Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA
Published: February 18, 2025

DOI: 10.46254/AN15.20250068

Unsupervised Learning-Based PHM for Predictive Maintenance and Optimal Configuration of Air Compressors in the Semiconductor Industry

Tai-Ching Hung
Master student at Department of Industrial Engineering and Engineering Management
National Tsing Hua University, Hsinchu, Taiwan, ROC
eddy112034558@gapp.nthu.edu.tw

Kuo-Hao Chang
Professor at Department of Industrial Engineering and Engineering Management
National Tsing Hua University, Hsinchu, Taiwan, and Deputy Director at National Science and Technology Center for Disaster Reduction, Taiwan, ROC
chang@mx.nthu.edu.tw

Jing-Wei Chen
Researcher at Industrial Technology Research Institute, Hsinchu, Taiwan, ROC
jingweichen@itri.org.tw

Jing-Yao Jhang
Researcher at Industrial Technology Research Institute, Hsinchu, Taiwan, ROC
itriB20638@itri.org.tw

Abstract

In smart manufacturing, companies continuously gather large volumes of data from machine sensors. Extracting value from this data requires advanced analytical methods to integrate, analyze, and predict potential issues, which is where Prognostic and Health Management (PHM) comes into play. PHM is a comprehensive framework that uses data mining, big data analytics, and artificial intelligence (AI) to perform condition monitoring, fault diagnosis, failure prediction, and lifespan tracking of industrial machines. The goal of PHM is to improve machine performance, reduce downtime, and optimize maintenance strategies. In this research, unsupervised learning techniques are employed to process the vast amounts of raw, unlabeled data commonly found in industrial environments. These techniques transform raw data into machine health scores, allowing real-time monitoring of machine conditions. By visualizing health scores of each machine, the system helps on-site personnel rapidly identify anomalies and determine whether predictive maintenance is required. This shifts the focus from conventional preventive maintenance, which is often time-based, to a more efficient predictive maintenance approach, which is driven by actual machine performance data. Moreover, this research integrates optimization algorithms to enhance the configuration of air compressors, a critical component in many industrial systems. By optimizing the operation and configuration of these compressors, the study aims to maximize their efficiency, reduce energy consumption, and ensure consistent production line performance. This combined approach not only supports predictive maintenance but also leads to significant cost savings and improved overall system reliability, making it an essential tool in modern smart manufacturing environments.

Keywords
Big data analytics, Artificial intelligence algorithms, PHM, Optimization and Air compressor.

1. Introduction
In the semiconductor manufacturing industry, energy consumption is immense, with a typical fabrication plant (fab) consuming as much electricity as 50,000 homes (Chen, Gautam, and Weig, 2013). This massive energy usage places significant strain on electrical grids, contributes to climate change, and incurs substantial operational costs. In fact, energy costs can account for anywhere from 5% to 30% of a fab's total operating expenses. Consequently, optimizing energy efficiency while maintaining production goals offers a crucial competitive advantage to semiconductor manufacturers. One of the most promising approaches to achieve these outcomes is through smart manufacturing and big data analytics.Smart manufacturing leverages digital technologies and automated processes to enhance adaptability and streamline production. The core of a smart factory lies in the cyber-physical system (CPS), which integrates physical and virtual spaces to enable real-time data exchange, feedback loops, and self-correcting abilities (Monostori, 2018). This real-time data, collected from smart devices such as sensors and SCADA systems, can be further analyzed for prediction and optimization through big data analytics (Li et al., 2018). In the semiconductor industry, big data analytics (BDA) has been widely applied to improve yield, enhance production efficiency, and reduce costs. BDA supports a wide range of applications, including fault detection, process optimization, and predictive maintenance, which help manufacturers achieve better competitive and business performance (Gunasekaran et al., 2017).Prognostics and Health Management (PHM), originally developed in the aerospace and defense sectors, has emerged as a critical framework in industries like manufacturing and energy, especially in predictive maintenance and failure prediction. PHM uses data mining, big data analysis, and AI to monitor and predict the Remaining Useful Life (RUL) of industrial machinery, enabling more efficient and cost-effective maintenance strategies (Huang et al., 2022). In particular, deep learning techniques in PHM extract non-linear patterns from raw data, offering robust tools for failure detection and health monitoring. Among the various methods, autoencoders (AE), a prominent unsupervised deep learning model, have proven effective in reducing the dimensionality of high-dimensional data and removing redundant information (Liu et al., 2023). AEs are particularly useful in analyzing complex thermographic data, identifying nonlinear features, and offering insights that can guide maintenance decisions.

Incorporating both PHM and big data analytics into smart manufacturing enables manufacturers to enhance equipment monitoring, predict failures, and optimize energy efficiency. By utilizing AI-driven models such as autoencoders for anomaly detection and deep learning-based prognostics, companies can transition from preventive to predictive maintenance, improving overall operational efficiency. Given the increasing competitiveness of the semiconductor industry, integrating advanced technologies to monitor machinery health, predict equipment failures, and reduce energy consumption is becoming a critical strategy for maintaining a competitive edge.

