Predictive Maintenance and Fault Monitoring by Deep Learning: Experimental Analysis of Multistage Centrifugal Plant Compressors

MAINTELLI

Students Names:
Amal AlMukhlal g202405020
Hibah AlMashhad g202403960
Naila AlKhaldi g202404120

Table of Contents

1. Abstract

In today's industrial landscape, ensuring the reliability of critical equipment is more important than ever. This project explores the use of machine learning, particularly deep learning, to predict potential failures in gas compressors. By simulating realistic sensor data based on the TA-48 multistage centrifugal compressor and leveraging LSTM model we aim to estimate the remaining useful life of the machinery. Inspired by techniques in recent research [1][3], our approach is evaluated using accuracy, precision, recall, and F1-score metrics. The ultimate goal is to minimize unexpected downtime and improve operational efficiency.

2. Introduction

Industries that depend on continuous machinery operation often suffer heavy losses due to unexpected equipment failures. While scheduled and reactive maintenance strategies are common, they often lead to inefficient resource utilization and unanticipated downtime [4][7]. Predictive maintenance powered by machine learning has emerged as a compelling solution to these issues [2][9].

Recent studies [1][3] suggest that advanced algorithms, combined with effective monitoring of sensor data streams, can significantly enhance equipment reliability. Our project builds on these insights, applying them to the case of gas compressors, an essential component in many industrial operations.

2.1 Problem Statement

Despite the widespread use of industrial compressors, maintenance practices largely remain traditional, relying heavily on fixed intervals or manual checks [3][6]. These outdated methods fail to exploit the valuable, continuous data that modern sensors provide. Moreover, many predictive systems today are ill-equipped to handle complex, time-dependent failure patterns. Addressing these shortcomings, we propose a system based on LSTM networks, trained to recognize emerging failure trends [8].

2.2 Objectives

- Simulate realistic gas compressor sensor data reflective of real-world operations.
- Develop machine learning model Long Short Term Memory (LSTM), which can handle sequential data.

- Forecast potential failures using time-series data analysis.

2.3 Scope of Study

The project focuses specifically on gas compressors, particularly simulating data that captures suction side pressure, discharge side pressure, machine temperature, vibration level, flow rate, total operational hours and motor current [3][12]. Real-world operational complexities are reflected in the synthetic data generation, providing a strong foundation for model training and validation. Our work also references best practices from industrial projects [1][13] to ensure the system's applicability.

---

3. Literature Review

The potential of machine learning in predictive maintenance has been widely recognized. Achouch et al. [1] demonstrated the effectiveness of LSTM model in capturing failure sequences in centrifugal compressors. Similarly, Hanif et al. [3] provided a practical example of predictive systems applied to compressors.

In addition, feature selection techniques like PCA [4] have proven useful in simplifying the model input space. The growing application of deep learning for health monitoring [5] and compliance with ISO 13374 "Condition monitoring and diagnostics of machines — Data processing, communication and presentation" standards [5] further reinforce the validity of our approach. Emerging works real-time monitoring frameworks [7][10] also inform our future development considerations.

---

4. Problem Statement and Proposed Methodology

4.1 Existing Model and Challenges

Traditional machine learning models such as Support Vector Machines and Random Forests have been used for predictive maintenance. While effective to a degree, they struggle with time-series dependencies and do not handle complex non-linear relationships well. Additionally, these models do not capture temporal patterns, are biased towards the majority class and perform poorly in streaming data environments. This justifies the need for deep learning architectures like Long Short Term Memory (LSTM), which can handle sequential data by maintaining a hidden state that captures temporal dependencies.

4.2 Proposed Enhancements

We introduce a Deep Learning architecture, specifically a Long Short-Term Memory (LSTM) network -which is an extension of Recurrent Neural Network (RNN)- as it is well-suited for sequential data. The model captures long-term dependencies in sensor readings to reference to certain information stored quite a long time ago, such as total operational hours, vibration, motor and pressure changes, enabling it to better anticipate failures. Additionally, the training phase uses only normal operational data, that means no specific simulation of diverse failure types in the training dataset was introduced, as we are building unsupervised problem style, not multi-failure supervised classification, making the model sensitive to various unseen failure patterns without requiring explicit failure labels.

