electronics

MDPI

Article

Industrial Internet of Things for Condition Monitoring and Diagnosis of Dry Vacuum Pumps in Atomic Layer Deposition Equipment

Yongho Lee, Chanyoung Kim and Sang Jeen Hong *

Department of Electronics Engineering, Myongji University, 116 Myongji-ro, Yongin-si 17058, Gyeonggi-do, Korea; cylinder12@naver.com (Y.L.); chan5863@naver.com (C.K.)
* Correspondence: samhong@mju.ac.kr; Tel.: +82-31-330-6374

Abstract: In the modern semiconductor industry, defective products occur with unexpected small variables due to process miniaturization. Managing the condition of each part is an effective way of preventing unexpected errors. The industrial internet of things (IIoT) environment, which can monitor and analyze the performance degradation of parts that affect process results, enables advanced process yield management. This paper introduces the IIoT concept-based data monitoring and diagnostic system construction results. The process of pump vibration data acquisition is explained to evaluate the effectiveness of this system. The target process is deposition. The purpose of the system is to detect degradation of pumps due to by-products of the atomic layer deposition (ALD) process. The system consists of three areas: a data acquisition unit using six vibration sensors, a Web access-based monitoring unit that can monitor vibration data, and an Azure platform that searches for outliers in vibration data.

Keywords: industrial internet of things; web-access; azure; accelerometer; preprocessing

check for updates

Citation: Lee, Y.; Kim, C.; Hong, S.J. Industrial Internet of Things for Condition Monitoring and Diagnosis of Dry Vacuum Pumps in Atomic Layer Deposition Equipment. *Electronics* **2022**, *11*, 375. https://doi.org/10.3390/electronics11030375

Academic Editors: Matteo Golfarelli and Enrico Gallinucci

Received: 6 January 2022
Accepted: 24 January 2022
Published: 26 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Internet of things (IoT) is a technological innovation that creates an environment of convergence around nature. Objects and conditions are digitally represented and remotely monitored over the networks. They are maintained and controlled in intelligent ways and conventional hard limits. Five layers of IoT technology were suggested to define its key technology components: (1) perception layer (data gathering) by a physical object, sensor, and actuator; (2) network layer (data transmission) via network technologies; (3) middleware layer (process information) for data analysis and decision making; (4) application layer (smart application) with graphic data representation; and (5) business layer (system management) for the benefits of business [1]. When IoT was conceptually developed, the network and middleware layers were more focused on building an IoT framework to utilize various IoT data from numerous perception layers. However, the utilization of the developed IoT has been applied within the limited scope of the domain. A great potential of the IoT convergence lies in industrial domains, where the entities for the IoT needs are many and a huge business innovation opportunity, and the evolution of IoT continued with industrial IoT (IIoT) [2].

With the advent of industry 4.0 strategy, factory automation with industry internet can be represented by a programmable logic controller based supervisory control and data acquisition, manufacturing excursion system, and converging information technology (IT) and operational technology of enterprise resource planning to achieve the goal of manufacturing technology to increase the productivity and flexibility to reduce production cost. To achieve one step closer to the success of IIoT, Chen et al. suggested a framework for collaborative sensing intelligence with a research challenge [3]; however, the actual data acquisition considering data quality remains a practical challenge in a perception layer. The

differences between the machine-oriented IIoT and human-centered consumer IoT include existing devices and standards for data acquisition, fixed connectivity with centralized network management, and high data volume [4]. One of the most promising applications of IIoT is facility or equipment maintenance by automated condition monitoring data acquisition and abnormal detection, which is called the predictive maintenance, or e-maintenance, in the manufacturing domain [5–8]. Rather simplified practice on facility management with temperature and humidity monitoring or standalone equipment with vibration sensors has been demonstrated in a capacity of characteristic condition sensing with existing sensors, wireless data transmission, and data visualization. However, a more realistic facility-level IIoT requires a very complicated framework with subject matter experts (SMEs). Recently, a comprehensive categorical framework for IIoT was proposed. Additionally, it is essential to set a practical classification scheme for the detailed and successful IIoT framework [9].

Semiconductor chip manufacturing facilities, called FAB, consists of multiple facility layers/areas: a fan filter unit layer for cleaning air supply, cleanroom layer for operating wafer production and inspection, service area for production equipment, and sub-fab layer of clean and facility sub-fab [10]. A plenum layer can be included, as shown in Figure 1. In current high-volume memory manufacturing, the facility expands four times larger than the soccer field to increase chip production. The manufacturing accordance of thousands of production tools in the cleanroom layer and several thousands of dry pumps and gas scrubbers in the clean sub-fab layer are essential to increase the production efficiency by extending the preventive maintenance (PM) schedule. The next-generation semiconductor devices require new materials to enhance chip performance. However, the formation of new materials with a gas-phase reaction leaves unexpected by-products to call maintenance [11,12].

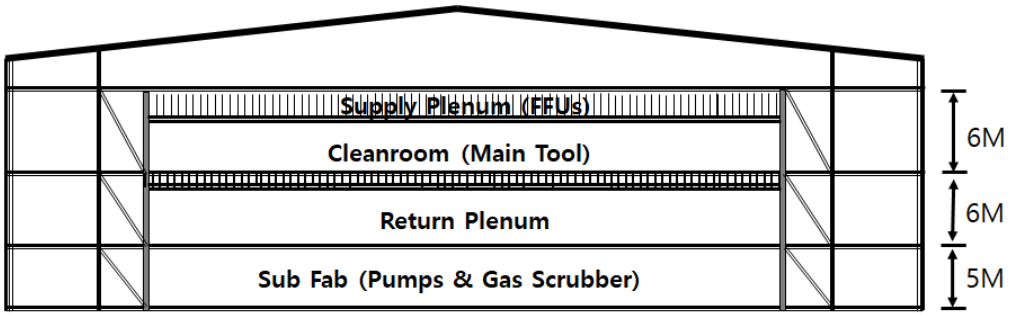


Figure 1. Semiconductor FAB structural layout.

Many processes in semiconductor fabrication require a vacuum environment to meet the chemical and physical requirements, and the use of dry vacuum pumps prevails in semiconductor manufacturing [13]. The dry vacuum pumps perhaps do keep the process low in the process chamber. The pumping efficiency drifts along the time of use, which can jeopardize the process quality; thus, several studies have been conducted on the diagnostics on dry pumps [14–17]. Current semiconductor fabrication technology employs diverse process gas to form or remove thin film materials in plasma. For instance, atomic layer deposition (ALD) of *TiN* thin films with $TiCl_4$ and NH_3 process gases forms the undesired TiO_2 by-product and coagulation of the undesired TiO_2 by-product under the vapor pressure in the foreline of the chamber and inside the dry vacuum pump. The scale, formed from TiO_2 power, degrades the pumping efficiency of dry vacuum pumps. It also shortens the remaining useful time for the pump. Indeed, semiconductor production is inevitable to hold for maintenance until the replacement and qualification for production are made. The role of a dry vacuum pump is to maintain a constant pressure level as an auxiliary facility; however, when it goes wrong, one should pay for the expensive production loss in semiconductor chip manufacturing. Therefore, the monitoring and

diagnosis of dry vacuum pumps in semiconductor manufacturing with the scope of IIoT give many benefits to avoid misprocessing from unexpected component drift or failure.

