Research Article

A Multimodal Deep Learning Framework for Predicting Machine Anomalies Using IoT-Enabled Vibration and Sound Data

Alper SAYLAM[1*], Mehmet Ayberk ÇAKAR[2*], Haluk ATLI[3*]

[1] Supply Chain Wizard, Orcid ID: https://orcid.org/0000-0002-8929-541X, E-mail: alper.saylam@supplychainwizard.com
[2] Supply Chain Wizard Orcid ID: https://orcid.org/0000-0002-1391-259X, E-mail: ayberk.cakar@supplychainwizard.com
[3] Supply Chain Wizard, Orcid ID: https://orcid.org/0000-0003-2228-6305, E-mail: haluk.atli@supplychainwizard.com
[*] Correspondence: alper.saylam@supplychainwizard.com; Tel.: (+90 507 491 22 55)

Received: 01 July 2025
Revised: 02 October 2025
2nd Revised: 05 November 2025
Accepted: 08 December 2025
Published: 15 December 2025

Reference: Saylam, A., Çakar, M. A., & Atlı, H. (2025). A multimodal deep learning framework for predicting machine anomalies using IoT-enabled vibration and sound data. The European Journal of Research and Development, 5(1), 494–501.

Abstract

Unplanned machine downtimes caused by component failures, overheating, or mechanical stress significantly impact manufacturing efficiency and profitability. Predicting such failures before they occur is a core objective of smart manufacturing and Industry 4.0. Leveraging recent advances in sensor technology and machine learning, this study proposes an anomaly detection architecture that predicts the operational state of manufacturing machines one step ahead, enabling early detection of potential downtime.

The system integrates two primary data sources: vibration signals collected by an IoT-enabled device and sound recordings obtained from a microphone positioned close to the manufacturing equipment. These complementary signals capture the machine's dynamic behaviour under varying operational conditions. While vibration and line status data are directly utilized, sound recordings undergo pre-processing using a low-pass filter to remove irrelevant background noise. The filtered *recordings are segmented into one-minute intervals, and statistical features are extracted in both time and frequency domains, including mean, standard deviation, skewness, and kurtosis. Since the available dataset covers only one day, a moving block bootstrap technique is employed to improve robustness and generalization.*

Two deep learning architecture, Long Short-Term Memory (LSTM) and Multi-Layer Perceptron (MLP), are implemented to forecast the machine state at time t + 1. The dataset, consisting of nine features and approximately 13,200 samples, is divided into training, validation, and test sets in a 70/15/15 ratio. Both models are trained using the Adam optimizer and binary cross-entropy loss. Performance is evaluated using precision, recall, and F1 score metrics.

Overall, the proposed approach demonstrates that combining vibration and acoustic data with deep learning can effectively predict machine anomalies in real time, contributing to proactive maintenance and reduced production downtime in smart manufacturing environments.

Keywords: Predictive Maintenance, Anomaly Detection, Vibration and Acoustic Analysis, Deep Learning Models, Smart Manufacturing

1. Introduction

In modern manufacturing environments, ensuring continuous machine operation is critical to maintaining productivity and minimizing production losses. Unexpected machine downtimes often caused by factors such as component failures, excessive temperature, or mechanical stress can lead to significant inefficiencies and financial losses. Therefore, the early detection and prediction of potential machine failures have become essential components of smart manufacturing and Industry 4.0 initiatives.

Recent advances in machine learning and sensor technologies have enabled data-driven approaches for predictive maintenance and anomaly detection. By leveraging real-time sensor data such as vibration, temperature, and acoustic signals, machine learning models can identify abnormal patterns that precede equipment failures. Among these modalities, sound and vibration signals are particularly valuable, as they directly reflect the dynamic behavior of machines under varying operational conditions.