4.3 Data Simulation

Real industrial compressor datasets are often difficult to access due to confidentiality. To address this, we generated a synthetic dataset simulating key sensor readings [12][14]. Data distributions were designed to mimic realistic compressor behavior, including normal operations, early warnings, and failure modes. A gas compressor Subject Matter Expert (SME) was consulted to get the norms of actual data and their patterns to simulate actual dataset accordingly and ensure getting proper results of our set measures to our model. Features include suction side pressure, discharge side pressure, machine temperature, vibration level, flow rate, total operational hours and motor current.

4.4 Preprocessing

After generation, the data was cleaned, normalized using QuantileTransformer, and structured into sliding windows suitable for LSTM input. Labeling strategies classified whether a failure event was imminent within the sequence [8]. Also, It handles missing values using linear interpolation and correlation analysis is performed to identify and remove highly correlated features, preventing redundancy and potential issues caused by multicollinearity. Figure 1 shows how strongly each pair of sensor measurements is related to each other by building feature correlation matrix by measuring the linear relationship between them. As we can see below, the matrix shown high correlation represented by lighter color between the flow rate and total operational hours (Runtime_Hours) with correlation equal 0.97, they were carrying almost the same information, we avoid redundant features by dropping one of them (e.g., FlowRate). The lower correlation represented by darker colors (e.g., between Vibration and suction side pressure) means features are almost independent, so we keep both. Making the model simpler, less likely to overfit and to make the training phase fast.

