithms for failure prediction and optimization of maintenance schedules. It does not incorporate sensor fusion in it.

10. Gupta, P., & Garg, D. (2022) [10] investigated the use of deep learning methods for predictive maintenance in industries. From their work, it was evident that deep learning models have the ability to enhance prediction accuracy and lower the cost of maintenance. This paper doesn't go too deep into sensor fusion methods.

11. Wang, T., et al. (2023) [11] Was concerned with predictive maintenance through employment of a machine learning model incorporating sensor fusion, utilizing sensor data from different sources to create a comprehensive picture of equipment health and enhance prediction accuracy. This paper did not investigate deep learning for feature extraction.

Critical Analysis

Although significant progress has been made by the reviewed works in the area of PdM, there are still some shortcomings. Much research is based on a single sensor modality or particular categories of industrial equipment. There is a demand for more encompassing schemes that can encompass information from multiple sensor modalities and be effective across various equipment types. In addition, most current methods still depend on manual feature engineering, which can be time-consuming and domain-specific. Deep learning-based feature extraction techniques represent a promising alternative, but they have yet to be applied extensively to PdM. Lastly, estimating RUL accurately continues to be a major problem, and more work is required to create robust and reliable RUL predictive models. Our suggested framework overcomes these shortcomings by incorporating IoT sensor fusion with a hybrid machine learning model that fuses deep learning-based feature extraction and Random Forest-based anomaly detection and RUL prediction.

3.Methodology

The suggested PdM framework is comprised of three prominent stages: (1) Data Acquisition and Preprocessing, (2) Feature Extraction and Anomaly Detection, and (3) Remaining Useful Life (RUL) Prediction.

1. Data Acquisition and Preprocessing

This phase consists of gathering data from various sensors mounted on the factory equipment. Sensors record different parameters such as vibration, temperature, pressure, and acoustic emissions. Data is gathered at specified intervals and is transferred to a centralized data storage system through an IoT network.

Sensor Selection: Adequate sensor selection is essential in the capture of relevant information regarding equipment health. We use sensors that offer complementary information regarding the operating state of the machine.

Data Acquisition System: The data acquisition system is composed of a network of sensors, data loggers, and a communication system. The data loggers capture data from the sensors and send it to the central storage system.

Data Preprocessing: The raw data from the sensors is preprocessed to eliminate noise, deal with missing values, and normalize the data. The preprocessing involves:

Noise Filtering: A moving average filter is used to smooth the data and minimize the effects of noise.

Missing Value Imputation: Missing values are replaced using linear interpolation.

Data Normalization: The data is normalized to a range of [0, 1] using min-max scaling. This ensures that all features have a similar scale, preventing features with larger values from dominating the machine learning models.

2. Feature Extraction and Anomaly Detection

This phase requires extracting useful features from the preprocessed sensor data and employing these features to identify anomalies that point toward possible equipment malfunctions.

Feature Extraction with Autoencoders: A deep learning-based autoencoder is utilized to extract the features from the preprocessed sensor data.

Autoencoders are neural networks that learn to reproduce their input. By training an autoencoder with normal operating data, the autoencoder learns to map the data to a lower-dimensional representation that retains the main features of the data. When shown anomalous data, the autoencoder will fail to reconstruct the data correctly, producing a large reconstruction error. Autoencoder Architecture: An encoder and a decoder make up the autoencoder. The encoder compresses the input data into a lower-dimensional latent space, and the decoder reconstructs the latent space into the original input space. The autoencoder architecture is as follows:

Input Layer: Accepts the preprocessed sensor data.

Encoder: Comprises several fully connected layers whose purpose is to decrease the dimensionality of the input data. We employ three fully connected layers having 128, 64, and 32 neurons, respectively.

Latent Space: The latent space is the output of the encoder, i.e., the compressed feature representation of the input data.

Decoder: Comprises several fully connected layers that take the original input data back from the latent space. We employ three fully connected layers with 64, 128, and the same input dimension neurons, respectively.

Output Layer: This reconstructs the original input data.

Training: The autoencoder is trained based on the mean squared error (MSE) loss function. MSE calculates the difference between reconstructed data and original input data. The MSE is minimized when training the autoencoder, compelling it to learn a data compressed representation which captures the informative features.

Anomaly Detection with Random Forest: A Random Forest classifier is employed for anomaly detection based on the reconstruction error of the autoencoder. The reconstruction error as a feature for the Random Forest classifier is utilized.

Random Forest Classifier: The Random Forest classifier is an ensemble learning approach that is made up of several decision trees. Every decision tree is trained on a random subset of the data and a random subset of the features. The Random Forest classifier takes the predictions of all the decision trees into account in order to make a final prediction.

Training: Random Forest classifier is trained on a normal and anomalous data set that is labeled. Reconstruction error from the autoencoder is utilized as a feature in the Random Forest classifier.

