DOI: 10.12962/j20882033.v36i1.22501 | Submitted 01-30-2025; | Reviewed 03-11-2025; | Accepted 03-18-2025

ORIGINAL RESEARCH

Predicting Failure Using Machine Learning and Statistical-Based Method: A Production Machine Case Study

Effi Latiffianti* | Stefanus Eko Wiratno | Samuel Aditya Christianta

[1]Dept. of Industrial and Systems Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Correspondence

*Effi Latiffianti, Dept of Industrial and Systems Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.
Email: effi_1@ie.its.ac.id

Present Address

Gedung Teknik Industri, Kampus ITS Keputih Sukolilo, Surabaya 60111, Indonesia

Abstract

This research investigates the applicability of failure detection models based on machine learning and statistical approaches to reduce unplanned downtime in a food production company. Sensor data is utilized to for identifying early failure symptoms. To capture temporal and sequential dependencies in time-series data, we employ one of potential network based method so called the Long Short Term Memory (LSTM) Autoencoder. Furthermore, we contrast the performance of the result with the traditional statistical method, the multivariate Exponentially Weighted Moving Average (EWMA). While both models successfully detected all failures, LSTM-AE demonstrated superior performance by reducing false alarms and providing true alarms with a longer time-to-failure. The findings highlight the potential of leveraging limited data for failure prediction, demonstrating the effectiveness of both models in detecting anomalies while emphasizing their role in enhancing productivity through early failure detection.

KEYWORDS:

Anomaly detection, failure, fault, Long Short Term Memory Autoencoder, Multivariate Exponentially Weighted Moving Average.

1 | INTRODUCTION

Maintenance is crucial for optimal performance and lifetime of industrial infrastructure and equipments. In general, it includes regular inspections, repairs, strategic asset management, and life cycle planning to reduce risks, prevent failures, and improve performance. As industries adopt new technologies, effective maintenance strategies are vital for maintaining operational efficiency and competitiveness. Unplanned downtime due to sudden failures and emergency maintenance affect around 90% of organizations in European companies [1], with traditional maintenance methods often failing to prevent these disruptions [2]. Even with planned maintenance, unexpected breakdowns can still occur, leading to inefficiencies.

Detecting a failure before its occurrence can prevent sudden machine breakdowns, which contribute to significant production losses. Early detection allows for proactive maintenance, reducing downtimes and managing costs more effectively. Data-driven failure detection models have been extensively applied to wind turbines, leveraging sensor data collected through Supervisory Control and Data Acquisition (SCADA) systems to enhance operational reliability [3–6]. Similarly, recent studies by [7] and [8] demonstrate the potential of predictive models in forecasting failures within manufacturing machinery. Building on these advancements, this study aims to implement and evaluate failure detection models within a selected case study to further explore their applicability and effectiveness.

Our case study focuses on a production machine of a leading poultry-based food industry in Indonesia. The maintenance team has observed several instances of abnormal machine behavior, including increased noise levels, reduced performance, and elevated temperatures, which have resulted in breakdowns. Some of these breakdowns cause the production line to come to a complete halt, leading to significant financial losses. In some cases, the unavailability of replacement parts further prolongs the downtime, exacerbating the impact on production. This highlights the critical importance of implementing effective failure detection mechanisms, such as anomaly detection models, which can monitor equipment conditions, provide early warnings of potential failures, and enable proactive interventions. By preventing unexpected downtimes and ensuring the availability of necessary parts, such systems can significantly enhance operational efficiency, reduce costs, and maintain the continuity of production. Anomaly detection identifies deviations from expected patterns in data, signaling potential equipment failures [9]. However, false alarms, where normal conditions are inaccurately flagged as anomalies, pose a significant challenge [3, 7]. A systematic approach is needed to filter out false alarms and improve model accuracy.

In light of the above details, this study covers three main objectives. The first objective is to apply one of machine learning techniques that has shown some excellent performance in failure detection – long short term memory (LSTM), which in our case will be combined with autoencoder and refer to as LSTM-AE. The second objective is to investigate whether the classical statistical-based approach, i.e multivariate exponentially weighted moving average (EWMA) can also work comparably well in the case study. Additionally, we would like to understand how much the two differ by comparing their performance.

The subsequent sections of this document are organized to provide a comprehensive understanding of the study. Section 2 focuses on the data used in the analysis. Section 3 provides the methodologies applied, detailing the approaches, techniques, and models employed to address the research objectives. Section 4 presents the results of the analysis and provides an in-depth discussion of the findings, highlighting their implications and relevance to the study. Finally, Section 5 reports the conclusion, the key insights summary, contributions, and potential directions for future research or practical applications.

2 | THE DATA

In this study, the dataset was collected from a production machine in the poultry-based food production line. The machine, a Townsend NL-17 sausage linker, has been recently equipped with sensors to enhance maintenance and operations. However, due to resource limitations, the collected data has not been adequately utilized. The sausage linker is a critical component of the production line, frequently failing and accounting for over 50% of the downtime in sausage production.

The dataset consists of four variables collected from four different types of sensors, with measurements recorded every 5 minutes over one month of machine operation. The dataset has a dimension of 10,397 rows and 5 columns. Table 1 outlines the variables included in the dataset.

TABLE 1 Available variables from sensor measurement in the dataset.

Variable measured	Description	Unit of measurement
Current	Flow of electrical charge in the linker conductor	Ampere (A)
Voltage	Electrical potential difference in the linker's electrical circuit	Volt (V)
Power	Rate of energy conversion in the linker component	Watt (W)
Power factor	Ratio of real power to total power supplied	Ratio

To validate the early failure detection results, failure log data was also utilized. Our analysis focuses on breakdowns caused by linker component failures in the sausage linker machine. Table 2 presents details of these breakdowns, including their duration, timestamp, and description.

TABLE 2 Failure record in the dataset.