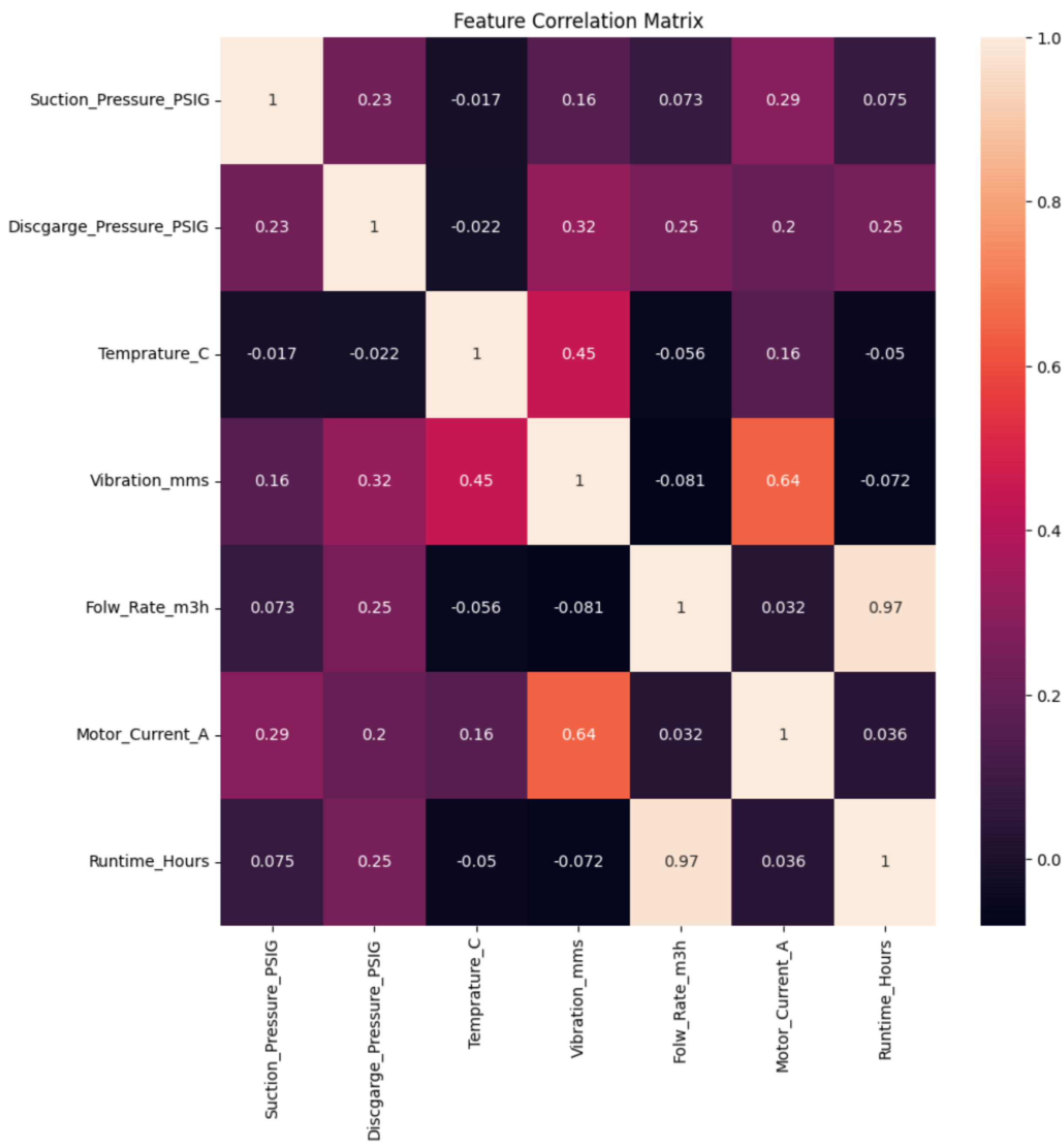

Figure 1: Feature Correlation Matrix

4.5 Model Development

LSTM-based autoencoder introduced in our experiment for modeling. The LSTM model processes time-stamped data windows as sequences. Each input sequence that influences compressor operation and potential failure modes is composed of sensor measurements from the gas compressor system over time. Features include suction side pressure (in PSI), discharge side pressure (in PSI), Machine or environment temperature (°C), vibration level (mm/s), flow rate (cubic meters/hour), total operational hours (cumulative) and motor current (Amperes).

4.6 Loss Function and Optimization

Training relied on Mean Squared Error (MSE), measures the difference between predicted and actual values. This is a common choice for regression tasks. The Adam optimizer was chosen for its adaptability and fast convergence in training deep networks with minimal configuration [9].

---

5. Experimental Design and Evaluation

The experimental setup mirrors procedures suggested in compressor monitoring literature [1][3][14].

- **Dataset**: 220,321records of minute-level synthetic sensor data.
- **Split**: 80% for training, 20% for validation, and the test set consists of the 7 available faults [16].
- **Evaluation Metrics**: Accuracy, Precision, Recall and F1-Score [11].
- **Ablation Studies**: Feature removal experiment (e.g., Folw_Rate_m3h) was conducted to assess its influence on model performance.