2. Literature Review
A literature review is an essential step in the research process, playing a critical role in summarizing prior studies, identifying gaps in knowledge, and determining the direction of new research. In this literature review, we will explore the academic contributions made to the research topic, evaluating previous methods, findings, and conclusions. This analysis helps establish the uniqueness and innovation of our study, providing a theoretical framework that enhances the reader's understanding of the research background. Additionally, it serves as a foundation for future researchers to explore this field more comprehensively. The literature review offers strong support for further investigation in our research. This section is divided into two parts: Section 2.1 delves into PHM (Prognostics and Health Management) technology, which not only provides real-time monitoring of equipment and system conditions but also offers predictive analytics to prevent potential failures. We will review how previous studies have applied PHM technology in industrial environments, examining its role in enhancing production efficiency and overall effectiveness. Section 2.2 focuses on the recent advancements in Autoencoders for feature extraction and reconstruction of data. In the context of PHM applications, Autoencoders have been widely utilized for dimensionality reduction and anomaly detection. This part of the review will investigate how Autoencoders have been applied in the literature, emphasizing their strengths in the PHM field and their ability to improve fault detection in manufacturing processes through effective learning techniques.

2.1 Fault Prognostics and Health Management (PHM)

PHM is a method that uses data to monitor and predict the condition of machinery and systems. Initially used in the aerospace and defense industries to ensure the safety and reliability of critical and expensive equipment (such as aircraft systems) (Batzel and Swanson, 2009; Che et al., 2019), PHM encompasses various tasks, including fault analysis, health indicator identification, diagnostics, and prognostics (Rezaeianjouybari and Shang, 2020). In recent years, PHM has gained popularity in other industries, such as smart manufacturing (Xia et al., 2018; Vrignat et al., 2022) and energy systems (Meng and Li, 2019; Yucesan et al., 2021). PHM provides health information that serves as an early warning for potential failures and offers opportunities to improve system reliability, safety, and efficiency, ultimately reducing unplanned downtime and maintenance costs (Dong et al., 2019; Biggio and Kastanis, 2020; Rezaeianjouybari and Shang, 2020). PHM can be categorized into three main areas: Fault Detection, Fault Diagnostics, and Fault Prognostics. As this research focuses on Fault Detection, we will further explore literature related to this area. Fault Detection methodologies are typically divided into Model-based and Data-driven approaches (Zio, E., 2022).

2.2 Autoencoders

Autoencoder (AE) networks are commonly used in anomaly detection. An AE consists of an encoder network that maps raw data to a lower-dimensional feature space and a decoder network that attempts to reconstruct the data from this lower-dimensional space. The parameters of both networks are learned through a reconstruction loss function. Typically, a bottleneck architecture is employed to capture critical features in a lower-dimensional space, retaining the most important information from the original data. The aim is to minimize overall reconstruction error, meaning that information retained must be highly relevant to the primary data (e.g., normal data). Consequently, anomalies, which deviate from the majority of the data, are poorly reconstructed, and their reconstruction error can be directly used as an anomaly score (Guansong Pang et al., 2021). For anomaly detection, especially using AE-based methods, the effectiveness of unsupervised learning can be explained as follows: AEs use healthy or normal operational data to create a low-dimensional representation of the system's healthy state with minimal error and then reconstruct it (Torabi et al., 2023). When unseen data is input into the trained model, the magnitude of the reconstruction error indicates anomalies or potential system faults (Ahmad et al., 2020; Torabi et al., 2023).

Many recent studies have employed AE for anomaly detection or combined it with other techniques such as clustering or convolutional neural networks (CNNs). For example, Castellani et al. (2020) introduced an innovative approach for detecting anomalies in industrial environments by creating a digital twin (DT) to simulate normal machine operations and combining it with a clustering-based method and Siamese Autoencoder (SAE) for effective weakly-supervised anomaly detection. Li et al. (2022) developed a new unsupervised anomaly detection method called "Variable Cumulative Error Anomaly Detection" (VCEAD), which uses AE for signal reconstruction error calculation and a temporal convolutional network for vibration signal prediction. Yan et al. (2023) created a hybrid robust convolutional autoencoder (HRCAE) for unsupervised anomaly detection in noisy machine tools. Bae et al. (2023) developed an innovative method using event log data and AE to identify irregular behaviors in automatic test equipment (ATE) during wafer testing.

3. Methods

The architecture diagram as shown in Figure 1 outlines a predictive maintenance system for industrial equipment, specifically focusing on two main operational phases: the System Machines Configuration and the Machine Fault Prediction. The process begins with 'Data Collection,' where relevant machine data is gathered. This data then undergoes 'Data Preprocessing' to clean and prepare it for further analysis. Next, the system 'Determines Seasonality' to identify any seasonal patterns in the data, which are then adjusted in the 'Deseasonalization' step. In the 'System Machines Configuration' phase, located within the purple box, the cleaned and adjusted data is used to 'Optimize Machines Configuration.' This optimization aims to fine-tune the settings and operations of the machinery based on the processed data to enhance overall performance and efficiency. Parallelly, the 'Machine Fault Prediction' phase within the orange box involves 'Training the Autoencoder model' with the preprocessed data. This model learns to identify normal and abnormal patterns in machine operation. Following the autoencoder model training, the 'Health Score Transformation Model' is trained. This model transforms the outputs from the autoencoder into actionable health scores, indicating the operational health of each machine. The final output of both phases feeds into the 'User Interface (Result Visualization),' where the health scores and other relevant data are visually presented. This interface enables easy monitoring and decision-making concerning maintenance strategies based on real-time data analytics.