Duration (in minutes)	Failure time	Failure description
480	2024-03-02 10:20	Linker chain dislodgement
300	2024-03-15 03:30	Bearing problem (wear and tear)
355	2024-03-22 02:30	Linker chain tension issue
480	2024-04-05 19:20	Linker chain problem (dullness)

3 | METHOD

We apply two methods in the analysis: the LSTM-AE and multivariate EWMA. The LSTM-AE combines the feature reconstruction capabilities of autoencoders with LSTM's ability to capture dynamic relationships in time-series data, enabling effective anomaly detection in sequential, unlabeled data. Because it is an unsupervised model, we do not actually need to split the data into training and testing sets. However, we will do the splitting to evaluate the performance on the training period to see the consistency of the results when it is applied on the test set. Multivariate EWMA, on the other hand, employs a multivariate time-series approach with exponentially decreasing weights to the historical data, enhancing its ability for detecting changes in the data especially in time series contexts. Although multivariate EWMA does not required labeled data, it is categorized as semi-supervised learning. In order to detect abnormality using multivariate EWMA, a baseline needs to be determined for the normal data. Because the sensor data is accompanied by the failure logs, the task of identifying normal data is less complicated.

The following subsections explain the LSTM-AE, multivariate EWMA, and the implementation details that we used in this study.

3.1 | Long Short Term Memory Autoencoder (LSTM-AE)

Long Short Term Memory (LSTM) networks is an advanced type of Recurrent Neural Network (RNN) that are particularly effective in capturing long-term dependencies in sequential data such as time series data. LSTMs use gates and memory cells to selectively store or discard information, enhancing their ability to model sequential data[10], which is similar to the case like predictive maintenance. The primary advantage of LSTM is its ability to identify long-term dependencies in sequential data. Figure 1 presents the structure of LSTM.

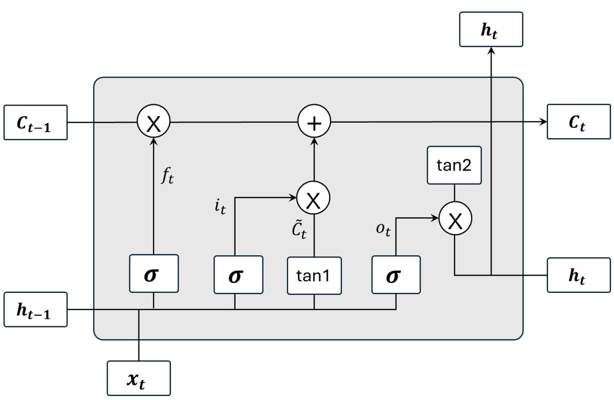

FIGURE 1 The structure of LSTM.

In Figure 1 , x_t is the input vector at timestamp t, h_t and h_{t-1} represent the current and previous output, respectively. C_t and C_{t-1} represent the new and previous cell states or memory at timestamp t and $t-1$, respectively. The $\times$ and $+$ in the circle are vector multiplication and vector addition, respectively; tan1 and tan2 are tanh neural networks and tangent function. The equations for the gates are provided in Equation 1 – 3.

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i), \tag{1}$$

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f), \tag{2}$$

$$o_t = \sigma(w_0[h_{t-1}, x_t] + b_0), \tag{3}$$

where f_t,i_t, and o_t represent forget gate, input gate, and output gate, respectively. The sigmoid function is represented by σ, w_x is relevant weight in respective gate x associated with each block, and b_x is bias neurons at gate x. The equations for the cell state, candidate cell state and the final output are presented in Equation 4–6 as follows:

$$\tilde{C_t} = \tanh\left(w_c\left[h_{t-1}, x_t\right] + b_c\right) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C_t}, \tag{5}$$

$$h_t = o_t * \tanh\left(C_t\right), \tag{6}$$

where $\tilde{C_t}$ refers to a tanh output and represents candidate for cell state at timestamp t, and * represents the element wise multiplication of the vectors.

LSTMs have demonstrated superior performance over traditional machine learning models in predicting the remaining useful life of equipment[11, 12], as well as fault detection in different application such as aircraft engines[13, 14], and wind turbine systems[15–17]. LSTM was also reported to perform better that other models like Decision Tree, Random Forest, Convolutional Neural Networks, and XGBoost in terms of accuracy and other metrics[18, 19].

Autoencoders are a specialized type of feedforward neural network designed to reproduce input data as accurately as possible in the output. They achieve this by transforming the input into a lower-dimensional representation, often called the latent space or code, and then reconstructing the original input from this compact representation[20]. The latent code serves as a condensed summary, capturing the most essential features of the input. An autoencoder is composed of three primary components: the encoder, the code, and the decoder as illustrated in Figure 2 . The encoder compresses the input data to generate the latent code, while the decoder reconstructs the input using only this compressed representation. This structure enables autoencoders to learn efficient data representations and uncover underlying patterns.

In manufacturing, combining Autoencoders with Long Short-Term Memory (LSTM) deep learning models is recommended for anomaly detection[18]. LSTM, with its Encoder-Decoder architecture, can process input sequences of varying lengths and generate corresponding output sequences. LSTM-Autoencoder (LSTM-AE) is particularly effective in capturing the dynamics and complex relationships within time-series data, making it well-suited for detecting anomalies in manufacturing processes[21].

Studies highlight LSTM-AE's effectiveness in different applications.[22] demonstrates its success in detecting anomalies in industrial gas turbines and CPU utilization. Similarly,[23] showcases its potential for fault detection in electric drives. Another study integrates LSTM and Autoencoder networks into a specialized encoder-decoder LSTM architecture, enhancing monitoring capabilities in complex dynamic processes[24].