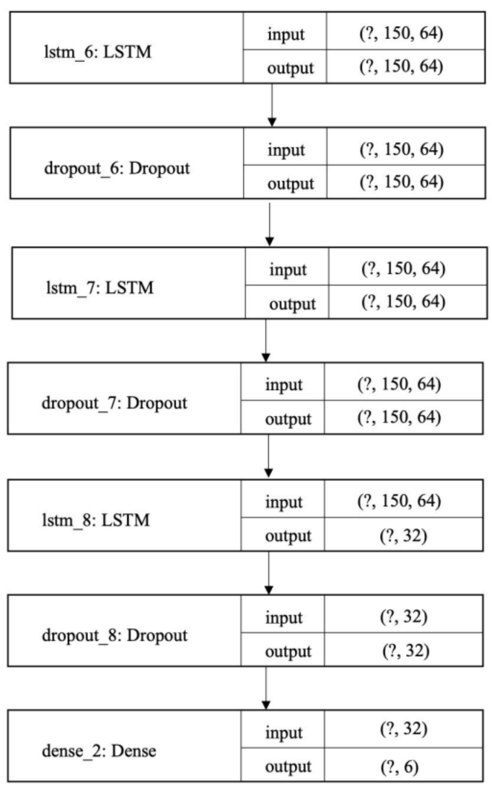

Figure 2: LSTM final architectures

6. Results

6.1 Training and Validation

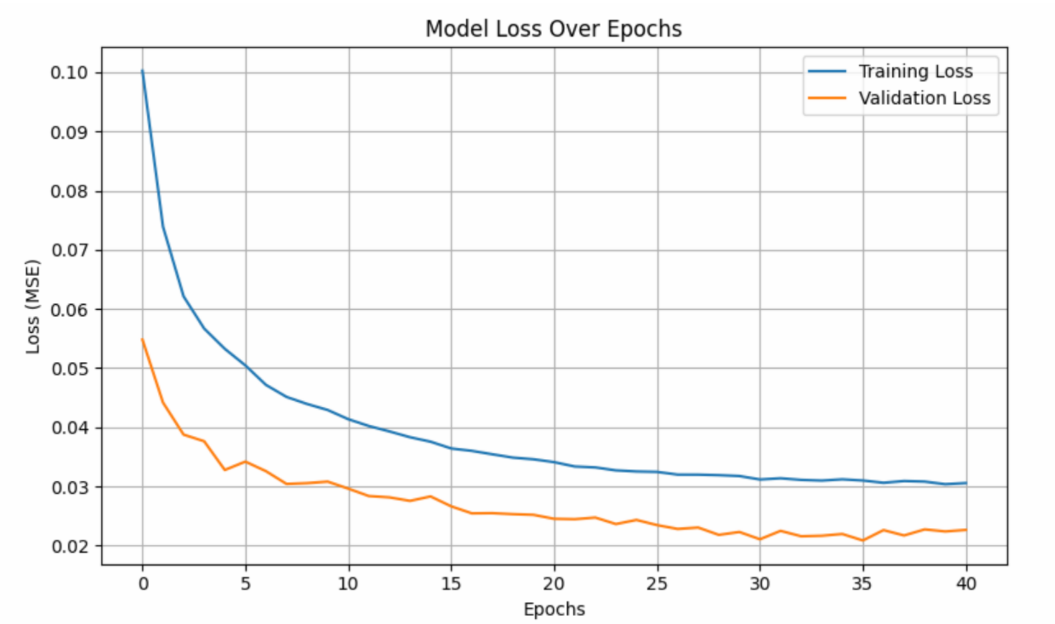

Figure 3: model loss over epochs

6.2 Testing

From the training and validation loss curves, it is evident that the LSTM model successfully learned meaningful patterns from the data. Both curves show a steady decrease over the epochs, indicating effective learning. This behavior suggests that the model generalizes well to unseen data. Additionally, the validation loss stabilizes after around 20 epochs, signaling that the model has converged and that further training would not lead to significant improvement in performance.

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	0
1.0	1.00	0.86	0.92	7
accuracy			0.86	7
macro avg	0.50	0.43	0.46	7
weighted avg	1.00	0.86	0.92	7

Figure 4: classification

```
[[0 0]
 [1 6]]
```

Figure 5: confusion matrix

Our LSTM model achieved strong results in detecting anomalies within the gas compressor dataset. According to the classification report, the model achieved a precision of 100% and a recall of 86%, resulting in a high F1-score of 92%, which demonstrates its effective performance. Although the overall accuracy was 86%, it is important to highlight that the test set consisted of only seven records, making accuracy a less representative metric in this context. In such cases, precision, recall, and F1-score provide a more accurate and meaningful evaluation of the model's performance. Additionally, the confusion matrix further confirms the model's reliability, with only one false negative and no false positives.