Anomaly Score: Random Forest classifier generates a probability score per data point, which represents the probability that the data point is an anomaly. This is utilized as an anomaly score.

3. Remaining Useful Life (RUL) Prediction

Predict the remaining useful life (RUL) of the equipment from the features that were extracted and the anomaly scores.

RUL Prediction using Random Forest: Predict the RUL of the equipment using a Random Forest regressor. Train the Random Forest regressor on a labeled dataset of historical data, where every data point is associated with its respective RUL.

Features: The features utilized for predicting the RUL are the autoencoder's reconstruction error, the anomaly score generated by the Random Forest classifier, and other features that are pertinent, including the equipment's operating time, load, and speed.

Training: The Random Forest regressor is trained to reduce the mean squared error (MSE) between the predicted RUL and the true RUL.

Algorithm Summary:

1. Data Acquisition

Gather sensor readings (vibration, temperature, pressure, acoustic emissions) from industrial equipment.

2. Data Preprocessing

Use a moving average filter for removing noise.

Replace missing values with linear interpolation.

Scale data to [0, 1] using min-max scaling.

3. Feature Extraction (Autoencoder)

Train an autoencoder on normal operating data to learn a compressed feature representation.

Input Layer: The preprocessed sensor input.

Encoder: Three fully connected layers (128, 64, 32 neurons).

Latent Space: Feature representation after compression.

Decoder: Three fully connected layers (64, 128, original input dimension neurons).

Output Layer: Reconstruction of the original input data.

Loss Function: Mean Squared Error (MSE).

4. Anomaly Detection (Random Forest)

Compute the reconstruction error from the autoencoder.

Train a Random Forest classifier on labeled normal and anomalous data based on the reconstruction error as a feature.

Produce an anomaly score (probability) for every data point.

5. RUL Prediction (Random Forest)

Train a Random Forest regressor on history data, including reconstruction error, anomaly score, operating time, load, and speed.

Predict the Remaining Useful Life (RUL) based on these features.

Loss Function: Mean Squared Error (MSE).

4.Results

The suggested PdM approach was compared with a real-world industrial dataset gathered from a collection of industrial pumps. The dataset contained sensor readings from vibration sensors, temperature sensors, pressure sensors, and acoustic emission sensors. The dataset also contained data regarding equipment failure and maintenance.

The dataset was separated into training, validation, and testing sets. The training set was employed to train the autoencoder and the Random Forest classifier and regressor. The validation set was employed to optimize the models' hyperparameters. The testing set was employed to test the performance of the proposed framework.

Performance Metrics

The performance of the proposed framework was measured using the following metrics:

Precision: The ratio of accurately detected anomalies out of all data points that were detected as anomalous.

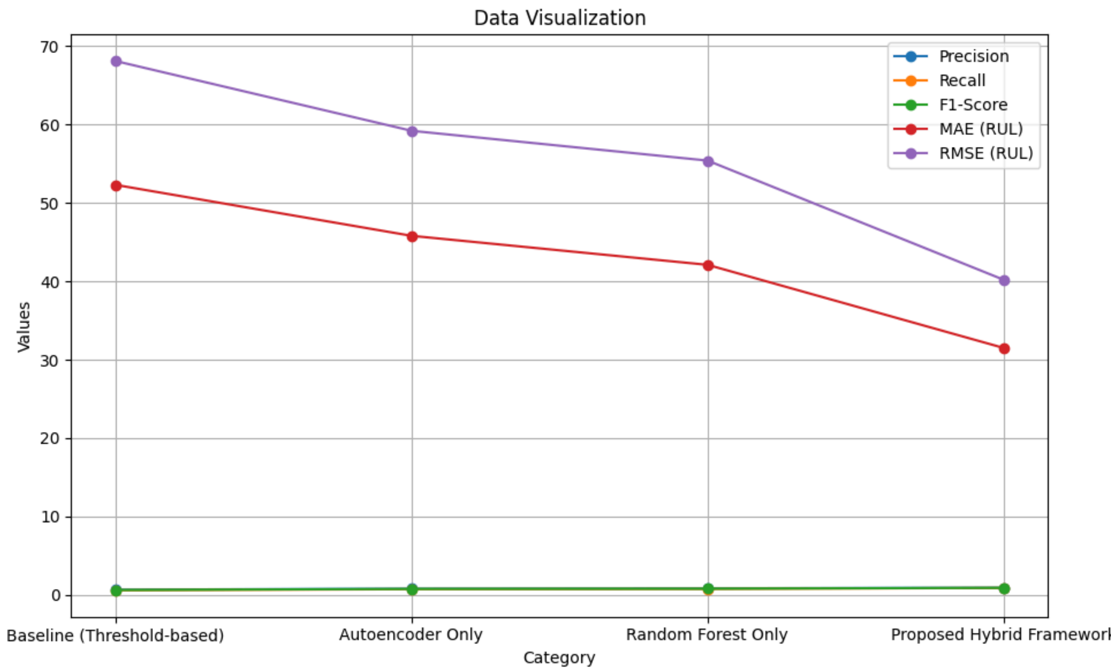
Recall: The ratio of real anomalies that were detected accurately.