In this study, we propose a machine learning–based anomaly detection system designed to predict the operational state of manufacturing equipment one step ahead. The system integrates two primary data sources: vibration signals captured by an IoT-enabled Hexbox [1] device and sound recordings collected via a microphone positioned above the machine. The extracted features from these data streams are used to train and evaluate two deep learning architectures: a Long Short-Term Memory (LSTM) [2] network and a Multi-Layer Perceptron (MLP) [3]. The performance of these models is assessed based on precision, recall, and F1 score metrics to determine their effectiveness in detecting anomalies that may lead to downtime.

2. Materials and Methods

In manufacturing environments, machine downtimes caused by unexpected reasons such as component failures, excessive temperature, or humidity are highly critical. Predicting the causes of downtime in a proactive manner is one of the prominent topics in the literature. In this paper, we refer to these unexpected machine downtimes as *anomalies* and propose a machine learning–based anomaly detection architecture. In our system, the downtime state is referred to as an *anomaly*, while the running state is considered *normal*.

In this section, we explain the proposed architecture, which predicts the anomaly state of machines in manufacturing environments. The architecture utilizes sensor data such as humidity and sound recordings captured by a microphone. As shown in Fig. 1, the model takes vibration and sound data at time *t* and predicts the anomaly state of the machine at *t + 1*.

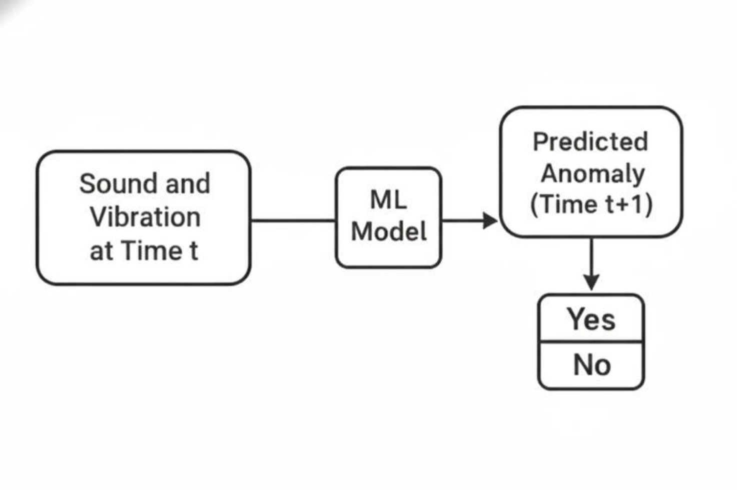


Figure 1: Overall Architecture

2.1. Data Collection

The designed system incorporates two distinct data sources. Vibration data are collected using a Hexbox, an IoT device that can be integrated with various sensors [1]. Sound recordings are captured by a microphone positioned above the CNC machine, as illustrated in Fig. 2. These audio recordings are stored in cloud in five-minute segments and saved in the opus format. In addition, the machine's operational status (running or down) is obtained through the integration of the Hexbox and the Andon light system.

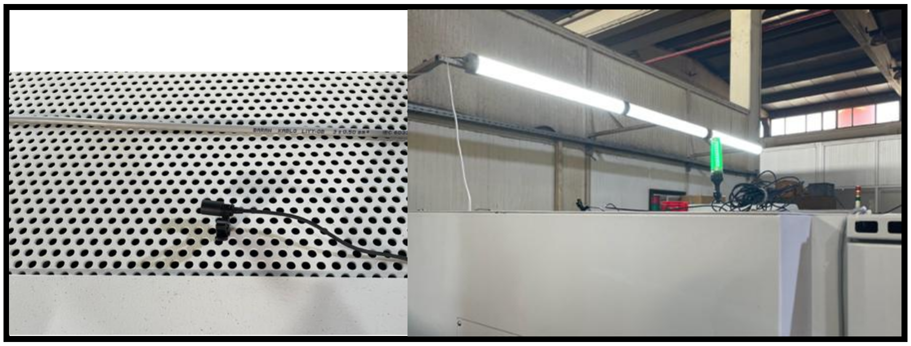
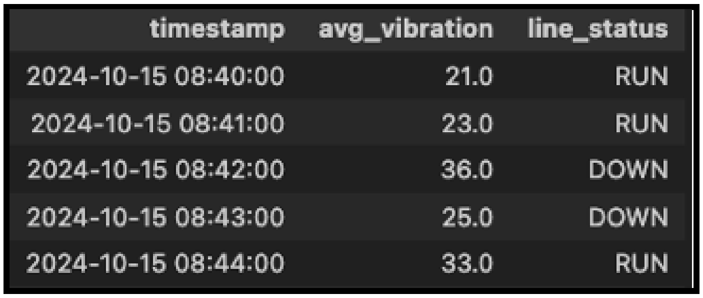