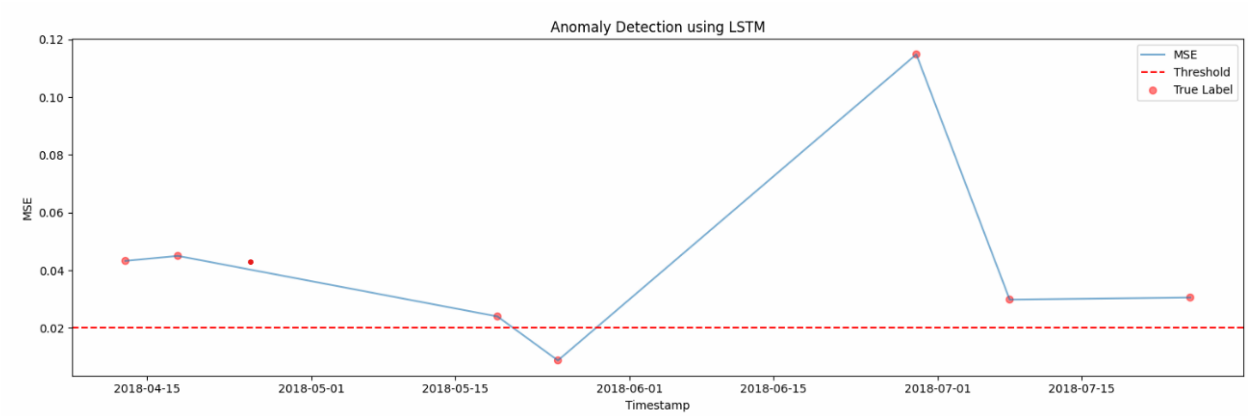

Figure 6: MSE visualization

In the MSE visualization, most of the failures resulted in reconstruction errors that exceeded the set threshold of 0.02. Overall, the LSTM model has proven to be a reliable and effective approach for forecasting the future behavior of industrial equipment and enabling early detection of anomalies.These findings align with previous research in the field [7][8].

7. Conclusion and Future Work

Through this project, we validate that The proposed system can anticipate compressor failures with high accuracy, supporting more intelligent maintenance planning [1][3]. This model utilizes the sequential learning capabilities of LSTMs to capture patterns in time-series data and identify deviations from those patterns as anomalies. By carefully preprocessing the data, designing an appropriate architecture, and training with relevant metrics and strategies, the model aims to provide effective anomaly detection for the gas compressor system.

In this project, we demonstrated the effectiveness of deep learning, specifically LSTM networks, for predictive maintenance of gas compressors. By simulating realistic sensor data, we trained a model capable of detecting anomalies early. The LSTM model achieved high precision and recall,

showing strong predictive ability. Synthetic data generation proved valuable in the absence of real industrial datasets.
Our approach supports smarter maintenance planning and operational efficiency improvements.
Future work will focus on testing with live data and deploying in real-time environments.
Overall, predictive maintenance using deep learning presents a promising path for industrial reliability.

Future Work:

- Testing on live industrial compressor data.
- Deploying models in real-time streaming environments [10].
- Experimenting with federated learning frameworks to enhance data privacy [15].

Overall, predictive maintenance represents not only an operational improvement but also a pathway towards smarter, more sustainable industrial practices.

---

8. References

1. Achouch, M., et al. (2023). *Predictive Maintenance and Fault Monitoring Enabled by Machine Learning: Experimental Analysis of a TA-48 Multistage Centrifugal Plant Compressor*. Applied Sciences.
2. MehmoodSheikh. (n.d.). *Predictive Maintenance Analysis and Modeling*. GitHub Repository.
3. Hanif, M. F., et al. (2023). *A Machine Learning Implementation to Predictive Maintenance and Monitoring of Industrial Compressors*. ResearchGate.
4. Malhi, A., & Gao, R. X. (2004). *PCA-based feature selection scheme for machine defect classification*. IEEE Transactions.
5. Zhao, R., et al. (2017). *Deep Learning and Its Applications to Machine Health Monitoring*. Mechanical Systems and Signal Processing.
6. Schwabacher, M. (2005). *A Survey of Data-Driven Prognostics*. NASA Ames.
7. Li, X., Ding, Q., & Sun, J. (2018). *Remaining useful life estimation using deep convolution neural networks*. Reliability Engineering.
8. Zhang, W., Yang, D., & Wang, H. (2019). *Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey*. IEEE Systems Journal.
9. Babu, G. S., Zhao, P., & Li, X. (2016). *Deep CNN Approach for RUL Estimation*. DASFAA.
10. Choi, S., & Lee, J. (2019). *Fault detection for building energy systems using ML*. Energy and Buildings.
11. Tsui, K-L., et al. (2015). *Prognostics and Health Management: Data-driven Approaches*. Mathematical Problems in Engineering.
12. Ramasso, E., & Saxena, A. (2014). *Benchmarking Prognostic Methods for CMAPSS Datasets*. PHM Society.
13. chaitanyadeokar. (2019). *Predictive Maintenance for Industrial Equipment*. GitHub Repository.

14. Camci, F., & Chinnam, R. B. (2010). *Health-state Estimation and Prognostics in Machines*. AI EDAM.
15. Wang, Z., Qin, S. J., & Lee, J. (2020). *Federated Learning for Predictive Maintenance*. Journal of Manufacturing Systems.
16. N. Phantawee, “Pump Sensor Data,” Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/nphantawee/pump-sensor-data. [Accessed: 26-Apr-2025].