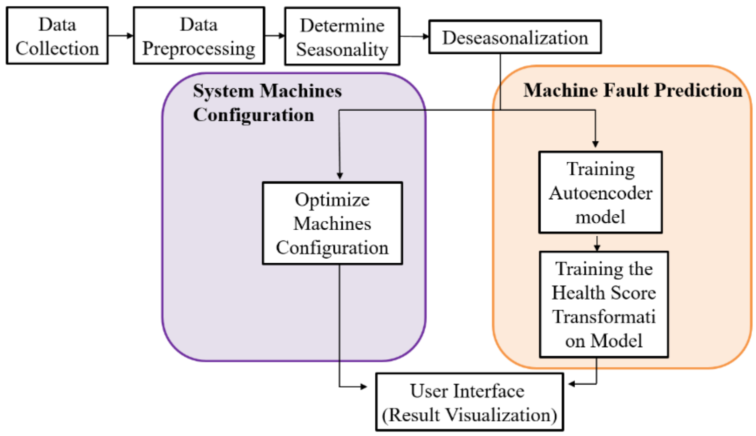

Figure 1. Proposed Framework

3.1 Machine Fault Prediction

In contemporary industrial and manufacturing contexts, a vast amount of sensor data is regularly gathered on various machine parameters or environmental conditions and stored in cloud databases such as SQL. This data includes measurements related to factors like temperature, pressure, humidity, vibration, proximity, lighting, and gas levels. Initially, in the offline phase, comprehensive data collection from these sensors, along with external parameter readings, is performed. During data transmission, some discrepancies such as signal irregularities might lead to missing values, and it may take some time for the equipment to stabilize post-maintenance. The first step in the data preprocessing phase involves eliminating these irregular data points. Since the scales of various parameters differ significantly, it becomes essential to standardize the data. This is typically achieved through normalization techniques that adjust the sensor data to have a zero mean and a unit standard deviation, enhancing the accuracy and efficacy of any learning models used later in the process.

In many manufacturing environments, sensor data can be highly influenced by external elements, such as outdoor temperatures. Taking air compressors as an example, several operational parameters might be temperature-dependent and therefore exhibit strong correlations with external air temperatures. Post-data preprocessing, our predictive health management (PHM) framework performs an analysis using Pearson correlation to identify which parameters have strong associations with outdoor temperatures. Parameters that demonstrate significant correlations are then adjusted using deseasonalization methods. This approach modifies data to account for seasonal variations, ensuring that these do not skew the anomaly detection capabilities of the PHM system.

More specifically, Pearson's correlation coefficients, symbolized as r, are calculated for each variable relative to outdoor air temperature using the equation 1, In the formula, $X_j$denotes the j-th measurement of a specific machine parameter, and Y_j is the corresponding j-th reading of the outdoor air temperature, with $\bar{X}$ and $\bar{Y}$ representing the average of all N readings of the machine parameter and outdoor temperature, respectively, within the dataset.

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \quad (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^{2} \sum_{i=1}^{n} (y_i - \bar{y})^{2}}} \tag{1}$$

Parameters are deemed highly correlated with outdoor air temperature if their Pearson correlation coefficients, r, reach or exceed 0.7. For parameters that do not display a strong correlation, the deseasonalization step in the offline phase is omitted. Conversely, for parameters that show significant correlation with outdoor air temperatures, deseasonalization is performed using the Centered Moving Average (CMA) method over a one-month smoothing period. The formula 2 for this calculation is CMA(m, s), where 'm' indicates the specific month and 's' the smoothing period. The formula 3 involves calculating CMA(m, s) for each month 'm', followed by a second centered moving average on these values to obtain CMA′(m, s). As shown in equation 4, the average outdoor air temperature for each month, denoted as Y(m), is then divided by CMA′(m, s) to derive the "Monthly Deseasonalization Factor" or MDF(m).

When the preprocessed data from any given month 'm' is divided by its corresponding MDF(m), the data is effectively deseasonalized. This deseasonalization is critical for each machine being evaluated within the PHM framework, ensuring that seasonal temperature variations do not affect the anomaly detection and health assessment processes.

$$CMA(m,s) = \frac{\Sigma_{j=1}^{\frac{s}{2}} Y\left(m - \left(s - \frac{2j-1}{2}\right)\right) + Y(m) + \Sigma_{j=1}^{\frac{s}{2}} Y\left(m + \left(s - \frac{2j-1}{2}\right)\right)}{s} \quad (2)$$

$$CMA'(m,s) = \frac{\Sigma_{j=1}^{\frac{s}{2}} CMA\left(m - \left(s - \frac{2j-1}{2}\right), 2s\right) + Y(m) + \Sigma_{j=1}^{\frac{s}{2}} CMA\left(m + \left(s - \frac{2j-1}{2}\right), 2s\right)}{s} \quad (3)$$

$$MDF(m) = \frac{Y(m)}{CMA'(m,s)} \quad (4)$$

Following preprocessing and deseasonalization, the Autoencoder (AE) model is trained using data from machines that are operating normally. This AE model includes an encoder which reduces the input data into a more compact lower-dimensional form, and a decoder that attempts to recreate the original data from this compressed form, as depicted in Figure 2. The primary objective of the model is to reduce the reconstruction error, which is utilized as the model's loss function. Consequently, the AE model masters a concise representation of the input data, which is advantageous for identifying anomalies and can also be applied to tasks like data compression and noise reduction. It is important to note that a distinct AE model is developed for each machine under analysis.