Figure 2: Data Collection Setup



timestamp	avg_vibration	line_status
2024-10-15 08:40:00	21.0	RUN
2024-10-15 08:41:00	23.0	RUN
2024-10-15 08:42:00	36.0	DOWN
2024-10-15 08:43:00	25.0	DOWN
2024-10-15 08:44:00	33.0	RUN

Figure 3: Vibration Values and the Line Status

Fig. 3 illustrates sample data from the vibration sensor and the machine's line status. The timestamp interval for these values is one minute. However, the timestamp interval for the sound recordings is five minutes, as shown in Fig. 4. In Section 2.2, we describe the process of converting the sound recordings into one-minute segments and extracting relevant features.

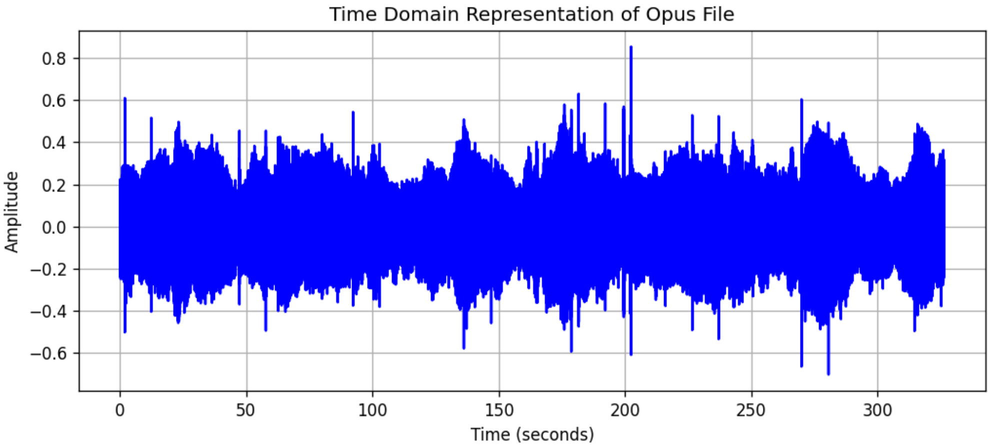
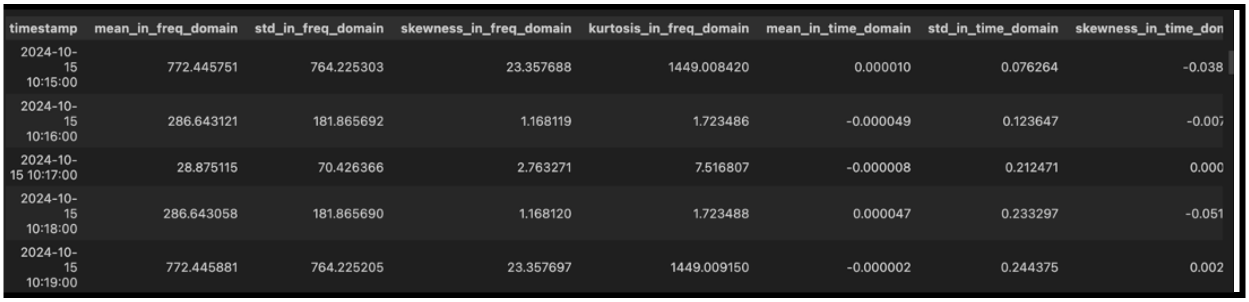