In the existing ALD environment, a pump is used as a method of conducting PM at regular intervals rather than finding and solving problems at the time of faulty occurrence. This is for the purpose of prevention, but it is because there is no means to feedback the abnormal state that has occurred in the pump. Therefore, the change of the pump was observed based on the amount of change in the vibration of the pump acquired in the environment where the by-product was accumulated. Through this, an IIoT monitoring system that can monitor changes in real time was created. By analyzing the vibration data of the pump in Azure, a cloud-based analysis platform, we created an analysis system that accurately classifies outliers. Finally, it was integrated to build an IIoT pump monitoring and diagnostic system. This new system can monitor the pump status in the ALD process environment and reflect the feedback of all major vibration points through ML, a multivariate analysis technique.

A literature survey on the condition monitoring and diagnosis of dry vacuum pumps shows a good technical improvement for characterizing the system using advanced signal processing techniques. Jiang et al. proposed a wavelet cluster-based frequency response analysis for bearing defects, employing a vibrational transducer. They suggested that methods of filtering and modulation of the acquired frequency spectral data were effective in fault diagnosis [18]. Similarly, fuzzy-model-based condition monitoring was suggested by analyzing the exhaust pressure signal [19]. Additionally, a prediction of vacuum pump degradation in semiconductor processing was proposed to estimate the remaining useful time of the pump [20]. This research also employed foreline pressure measurement to acquire pump status data in the perception layer of IIoT and data analysis for decision making using an artificial neural network as a middleware layer. An interesting aspect of this research includes gas data as chamber data, whether the pump is idling or processing modes. This is one aspect of SME. Along with the technology development of IT and computational intelligence technology, the development of the monitoring and diagnostic method of dry vacuum pumps can be found [21–23].

In the industrial field, machine learning has been utilized in various aspects to increase the efficiency of the manufacturing process. A machine learning concept was used to detect anomalies based on data acquired by IoT in a Wi-Fi indoor localization environment [24]. In addition, there has been a study using a virtual metrology model based on a tree ensemble method, such as a random forest to reduce the measurement time and cost [25]. In the semiconductor field, using plasma measurement sensors such as Optical Emission Spectroscopy, machine learning analysis was performed as a means of linking the complex relationship between sensor data and the process results [26].

In this study, we present a realistic practice of the in situ monitoring and diagnosis of the dry vacuum pump installed at the semiconductor fabrication equipment within the scope of the practical application of IIoT. The perception layer is established with six vibrational transducers to collect in situ pump status data. The collected analog vibrational datasets are then converted to digital signals using a data acquisition module. For this, 2.4 GHz Wi-Fi wireless data transmission is used. The data acquisition environment can use wired instead of wireless to prevent data loss and enhance the security [27]. The data visualization application layer is constructed with Wise-PaaS provided by Advantech Korea. The average deployment cost for data acquisition and monitoring system configuration is approximately 3354 USD (Including six accelerometers and cables, data acquisition card, and Wise-PaaS). Finally, Microsoft Azure is employed as a middleware layer for data analysis considering the university research environment. There exists a virtual machine and storage usage fee for machine learning computation in the Azure environment. A virtual machine with 4 vCPUs with 14 GB of RAM and 200 GB temporary storage costs about 0.229 USD per hour. Figure 2 shows the cloud diagnostic system structure and edge monitoring system. By demonstrating the four layers of IIoT with the application

in a semiconductor manufacturing facility, the final business layer can easily be tried by industry, confirming the potential of IIoT in the near future.

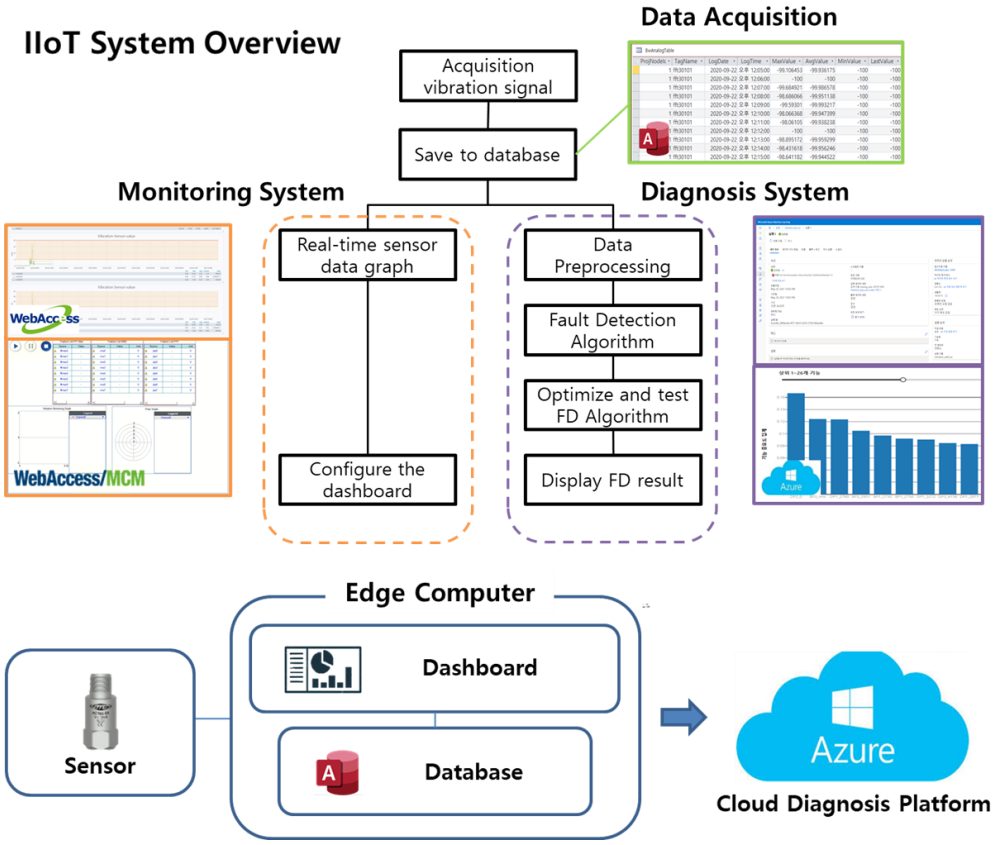


Figure 2. Edge monitoring system and cloud diagnostic system.

2. Subject Matter

2.1. Equipment Degradation Mechanism

By-products generated during the semiconductor process may degrade the functionality of the dry vacuum pump and shorten the remaining useful time before breakdown maintenance (BM). TiO_2 powder combined in the exhaust process after ALD of *TiN* thin film is essential to the performance and lifespan of the pump. The combination of TiO_2 powder and NH_4Cl particles can be hardened inside the exhaust line and dry pump. The solidified gas-phase by-products lower the conduction performance, creating an alarm on the pressure controller of the main process equipment. When it leaches to the equipment operational limits, a warning alarm from the pressure gauge inside the dry pump shuts down the pump operation via interlock functions of the high-end fabrication equipment. The dry pumps in the semiconductor manufacturing facility operate 24 h a day. It is desired that they can be operated for more than a year without maintenance. However, in reality, the lifetime of the dry pumps is as short as a month due to the undesired particles in some processes.

Failure to properly treat process by-products in the IC manufacturing process can lead to clogging of the exhaust system, damage to vacuum pumps and valves, and unexpected equipment failure. Foreline clogging decreases the flow conductance and increases particles inside the chamber. It also causes a safety problem due to the accumulation of by-products. For example, in the ZrO_2 ALD process, by-products generated in the exhaust pipe have a high risk of dust explosion [28].

