Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

SEMICONDUCTOR EQUIPMENT HEALTH MONITORING WITH MULTI-VIEW DATA

Jeongsun Ahn
Hong-Yeon Kim
Sang-Hyun Cho
Hyun-Jung Kim

Department of Industrial and Systems Engineering
Korea Advanced Institute of Science and Technology
291 Daehak Street
Daejeon, 34141, REPUBLIC OF KOREA

Hongyeon Kim
Hyeonjeong Choi
Dain Ham

Software Team
Wonik IPS
75 Jinwisandan Street
Pyeongtaek, 17709, REPUBLIC OF KOREA

ABSTRACT

Monitoring the state of semiconductor equipment is crucial for ensuring optimal performance and preventing downtime. In previous studies, researchers have attempted to derive a health index that represents the overall condition of the equipment as a single index. However, these studies have often relied solely on time-series data from each sensor, neglecting other important viewpoints engineers consider when monitoring the equipment. To address this limitation, we propose a multi-view data set specifically designed for semiconductor equipment, which incorporates process, trend, and spatial data. In addition, we present a framework for deriving a hierarchical health index based on a multi-view data set. The hierarchical structure is derived using a hierarchical spectral clustering method, and an autoencoder-based health index is used. We have verified the effectiveness of our approach with real data sets, demonstrating its potential as a valuable tool for monitoring the condition of semiconductor equipment.

1 INTRODUCTION

Semiconductor manufacturing plays a critical role in our modern life, particularly with the development of artificial intelligence and the increasing demand for semiconductors. However, this industry faces several challenges, including short product life cycles, diverse customer demands, and increasing demand uncertainty (Chien et al. 2010). The semiconductor manufacturing process involves using highly advanced equipment and technology. It is crucial to ensure that this equipment is functioning at optimal levels at all times to maintain high-quality production. Any abnormal operation or machine failure can result in reduced production volume and poor wafer quality, which can be detrimental to the manufacturing process. Therefore, monitoring the state of equipment and detecting any machine state changes as early as possible is of utmost importance.

This paper focuses on the equipment used in the deposition process, which deposits chemical gases to create a thin film on a wafer with specific electrical characteristics. During the deposition process, chemical substances accumulate inside the chamber, requiring periodic cleaning to maintain the equipment in a normal state. The current practice of timed-based cleaning decisions is not based on actual equipment conditions. As a result, there is a need for an approach that can monitor machine conditions in real-time. This approach can achieve condition-based cleaning or identifying malfunctions. To monitor machine conditions in real-time, various sensor information is collected. However, monitoring all the sensors can increase the workload of engineers. Therefore, an efficient indicator is required to represent the overall health of the machine, called a health index.

979-8-3503-6966-3/23/$31.00 ©2023 IEEE

There has been extensive research on real-time monitoring and developing a health index for semiconductor equipment. One approach involves the use of statistical control charts (Skinner et al. 2004; Jeong et al. 2013; Lee et al. 2011; Chao et al. 2008), while another involves monitoring nonlinear sensor data through autoencoder (AE)-based algorithms (Blázquez-García et al. 2021; Li et al. 2021; Deng et al. 2023; Lee et al. 2019; Yang et al. 2019). Jeong et al. (2013) proposed a weighted moving average-based control chart for the plasma etching process, and Chao et al. (2008) developed a health index for batch process in semiconductor manufacturing through Hotelling T2 analysis. These methodologies have the advantages of being easy to apply and having convenient result interpretation. However, as the complexity of the data relationships increases, they become more challenging to use. Machine learning and deep learning methodologies are being developed to address these limitations. Ge and Song (2010) and Li and Zhang (2014) proposed adaptive principle component analysis (PCA) and k-nearest neighbor (KNN)-based methodologies, respectively, for semiconductor process monitoring. Additionally, the most widely utilized methodologies are based on AEs. Li et al. (2021) employed a stacked AE to effectively handle non-linear relationships. Furthermore, Zhang et al. (2019) introduced a convolutional neural network (CNN)-based AE, while Deng et al. (2023) proposed a recurrent neural network (RNN)-based AE. While these approaches demonstrate promising performance, one drawback is the difficulty in performing root cause analysis when issues arise.

In the field of monitoring, monitoring algorithms not only demonstrate the ability to effectively detect issues but also emphasize the importance of root cause analysis to identify the source of the problem. Consequently, there has been a growing interest in hierarchical health indexes, which derives a single health index based on a hierarchical structure formed by grouping similar sensors together (Hu and Su 2004; Yu et al. 2014; Lee and Dong 2019; Krauel and Weishäupl 2016). To enhance efficiency and usability, Yu et al. (2014) proposed a hierarchical structure for the health index. They developed a control chart for each sensor and labeled the health index of each sensor as 0 or 1 depending on whether its current value fell outside or within the confidence level of the control chart, respectively. The overall health index was then derived by calculating the weighted sum of the indices of all sensors. Another approach was proposed by Krauel and Weishäupl (2016), which extracted useful features based on the shape of sensor data and calculated a health index as the average of all the sensor health indices based on the control chart. Lee and Dong (2019) also proposed a hierarchical health index based on multiple control charts and an automated extraction method of weight values using the analytic hierarchy process (AHP).