Figure 4: Time Domain Representation of Sound File

2.2. Feature Selection and Preprocessing

While vibration data and line status information are directly utilized, several preprocessing steps are applied to the sound recordings before feature extraction. A low-pass filter is first applied to the audio signals to eliminate human voices and other background noises, preserving only the relevant machine sounds. Subsequently, each sound file is segmented into one-minute windows. For each window, statistical features are computed in both the time and frequency domains, including the mean, standard deviation, skewness, and kurtosis. Some of the extracted features from the sound recordings are illustrated in Fig. 5. Since the dataset covers only a single day of operation, a moving block bootstrap method is employed to increase the robustness and generalizability of the model.



timestamp	mean_in_freq_domain	std_in_freq_domain	skewness_in_freq_domain	kurtosis_in_freq_domain	mean_in_time_domain	std_in_time_domain	skewness_in_time_dom
2024-10-15 10:15:00	772.445751	764.225303	23.357688	1449.008420	0.000010	0.076264	-0.038
2024-10-15 10:16:00	286.643121	181.865692	1.168119	1.723486	-0.000049	0.123647	-0.007
2024-10-15 10:17:00	28.875115	70.426366	2.763271	7.516807	-0.000008	0.212471	0.000
2024-10-15 10:18:00	286.643058	181.865690	1.168120	1.723488	0.000047	0.233297	-0.051
2024-10-15 10:19:00	772.445881	764.225205	23.357697	1449.009150	-0.000002	0.244375	0.002

Figure 5: Features extracted from sound data

2.3. Architectures of Selected Models

In this study, the proposed models aim to predict the machine line's state one step ahead (one minute into the future) using the features described in the data preprocessing section. By detecting potential failures just before the transition to a downtime state, the system can help prevent production losses. To achieve this objective, two deep learning architectures are employed: Long Short-Term Memory (LSTM) and Multi-Layer Perceptron (MLP) networks [2-3].

As shown in Table 1, the LSTM model consists of three layers. The first layer includes 50 LSTM units, followed by two fully connected layers with 10 neurons each. Since the model performs binary classification of the line status, the output layer contains a single neuron. The model is trained for 100 epochs using the Adam optimizer.

Similarly, as presented in Table 1, the MLP model comprises three layers with 64, 32, and 1 neurons, respectively. The ReLU activation function is applied to the input and hidden layers. The training parameters are kept identical to those of the LSTM model.

Table 1: Model Architectures

Model	Parameters
LSTM	Number of layers: 1 LSTM layer, 2 fully connected layers Neurons in layers in vector notation: [50,10,1]
MLP	Number of layers: 5 Neurons in layers in vector notation: [64,32,1]

3. Results

The dataset consists of nine features and 13,200 samples, which, after applying bootstrapping, correspond to approximately nine days of data. The dataset is divided into training, validation, and test subsets, with 70%, 15%, and 15% of the data allocated to each, respectively.

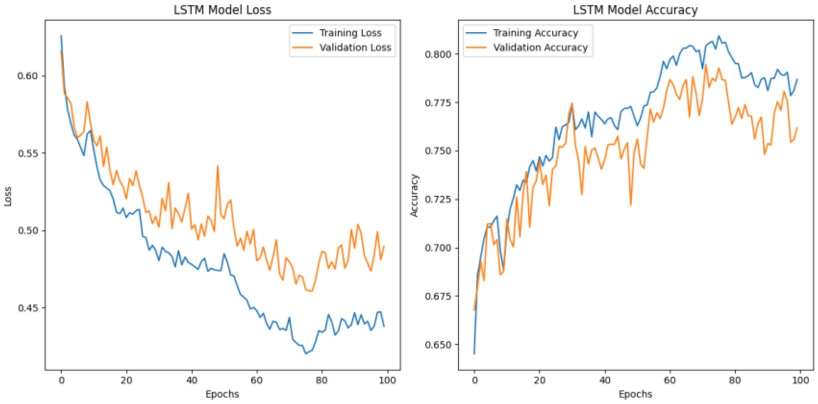


Figure 5: LSTM model Training Loss and Accuracy

Fig. 5 illustrates the training loss (Binary Cross-Entropy) and accuracy of the LSTM model during the training phase, while Fig. 6 presents the same metrics for the MLP model.