Semiconductor by-products depend on the type of process gas and condition under which the chemical reactions can occur. In the ALD process, the unreacted process gas is deposited inside the vacuum pump and causes pump failure [29]. TiN thin film ALD

process using NH_3, N_2, and $TiCl_4$ gases generate TiO_2 and NH_4Cl by a chemical bond. The chemical bonds of semiconductor by-products that occur in the ALD process are as follows:

$$TiCl_4 + O_2 \rightarrow TiO_2 + 2Cl_2, \quad (1)$$

$$NH_3 \rightarrow {NH_2}^- + H^+, \quad (2)$$

$$NH_3 + H^+ \rightarrow {NH_4}^+, \quad (3)$$

$${NH_4}^+ + Cl^- \rightarrow NH_4Cl. \quad (4)$$

According to Equation (1), the $TiCl_4$ gas exiting through the exhaust stage after the process meets O_2 in the exhaust process and chemically bonds to TiO_2, leaving $2Cl_2$. Then, NH_3 gas is dissociated into ${NH_2}^-$ and H^+ during the ionization process. The separated H^+ recombines with the existingNH_3, forming ${NH_4}^+$, as shown in Equation (3). Finally, NH_4Cl (Ahmonium Chloride; an organic compound similar to white crystalline salt) and TiO_2 is formed as chemical reaction by-products.

The $TiCl_4$ gas reacts with O_2 in the exhaust process to form TiO_2 powder; NH_3 and $TiCl_4$ combine to form NH_4Cl. The basic nature of NH_4Cl causes internal corrosion of the pump. One may concern a disruption of the flow inside the pump and the occurrence of pump down due to the combination with TiO_2. In fact, a more significant concern is the repeated metal surface corrosion of the dry pump rotor and shaft from NH_4Cl and the coagulation of TiO_2 nanoparticles inside the dry pump parts, which affect the long-term reliability of the equipment.

Therefore, to prevent an undesired generation of corrosion of metal parts and generation of by-product particles, the atmosphere of the pump is operated in a condition comparable to the melting point of the gas-phase by-products. Then, a cold trap is installed to convert gas-phase particles to powder to be captured in a filter. All these engineering effects may provide an indirect treatment, but more informative data for the pump condition can be used for scheduling PM and preventing unnecessary BM. An in situ monitoring system for determining the changes inside the pump and detecting abnormal conditions before the pump goes down is essential in a manufacturing environment. Therefore, this paper presents a realistic construction of an in situ pump status monitoring system capable of a cloud computing-based IoT system. The suggested proactive aims are for data-driven maintenance efficiency for the clustered dry pump diagnostic system in a high-volume semiconductor manufacturing environment.

2.2. Mechanical Vibration Analysis

Vibration is a phenomenon in a physical quantity that fluctuates around the average value. Automated machines consisting of various moving and rotating parts vibrate in response to the force fluctuation during operation. Analyzing the change in vibration characteristics due to a physical change can diagnose a defect of the specific part. The ISO 18436-2 standard presents the effectiveness and rationale of analyzing such component defectives as a vibration mechanism [30]. By-product issues of the pump down occur due to the pumping process load caused by the internal rotational limitation of the pump. Accumulated by-product particles inside the vacuum pump during the process can decrease the pumping speed [31]. Time-domain and frequency-domain analyses are the most well-known methods used in vibration-based equipment diagnosis [32]. The time-domain analysis is based on the amplitude or temporal fluctuation, impact, and symmetricity of the waveform, using the probability density function of the amplitude and the peek-to-peek. The frequency-domain analysis method decomposes time-domain sinusoidal waves into frequency components using fast Fourier transform (FFT). The frequency included in the sinusoidal wave and the amplitude of the frequency spectra can be analyzed. Figure 3 shows an example of the decomposed complex waveform to the frequency by applying FFT in this experiment.

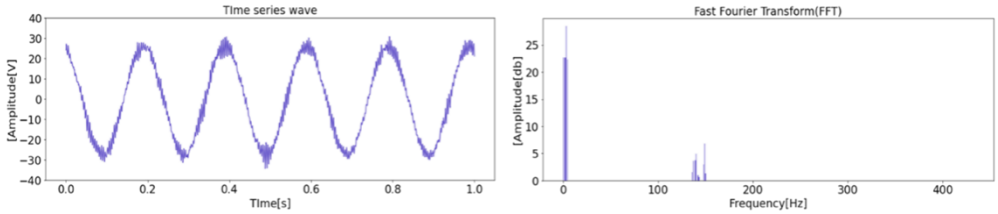


Figure 3. Fast Fourier transform (FFT) example acquired from the experiment.

3. In Situ Vibration Data Acquisition

Finding a health index of the dry pump can help the condition monitoring and diagnosis to prevent undesired machine breakdowns. Feature extraction of the vibrational data from the dry vacuum pump is the most well-known method. However, the structures of the pumping mechanism and mechanical supports with the housing vary from the pump manufacturer. Thus, the acquired data can vary from the different types of pumps and pump manufacturers. A designated dry pump of AA220W manufactured by Ebara, located at the semiconductor process diagnostic research center (SPDRC) at Myongji University was employed to collect the vibrational data in the experimental stage for the easiness of IIoT application development. The AA200W for the high-speed exhaust of large chambers has two-stage pumps, consisting of a screw-type dry pump and a root-type blower pump. The designated dry pump is connected with the ICP type 300 mm production etch chamber for the IIoT system development. However, it still conceptually holds the purpose of the suggested research. Once the suggested system is ready, the developed IIoT system is attached to the dry pump connected to the *TiN* deposition chamber for the application test, which will be described in the later section.

The vibration sensor converts the vibration detected from the machine into a digital signal, and their types are typically a velocity sensor, proximity probe, and accelerometer. The sensitivities in the frequency range are different; the type of sensor should be considered with respect to the frequency response characteristics, signal-to-noise ratio, and sensitivity. In this paper, we used a piezoelectric vibration acceleration sensor with a higher frequency response and higher sensitivity than other vibration sensors to investigate two different models of dry pumps. The employed vibration sensor has a frequency response of about 10 kHz and a sensitivity of 100 mV/g. The noise density over the 10, 100, and 1000 Hz ranges are 14, 2.3, and 2 $\mu g/\sqrt{Hz}$, respectively. A suggested system sampling rate is 20.08 Hz (20.08 data per 1 s). It is also crucial for the data acquisition from the vibration sensors to have concrete physical contact with the entity that affects the contact resonance. The measurable frequency range reduces at low installation resonant frequencies. A magnetic stud that can be fixed with screws is used to minimize external vibrations that occur during data acquisition. The closer and more fixed to the pump surface, the more advantageous it is to acquire high-frequency region data.

For accurate vibration measurement, it is necessary to select an installation location that best reflects the vibration characteristics of the pump. When the pump performance decreases due to by-products being adsorbed inside, the load on the internal bearing increases. Based on the domain knowledge and experience, TiO_2 tends to be adsorbed near the dry pump inlet and internal screw. We assumed that the vibration analysis of the relevant area would also change significantly (perhaps, this is not always true). The vibration sensor measures only the vibration in one direction; therefore, a single sensor cannot measure the lateral vibration caused by the rotational motion of the facility. The lateral vibration data can be used together for a more versatile analysis. Four acceleration sensors are attached to the bearings of dry and booster pumps in the horizontal and vertical directions. The remaining two sensors are attached to the inlet part and screw end of the dry pump. Figure 4 shows a photograph of the pump (taken by the authors) and the set-up points of the accelerometers for the vibrational data acquisition in this exercise.