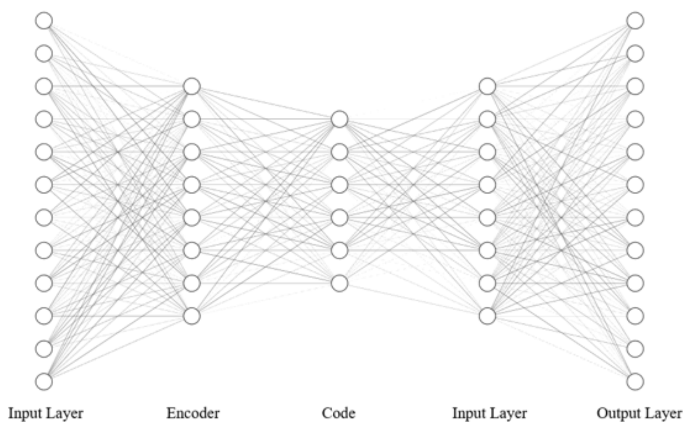

Figure 2. Illustration of autoencoder architecture

Once the AE model is trained, a dataset comprising both healthy and faulty data is fed into the model to compute the loss values for each data point across the machines. These loss values are later used to construct each machine's health score transformation model. While the mean squared error (MSE) is commonly employed as a loss function, it might not be suitable for all manufacturing scenarios due to the varying levels of importance that different machine parameters hold in determining overall machine health. To address this issue, we use a weighted mean squared error (WMSE) as the loss function as equation 5, which allows us to assign greater importance to key parameters. The weights in the WMSE equation can be determined in collaboration with domain experts, equipment manufacturers, or by using a clustering algorithm such as k-means. For any given time point t, where $t \in \{1, 2, \ldots, N\}$, the WMSE at that time point is represented as *WMSEt*, and it is calculated using the equation provided below.

$$
\begin{aligned}
WMSE_t &= \sum_{i=1}^{n} \frac{w_i}{\sum_{i=1}^{n} w_i} (x_i - y_i)^2 \\
&= \sum_{i=1}^{n} \frac{w_i}{\sum_{i=1}^{n} w_i} \left(x_i - d_\phi(e_\theta(x_i))\right)^2
\end{aligned}
\tag{5}
$$

In this equation, w_i represents the weight assigned to machine parameter *i*, while x_i and y_i denote the input and output data, respectively, as illustrated in Figure 2. The index *i* ranges from 1 to *n,* where *n* is the total number of machine parameters for the specific machine. For simplicity, the $1 \times n$ vectors for input data, output data, and weights are referred to as *x*,*y* and *w,* respectively. Additionally, e_θ stands for the encoder function, and $d_\emptyset$ refers to the decoder function, both of which were previously established during the AE model training. After calculating a WMSE value for each of the N data points in the training dataset, we proceed to construct a health score linear transformation model for each machine. Specifically, a piecewise linear transformation is employed to map each WMSE value, denoted as $WMSE_t$, to its corresponding health score, h_t. This transformation adjusts higher WMSE values to correspond with lower health scores, and lower WMSE values to higher health scores. The formula for this piecewise linear transformation of health scores is given as follows:

$$
h_t = (a_k * WMSE_t + b_k) \cdot I\{L_k \leq WMSE_t \leq U_k\} \tag{6}
$$

In formula 6, I{·} represents the indicator function, which equals 1 if the condition inside the curly brackets is satisfied and 0 otherwise. The condition being evaluated by the indicator function is whether the WMSE value lies between the endpoints of the *k*-th interval of observed WMSE values, where L_k is the lower endpoint and U_k is the upper endpoint. Additionally, a_k and b_k represent the slope and intercept, respectively, of the piecewise linear transformation function for the *k*-th interval.As depicted in Figure 3, we define six intervals of WMSE values for constructing the piecewise linear transformation. These intervals are outlined as follows:

$$
(L_k, U_k) = \begin{cases}
(0, Q1), & \text{for } k = 1 \\
(Q1, Q2), & \text{for } k = 2 \\
(Q2, Q3), & \text{for } k = 3 \\
(Q3, Q3 + 1.5 \cdot IQR), & \text{for } k = 4 \\
(Q3 + 1.5 \cdot IQR, Q3 + 3 \cdot IQR), & \text{for } k = 5 \\
(Q3 + 3 \cdot IQR, Max), & \text{for } k = 6
\end{cases}
\tag{7}
$$

In formula 7, Q1, Q2 and Q3 refer to the first, second, and third quartiles of the WMSE values, while "Max" represents the maximum WMSE value among all N data points, and the interquartile range (IQR) is defined as Q3−Q1. Thus, k=1 pertains to the bottom 25% of WMSE values (those up to Q1), k=2 includes values between Q1 and Q2, and k=3 covers values between Q2 and Q3. For k=4, this corresponds to WMSE values above Q3 but not considered outliers. In line with standard practice, k=5 is assigned to mild outliers, and k=6 applies to extreme outliers. For each machine, a specific set of intervals (Lk,Uk) is constructed to define the six intervals on the x-axis in the piecewise linear transformation model.The transformed health score's upper and lower bounds for each of these intervals are displayed on the y-axis in Figure 3. These limits are determined in consultation with domain experts and can be tailored to fit the particular industrial application or scenario being examined.