In our case study, we set the sequence length of data used in the model to 5, which is equal to 25 minutes based on the nature of the data The selection was based on the result from autocorrelation analysis. Furthermore, we used Mean Squared Error (MSE) as the loss function in training and testing of the model. MSE is particularly suited for this study because it squares the error term, making it more sensitive to large deviations or outliers. This means that large errors (anomalies) disproportionately impact the final MSE value, making significant deviations easier to identify[25]. Anomalous data points are identified as those

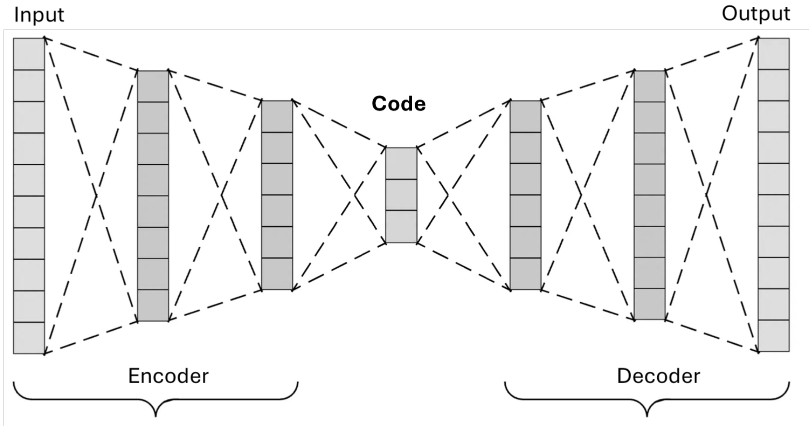

FIGURE 2 Autoencoder illustration (Reprinted from Wolpe and Waal, 2019).

with higher reconstruction errors compared to the majority of the data. Finally, a threshold was selected by targeting the need for proper lead time in detecting failures in the training data and acceptable number of false alarms justified by the maintenance engineers in the observed production line.

3.2 | Multivariate EWMA

The exponentially weighted moving average (EWMA) control chart is a well known statistical process control (SPC) method renowned for its effectiveness in identifying small shifts in process parameters. In contrast to traditional Shewhart control charts, which analyze individual data points, EWMA prioritizes trend detection by applying exponentially decreasing weights to older observations. This characteristic enhances its sensitivity to subtle deviations, often serving as early indicators of potential failures. As [26] notes, the cumulative approach of EWMA offers a distinct advantage in identifying gradual changes in processes, making it particularly valuable for applications that demand timely failure detection. EWMA tracks the mean value of a process by giving recent observations more importance through exponentially decreasing weights. This method provides a smoothed estimate of the process mean, making it sensitive to subtle changes in data that the traditional control chart method might miss [27].

In multivariate anomaly detection, EWMA has been adapted from its original use with single variable to handle complex datasets with multiple interconnected variables. This adaptation brings several advantages. For instance, multivariate EWMA can effectively capture relationships between variables by incorporating covariance matrices. This capability allows it to detect anomalies that affect multiple variables simultaneously, which is crucial and has become the most rapidly developing area in industries like manufacturing and complex systems monitoring [28].

The application of EWMA in detecting anomalous data points has been widely reported to be effective. [29] highlights that EWMA excels in the early detection of anomalies across interconnected variables, providing a proactive approach to system monitoring, particularly in scenarios where anomalies can impact multiple parameters simultaneously. When combined with other methods, EWMA performs well in simulated data for failure detection [30, 31].

Successful implementations of EWMA, as well as its modified version, the Unscented Kalman Filter (UKF)-based EWMA, in fault detection for wastewater treatment plants further demonstrate its applicability across various domains [32, 33]. The study in [32] also compares EWMA with m-CUSUM, revealing that EWMA outperforms its counterpart in fault detection for wastewater treatment plants.

In photovoltaic systems, the EWMA control chart effectively monitors abnormal states and accurately identifies the specific string where a fault occurs [34]. Its detection power can be further enhanced when integrated with other methods [35]. Similarly, in wind turbine systems, EWMA has successfully detected outliers up to one week before bearing failure [36], allowing for early fault alarms and proactive maintenance.

Based on its capability and the nature of the case study, we consider multivariate EWMA model as a potential alternative of solution for early failure detection. Unlike machine learning approaches, there is no iterative training on the multivariate EWMA approach. This study uses the basic form of the multivariate EWMA chart to monitor shifts in the mean vector of a multivariate process. The multivariate EWMA score per data point is the output of the multivariate EWMA approach and can be obtained using Equation 7 and 8.

$$\begin{aligned} \boldsymbol{z}_t &= \lambda.\boldsymbol{x}_t + (1-\lambda)\boldsymbol{z}_{t-1} \\ &= \sum_{j=1}^{i-1} \lambda(1-\lambda)^j \boldsymbol{x}_{i-j} + (1-\lambda)^i \boldsymbol{z}_0, \end{aligned} \tag{7}$$

$$T_i^2 = [\boldsymbol{z}_t - \boldsymbol{\mu}_0]^t . \boldsymbol{\Sigma}_{z_i}^{-1} [\boldsymbol{z}_t - \boldsymbol{\mu}_0] \tag{8}$$

where $\boldsymbol{\mu}_0$ is the mean vector of $\boldsymbol{x}_i$ and $\boldsymbol{\Sigma}_{z_i}$ are the the covariance matrix of $\boldsymbol{z}_i$. When i is large, i.e. $i \to \infty$, $\boldsymbol{\Sigma}_{z_i}$ can also be estimated as: $\frac{\lambda}{2-\lambda}\boldsymbol{\Sigma}_0$, where $\boldsymbol{\Sigma}_0$ is the covariance matrix of $\boldsymbol{x}_i$. In our case, $\boldsymbol{x}_i$ was defined as the vector of variables in data points that are considered normal, i.e. at least 24 hours away from the failure events.

Furthermore, the implementation of EWMA requires two parameters to be determined: the smoothing constant λ, and the threshold or the control limit that acts as a boundary between the anomalous and normal.[37] proposed the use of a smoothing constant (λ) that ranges between 0.05 and 0.25. In our study, we used $\lambda = 0.20$ as it has been proven robust in anomaly detection[38] and it also produced best result in our case study. The selection of the control limit was based on the training data by targeting the true detection of the actual failure. Any multivariate EWMA value higher than the threshold indicates anomalous data behavior and to be flagged as an alarm.