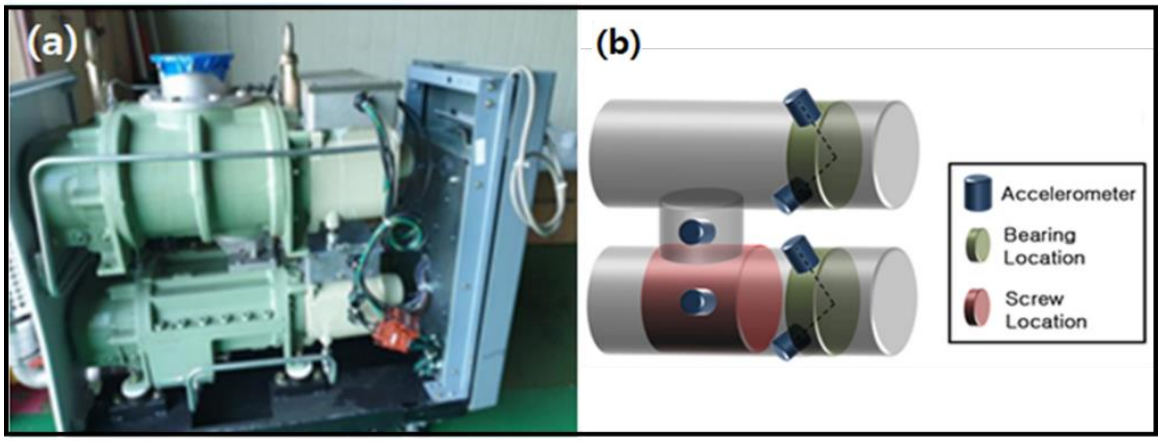


Figure 4. Set-up points of the accelerometer. (**a**) photograph of the pump (taken by the authors); (**b**) the set-up points of the accelerometers.

Edge computing is a computing environment that can operate close to users or data sources. When the instantaneous data management operation is possible, it can effectively reduce the delay time in the conventional data transmission process and prevent the data computation overload. The vibration data acquired from six vibration sensors in this experiment is processed digitally using the data acquisition (DAQ) card installed in the edge computer. The acquired vibration data are converted into digital signals through Advantech's machine condition monitor (MCM), a data analysis tool configured in edge computing, and stored in sinusoidal raw data with root mean square (RMS) and peak-to-peak (PTP) functions applied, and frequency data applied using FFT. The acquired data can be transmitted wirelessly to the 2.4 GHz Wi-Fi wireless I/O module built into the DAQ card. The wired environment used in this paper is limited to the data acquisition stage. The monitoring system was built in the 2.4 GHz Wi-Fi environment, and the Diagnosis system was built in the Azure cloud environment. Additionally, real-time data monitoring sends data to an external computer instead of the measurement site using Wise-PaaS provided by Adventech Korea.

To apply the acquired vibration data to anomaly analysis, it is necessary to perform signal processing suitable for analysis. The signal processing discussed in this paper includes sampling to frequency conversion, filtering, and windowing. Increasing the data resolution in the data acquisition process can increase the accuracy of the data analysis. Since the analyzable frequency changes depending on the measure location, it is possible to set the ideal data resolution by minimizing the max frequency within the analyzable range and maximizing the sampling rate according to the sensor performance. The frequency-domain analysis uses an FIR filter to limit specific frequencies. Time windows, such as Hanning and flat-top windows, can be applied to make it easier to distinguish the number of changes in the spectral data. However, both methods are unsuitable for data-driven analysis. The Time Window function changes the original data by dividing the spectral energy. The time window function is not suitable for data-driven analysis because it affects the characteristics of actual data. Therefore, the edge computing system constructed in this paper acquires the frequency change within 8 kHz of the maximum frequency with 20 Hz of sampling rate without applying filtering or time window.

To conduct time-domain analysis on such a complex waveform, it is estimated that the amount of change in amplitude expressed using RMS and PTP functions can reflect the change in the vibration of the pump generated as a semiconductor by-product. Twelve major variables were selected, including RMS and PTP values obtained from six sensors.

When semiconductor by-products accumulate in the pump, it is assumed that the amplitude change amount of the observed frequency in the steady-state can be expressed as the amplitude change amount of the observed frequency. Therefore, 364 of the frequency

data showing effective changes in each sensor were selected as the main variables of further data processing.

For the classification algorithm to be applied to the experimental data, the top seven algorithms with the highest performance for the data were selected based on the algorithm database in Azure. Dry pump vibration data of 300 mm ICP etcher were classified using MinMaxScalerSVM [33], StackEnsemble [34], RobustScalerSVM, MaxAbsScalerLightGBM [35], StandardScalerWrapperLightGBM, and SparseNormalizerXGBoostClassifier [36].

As the main process of this paper, dry pump vibration data of the TiN deposition process were classified using MinMaxScaler_RandomForest [37], StandardScalerWrapperRandomForest, SparseNormalizerXGBoostClassifier, MaxAbsScalerXGBoostClassifier, MinMaxScalerExtremeRandomTrees [38], MaxAbsScalerLightGBM, and StandardScalerWrapperExtremeRandomTrees. The pseudo-code of the algorithm used as the main structure of each algorithm, such as (a) SVM, (b) StackEnsemble, (c) LightGBM, (d) XGBoostClassifier, (e) RandomForest, and (f) ExtremeRandomTrees, are presented in Appendix A.

4. Monitoring and Diagnostic System

The data collected from the edge computer is stored in an external Web access PC as a 2.4 GHz Wi-Fi wireless I/O module. Data are transmitted according to the predefined Modbus communication value. The transmission period can be set-up to a 0.1 s period. The calculations of functions for time-domain and frequency-domain values are performed on the edge computer. Web access is designed to store the calculated function values in a database. The data stored in the database are displayed in real-time on the dashboard configured in Web access. As shown in Figure 5, the dashboard visualizes the maximum value of RMS, PTP, and FFT of each sine wave sensor with a control chart and single status warning alarm. The control chart was configured to check the occurrence of outliers based on the upper and lower levels of 10%, and the single status warning alarm was configured to display a warning when each sensor deviates from a certain level. Through this, it is possible to review the entire data in addition to real-time monitoring. The IIoT based in situ monitoring system can quickly provide equipment status information to users. Users can operate the process stably and find abnormal conditions based on the status information. In addition, the time-series database can be used as a variable when analyzing process results.

The diagnostic system acts as a middle layer for machine learning analysis with the data acquired from the pump. This system was built with Microsoft Azure, a cloud platform with various advantages, such as algorithm optimization function, evaluation of multi-faceted model performance, and model variable validation evaluation. Vibration data preprocessed through a data acquisition system configured on the MCM PC are applied to a machine learning algorithm configured on the cloud. A rather high initial installation cost is required to configure the proposed system. The cost for storing and calculating data in the cloud environment is added. Besides, proposed systems cannot give feedback to each other because of differences in the Monitoring and Diagnosis systems operating environments. However, machine learning analysis can use multivariate analysis to calculate relationships between variables that are more complex than conventional statistical process control methods. Machine learning learned from accumulated data can show very high analytical results for multivariate data. In particular, higher diagnostic accuracy can be expected if an algorithm optimized for the acquired data is found among various algorithms. The cloud diagnosis system utilizes IIoT equipment data to optimize PM.

Vibration data obtained from the vacuum pump were used to test the diagnostic system. In the AA200W connected to a 300 mm ICP type etching chamber, it is not easy to acquire error data due to by-products. Therefore, the abnormal data extracted from normal data were used for temporary classification accuracy evaluation. Anomalous labels were selected using an isolated forest algorithm with a contamination value of 10%. Figure 6

shows a *t*-Stochastic Neighbor Embedding (*t*-SNE) graph that visualizes multivariate data in three-dimensional form to express the outlier distribution results.