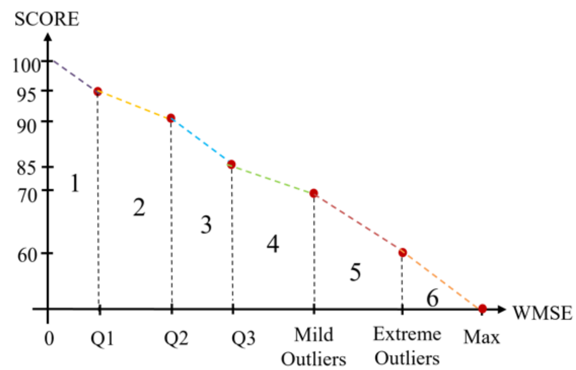

Figure 3. Health score linear transformation Illustration

3.2 System Machines Configuration

The primary goal of this research framework is to optimize overall system energy efficiency while ensuring that production requirements are met. In this section, we explain in detail how this is accomplished using a dynamic programming-based optimization approach. However, before we frame the problem within the context of the empirical study, it is essential to first clarify several key terms.

- F: Total flow demand, as specified by the user
- N: The number of machines in the system
- $f_{(i)}$ for i ∈ [1, N]: The output flow produced by each machine
- $L_{(i)}$ for i ∈ [1, N]: The maximum allowable energy flow for each machine, provided by the user
- $KW_{(i)}$ for i ∈ [1, N]: The power consumption of each machine

The decision variable in this problem is a binary indicator representing the machine configuration settings, and it is defined as follows:

$$I_{(i)} \forall i \in [1, \mathrm{N}] = \begin{cases} 0 & if\ machine\ i\ is\ turned\ of \\ 1 & if\ machine\ i\ is\ turned\ on \end{cases} \tag{8}$$

The problem definition is then

$$Min \sum_{i=1}^{N} KW_{(i)} \times I_{(i)} \tag{9}$$

$$Subject\ to \sum_{i=1}^{N} f_{(i)} \times I_{(i)} \geq F \tag{10}$$

In other words, the goal is to minimize the total power consumption across all machines that are active, while ensuring that the combined output flows from all machines that are turned on meet or exceed the total required flow demand. We define the operational efficiency of the machine as the output of the machine divided by its energy consumption, as shown in Equation 11, which represents how much output is generated per unit of energy consumed. The efficiency of each machine is then used as an input for the next phase of machine scheduling optimization. Machines with higher efficiency are kept running, while those with lower efficiency are shut down. The focus of optimization is on machines with intermediate efficiency, determining whether they should be turned on or off. Additionally, the calculated machine efficiency can be compared with the previously determined machine health scores, providing a dual validation function. The final output will provide optimized machine scheduling recommendations based on the required airflow at the site.

$$\text{Efficiency} = \frac{f_{(i)}}{KW_{(i)}} \tag{11}$$

We employ both Principal Component Regression (PCR) and Stepwise Regression models to predict machine operational efficiency. First, Principal Component Regression is a statistical technique that combines Principal Component Analysis (PCA) with linear regression, primarily used to address multicollinearity issues in high-dimensional data. During Principal Component Regression, PCA is first applied to reduce the dimensionality of the dataset and extract the most significant features. PCA is a dimensionality reduction technique that transforms the original variables by identifying the main directions of variation in the data, known as principal components. Each principal component is a linear combination of the original variables and is orthogonal (i.e., independent) to the others. The first principal component captures the largest amount of variation, the second captures the largest part of the remaining variation, and so on, until a sufficient amount of variation is covered. Assuming there are originally n parameters, PCA can produce n principal components through linear combinations, as shown in Equations 12, 13, and 14.

$$PC_1 = c_{11}x_1 + c_{12}x_2 + \cdots + c_{1n}x_n \tag{12}$$

$$PC_2 = c_{21}x_1 + c_{22}x_2 + \cdots + c_{2n}x_n \tag{13}$$

$$PC_3 = c_{31}x_1 + c_{32}x_2 + \cdots + c_{3n}x_n \tag{14}$$

Stepwise Regression is a statistical method used for variable selection, with the main objective of identifying the subset of predictor variables that most significantly contribute to predicting the target variable, while improving model interpretability and reducing the risk of overfitting. This approach iteratively adds or removes variables to build the regression model, with selection criteria typically based on statistical tests or information criteria. Stepwise regression includes three main types: Forward Selection, Backward Elimination, and Bidirectional Selection. Since Backward Elimination only considers the individual effect of each feature on the model, it may overlook interactions and combined effects among multiple features. Bidirectional Selection, while accounting for both adding and removing variables, tends to be less computationally efficient and time-consuming. Therefore, this project adopts the Forward Selection method, which progressively adds variables. The order in which variables are selected provides an indication of the importance of each predictor, offering useful insights for decision-makers.

The selection process in stepwise regression is iterative, and at each step, variables are evaluated based on certain criteria (e.g., the variable's p-value must fall below a specified threshold) to determine whether they should be included or excluded from the model. This approach simplifies the model, making it easier to interpret while avoiding excessive complexity that could result from including too many unnecessary variables. By combining both models (Principal Component Regression and Stepwise Regression), we can filter out the variables that have the most significant impact on energy efficiency, thereby optimizing the prediction model to accurately forecast the energy efficiency of each machine. These prediction results are then used as inputs to match the most energy-efficient machine combinations in the system, with the goal of minimizing total system energy consumption while meeting required flow demands.

By employing the Random Search Algorithm for optimizing fleet scheduling, a global optimization method based on random selection, this approach aims to identify the optimal machine configuration. The fundamental concept of the Random Search Algorithm is to generate candidate solutions randomly within the parameter space, and then evaluate their performance iteratively to gradually find the best solution. Specifically, the algorithm operates by first initializing the parameter space and generating candidate solutions at random. These solutions are evaluated based on total energy consumption, which serves as the performance metric. This approach not only enhances production efficiency but also significantly reduces energy consumption and operational expenditures, offering a win-win scenario for both economic and environmental benefits.