Previous studies on hierarchical health indexes often assume prior knowledge of the hierarchical structure or only utilize real-time time-series data collected from sensors when forming monitoring groups. However, in reality, the construction of this hierarchical structure is a manual process that depends on the expertise of engineers. This not only increases the workload of engineers but also requires a new design each time it is applied to other equipment. Furthermore, in practical scenarios, engineers need to consider not only the configuration of the collected sensors but also comprehensive factors such as the physical location of the sensors and long-term trend changes. These aspects should be taken into account when forming monitoring groups.

Therefore, we propose a novel hierarchical health index framework that utilizes multi-view data obtained from semiconductor equipment. To determine the criteria for forming the most basic group of sensors for machine monitoring, we conducted interviews with equipment control engineers to identify three key perspectives and develop a method for extracting a similarity graph for each. These graphs are merged into one, and the minimum sensor groups that serve as the fundamental monitoring units and a hierarchical structure are derived through spectral clustering, a graph-based clustering method known for its high performance. To calculate the health index of a minimum sensor group, we use a long short-term memory (LSTM)-based AE because traditional statistical control charts may struggle to capture non-linear relationships between sensors. To the best of our knowledge, the existing studies on hierarchical structure have solely utilized statistical control charts. Therefore, the structure we propose aims to leverage the strong anomaly detection capabilities of existing deep learning models while overcoming their limitation

of challenging root cause analysis through the utilization of a hierarchical structure. Finally, we apply our proposed framework to real-world data sets from a semiconductor equipment company in South Korea and verify its effectiveness.

2 HIERARCHICAL STRUCTURE BASED ON SPECTRAL CLUSTERING ALGORITHM

In this section, we discuss the process of building a hierarchical structure, as illustrated in Figure 1. First, we collect semiconductor equipment data from various perspectives and construct a similarity graph for each viewpoint, known as a multi-view data. These undirected and fully connected graphs consist of nodes representing each sensor and edges weighted to indicate the degree of similarity between them, ranging from 0 to 1. We then merge these individual graphs into a unified graph and apply a hierarchical spectral clustering algorithm.

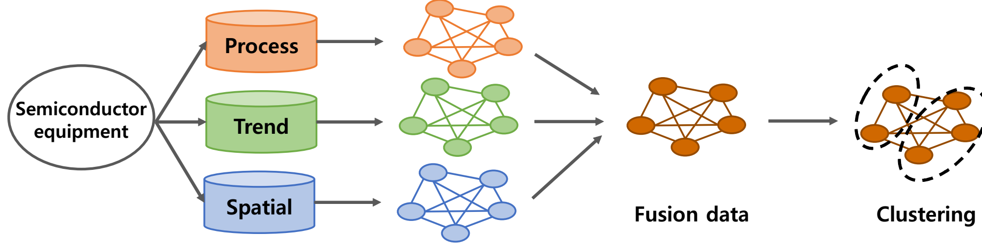


Figure 1: The overall framework of the clustering algorithm.

2.1 A Multi-View Data for Semiconductor Equipment

A multi-view data is a collection of multiple distinct views, commonly used in various fields such as image classification and clustering (Yan et al. 2021; Zhan et al. 2018). In the context of a health index of semiconductor equipment, previous studies have primarily focused on using only one data source, namely real-time time-series data. However, in real-world scenarios, engineers often monitor certain groups of sensors together based on their experience, indicating that there may be multiple views of the data. Thus, it is desirable to utilize the expertise of engineers to formalize the data and build the structure. To construct the multi-view data, we conducted interviews with engineers and collected three different data sources from the views of process, trend, and spatial. For each view, we propose a method to build a similarity graph. It should be noted that there is no clear label for the normal state of the equipment, but the data collected immediately after cleaning can be considered as the closest to the normal state state without loss of generality. Therefore, these data are referred to as ‘normal data’ in this paper, and a multi-view data is considered only within the normal data sets. A detailed explanation of each data view is provided below.

First, ***process data*** of each sensor represents the time-series data collected during deposition. If the process data from two sensors have similarities, it suggests that the information they captured has some correlation or follows a similar control method. Also, if the condition of the machine deteriorates, it may show an abnormal shape different from the normal process data. Therefore, it is recommended to group and monitor sensors that exhibit such similarities. Since not all normal data have identical shapes, we extract representative shapes for each sensor using dynamic time warping (DTW) (Sakoe and Chiba 1978).