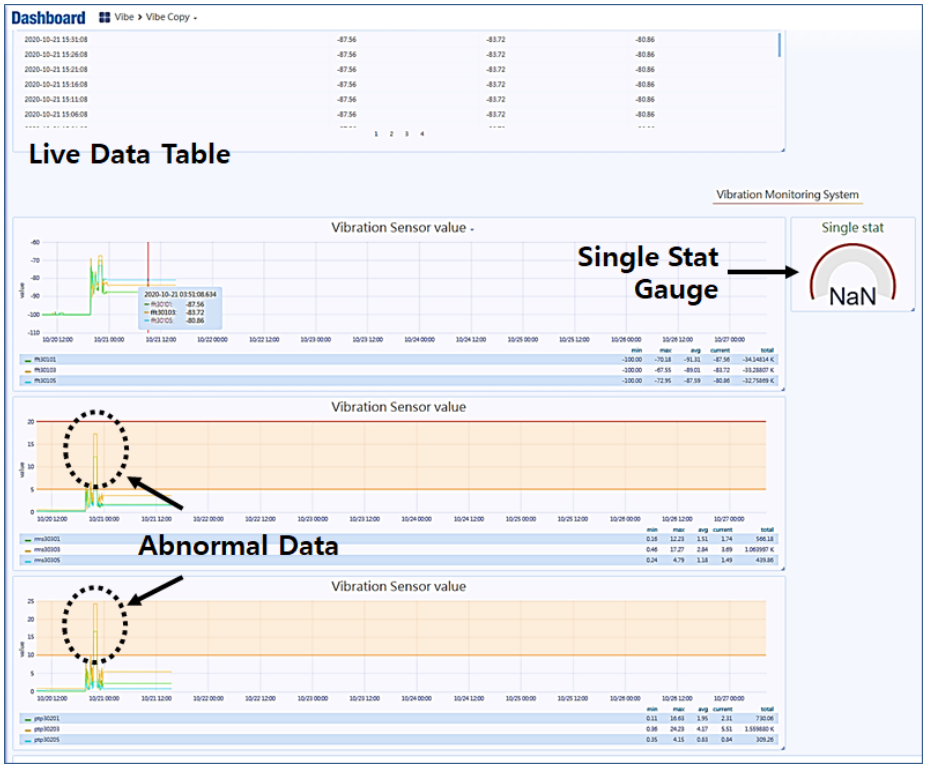


Figure 5. Dry pump vibration data monitoring system.

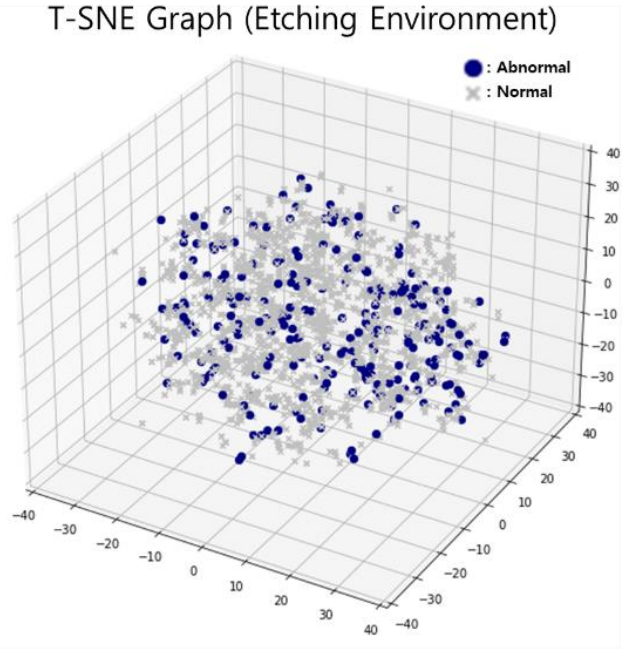


Figure 6. *t*-SNE graph in an etching environment.

Various data processing algorithms can be selected from the established dashboard based on high classification accuracy when using the acquired data. Due to the recently increased study of machine learning, eight machine learning algorithms were tested. The SVM model using MinMaxScaler achieved the highest classification accuracy. Classification accuracy alone cannot be evaluated by modeling errors of the algorithms; multi-faceted

model performance evaluation is essential. Therefore, we employ more categories of performance evaluation, as shown in Table 1. We observed that the Balanced_Accurancy of all models is quite low, whereas other criteria are moderately accepted. The result of the confusion matrix in Figure 7 explains the low Balanced_Accurancy, and it falls into the data imbalance problem in machinery malfunction diagnosis. Suppose the equipment malfunction dataset can be compared to the number of nominal data available from the manufacturing site. In that case, we believe that the suggested methodology will work better for equipment diagnosis in real-time.

Table 1. Etching environment algorithm performance evaluation.

Experiment Condition	Accuracy	Balanced Accuracy	F1_Score	Recall Score Micro	Log_Loss
MinMaxScaler, SVM	0.918	0.672	0.918	0.918	0.217
StackEnsemble	0.916	0.661	0.916	0.916	0.222
RobustScaler, SVM	0.909	0.609	0.909	0.909	0.235
MaxAbsScaler, LightGBM	0.907	0.552	0.907	0.907	0.250
StandardScalerWrapper, LightGBM	0.908	0.563	0.908	0.908	0.248
SparseNormalizer, XGBoostClassifier	0.906	0.570	0.906	0.906	0.354

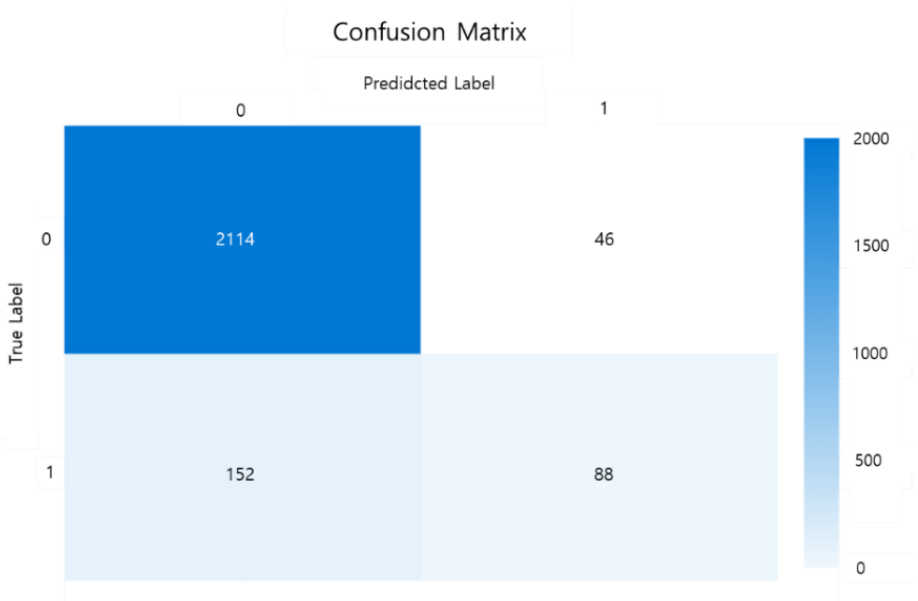


Figure 7. MinMaxScaler.SVM confusion matrix.

Based on the above study, vibration analysis was conducted to evaluate whether the *TiN* deposition affects the performance of the pump. Vibration data to be used for analysis was acquired in an environment where *TiN* can be deposited inside a dry pump connected by supplying NH_3, N_2, and $TiCl_4$ gases to a temporary chamber manufactured by simulating the ALD equipment. The location of the sensor for the acquisition was the same. The acquired data were stored as a database through MCM. The data acquisition period took two months, and the data acquired at the first and last moments were classified as normal and abnormal, respectively. Figure 8 shows the distribution of pump vibration data obtained from the deposition process as a *t*-SNE graph. It was confirmed that normal and abnormal data were classified to some extent on the graph.