4. Results and Discussion

To demonstrate the effectiveness of the proposed PHM framework, this study generated simulated data for seven air compressors, each equipped with a different number of sensors representing distinct machine parameters.

4.1 Correlation Analysis and Deseasonalization

Through analysis, we identified three parameters that were highly correlated with external air temperature: motor front bearing temperature, motor rear bearing temperature, and motor coil temperature, as illustrated in Figure 4. Since these parameters are influenced by external temperature and exhibit seasonal variations, it is necessary to apply deseasonalization to mitigate the fluctuations caused by changes in temperature across different seasons. This step ensures that the PHM framework can accurately assess machine health without being skewed by external environmental factors.

Correlation Coefficient	Second Stage Inlet Temperature	Post-Cooler Temperature	Oil Temperature	Intake Temperature	motor front bearing temperature	motor rear bearing temperature	motor coil temperature
External Air Temperature	0.3783	0.2208	0.1906	0.2405	0.8458	0.8619	0.7873

Figure 4. Correlation Analysis Results

After performing deseasonalization for seasonal variations, as shown in Figure 5, we observed that the three parameters exhibited unstable fluctuations at different time points before deseasonalization. However, after deseasonalization, the curves became smoother, and there were no significant differences at various time points. This process reduces the likelihood of misjudgments when inputting the data into the model.

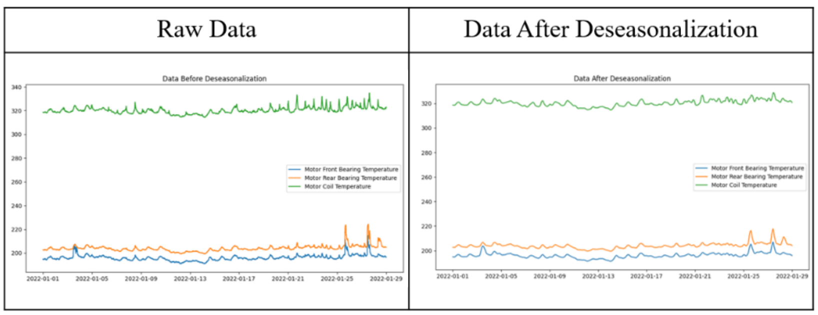

Figure 5. Deseasonalization Results

4.2 Health score transformation

In the left image of Figure 6, we display the reconstruction error results calculated for January. Due to the high density of data points and the large variance in reconstruction errors, directly converting them into health scores would result in significant variations, making the model overly sensitive. To address this, the study further applied the moving average method to smooth the health scores, using a 30-minute sliding window to calculate the average and observe the overall trend, represented by the orange line in Figure 6. The right image shows the health score results after the transformation. As seen in comparison to the left image, larger reconstruction errors are indeed converted into lower health scores. The moving average method reduces the model's sensitivity while retaining key trends, allowing simultaneous observation of both current and past machine health states.

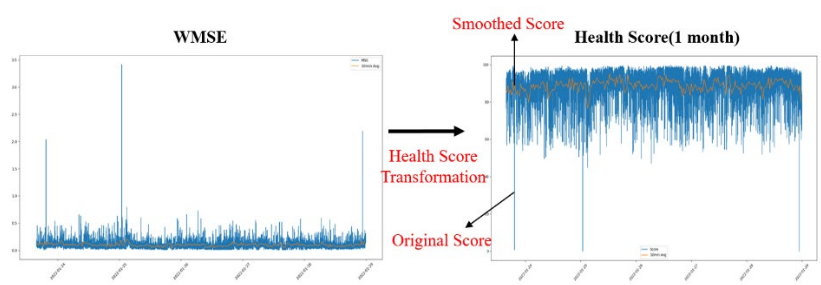

Figure 6. Health score transformation Results

4.3 Machine Operational Efficiency Prediction

In Figure 7, we can observe the scatter plot of actual values versus predicted values using Principal Component Regression (PCR). The correlation coefficient (R^2) between the actual and predicted values reaches 0.91, with a mean squared error (MSE) of only 0.16. Similarly, in Figure 8, the Stepwise Regression results also show a high correlation coefficient (R^2) of 0.91 and an MSE of 0.16. This demonstrates that both methods provide highly accurate predictions. We choose to adopt Stepwise Regression as the primary model, as it eliminates the need to determine the number of principal components, making it more convenient and flexible in application. Furthermore, from the scatter plot results, we can see that Stepwise Regression performs better in terms of both the correlation coefficient (R^2) and MSE,

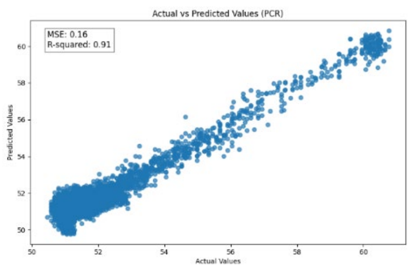

Figure 7 Principal Component Regression Prediction Results

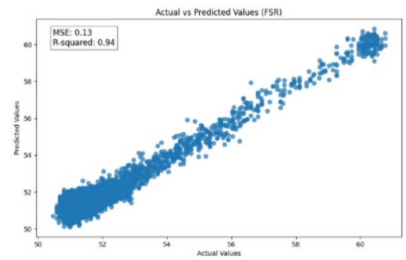

Figure 8. Stepwise Regression Prediction Results

indicating more accurate predictions of machine operational efficiency. The calculated efficiency for each machine can then be used as input for the next phase of machine scheduling optimization. Additionally, the machine efficiency values can be compared with the previously derived health scores to ensure the feasibility and condition of the machines, providing a dual-validation mechanism.