DTW is a widely used time-series similarity measure that finds the optimal alignment between two sequences of numerical values based on their shapes. If there are two sequences, $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_n)$, the cost of the optimal alignment can be recursively computed by

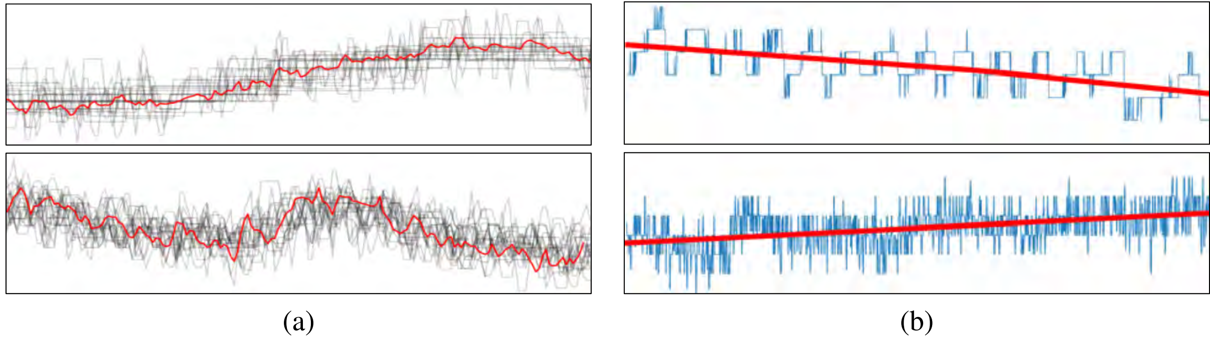
$$D(i, j) = d(x_i, y_j) + \min\{D(i-1, j-1), D(i-1, j), D(i, j-1)\}$$

where $D(i, j)$ indicates the distance between $(x_1, \ldots, x_i)$ and $(y_1, \ldots, y_j)$ with the best alignment and $d(x_i, y_j) = \sqrt{(x_i - y_j)^2}$. The time-series extraction method based on DTW proposed in Petitjean et al. (2011) is used

in this paper. Figure 2 (a) shows the process data of two sensors from atomic layer decomposition (ALD) equipment. The black and red lines correspond to the normal data and the process data extracted via DTW, respectively. The x-axis and y-axis indicate the time and the value, but the indices are omitted due to a confidential issue. It can be seen that the DTW-based extraction method efficiently derives the representative shape of several normal data. Using process data p_i of sensor s_i, the similarity between two process data p_i and p_j can be defined by a radial basis function (RBF) designed for time-series data (Langone et al. 2016) as follows.

$$K_1(p_i, p_j) = exp(-||p_i - p_j||^2_{cd} / \sigma^2_{cd})$$

where $||p_i - p_j||^2_{cd} = \sqrt{1/2(1 - R(p_i, p_j))}$ and $R(p_i, p_j)$ is the Pearson correlation coefficient between process data p_i and p_j.



(a) (b)

Figure 2: Multi-view data: (a) process data, (b) trend data.

Next, ***trend data*** indicates the inherent tendency of each sensor according to the cumulative usage time. Equipment deterioration can occur despite periodic cleaning as the time of equipment use accumulates. As a result, even the data collected immediately after cleaning can have a pattern in which the value increases or decreases slightly over time. The trend is the primary indicator that engineers consider for monitoring, and sensors with similar trends can be interpreted as having similar effects due to aging. Thus, trend data is selected as the second data. To obtain the trend data, the normal data is concatenated chronologically, and then ℓ_1 trend filtering is applied (Kim et al. 2009), which provides piece-wise linear trend estimates, to the concatenated data. The objective of ℓ_1 trend filtering is to minimize the following equation:

$$\frac{1}{2}\sum_{t=1}^{M}(x_t - z_t)^2 + \lambda \sum_{t=2}^{M-1} |z_{t-1} - 2z_t + z_{t+1}|$$

where x_t and z_t are the concatenated normal data and the trend estimate at time t, respectively, and M is the length of the concatenated data. In the equation, the size of the residual, $x_t - z_t$, is computed in the first term, while the second term regulates the smoothness of the estimated trend. When three estimated points (z_{t-1}, z_t, z_{t+1}) lie on a line, the second term becomes 0, and the trade-off between the residual and smoothness is controlled by the parameter λ. A smoother trend can be obtained as the value of λ increases. Figure 2 (b) illustrates the examples of the trend data. The blue line indicates the concatenated data, while the red line shows the trend data when λ is set to 5000. It can be seen that the trend of the above sensor data is decreasing while the below sensor has an increasing trend.