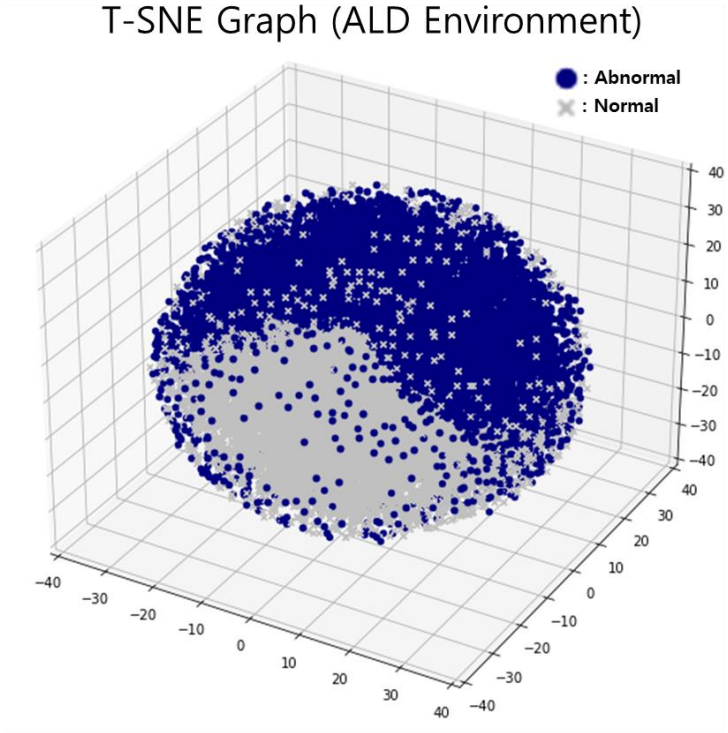


Figure 8. T-SNE graph in ALD environment.

The obtained vibration data were preprocessed for analysis, and a classification algorithm was applied and evaluated through Azure. As a result of the evaluation, the accuracy of all algorithms was calculated as 100%. Table 2 presents six algorithms selected among the algorithms applied by Azure. All algorithms classified abnormal data with 100% accuracy. Algorithm ranking is listed as a log loss parameter, but there is no significant difference because all values converge to 0. To prevent class imbalance in the training process, a dataset with a ratio of outliers to normal observations of 5:5 was used. Figure 9 shows that the MinMaxScalerRandomForest confusion matrix has no imbalance in the data and no misclassification in the analysis process. Figure 10a shows that the local feature importance graph can check the value at the time point with the greatest importance being among one variable in the overall learning. Figure 10b shows the affected feature to model prediction list, showing all variables in the order of the highest feature importance in the algorithm learning process. The key features found by analyzing the two results indicate which variables to focus on in the subsequent algorithm development process.

Table 2. ALD environment algorithm performance evaluation.

Experiment Condition	Accuracy	Balanced Accuracy	F1_Score	Recall Score Micro	Log_Loss
MinMaxScaler, RandomForest	1.000	1.000	1.000	1.000	0.0000005
StandardScalerWrapper, RandomForest	1.000	1.000	1.000	1.000	0.00001
SparseNormalizer, XGBoostClassifier	1.000	1.000	1.000	1.000	0.00004
MaxAbsScaler, XGBoostClassifier	1.000	1.000	1.000	1.000	0.00005
MaxAbsScaler, LightGBM	1.000	1.000	1.000	1.000	0.0003
StandardScalerWrapper, ExtremeRandomTrees	1.000	1.000	1.000	1.000	0.001

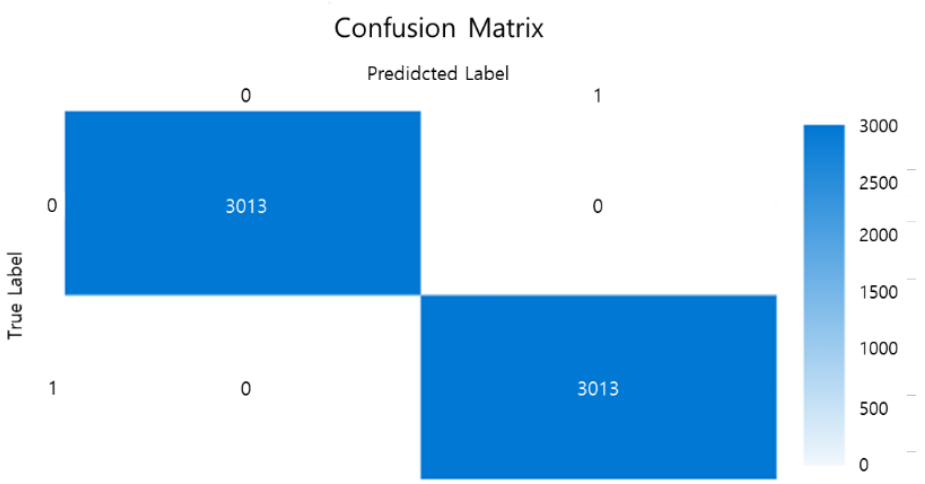


Figure 9. MinMaxScalerRandomForest confusion matrix.

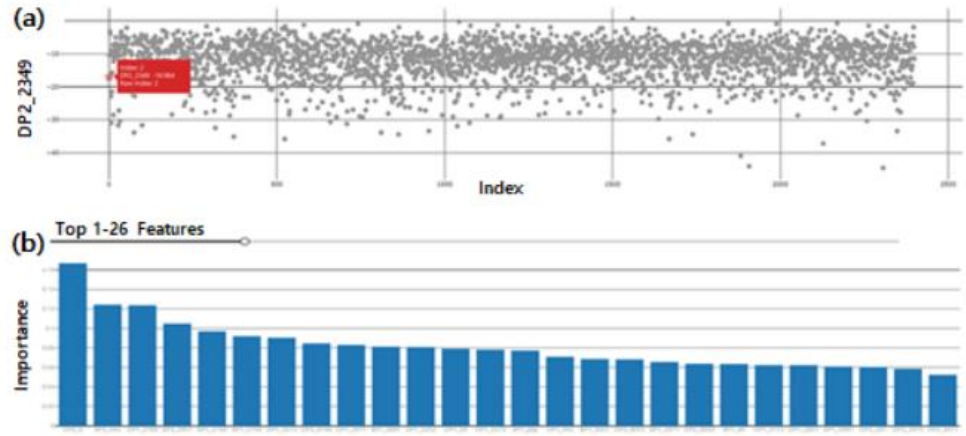


Figure 10. Model variable validation evaluation process; (**a**) local feature importance graph and (**b**) affected feature to model prediction list.

5. Conclusions

This paper introduced a system to monitor and analyze pump failure as a by-product of semiconductors generated in CVD in the ALD process. The monitoring system acquires and accumulates vibration data generated by the vibration pump when semiconductor by-products accumulate. In addition, the process operator can intuitively evaluate the internal state of the pump in real time. The system of this paper, which adds the ML analysis environment, performs multivariate analysis using all significant frequency bands as variables. Multivariate analysis algorithm has higher classification accuracy and result recall than single-variable analysis.

As a result of applying the vacuum pump vibration change amount to the data diagnostic system in the ALD process environment, the top six algorithms showed high classification accuracy. This paper introduced a system to monitor and analyze pump failure as a by-product of semiconductors generated in CVD during ALD. Since the abnormal state of the vibration data is accumulated and recorded, it is expected that it will be easy to prepare an analysis report through this. Additionally, the Azure cloud platform, to which a machine learning algorithm is applied, can detect data-based outliers using accumulated vibration data. Since it is based on multivariate analysis, it is expected to detect more various types of anomalies. If this is applied in an environment close to mass production for a longer period, it is expected that an outlier detection system with higher reliability can be constructed.