3.3 | Implementation Details

We begin our analysis with data preprocessing, where we normalized the variables using a robust scaling approach. This technique adjusts features by subtracting the median and scaling them based on the interquartile range (IQR). Unlike methods that rely on the mean and standard deviation, robust scaling is less sensitive to outliers and extreme values[39], making it particularly suitable for anomaly detection context. The scaled values were using the formula in Equation 9.

$$x_i' = \frac{x_i - Q_2(x)}{Q_3(x) - Q_1(x)} \tag{9}$$

where x_i' is the transformed value of the i^{th} variable, x_i is the original value of the i^{th} variable, and $Q_1(x), Q_2(x), Q_3(x)$ represent the first, second, and third quartile of the data, respectively.

After the transformation, we employ the autocorrelation function (ACF) to analyze the relationships between past and current data points for each feature, determining if the data sequence should be treated in a time-series context. The ACF results showed significant short-term dependencies, especially up to five lags, indicating that recent past values influence future values, which informed our decision to set the LSTM sequence length to five. While the voltage did not show time-dependent behavior, it was still included as a predictor due to its relevance in indicating potential failures. The data exhibited short-term temporal dependencies, underscoring the need for a time-series approach in our predictive modeling to capture these patterns effectively.

Next, we divided the data into two subsets: 70% for the training set and 30% for the testing set. While the two methods primarily rely on unsupervised and semi-supervised learning techniques, we performed the split in the same way as in supervised learning. This approach allows us to simulate real-world conditions, where future observations are unavailable during model training. Consequently, we tuned the model parameters based solely on the training set.

We applied two models, LSTM-AE and multivariate EWMA, to the subsets. The LSTM-AE model was trained using the training set and evaluated on the test set. The multivariate EWMA approach, however, was implemented in two distinct phases. In Phase I, normal data was selected to establish $\boldsymbol{mu}_0$, $\boldsymbol{\Sigma}_0$, and the control limit (H) tailored for anomaly detection. In Phase II, the entire dataset was analyzed using all those parameters defined in Phase I to perform the fault detection.

Finally, we evaluated the performance of both approaches using two metrics: the total number of false alarms and the failure detection lead time. The failure detection lead time refers to the time interval between the first alarm or detection and the actual machine failure. For this case study, we established a maximum lead time of 24 hours, based on input from maintenance experts familiar with the production machines and processes. This threshold was chosen to reflect the machine's historical behavior and the characteristics of the components monitored by the sensors.

Furthermore, since production operates on three shifts per day, a lead time exceeding 8 hours was considered preferable in practice. This would provide the maintenance team with sufficient time to plan and execute maintenance activities during shift changes. Consequently, we defined a false alarm as any alarm triggered more than 24 hours before a failure or less than 8 hours before a failure.

The numerical analysis was conducted using Python, with PyTorch specifically used for the LSTM-AE model. Since the dataset is not large, the implementation runs efficiently on any computer with at least 8 GB of memory.

4 | RESULT AND DISCUSSION

Once the models are implemented, we calculate a score that is used for failure detection. To trigger an alarm, a threshold is established. If the score exceeds this threshold, an alarm is raised. The alarm will remain active until the score drops back below the threshold. The tricky part is where to set the threshold, considering the balance between the true detection and false alarms. Naturally, lowering the threshold increases the likelihood of false alarms. However, it also reduces the chances of missing a true failure, resulting in a higher true positive rate. Conversely, raising the threshold decreases the number of false alarms but increases the risk of failing to detect actual failures. Practitioners often look for the balance between the two, aiming at the highest detection rate with considerably acceptable number of false alarms.

In practice, missed detections are the most critical issue to avoid, as their impact is significantly more severe. Missed detection results in an unscheduled production line stoppage, leading to lost production and reduced productivity for all other resources that depend on the production process. Furthermore, unexpected failures often create additional challenges, such as the unavailability of necessary inventory to restore operations, further compounding delays and inefficiencies. For this reason, sometimes the performance of a model can be measured from the number of miss detections produced from the model.

In many machine learning applications, model performance is commonly evaluated using metrics such as F1 score, accuracy, precision, and recall. While these metrics are effective in many contexts, they are not always ideal for failure detection. In practice, production managers often prioritize the practical benefits of a model, such as the cost savings or financial impact it brings to real-world operations. Although this type of metric is less commonly used, it has been introduced in certain studies. For example, [3] calculated the total savings generated by a method based on real-world maintenance costs, including replacement costs, inspection costs, and repair costs.

In our case, we lack the necessary data on these costs to perform a similar analysis. However, we can still evaluate the models using alternative metrics, such as average lead time, the number of false alarms, and the number of missed detections, which can provide comparable insights into the practical effectiveness of the models.

Table 3 and Table 4 present a small range of threshold, denoted as H, that were implemented on the two methods. From these two tables, we can see how different values of H affect the results of the detection. Suppose our goal is to detect all failures within 8 to 24 hours of their occurrence. Based on Table 3, we have two options to consider for detection: setting the threshold at $H = 4.5$ or $H = 4.0$. Considering the nature of the case, where the production line operates in three working shifts, we can conclude that the average lead time between the two thresholds is not significantly different. In other words, detections will typically occur within one or two shifts of the actual failure. This lead time should provide sufficient opportunity to arrange maintenance actions during shift changes, when the machine is normally shut down. Therefore, $H = 4.5$ is the more favorable option to be used with LSTM-AE scores, as it leads to fewer false alarms. In a similar fashion, using information presented in Table 4, $H = 7$ or $H = 8$ are two very good options for the multivariate EWMA scores.

With the thresholds set for both methods, we can now compare their performance in terms of the number of false alarms and the average lead time. There are two approaches to this comparison:

1. Set H to values that result in a comparable average lead time for both methods, then evaluate the number of false alarms.
2. Set H to values that produce a similar number of false alarms for both methods, then assess the average lead time.