Based on the trend data, we calculate the similarity between sensors as follows. First, we classify trends into four types: increasing, decreasing, constant, and curve. Then, we apply the k-means clustering algorithm with a value of k set to four to find similar trend groups. The DTW distance measure is used for the clustering algorithm. After clustering, each sensor is assigned to a trend cluster $C(t_i)$ based on its

trend data t_i. Finally, the similarity between sensors s_i and s_j is defined as the proportion of trend clusters in which the two sensors are assigned to the same cluster as follows.

$$K_2(t_i,t_j) = \begin{cases} 1, & \text{if } C(t_i) == C(t_j) \\ 0, & \text{otherwise} \end{cases}$$

Note that the process data is obtained during one wafer processing, whereas trend data is collected during the period between cleaning operations.

Finally, ***spatial data*** of sensors is used for clustering. The schematic layout of the semiconductor equipment is presented in Figure 3. It comprises a chamber, a pump for residual discharge, a mass flow controller (MFC) responsible for the management of chemical gases, and pipes that are connected to the MFC, and there are 12 sensors (s_1, s_2,...,s_{12}) in the figure. The information being observed by the sensors varies depending on their location, and the same information is being observed in several other locations. For example, s_1 and s_4 in Figure 3 monitor the operation of the chemical gas related to MFC1 at different locations.

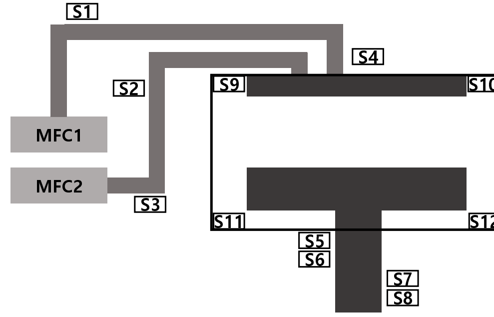


Figure 3: The schematic layout of the semiconductor equipment.

Monitoring techniques based on the similarity of sensor observations are preferred in the field. Therefore, we formulate new spatial data as follows. First, sensors are classified according to the type of information being observed, denoted as $G(s_i)$. In Figure 3, sensors s_1 and s_4, sensors s_2 and s_3, sensors s_5, s_6, s_7, and s_8, and sensors s_9, s_{10}, s_{11} and s_{12} have the identical $G(s_i)$ value, respectively. Next, the Manhattan distance is employed to measure the distance between sensors with the same $G(s_i)$ value. A weighted version is used to assign different weights to each axis, placing greater emphasis on more significant axes. This is because, in the case of pipes or pumps, the distance along the z-axis is more critical than the distance along the other two axes due to their shapes. Let the weight of axis q be defined as α_q, and let the coordinates of sensor s_i be (i_x, i_y, i_z). The distance $d(s_i, s_j)$ between two sensors s_i and s_j and the similarity can be defined by

$$d(s_i,s_j) = \sum_{q \in A} \alpha_q |i_q - j_q|$$

$$K_3(s_i,s_j) = \begin{cases} 1 - \frac{d(s_i,s_j)^2}{Q} & \text{if } G(s_i) == G(s_j) \\ 0, & \text{otherwise} \end{cases}$$

where $Q = \max_{s_i,s_j \in S, s_i \neq s_j} d(s_i,s_j)^2$, and S and A are the set of sensors and axes, respectively. The square term in $K_3(s_i,s_j)$ enables it to effectively consider close distances to be closer and far distances to be far. We use 0.25, 0.25, and 0.5 for α_x, α_y, and α_z, respectively.

After creating graphs, they are fused into a single graph for use in clustering. The weighted sum method, a commonly used technique in graph fusion, is employed for this purpose (Shi et al. 2020). Let the weight of each view's graph be denoted by w_v: w_1 for process data, w_2 for trend data, and w_3 for

spatial data. These weights are parameters that the user can choose to reflect the importance of each view depending on the types of equipment and their expertise. The final weight $w(s_i, s_j)$ of the edge connecting nodes s_i and s_j in the final graph is defined by the following formula.

$$w(s_i, s_j) = w_1 K_1(p_i, p_j) + w_2 K_2(t_i, t_j) + w_3 K_3(s_i, s_j)$$

2.2 Hierarchical Spectral Clustering

Spectral clustering (Von Luxburg 2007) is a graph-based clustering method that uses a similarity graph to capture the relative relationships and connections between data points. The objective is to identify a cut that maximizes the sum of edges within a cluster while minimizing the sum of edges between clusters. As it is difficult to find such a cut directly, spectral clustering relies on the eigenvalue decomposition of the Laplacian matrix as its main approach.