Author Contributions: Conceptualization, principal investigation and supervision, funding acquisition, writing—review and editing, S.J.H.; methodology, visualization, data curation, writing—original draft preparation, Y.L. and C.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Korean Ministry of Trade, Industry and Energy (MOTIE) via Advanced Institute of Convergence Technology (GID: AICT 006-T1), and Korea Institute for Advancement of Technology (KIAT) grant funded by Korea Government(MOTIE). (P0008458, The Competency Development Program for Industry Specialist).

Acknowledgments: Authors are grateful to Advantech Korea for their technical support on WISE-PaaS, and Staff of Semiconductor Process Diagnosis Research Center.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Algorithm A1 SVM Pseudo-Code

Input: parameters:
  out = array of SVM outputs
  target = array of booleans: is i^{th} example a positive example?
  prior1 = number of positive examples
  prior0 = number of negative examples
Output: Parameters of sigmoid
A = 0
B = log((*prior*0 + 1)/(*prior*1 + 1))
log((prior0 + 1)/(prior1 + 1))
    hiTarget = (prior1 + 1)/(prior1 + 2)
    loTarget = 1/(prior0 + 2)
    lambda = $1e^{-3}$
    olderr = $1e^{300}$
    pp = temp array to store current estimate of probability of examples
set all pp array elements to (prior1 + 1)/(prior0 + prior1 + 2)
count = 0
for it = 1 to 100
  a = 0, b = 0, c = 0, d = 0, e = 0
//First, compute Hessian & gradient of error function
//with respect ro A & B
  for i = 1 to len {
  if (target[i])
  t = hiTarget
  Else
  t = loTarget
  d1 = $pp[i] - t$
  d2 = $pp[i]*(1 - pp[i])$
  a + = $out[i]*out[i]*d2$
  b + = $d2$
  c + = $out[i]*d2$
  d + = $out[i]*d1$
  e + = $d1$
    //If gradient is really tiny, then stop
  if ($abs(d) < 1e{-}9$ && $abs(e) < 1e{-}9$)
   Break
  oldA = A
  oldB = B
  err = 0
    //Loop until goodness of fit increases
  while (1) {

```
det = (a + lambda)*(b + lambda) − c*c
if (det == 0) {//if determinant of Hessian is zero,
//increase stabilizer
   lambda* = 10
     Continue
}
A = oldA + ((b+lambda)*d − c*e)/det
B = oldB + ((a + lambda)*e − c*d)/det
//Now, compute the goodness of fit
for i = 1 to len {
   p = 1/(1 + exp(out[i]*A + B ))
   pp[i] = p
   //At this step, make sure log(0) returns -200
   err − = t*log(p) + (1 − t)*log(1 − p)
      }
if (err < olderr*(1 + 1e − 7)) {
   lambda* = 0.1
      Break
      }
//error did not decrease: increase stabilizer by factor of 10
//& try again
lambda* = 10
if (lambda ≥ 1e6)//something is broken, Give up
      Break
   }
   diff = err-older
   scale = 0.5*(err + olderr + 1)
   if (diff > − 1e − 3* scale&&diff < 1e − 7*scale)
   count ++
   Else
   count = 0
   older = err
   if (count == 3)
Break
}
```

Algorithm A2 Stacked Ensemble Classifier Pseudo-Code

Input: Dataset $D = \{(x_1,y_1),(x_2,y_2),\cdots,(x_m,y_m)\}$;
  Base-classifier $f_1 = RF_1, f_2 = RF_2, f_3 = ET_1, f_4 = ET_2$
For $t = 1,\cdots,4$;
  Train the base-classifiers in the first stage
$h_t = ft(D)$;
End
For i = 1,2,$\cdots$,m:
  for $t = 1,\cdots,4$:
    Generate new feature vector for each sample
    $z_{it} = ht(xi)$;
  End
  $D' = D' \cup ((z_{i1},\cdots,z_{i4}),y_i)$;
End
Train the meta-classifier in the second stage
$h' = f(D')$;
Output: $H(x) = h'(h_1(x), \ldots, h_4(x))$

Algorithm A3 LightGBM Pseudo-Code

```
Input: I: training data, d: max depth
Input: m: feature dimension
nodeSet←{0}.ree nodes in current level
rowSet ← {{0, 1, 2, . . . }} ▷ data indices in tree nodes
for i = 1 todo
   for node in nodeSet do
      usedRows ←rowSet[node]
      fork = 1 tomdo
         H ← new Histogram()
         ▷ Build histogram
         forjinusedRowsdo
            bin ← I.f[k][j].bin
            H [bin].y H [bin].y + I.y[j]
            H [bin].n H [bin].n + 1
         Find the best split on histogram H.
      . . .
Update rowSet and nodeSet according to the best split points.
 . . .
Input: I: training data, d: iterations
Input: a: sampling ratio of large gradient data
Input: b: sampling ratio of small gradient data
Input: loss: loss function, L: weak learner
models ← { }, fact ← (1−a)/b
topN ←a × len(I),randN← b ×len(I)
for i = 1 todo
   preds ← models.predict(I)
   g ←loss(I, preds), w ← {1,1, . . . }
   sorted ← GetSortedIndices(abs(g))
   topSet ← sorted [1:topN]
   randSet ← RandomPick(sorted[topN:len(I)], randN)
   usedSet ← topSet + randSet
   w[randSet] × = fact ▷ Assign weight fact to the small gradient data.
   newModel ← L(I [usedSet],− g[usedSet], w[usedSet])
   models.append(newModel)
```

Algorithm A4 XGboost Pseudo-Code

```
Input: I, instance set of current node
Input: d, feature dimension
gain← 0
G ← ∑_{i∈I} g_i, H ← ∑_{i∈I} h_i
for q = 1 to Q do:
   G_L←0,H_L←0
   for jinsorted (I, byx_jq) do
      G_L← G_L + g_j, H_L← H_L + h_j
      G_R←G−G_L, H_R←H − H_L
   score ← max(score, G_L^2/(H_L+λ) + G_R^2/(H_R+λ) − G^2/(H+λ))
   End for
End for
Output: Split with max score
```

Algorithm A5 Random Forest Pseudo-Code

Require: IDT(a decision tree inducer),
T(the number of iterations), S (the training set),
μ (the subsample size), N (number of attributes used in each node)
Ensure: Mt;$t = 1, \ldots, T$
$t \leftarrow 2$
Repeat
  $St \leftarrow$ Sample μ instances from S with replacement
  Build classifier Mt using $IDT(N)$ on St
  t ++
until $t > T$

Algorithm A6 Extra-Trees Splitting Algorithm Pseudo-Code

Split_a_node(S)
Input: the local learning subset S corresponding to the node we want to split
Output a split $[a < ac]$ or nothing
If **Stop_split**(S) is TRUE then return nothing
Otherwise select Kattributes $\{a1, \ldots a_K\}$
among all non constant (in S) candidate attributes;
Draw Ksplits $\{s_1, \ldots s_K\}$, where $s_i =$ **Pick_a_random_split**$(S, a_i), \forall i = 1, \ldots K$;
Return a split s_* such that Score(s_*, S) = $\max_{i=1, \ldots, K}$ Score(s_i, S).

Pick_a_random_split(S,a)
Inputs: a subset S and an attribute a
Output: a split
Let a^S_{min} and a^S_{max} denote the maximal and minimal value of a in S;
Let a^S_{min} and a^S_{max} denote the maximal and minimal value of a in S;
Draw a random cut-point $a_c$uniformly in $[a^S_{min}, a^S_{max}]$;
Return the split $[a < a_c]$.