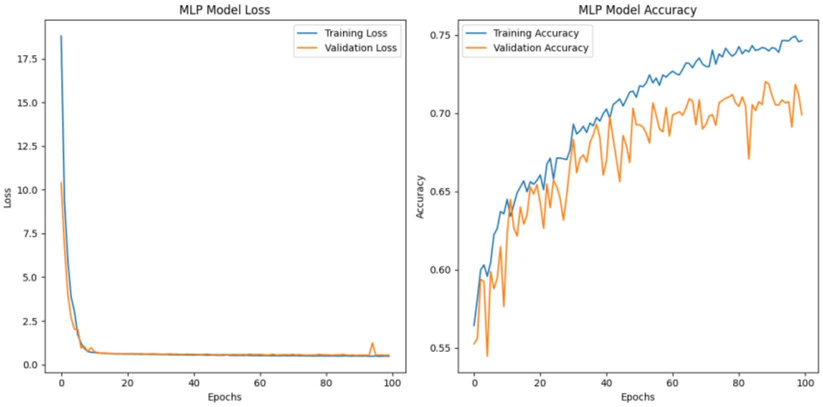


Figure 6: MLP model Training Loss and Accuracy

Since the task is formulated as a binary classification problem, precision, recall, and F1 score metrics were used to evaluate and compare the models. As presented in Table 2, the LSTM model outperforms the MLP model in terms of F1 score (0.78 vs. 0.75), indicating a more balanced trade-off between precision and recall. However, the MLP model achieves a higher recall value (0.91), suggesting that it is more effective in identifying positive (anomalous) instances.

Table 2: Model Comparison based on F1 score and Binary Cross Entropy

Model	Precision	Recall	F1 Score
LSTM	0.76	0.8	0.78
MLP	0.64	0.91	0.75

4. Discussion and Conclusion

The results demonstrate that integrating multimodal sensor data, specifically vibration and sound, provides a reliable foundation for predicting machine anomalies in manufacturing environments. Both LSTM and MLP models successfully identified abnormal operating conditions that preceded downtime, confirming the potential of data-driven approaches for predictive maintenance. The LSTM model, with its ability to capture temporal dependencies, achieved a more balanced precision–recall trade-off, making it suitable for real-time monitoring where both false positives and false negatives must be minimized. Meanwhile, the MLP's higher recall indicates its sensitivity to detecting anomalous patterns, which can be valuable in applications prioritizing early warnings. Broader datasets covering multiple machines, operating conditions, and longer durations would enhance model robustness and generalizability. Future work will focus on extending this research by incorporating additional sensor modalities such as temperature, current, and acoustic emission data to further improve anomaly detection accuracy. Moreover, deploying the trained models within an edge-computing framework could enable real-time diagnostics directly on IoT devices.

5. Acknowledge

We express our gratitude to our colleagues at Supply Chain Wizard, who contributed to the completion of this research on predictive analytics for production line downtime. Their support and collaboration have been instrumental in advancing our understanding of efficient manufacturing processes and the application of machine learning in industrial settings.

References

[1] Next Generation Digital Factory Platform. SCW.AI. (2023, September 1). https://scw.ai

[2] Hochreiter, S.&Schmidhuber, J. (1997). Long short-term memory. Neural computation 9, no. 8, 1735-1780.

[3] Rumelhart, D.E., Hinton,G. E.&Ronald,J. W. (1986). Learning representations by back-propagating errors." Nature 323, no. 6088: 533-536