More specifically, a Laplacian matrix is constructed from a similarity graph. Let W be the similarity matrix, which represents the weight of each edge in the graph, and then the Laplacian matrix, denoted by L, is defined as

$$L = D - W$$

where D is a diagonal matrix representing the degrees of each vertex. After forming the Laplacian matrix, the eigenvalues and eigenvectors of L are computed, and the eigenvectors are used to embed the original data into a lower-dimensional space to make it easier to perform clustering. Specifically, the k eigenvectors corresponding to the smallest k eigenvalues (i.e., the first k eigenvectors after sorting the eigenvalues in ascending order) are used to form a new k-dimensional representation of the data. While various clustering methods can be employed on the low-dimensional representation of the data, this paper utilizes hierarchical clustering with the Ward distance metric (Murtagh and Legendre 2011) to create the hierarchical structure automatically. By using different numbers of thresholds, it is possible to adjust the level of detail in the hierarchical structure, and identify clusters of sensors at different levels of granularity. For instance, using two thresholds can generate a structure with three levels, and the clusters obtained at the lowest threshold can serve as the minimum sensor sets for computing a health index.

3 HIERARCHICAL HEALTH INDEX FRAMEWORK

3.1 AutoEncoder

The AE (Bank et al. 2020), widely adopted for anomaly detection and monitoring, is used to derive a health index. The AE consists of an encoder network (g_ϕ) that maps the input data to a lower-dimensional latent space representation (z) and a decoder network (f_θ) that reconstructs the original input from this representation, as shown in Figure 4.

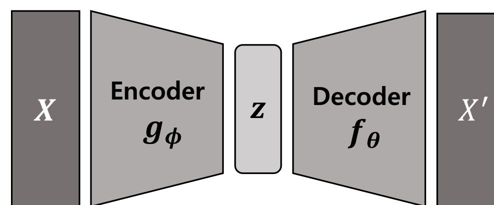


Figure 4: The structure of the AE.

The goal of the AE is to minimize the reconstruction error (RE), which is the difference between the input X and the output $X^{'}$. Because abnormal data has different values and shapes from normal data, the AE trained with normal data cannot adequately reconstruct the abnormal data, leading to the high RE value. Hence, the RE value is a suitable metric for indicating the degree of divergence between the present state and the normal state. To derive a health index, we use AEs and construct multiple independent AEs for

each minimum sensor set, training them only on normal data. For instance, if there are three minimum sensor sets, we create three AEs using different sensor sets. To effectively capture the time-dependent properties, we use the LSTM architecture for both the encoder and decoder. LSTM is a type of recurrent neural network that uses memory cells to selectively store or discard information over time. This enables it to retain long-term dependencies and manage input sequences of variable length.

3.2 Calculate the Health Index

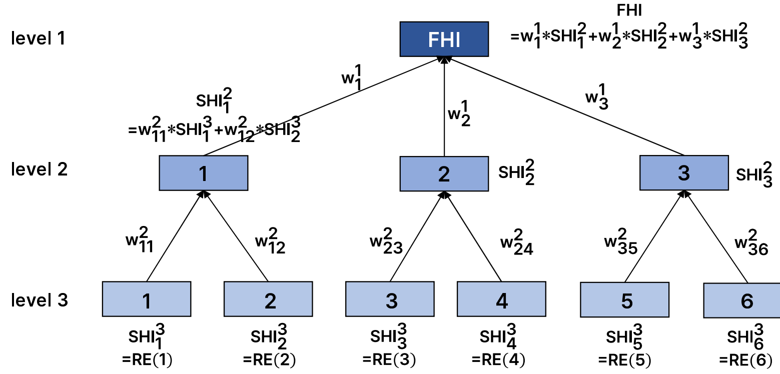


Figure 5: The FHI calculation in a bottom up manner.

After computing the RE values for each minimum sensor set, we calculate the semi-health index (SHI) for each level by taking a weighted sum of the RE values from the immediately preceding lower level. Let us assume that there are l levels in the hierarchical structure, where level 1 represents the final health index (FHI) and level l represents the minimum sensor sets. To calculate the FHI, we traverse the levels from l to 1 in a bottom-up manner, computing the SHI for each level and aggregating the information from the previous level with a weighted sum to obtain the SHI for the current level. Finally, the FHI is obtained as the weighted sum of the SHI values for all levels, which provides a comprehensive assessment of the system's health.

There are n_k clusters in level k, and SHI_v^k denotes the SHI of cluster v at level k. Also, P_v^k is the set of child clusters at level $k+1$ of cluster v at level k. Figure 5 shows a simple example that $l = 3$ and three and six clusters are in level 2 and level 3, respectively. For cluster 1 at level 2, there are 2 clusters (cluster 1 and cluster 2 at level 3) in P_1^2. The weight value w_{vu}^k exists between cluster u of level $k+1$ and cluster v of level k if cluster u is in P_v^k. For $k = 1$, there is only one cluster, so we express the weight as w_u^1. The sum of w_{vu}^ks for cluster v is equal to one. Then, SHI_v^k can be calculated by

$$SHI_v^l = RE(v) \quad \forall v = 1, ..., V_l$$

$$SHI_v^k = \sum_{u \in P_v^k} w_{vu}^k * SHI_u^{k+1} \quad \forall v = 1, ...V_k \;\; \forall k = 2, ..., l-1$$

$$FHI = \sum_{u \in P_v^1} w_u^1 * SHI_u^2$$

where $RE(v)$ and V_k are the RE value of cluster v and the number of clusters in level k, respectively. By computing several SHIs at varying levels, the FHI can offer a more comprehensive assessment of the equipment's overall health, empowering engineers to make more informed decisions regarding repairs and maintenance.