TABLE 3 LSTM-AE performance based on selected threshold.

Threshold	Average lead time (hour)	# of false alarms	# of true alarm	# of total alarm	# of miss detection
$H = 5.0$	6.67	8	3	11	2
$H = 4.5$	14.54	10	8	18	0
$H = 4.0$	16.92	18	17	35	0

TABLE 4 Multivariate EWMA performance based on selected threshold.

Threshold	Average lead time (hour)	# of false alarms	# of true alarm	# of total alarm	# of miss detection
$H = 12$	6.00	15	5	20	2
$H = 11.5$	7.90	17	5	22	2
$H = 10$	10.23	24	6	30	1
$H = 9$	10.25	28	7	35	1
$H = 8$	13.67	28	10	38	0
$H = 7$	16.19	30	11	41	0
$H = 6$	16.67	35	13	48	0

Both approaches are expected to lead to the same conclusion.

On the first approach, let $H = 4.5$ for the LSTM-AE and $H = 8$ for multivariate EWMA. Based on Table 3 and Table 4, this will result in the average lead time of 14.54 hours and 13.67 hours for LSTM-AE and multivariate EWMA, respectively. With these thresholds, the total number of false alarms are 10 and 28 for LSTM-AE and multivariate EWMA, respectively, which means LSTM-AE outperformed multivariate EWMA.

On the second approach, let $H = 4.0$ for the LSTM-AE and $H = 11.5$ for multivariate EWMA. This will give us 18 and 17 false alarms for LSTM-AE and multivariate EWMA, respectively. In terms of average lead time, LSTM-AE was able to detect all 4 failures with average lead time of 16.92 hours, while multivariate EWMA could not detect the first and second failures, resulting in only 7.90 hours lead time in average. Again, LSTM-AE outperformed multivariate EWMA.

Table 5 presents detection time and time to failure produced from LSTM-AE with $H = 4.5$ and MEWMA with $H = 8$. The detection plots for both methods are presented in Figure 3 .

TABLE 5 Failure detection times resulted from LSTM-AE and multivariate EWMA

Failure time	LSTM-AE		EWMA	
	Earlierst alarm	Lead time (hour)	Earlierst alarm	Lead time (hour)
2024-03-02 10:20	2024-03-01 15:40	11.66	2024-03-01 16:10	11.16
2024-03-15 03:30	2024-03-14 08:00	12.50	2024-03-14 11:25	9.08
2024-03-22 02:30	2024-03-21 05:05	14.42	2024-03-21 06:25	13.08
2024-04-05 19:20	2024-04-04 16:45	19.58	2024-04-04 15:00	21.33

Both the LSTM-AE, a machine learning-based method, and EWMA, a statistical approach, effectively detect anomalies to provide early failure warnings. However, the LSTM-AE outperforms the multivariate EWMA model by detecting failures earlier, i.e. about an hour or 5.05% earlier. Furthermore, when the multivariate EWMA threshold is adjusted to match the failure detection lead time of the LSTM-AE, the multivariate EWMA model generates more false alarms, i.e. two times higher (or 211%). This highlights the LSTM-AE's superiority over the multivariate EWMA model, both in terms of earlier failure detection and reduced false alarms.

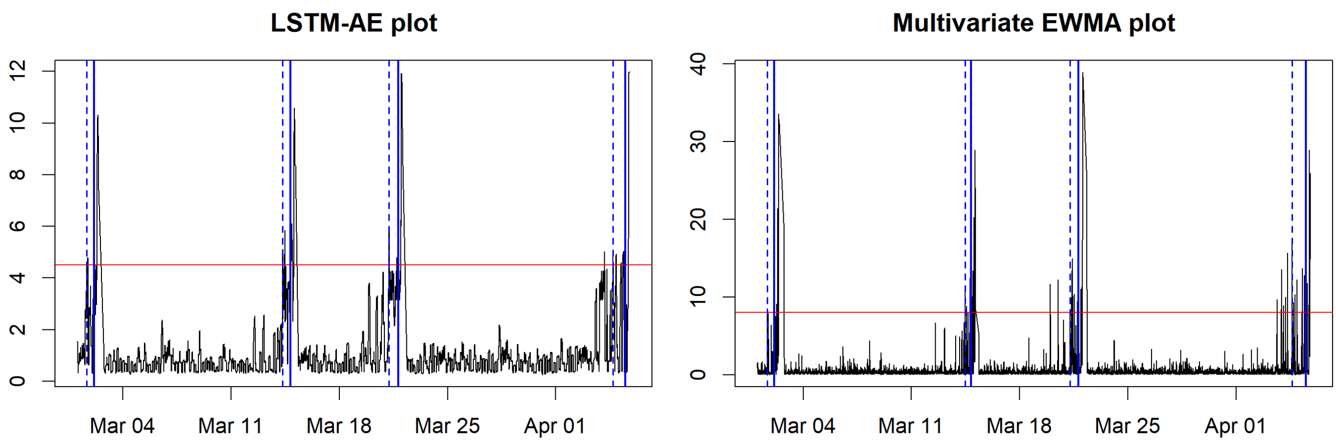

FIGURE 3 Detection plots of multivariate EWMA (left) and LSTM-AE (right). The solid horizontal line represents the threshold H, the vertical solid lines mark the time of the actual failures' occurrence, and the vertical dashed lines are the time of the earliest alarms.

In practice, an early warning system enables management to anticipate failures more effectively. For instance, if the warning provides a lead time of more than eight hours, the maintenance team can schedule an inspection during the next shift change when the machine is normally stopped, minimizing production loss.

Additionally, if maintenance personnel are unavailable when the alarm is triggered, the advance notice allows sufficient time to bring them on-site for the scheduled inspection. Based on the inspection, appropriate maintenance actions can be determined. If immediate repairs are not possible, preventive measures can be taken to avoid failure until the necessary maintenance is performed. This proactive approach helps prevent prolonged downtime, production losses, and productivity declines.