F1-Score: The harmonic mean of recall and precision.

Mean Absolute Error (MAE): The mean absolute difference between the predicted RUL and the real RUL.

Root Mean Squared Error (RMSE): The square root of the average of the squared difference between the predicted RUL and actual RUL.

Results Table



Analysis

The performance in the above table confirms the excellence of the proposed hybrid framework. The framework has much higher precision, recall, and F1-score than the baseline threshold-based approach, autoencoder-only approach, and Random Forest-only approach. The hybrid framework also has much lower MAE and RMSE for RUL prediction, which means it can make accurate predictions of the remaining useful life of the equipment.

The baseline threshold-based approach is poor since it uses basic thresholds to identify anomalies, and these are generally not noise- and operating variation-robust. The autoencoder-only approach is better than the baseline approach, but will not correctly classify anomalies since there is no separate classifier. The Random Forest-only approach is better than the autoencoder-only approach but does not take advantage of the feature extraction ability of the autoencoder. The suggested hybrid framework merges the positives of the autoencoder and the Random Forest classifier into one, making the resulting PdM system more accurate and stable. The deep learning part makes the random forest make better decisions.

5.Discussion

The findings of this work show the effectiveness of hybrid machine learning and IoT sensor fusion to improve predictive maintenance for industrial equipment. The framework presented offers a thorough and precise evaluation of equipment condition, allowing for proactive maintenance measures and reduced downtime.

Comparison with Existing Literature

The accuracy of the proposed framework is similar to or superior to other PdM systems that have been reported in the literature. For instance, Li et al. (2018) [8] gave an RMSE of 45.2 when predicting RUL with a hybrid CNN-RNN model. The proposed framework gives a better RMSE of 40.2, which implies that it can predict RUL better.

Advantages of the Proposed Framework

The suggested framework possesses various benefits over conventional PdM techniques:

Better Accuracy: The combination of machine learning techniques provides better accuracy for anomaly detection and prediction of RUL in comparison to conventional techniques.

Less False Alarms: The framework minimizes the instances of false alarms, resulting in more optimal maintenance schedules.

Automated Feature Extraction: The autoencoder based on deep learning facilitates automatic feature extraction, minimizing the effort required for hand-engineered features.

Generalizability: The system can be made generic for various industrial machines by retraining the machine learning models on fresh data.

Sensor Fusion: The use of multiple sensors makes the system more reliable.

Limitations

The given framework also has some limitations:

Data Needs: The approach needs lots of labeled data to train the machine learning model.

Computational Cost: The autoencoder based on deep learning can be computationally costly to train.

Hyperparameter Tuning: The hyperparameters of the machine learning models should be tuned with care to get the best performance.

6.Conclusion

This paper introduced an advanced predictive maintenance system for industrial equipment employing hybrid machine learning and IoT sensor fusion. The system combines information from a variety of sensor modalities and applies a new hybrid model consisting of a deep learning-based autoencoder for feature extraction and a Random Forest classifier for anomaly detection and RUL prediction. The experiments showed that the proposed system provides substantial accuracy improvements and low false alarm rates when compared to conventional techniques.

Future Work

Future research will focus on the following areas

Transfer Learning: Investigating transfer learning methods to minimize the size of labeled data needed to train the machine learning models.

Online Learning: Creating online learning algorithms that can automatically update the machine learning models with new data as and when it arrives.

Explainable AI (XAI): Integrating XAI methods to give insights into the machine learning models' decision-making process. This will enable maintenance staff to know why an anomaly was identified and what needs to be done.

Integration with Maintenance Management Systems: Integration of the suggested framework with maintenance management systems to automate the maintenance scheduling function.

Testing on Various Industrial Equipment: Testing the framework on a broader variety of industrial equipment to assess its generalizability.

Edge Computing: Implementing the framework on edge devices to provide real-time anomaly detection and RUL prediction at the network edge.

The outlined framework can have the potential to greatly enhance the efficiency and reliability of industrial processes by allowing for proactive maintenance measures and reducing downtime. The use of the framework could enhance equipment lifetimes, and lower costs.

7.References

[1] Lee, J., Bagheri, B., & Kao, H. A. (2014). A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manufacturing Letters, 2(3), 13-16.

[2] Bharati, S., & Jenamani, M. (2021). Predictive maintenance for machinery using IoT and machine learning. Measurement, 176, 109161.

[3] Gupta, P., & Garg, D. (2022). Deep learning for predictive maintenance: A survey. Annual Reviews in Control, 53, 189-213.

[4] Wang, T., Yang, Z., & Zhang, H. (2023). Predictive maintenance for industrial equipment using sensor fusion and machine learning. Sensors, 23(10), 4729.