4 REAL WORLD APPLICATION

4.1 Data Description

We test our framework using real data sets obtained from an ALD machine in a semiconductor equipment company located in South Korea. The data set comprises 189 sensors, out of which only 54 sensors were selected based on the expertise of engineers and the variance value. These sensors include temperature, pressure, vaporizer, and others. The data set was collected every 0.2 seconds over a period of four months, from May 17th, 2021, to September 4th, 2021. Due to confidentiality concerns, one batch of data per day was collected. Moreover, the machine was cleaned every 3 to 5 days to maintain its optimal operational state.

4.2 Experimental Results

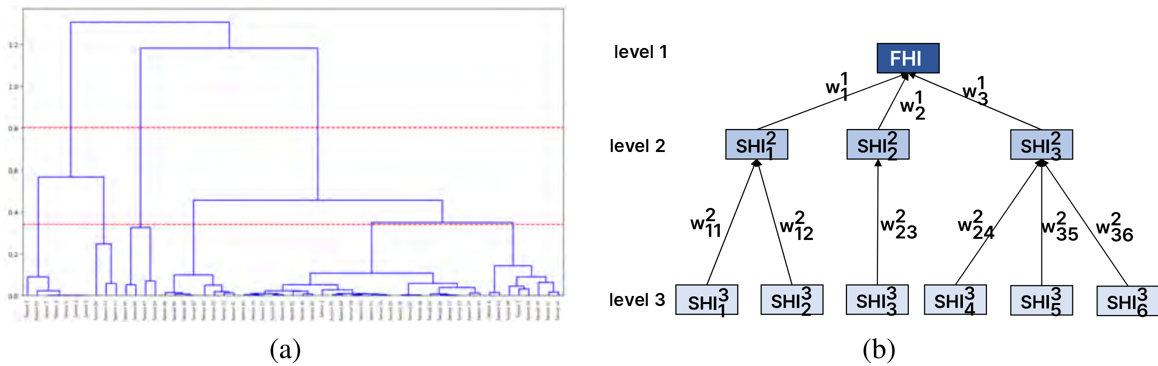


Figure 6: Hierarchical spectral clustering : (a) clustering result with two thresholds, (b) hierarchical structure used for calculating the FHI.

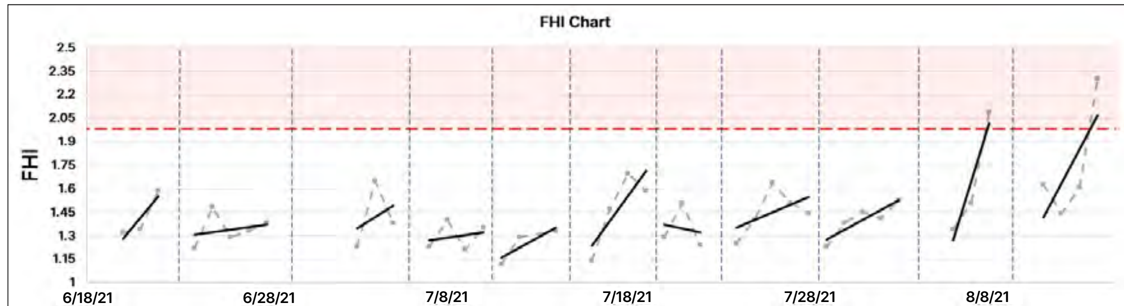


Figure 7: The FHI chart.

The process, trend, and spatial data sets were constructed based on the collected data set. The weights of 0.3, 0.5, and 0.2 were used for w_1, w_2, and w_3, respectively, to obtain the final graph. During the clustering process, the value of k that maximizes an eigengap, which is the difference between two consecutive eigenvalues, is selected, so k of 2 is used. The result of clustering is shown in Figure 6 (a). The red dotted lines indicate the thresholds to determine clusters, which were empirically set at 0.38 and 0.8. Figure 6 (b) shows the hierarchical structure of FHI based on the corresponding thresholds. The minimum sensor sets were obtained by a threshold of 0.38, which results in six clusters in level 3. Additionally, a threshold of 0.8 yields three clusters in level 2. These thresholds and the clustering results have been confirmed by engineers.

For hyper-parameters of AEs, we use a batch size of 256, a learning rate of 0.0005 with ADAM optimizer, three LSTM layers, and a time window size of 10 for the input data. Also, the weights w^2_{11}, w^2_{12}, w^2_{23}, w^2_{34}, w^2_{35}, and w^2_{36} were set to 0.7, 0.3, 1, 0.25, 0.625, and 0.175, respectively. These values were derived by taking the ratio of the number of sensors. Similarly, the values of w^1_1, w^1_2, and w^1_3 were derived using the same approach, resulting in the weights of 0.25, 0.02, and 0.73, respectively.