To understand what happened during the period after the earliest alarms, we perform additional analysis on the measured variables. The begin by dividing the data into five subsets: the normal data points, and four subsets of data points where the score from one or both methods are greater than the control limit. We plotted the density of each variable from those subsets in Figure 4 . We differentiated the plots from each subset using different colors. Based on Figure 4 , we can clearly see that voltage does not provide pronounced clue about the failure except for Failure #1 and Failure #3 where the sensor readings yield wider density, with spikes on the right side, indication some higher values compared to the rest of the data. The current and the power on the other hand, present similar patterns across different failures, showing sort of bimodal distribution with some density raising on the opposite direction of the normal ones. Finally, on the power factor, small portions of failure signatures show larger values, but at the same time, the density from all failures also shifted to the left, indicating that all four failures also present nearly zero power factor as the one of the symptoms.

In summary, the exploration of the sensors' reading values reveals that all four variables we used in the model contributes to the detection resulted by the two models. However, we cannot isolate nor differentiate the exact cause of each of the failures solely based on this analysis.

The early failure warning capability demonstrated in this study is based on a limited dataset. Nevertheless, the results serve as a validation of the anomaly detection model, showing that it can successfully identify anomalies as early indicators of failure and provide timely warnings. While these results may vary when tested on larger datasets, the findings presented here offer an initial understanding of the model's ability to provide early failure warnings.

5 | CONCLUDING REMARKS

This research highlights the benefits of using machine learning and statistical models for early anomaly detection in predictive maintenance. By analyzing time-series data, the LSTM-AE and multivariate EWMA models effectively identified early signs

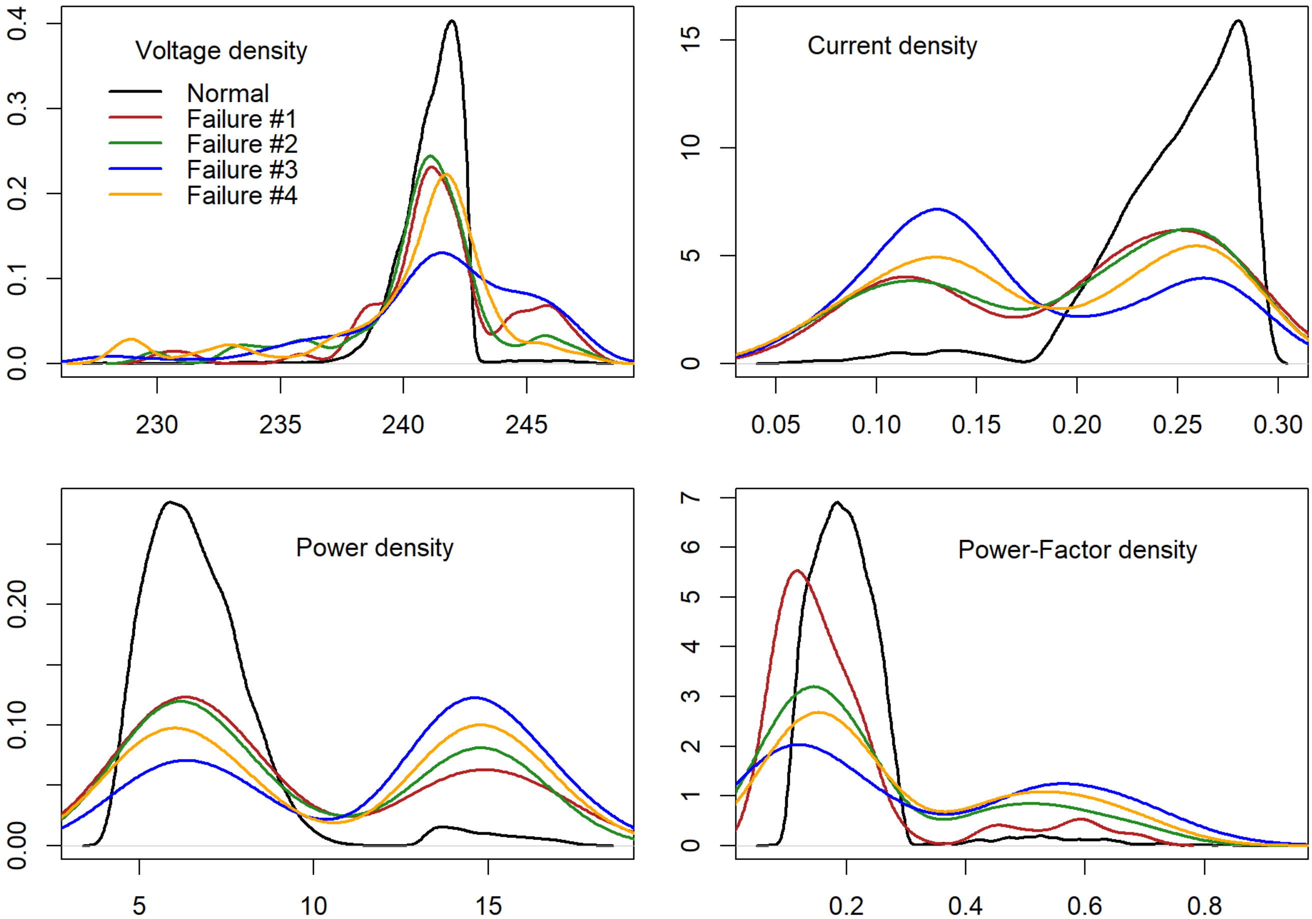

FIGURE 4 Variable exploration based on combined subset from both LSTM-AE and multivariate EWMA.

of equipment failure. These models detected failure patterns that were not apparent in the raw data, with the LSTM-AErelying on reconstruction errors and the multivariate EWMA model using statistical measures. While the results may vary with larger datasets, they provide an initial understanding of the models' ability to deliver early failure warnings.