4.4 Optimized Machine Combination

After determining the operational efficiency of each machine, we use the Random Search Algorithm and Dynamic Programming methods to optimize the configuration of the machines. Machines with higher efficiency are kept running, while those with lower efficiency are shut down. The main optimization focuses on machines with intermediate efficiency, deciding whether to turn them on or off. Table 1 presents the results of the machine configuration optimization using the Random Search Algorithm.

Table 1. Optimized Machine Configuration Results

	Random Search Optimization
Total Demand Flow	245CMM
Total Number of Machines	10
Total Supply Flow	246.42CMM
Energy Consumption	1525.01kW
Machine ID Activated	No.1,No3,No5

5. Conclusion

This study successfully developed a system for industrial process quality and energy anomaly detection, leveraging big data analytics and artificial intelligence technologies. By applying unsupervised learning methods, the autoencoder model demonstrated outstanding performance in detecting anomalies in equipment operations, effectively identifying abnormal situations. Additionally, the health score transformation model quantified the health status of the equipment. The application of optimized machine scheduling algorithms further enhanced the system's energy efficiency, achieving the goal of reducing energy consumption and operational costs.

The key outcomes of this research include:

- Anomaly Detection: Utilizing the autoencoder model, the system successfully detected anomalies in unlabeled data and quantified equipment status through health scores.
- Data Preprocessing: The deseasonalization process effectively eliminated the impact of seasonal fluctuations on the data, improving model accuracy.
- Energy Efficiency Optimization: Through Principal Component Regression and the Random Search Algorithm, the machine fleet scheduling was optimized, significantly reducing energy consumption.
- User Interface Design: A health score dashboard UI was implemented, allowing users to monitor equipment status and system energy consumption in real-time.

To further enhance the practical value of this research, the following recommendations are proposed:

1. Expand Application Scope: In the future, this technology can be applied to more industrial sectors, such as manufacturing and energy industries, to validate its generalizability and reliability.
2. Diversify Data Sources: Incorporate data from various types of equipment and operating environments to improve the model's generalization and adaptability.
3. Real-time Monitoring and Feedback: Strengthen the system's real-time monitoring capabilities, and integrate a real-time feedback mechanism to enable timely responses to anomalies.
4. Continuous Algorithm Optimization: Continuously refine and improve the anomaly detection and scheduling optimization algorithms to increase the system's precision and efficiency.
5. Enhance User Training: Provide comprehensive training for users to ensure they fully understand and utilize the system's functions, thereby maximizing the technical benefits.

In summary, the developed technology has demonstrated significant effectiveness in improving industrial process quality and energy management. With ongoing optimization and expansion of its applications, this system is expected to play a vital role in a broader range of industrial fields in the future.

References

Abela, Kyle, Paul Refalo, and Emmanuel Francalanza. "Analysis of pneumatic parameters to identify leakages and faults on the demand side of a compressed air system." *Cleaner Engineering and Technology* 6 : 100355. 2022.

Ahmad, S., K. Styp-Rekowski, S. Nedelkoski, O. Kao.Autoencoder-based condition monitoring and anomaly detection method for rotating machines. *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 4093–4102, 2020.

Bae, Y.M., Y.G. Kim, J.W. Seo, H.A. Kim, C.H. Shin, J.H. Son, G.H. Lee, K.J. Kim.Detecting abnormal behavior of automatic test equipment using autoencoder with event log data. *Computers & Industrial Engineering* 183 109547, 2023.

Batzel, T.D., D.C. Swanson. Prognostic health management of aircraft power generators. *IEEE Transactions on Aerospace and Electronic Systems* 45(2) 473–482, 2009.

Biggio, L., I. Kastanis. Prognostics and health management of industrial assets: Current progress and road ahead. *Frontiers in Artificial Intelligence* 3 578613, 2020.

Barros, M., Veloso, B., Pereira, P.M., Ribeiro, R.P., Gama, J. Failure Detection of an Air Production Unit in Operational Context. In: Gama, J., et al. IoT Streams for Data-Driven Predictive Maintenance and IoT, Edge, and Mobile for Embedded Machine Learning. ITEM IoT Streams 2020. *Communications in Computer and Information Science*, vol 1325, 2020.

Bousdekis A, Lepenioti K, Apostolou D, Mentzas G. A Review of Data-Driven Decision-Making Methods for Industry 4.0 Maintenance Applications. *Electronics*. 2021.

Castellani, A., S. Schmitt, S. Squartini. Real-world anomaly detection by using digital twin systems and weakly supervised learning. *IEEE Transactions on Industrial Informatics* 17(7) 4733–4742, 2020.

Chandola, V., A. Banerjee, V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys* (CSUR) 41(3) 1–58, 2009.

Che, C., H. Wang, Q. Fu, X. Ni. Combining multiple deep learning algorithms for prognostic and health management of aircraft. *Aerospace Science and Technology* 94 105423, 2019.

Chang, Kuo-Hao, et al. "Big data analytics energy-saving strategies for air compressors in the semiconductor industry–an empirical study." *International Journal of Production Research* 60.6 (2022): 1782-1794.