Figure 7 illustrates the FHIs for the period between June 19th, 2021, and August 13th, 2021. The blue dotted line indicates the point at which cleaning was performed, while the red dotted line represents the sum of the average RE and three standard deviations of the normal data, which is commonly used for the threshold of anomaly situations. The FHI values and their respective trend lines are represented by the gray points and black lines, respectively. It can be seen that most of the FHI trends display an increasing tendency immediately after cleaning until just before the subsequent cleaning process, with the FHI values increasing significantly in August.

Although there is no clear state label, records indicate several issues occurring during the cleaning process in August, which suggest that the proposed FHI effectively represents degradation due to cumulative usage. Furthermore, a hierarchical structure can provide several indices by dividing the equipment into subsystems through level 2 or level 3 SHI values. This approach enables the checking of only the related sensors rather than all sensors, which can be more efficient and effective.

5 CONCLUSION

This paper has proposed a novel framework for deriving a hierarchical health index for semiconductor equipment. The framework involves defining a multi-view data set obtained through interviews with engineers and developing a method for constructing a similarity graph for each data. Furthermore, an automated hierarchical structure has been developed through spectral clustering based on a multi-view data set. Finally, a health index has been calculated based on the RE, and the experimental results with real data sets have demonstrated that our index efficiently represents the machine state over time.

In the future, we will improve the accuracy of the framework by automating the weights used for the FHI calculation. Additional data has been collected for more precise performance verification, and discussions with field engineers to implement in the real machine are ongoing. Comparison with other methods is also required.

REFERENCES

Bank, D., N. Koenigstein, and R. Giryes. 2020. "Autoencoders". *arXiv preprint arXiv:2003.05991*. https://arxiv.org/abs/2003.05991, accessed 26th June 2023.

Blázquez-García, A., A. Conde, U. Mori, and J. A. Lozano. 2021. "A Review on Outlier/Anomaly Detection in Time Series Data". *ACM Computing Surveys (CSUR)* 54(3):1–33.

Chao, A.-G., S. Tseng, D. S.-H. Wong, S.-S. Jang, and S.-P. Lee. 2008. "Systematic Applications of Multivariate Analysis to Monitoring of Equipment Health in Semiconductor Manufacturing". In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, 2330–2334. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Chien, C.-F., Y.-J. Chen, and J.-T. Peng. 2010. "Manufacturing Intelligence for Semiconductor Demand Forecast based on Technology Diffusion and Product Life Cycle". *International Journal of Production Economics* 128(2):496–509.

Deng, W., Y. Li, K. Huang, D. Wu, C. Yang, and W. Gui. 2023. "LSTMED: An Uneven Dynamic Process Monitoring Method based on LSTM and Autoencoder Neural Network". *Neural Networks* 158:30–41.

Ge, Z., and Z. Song. 2010. "Semiconductor Manufacturing Process Monitoring based on Adaptive Substatistical PCA". *IEEE Transactions on Semiconductor Manufacturing* 23(1):99–108.

Hu, C.-H., and S.-F. Su. 2004. "Hierarchical Clustering Methods for Semiconductor Manufacturing Data". In *IEEE International Conference on Networking, Sensing and Control, 2004*, Volume 2, 1063–1068. IEEE.

Jeong, Y.-S., B. Kim, and Y.-D. Ko. 2013. "Exponentially Weighted Moving Average-based Procedure with Adaptive Thresholding for Monitoring Nonlinear Profiles: Monitoring of Plasma Etch Process in Semiconductor Manufacturing". *Expert Systems with Applications* 40(14):5688–5693.

Kim, S.-J., K. Koh, S. Boyd, and D. Gorinevsky. 2009. "ℓ_1 Trend Filtering". *SIAM Review* 51(2):339–360.

Krauel, C., and L. Weishäupl. 2016. "Multivariate Approach for Equipment Health Monitoring". *IFAC-PapersOnLine* 49(12):716–720.

Langone, R., R. Mall, C. Alzate, and J. A. Suykens. 2016. "Kernel Spectral Clustering and Applications". *Unsupervised Learning Algorithms*:135–161.

Lee, C.-Y., and Z.-H. Dong. 2019. "Hierarchical Equipment Health Index Framework". *IEEE Transactions on Semiconductor Manufacturing* 32(3):267–276.

Lee, S., M. Kwak, K.-L. Tsui, and S. B. Kim. 2019. "Process Monitoring Using Variational Autoencoder for High-Dimensional Nonlinear Processes". *Engineering Applications of Artificial Intelligence* 83:13–27.