A crucial aspect of effective anomaly detection is setting appropriate detection thresholds. Lower thresholds improve precision but result in shorter lead times, whereas higher thresholds allow for longer lead times but generate more false alarms. Finding the right balance between early detection and reliability is the key. Among the two models, the LSTM-AE showed better performance, particularly in detecting subtle anomalies and providing earlier warnings, while also generating fewer false alarms. However, the multivariate EWMA model remains valuable for its ability to effectively handle short-term patterns.

In summary, the LSTM-AE demonstrates strong potential for enhancing early failure detection by providing accurate and timely warnings. Future research should aim to test these models with larger datasets and refine threshold optimization to further improve their reliability and effectiveness. Another promising direction is the integration of machine learning-based anomaly detection with real-time industrial monitoring systems, such as Industrial IoT platforms, to enable automated decision-making and predictive maintenance. While this study is centered on a food production company, further research could investigate the applicability of these models in other industries, such as automotive manufacturing, pharmaceuticals, and energy production, to assess their robustness in different operational environments. Additionally, conducting a cost-benefit analysis of implementing machine learning-based predictive maintenance would provide decision-makers with valuable insights into cost savings, return on investment, and overall productivity improvements.

CREDIT

Conceptualization, Methodology, Writing - literature review, formatting, reviewing, editing, Formal analysis and investigation, finalizing draft: Effi Latiffianti; Conceptualization, Writing - reviewing, Data collection, and Research supervision: Stefanus Eko Wiratno; Data collection, Formal analysis and investigation, Writing - first draft: Samuel Aditya Christianta.

References

1. Infraspeak. Maintenance statistics and trends 2025. Infraspeak Team; 2024.
2. Ensafi M, Thabet W. Challenges and Gaps in Facility Maintenance Practices. In: Leathem T, Perrenoud A, Collins W, editors. ASC 2021. 57th Annual Associated Schools of Construction International Conference, vol. 2 of EPiC Series in Built Environment EasyChair; 2021. p. 237–245. /publications/paper/tZ57.
3. Latiffianti E, Sheng S, Ding Y. Wind Turbine Gearbox Failure Detection Through Cumulative Sum of Multivariate Time Series Data. Frontiers in Energy Research 2022;10. https://www.frontiersin.org/journals/energy-research/articles/10.3389/fenrg.2022.904622.
4. Khan PW, Yeun CY, Byun YC. Fault detection of wind turbines using SCADA data and genetic algorithm-based ensemble learning. Engineering Failure Analysis 2023;148:107209. https://www.sciencedirect.com/science/article/pii/S1350630723001632.
5. Meyer A. 1 - SCADA-based fault detection in wind turbines: data-driven techniques and applications. In: Garcia Marquez FP, Papaelias M, Junior VLJ, editors. Non-Destructive Testing and Condition Monitoring Techniques in Wind Energy Wind Energy Engineering, Academic Press; 2023.p. 1–13. https://www.sciencedirect.com/science/article/pii/B9780323996662000010.
6. Chen J, Li J, Chen W, Wang Y, Jiang T. Anomaly detection for wind turbines based on the reconstruction of condition parameters using stacked denoising autoencoders. Renewable Energy 2020;147(P1):1469–1480. https://EconPapers.repec.org/RePEc:eee:renene:v:147:y:2020:i:p1:p:1469-1480.
7. Hermansa M, Kozielski M, Michalak M, Szczyrba K, Wróbel Sikora M. Sensor-Based Predictive Maintenance with Reduction of False Alarms—A Case Study in Heavy Industry. Sensors 2022;22(1). https://www.mdpi.com/1424-8220/22/1/226.
8. Yadav DK, Kaushik A, Yadav N. Predicting machine failures using machine learning and deep learning algorithms. Sustainable Manufacturing and Service Economics 2024;3:100029. https://www.sciencedirect.com/science/article/pii/S2667344424000124.
9. Carrasco J, López D, Aguilera-Martos I, García-Gil D, Markova I, García-Barzana M, et al. Anomaly detection in predictive maintenance: A new evaluation framework for temporal unsupervised anomaly detection algorithms. Neurocomputing 2021;462:440–452. https://www.sciencedirect.com/science/article/pii/S0925231221011826.
10. Di Persio L, Garbelli M, Wallbaum K. Forward-Looking Volatility Estimation for Risk-Managed Investment Strategies during the COVID-19 Crisis. Risks 2021;9(2). https://www.mdpi.com/2227-9091/9/2/33.
11. Boujamza A, Elhaq SL. Attention-based LSTM for Remaining Useful Life Estimation of Aircraft Engines. IFAC-PapersOnLine 2022;https://api.semanticscholar.org/CorpusID:251366314.
12. Zheng S, Ristovski K, Farahat A, Gupta C. Long Short-Term Memory Network for Remaining Useful Life estimation. In: 2017 IEEE International Conference on Prognostics and Health Management (ICPHM); 2017. p. 88–95.
13. Haipeng C, Jie Y, Wenxing F. Two-layer fault diagnosis model of aircraft based on LSTM. Aerospace Science and Technology 2025;158:109756. https://www.sciencedirect.com/science/article/pii/S127096382400885X.

14. Du W, Zhang J, Meng G, Zhang H. Aero-Engine Fault Detection with an LSTM Auto-Encoder Combined with a Self-Attention Mechanism. Machines 2024;12(12). https://www.mdpi.com/2075-1702/12/12/879.

15. Yang T, Teng J, Li C, Feng Y. Wind turbine fault detection and diagnosis using LSTM neural network. In: 2020 39th Chinese Control Conference (CCC); 2020. p. 4042–4047.

16. Qin S, Tao J, Zhao Z. Fault diagnosis of wind turbine pitch system based on LSTM with multi-channel attention mechanism. Energy Reports 2023;10:4087–4096. https://www.sciencedirect.com/science/article/pii/S2352484723015081.

17. Xiang L, Wang P, Yang X, Hu A, Su H. Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism. Measurement 2021;175:109094. https://www.sciencedirect.com/science/article/pii/S026322412100124X.