Chang, K. H., Hsieh, T. C., Kuo, C. M., & Lin, H. P. Unsupervised Learning-Based Phm Framework to Inform Maintenance of Air Compressors in the Semiconductor Industry and an Empirical Study. Available at *SSRN* 4662600.

Castellani, A., S. Schmitt, S. Squartini. Real-world anomaly detection by using digital twin systems and weakly supervised learning. *IEEE Transactions on Industrial Informatics* 17(7) 4733–4742, 2020.

Dong, Y., T. Xia, X. Fang, Z. Zhang, L. Xi. Prognostic and health management for adaptive manufacturing systems with online sensors and flexible structures. *Computers & Industrial Engineering* 13357–68, 2019.

Desmet, A., & Delore, M. Leak detection in compressed air systems using unsupervised anomaly detection techniques. *Annual Conference of the PHM Society*, 9(1), 2017.

Hansen, Emil Blixt, et al. "A data-driven modular architecture with denoising autoencoders for health indicator construction in a manufacturing process." *arXiv preprint* arXiv:2208.05208 (2022).

Han, Peihua, et al. "Fault detection with LSTM-based variational autoencoder for maritime components." *IEEE Sensors Journal* 21.19 (2021): 21903-21912.

James, G., D. Witten, T. Hastie, R. Tibshirani, et al. An introduction to statistical learning, vol. 112.

Springer, 2013.
Khan, S., T. Yairi. A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing* 107 241–265, 2018.
Li, P., Y. Pei, J. Li. A comprehensive survey on design and application of autoencoder in deep learning. *Applied Soft Computing* 110176, 2023.
Li, Z., Y. Sun, L. Yang, Z. Zhao, X. Chen. Unsupervised machine anomaly detection using autoencoder and temporal convolutional network. *IEEE Transactions on Instrumentation and Measurement* 71 1–13, 2022.
Lei, Qiang, et al. "Machine learning based refrigerant leak diagnosis for a vehicle heat pump system." *Applied Thermal Engineering* 215 (2022): 118524.
Meng, H., Y.F. Li. A review on prognostics and health management (PHM) methods of lithium-ion batteries. *Renewable and Sustainable Energy Reviews* 116 109405, 2019.
Principi, E., D. Rossetti, S. Squartini, F. Piazza. Unsupervised electric motor fault detection by using deep autoencoders. *IEEE/CAA Journal of Automatica Sinica* 6(2) 441–451, 2019.
Principi, E., D. Rossetti, S. Squartini, F. Piazza. Unsupervised electric motor fault detection by using deep autoencoders. *IEEE/CAA Journal of Automatica* Sinica 6(2) 441–451, 2019.
Sahli, A., R. Evans, A. Manohar. Predictive maintenance in industry 4.0: Current themes. *Procedia CIRP* 104 1948–1953, 2021.
Sun, Y., G.G. Yen, Z. Yi. Evolving unsupervised deep neural networks for learning meaningful representations. *IEEE Transactions on Evolutionary Computation* 23(1) 89–103, 2018.
Torabi, H., S.L. Mirtaheri, S. Greco. Practical autoencoder based anomaly detection by using vector reconstruction error. *Cybersecurity* 6(1) 1, 2023.
Tsuda, T., S. Inoue, A. Kayahara, S.-I. Imai, T. Tanaka, N. Sato, and S. Yasuda. “Advanced Semiconductor Manufacturing Using Big Data.” *IEEE Transactions on Semiconductor Manufacturing*28 (3): 229–235, 2015.
Vrignat, P., F. Kratz, M. Avila. Sustainable manufacturing, maintenance policies, prognostics and health management: A literature review. *Reliability Engineering & System Safety* 218 108140, 2022.
Xia, T., Y. Dong, L. Xiao, S. Du, E. Pan, L. Xi. Recent advances in prognostics and health management for advanced manufacturing paradigms. *Reliability Engineering & System Safety* 178 255–268, 2018.
Xiaolin Li and Imin Kao, "Analytical fault detection and diagnosis (FDD) for pneumatic systems in robotics and manufacturing automation," *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, Alta.*, 2005, pp. 2517-2522, doi: 10.1109/IROS.2005.1545573.
Yan, S., H. Shao, Y. Xiao, B. Liu, J. Wan. Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises. *Robotics and Computer-Integrated Manufacturing* 79 102441, 2023.
Yu, Chih-Min, Chen-Fu Chien, and Chung-Jen Kuo. “Exploit the Value of Production Data to Discover Opportunities for Saving Power Consumption of Production Tools.” *IEEE Transactions on Semiconductor Manufacturing* 30 (4): 345–350, 2017.
Yucesan, Y.A., A. Dourado, F.A. Viana. A survey of modeling for prognosis and health management of industrial equipment. *Advanced Engineering Informatics* 50 101404, 2021.
Zemouri, R., M. L´evesque, E. Boucher, M. Kirouac, F. Lafleur, S. Bernier, A. Merkhouf. Recent ´ research and applications in variational autoencoders for industrial prognosis and health management: A survey. *2022 Prognostics and Health Management Conference (PHM-2022 London)*. IEEE, 193–203, 2022.
Zipfel, J., F. Verworner, M. Fischer, U. Wieland, M. Kraus, P. Zschech. Anomaly detection for industrial quality assurance: A comparative evaluation of unsupervised deep learning models. *Computers & Industrial Engineering* 177 109045, 2023.