Lee, S.-P., A.-K. Chao, F. Tsung, D. S. H. Wong, S.-T. Tseng, and S.-S. Jang. 2011. "Monitoring Batch Processes with Multiple ON–OFF Steps in Semiconductor Manufacturing". *Journal of Quality Technology* 43(2):142–157.

Li, Y., and X. Zhang. 2014. "Diffusion Maps based K-Nearest-Neighbor Rule Technique for Semiconductor Manufacturing Process Fault Detection". *Chemometrics and Intelligent Laboratory Systems* 136:47–57.

Li, Z., L. Tian, Q. Jiang, and X. Yan. 2021. "Distributed-Ensemble Stacked Autoencoder Model for Non-Linear Process Monitoring". *Information Sciences* 542:302–316.

Murtagh, F., and P. Legendre. 2011. "Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm". *arXiv preprint arXiv:1111.6285*. https://arxiv.org/abs/1111.6285, accessed 26th June 2023.

Petitjean, F., A. Ketterlin, and P. Gançarski. 2011. "A Global Averaging Method for Dynamic Time Warping, with Applications to Clustering". *Pattern Recognition* 44(3):678–693.

Sakoe, H., and S. Chiba. 1978. "Dynamic Programming Algorithm Optimization for Spoken Word Recognition". *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26(1):43–49.

Shi, S., F. Nie, R. Wang, and X. Li. 2020. "Auto-Weighted Multi-View Clustering via Spectral Embedding". *Neurocomputing* 399:369–379.

Skinner, K. R., D. C. Montgomery, and G. C. Runger. 2004. "Generalized Linear Model-based Control Charts for Discrete Semiconductor Process Data". *Quality and Reliability Engineering International* 20(8):777–786.

Von Luxburg, U. 2007. "A Tutorial on Spectral Clustering". *Statistics and Computing* 17:395–416.

Yan, X., S. Hu, Y. Mao, Y. Ye, and H. Yu. 2021. "Deep Multi-View Learning Methods: A Review". *Neurocomputing* 448:106–129.

Yang, C., J. Liu, Y. Zeng, and G. Xie. 2019. "Real-Time Condition Monitoring and Fault Detection of Components based on Machine-Learning Reconstruction Model". *Renewable Energy* 133:433–441.

Yu, H.-C., K.-Y. Lin, and C.-F. Chien. 2014. "Hierarchical Indices to Detect Equipment Condition Changes with High Dimensional Data for Semiconductor Manufacturing". *Journal of Intelligent Manufacturing* 25:933–943.

Zhan, K., C. Niu, C. Chen, F. Nie, C. Zhang, and Y. Yang. 2018. "Graph Structure Fusion for Multiview Clustering". *IEEE Transactions on Knowledge and Data Engineering* 31(10):1984–1993.

Zhang, C., D. Song, Y. Chen, X. Feng, C. Lumezanu, W. Cheng, J. Ni, B. Zong, H. Chen, and N. V. Chawla. 2019. "A Deep Neural Network for Unsupervised Anomaly Detection and Diagnosis in Multivariate Time Series Data". In *Proceedings of the AAAI Conference on Artificial Intelligence*, Volume 33, 1409–1416.

AUTHOR BIOGRAPHIES

JEONGSUN AHN is a Ph.D. student in Department of Industrial & Systems Engineering, Korea Advanced Institute of Science and Technology (KAIST). She received a M.S. in industrial and systems engineering from KAIST. She is interested in discrete event systems modeling and scheduling. Her e-mail address is jeongsun@kaist.ac.kr.

HONG-YEON KIM is a M.S. student in Department of Industrial & Systems Engineering, Korea Advanced Institute of Science and Technology (KAIST). He is interested in discrete event systems modeling and scheduling. His e-mail address is ghd4567@kaist.ac.kr.

SANG-HYUN CHO is a Ph.D. student in Department of Industrial & Systems Engineering, Korea Advanced Institute of Science and Technology (KAIST). He received a M.S. in industrial and systems engineering from KAIST. He is interested in scheduling methodologies and applications. His e-mail address is ie02002@kaist.ac.kr.

HONGYEON KIM is an engineer of software team in Wonik IPS. He is interested in automated control of semiconductor equipment. His e-mail address is kimhy4@wonik.com.

HYEONJEONG CHOI is an engineer of software team in Wonik IPS. She is interested in automated control of semiconductor equipment. Her e-mail address is hjchoi1020@wonik.com.

DAIN HAM is an engineer of software team in Wonik IPS. She is interested in automated control of semiconductor equipment. Her e-mail address is hamdi@wonik.com.

HYUN-JUNG KIM is an Associate Professor with the Department of Industrial & Systems Engineering, Korea Advanced Institute of Science and Technology (KAIST). She received B.S., M.S., and Ph.D. in industrial and systems engineering from KAIST. Her research interests include discrete event systems modeling, scheduling, and control. Her email address is hyunjungkim@kaist.ac.kr. Her website is https://msslab.kaist.ac.kr.