18. Kamat P, Sugandhi R. Anomaly Detection for Predictive Maintenance in Industry 4.0 – A survey. In: E3S Web Conf., vol. 170; 2020. .

19. Gong L, Chen Y. Machine Learning-enhanced loT and Wireless Sensor Networks for predictive analysis and maintenance in wind turbine systems. International Journal of Intelligent Networks 2024;5:133–144. https://www.sciencedirect.com/science/article/pii/S2666603024000083.

20. Guo J, Liu G, Zuo Y, Wu J. An Anomaly Detection Framework Based on Autoencoder and Nearest Neighbor. In: 2018 15th International Conference on Service Systems and Service Management (ICSSSM); 2018. p. 1–6.

21. Chalapathy R, Chawla S. Deep learning for anomaly detection: A survey. arXiv 2019;https://arxiv.org/abs/1901.03407.

22. Maleki S, Maleki S, Jennings NR. Unsupervised anomaly detection with LSTM autoencoders using statistical data-filtering. Applied Soft Computing 2021;108:107443. https://www.sciencedirect.com/science/article/pii/S1568494621003665.

23. Demidova G, Privalov D, Semenov D, Lukichev D, Liu Z, Anuchin A. Fault Detection in Electric Drives Based on LSTM Autoencoder Model Machine Learning Approach. In: 2024 IEEE 25th International Conference of Young Professionals in Electron Devices and Materials (EDM); 2024. p. 1670–1675.

24. Zeng L, Jin Q, Lin Z, Zheng C, Wu Y, Wu X, et al. Dual-attention LSTM autoencoder for fault detection in industrial complex dynamic processes. Process Safety and Environmental Protection 2024;185:1145–1159. https://www.sciencedirect.com/science/article/pii/S0957582024001721.

25. Lachekhab F, Benzaoui M, Tadjer SA, Bensmaine A, Hamma H. LSTM-Autoencoder Deep Learning Model for Anomaly Detection in Electric Motor. Energies 2024;17(10). https://www.mdpi.com/1996-1073/17/10/2340.

26. Montgomery DC. Introduction To Statistical Quality Control, 4Th Ed. Wiley India Pvt. Limited; 2007. https://books.google.co.id/books?id=rcu71DvkjX0C.

27. Kadri F, Harrou F, Chaabane S, Sun Y, Tahon C. Seasonal ARMA-based SPC charts for anomaly detection: Application to emergency department systems. Neurocomputing 2016;173:2102–2114. https://www.sciencedirect.com/science/article/pii/S0925231215014654.

28. Woodall WH, Montgomery DC. Research Issues and Ideas in Statistical Process Control. Journal of Quality Technology 1999;31(4):376–386.

29. Bersimis S, Psarakis S, Panaretos J. Multivariate Statistical Process Control Charts: An Overview. Quality and Reliability Engineering International 2007 08;23:517 – 543.

30. Lee JM, Yoo C, Lee IB. Efficient fault detection using multivariate exponentially weighted moving average and independent component analysis. In: Chen B, Westerberg AW, editors. Process Systems Engineering 2003, 8th International Symposium on Process Systems Engineering, vol. 15 of Computer Aided Chemical Engineering Elsevier; 2003.p. 916–921. https://www.sciencedirect.com/science/article/pii/S1570794603804248.

31. Harrou F, Nounou M, Nounou H. A statistical fault detection strategy using PCA based EWMA control schemes. In: 2013 9th Asian Control Conference (ASCC); 2013. p. 1–4.

32. Marais HL, Zaccaria V, Odlare M. Comparing statistical process control charts for fault detection in wastewater treatment. Water Science & Technology 2022;85(4):1250–1262.

33. Baklouti I, Mansouri M, Nounou H, Ben Slima M, Ben Hamida A. Fault detection in a wastewater treatment plant. In: 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP); 2017. p. 1–5.

34. Su J, Zeng Z, Tang C, Liu Z, Li T. A Photovoltaic Fault Diagnosis Method Integrating Photovoltaic Power Prediction and EWMA Control Chart. Energies 2024;17(17). https://www.mdpi.com/1996-1073/17/17/4263.

35. Mansouri M, Al-khazraji A, Hajji M, Harkat MF, Nounou H, Nounou M. Wavelet optimized EWMA for fault detection and application to photovoltaic systems. Solar Energy 2018;167:125–136. https://www.sciencedirect.com/science/article/pii/S0038092X18303050.

36. Li H, Deng J, Yuan S, Feng P, Arachchige DDK. Monitoring and Identifying Wind Turbine Generator Bearing Faults Using Deep Belief Network and EWMA Control Charts. Frontiers in Energy Research 2021;9. https://www.frontiersin.org/journals/energy-research/articles/10.3389/fenrg.2021.799039.

37. Lucas JM, Saccucci MS. Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. Technometrics 1990;32(1):1–12. https://www.tandfonline.com/doi/abs/10.1080/00401706.1990.10484583.

38. Jurjevčič B, Senegačnik A, Kuštrin I. A Surveillance of Direct-Firing System for Pulverized-Coal Using Statistically Treated Signals from Intrusive Electrostatic Sensors. Strojniški vestnik - Journal of Mechanical Engineering 2017;63(4):265–274. https://www.sv-jme.eu/article/a-surveillance-of-direct-firing-system-for-pulverized-coal-using-statistically-treated-signals-from-intrusive-electrostatic-sensors/.

39. de Amorim LBV, Cavalcanti GDC, Cruz RMO. The choice of scaling technique matters for classification performance. Applied Soft Computing 2023;133:109924. https://www.sciencedirect.com/science/article/pii/S1568494622009735.

How to cite this article: Latiffianti E., Wiratno S. E., Christianta S. A, (2025), Predicting Failure using Machine Learning and Statistical Based Method: a Production Machine Case Study, *IPTEK The Journal of Technology and Science*, *36(1):1–13*.