Stop_split(S)
Input: a subset S
Output: a Boolean
If $|S| < n_{\min}$, then return TRUE;
If all attributes are constant in S, then return TRUE;
If the output is constant in S, then return TRUE;
Otherwise, return FALSE.

References

1. Antao, L.; Pinto, R.; Reis, J.; Goncalves, G. Requirements for Testing and Validating the Industrial Internet of Things. In Proceedings of the 2018 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW), Västerås, Sweden, 9–13 April 2018; pp. 110–115. [CrossRef]
2. Serpanos, D.; Wolf, M. *Internet-of-Things (IoT) Systems*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 37–54. [CrossRef]
3. Chen, Y.; Lee, G.M.; Shu, L.; Crespi, N. Industrial Internet of Things-Based Collaborative Sensing Intelligence: Framework and Research Challenges. *Sensors* **2016**, *16*, 215. [CrossRef] [PubMed]
4. Sisinni, E.; Saifullah, A.; Han, S.; Jennehag, U.; Gidlund, M. Industrial Internet of Things: Challenges, Opportunities, and Directions. *IEEE Trans. Ind. Inform.* **2018**, *14*, 4724–4734. [CrossRef]
5. Civerchia, F.; Bocchino, S.; Salvadori, C.; Rossi, E.; Maggiani, L.; Petracca, M. Industrial Internet of Things monitoring solution for advanced predictive maintenance applications. *J. Ind. Inf. Integr.* **2017**, *7*, 4–12. [CrossRef]
6. Razmi-Farooji, A.; Kropsu-Vehkaperä, H.; Härkönen, J.; Haapasalo, H. Advantages and potential challenges of data management in e-maintenance. *J. Qual. Maint. Eng.* **2019**, *25*, 378–396. [CrossRef]
7. Lade, P.; Ghosh, R.; Srinivasan, S. Manufacturing Analytics and Industrial Internet of Things. *IEEE Intell. Syst.* **2017**, *32*, 74–79. [CrossRef]
8. Hong, S.J.; Lim, W.Y.; Cheong, T.; May, G.S. Fault Detection and Classification in Plasma Etch Equipment for Semiconductor Manufacturing e-Diagnostics. *IEEE Trans. Semicond. Manuf.* **2011**, *25*, 83–93. [CrossRef]

9. Boyes, H.; Hallaq, B.; Cunningham, J.; Watson, T. The industrial internet of things (IIoT): An analysis framework. *Comput. Ind.* **2018**, *101*, 1–12. [CrossRef]
10. Suh, Y.J.; Choi, J.Y. Efficient Fab facility layout with spine structure using genetic algorithm under various material-handling considerations. *Int. J. Prod. Res.* **2021**, 1–14. [CrossRef]
11. Kumar, A.; Bhattacharjee, N.; Patel, B.; Laloë, J.-B.; Famodu, O.O.; Ferain, L. Strategе for Reducing Particle Defects in Ti and TiN Thin-Film Deposition Process. *IEEE Trans. Semi. Manufac.* **2018**, *32*, 48–53. [CrossRef]
12. Jones, C.P. Thin Film Processes: Abatement of Waste Gases from Plasma Assisted Material Processes. In Proceedings of the 2020 China Semiconductor Technology International Conference (CSTIC), Shanghai, China, 26 June–17 July 2020; pp. 1–3. [CrossRef]
13. Lessard, P.A. Dry vacuum pumps for semiconductor processes: Guidelines for primary pump selection. *J. Vac. Sci. Technol. A* **2000**, *18*, 1777–1781. [CrossRef]
14. Zhan, H.; Li, N.; Wang, Y.; Wang, X.; Lin, S. Fault Diagnosis Method for Vacuum Pump of Space Environment Simulator. In Proceedings of the 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), Hefei, China, 5–7 June 2016; pp. 1703–1707. [CrossRef]
15. Ainapure, A.; Li, X.; Singh, J.; Yang, Q.; Lee, J. Deep Learning-Based Cross-Machine Health Identification Method for Vacuum Pumps with Domain Adaptation. *Procedia Manuf.* **2020**, *48*, 1088–1093. [CrossRef]
16. Rui, X.; Liu, J.; Li, Y.; Qi, L.; Li, G. Research on fault diagnosis and state assessment of vacuum pump based on acoustic emission sensors. *Rev. Sci. Instrum.* **2020**, *91*, 025107. [CrossRef]
17. LaRose, J.; Barker, J.; Finlay, B.; Trinidad, A.; Guyer, C.; Weinstein, J.; Conerney, B.; Ray, D.; Perry, J.; Tarnawskyj, W.; et al. Predictive Maintenance of Pump and Abatement Equipment in a 300mm Semiconductor Fab. In Proceedings of the 2021 32nd Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC), Milpitas, CA, USA, 10–12 May 2021; pp. 1–4. [CrossRef]
18. Jiang, W.; Spurgeon, S.K.; Twiddle, J.A.; Schlindwein, F.S.; Feng, Y.; Thanagasundram, S. A wavelet cluster-based band-pass filtering and envelope demodulation approach with application to fault diagnosis in a dry vacuum pump. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2007**, *221*, 1279–1286. [CrossRef]
19. Twiddle, J.A.; Jones, N.B.; Spurgeon, S.K. Fuzzy model-based condition monitoring of a dry vacuum pump via time and frequency analysis of the exhaust pressure signal. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2008**, *222*, 287–293. [CrossRef]
20. Butler, S.W.; Ringwood, J.V.; MacGearailt, N. Prediction of Vacuum Pump Degradation in Semiconductor Processing. *IFAC Proc. Vol.* **2009**, *42*, 1635–1640. [CrossRef]
21. Hsieh, F.-C.; Chou, S.-C.; Lin, P.-H.; Chang, S.-W.; Chen, F.-Z.; Liao, T.-S.; Liu, D.-R.; Chiang, D. Monitoring on dry vacuum pump characteristics by mobile device. In Proceedings of the 2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings, Graz, Austria, 13–16 May 2012; pp. 2444–2447. [CrossRef]
22. Qiu, G.Q.; Gu, Y.K.; Chen, J.J. Selective health indicator for bearings ensemble remaining useful life prediction with genetic algorithm and Weibull proportional hazards model. *Measurement* **2020**, *150*, 107097. [CrossRef]
23. Lee, K.; Song, D.; Lee, J.; Lee, C.-G.; Shin, G.-A.; Jung, S. Evaluating effectiveness of dust by-product treatment with scrubbers to mitigate explosion risk in ZrO2 atomic layer deposition process. *J. Hazard. Mater.* **2020**, *400*, 123284. [CrossRef]
24. Bhatti, M.A.; Riaz, R.; Rizvi, S.S.; Shokat, S.; Riaz, F.; Kwon, S.J. Outlier detection in indoor localization and Internet of Things (IoT) using machine learning. *J. Commun. Netw.* **2020**, *22*, 236–243. [CrossRef]
25. Kim, D.H.; Choi, J.E.; Ha, T.M.; Hong, S.J. Modeling with Thin Film Thickness using Machine Learning. *J. Semicond. Disp. Technol.* **2019**, *18*, 48–52. Available online: https://www.koreascience.or.kr/article/JAKO201919761177833.page (accessed on 6 October 2021).
26. Puggini, L.; McLoone, S. An enhanced variable selection and Isolation Forest based methodology for anomaly detection with OES data. *Eng. Appl. Artif. Intell.* **2018**, *67*, 126–135. [CrossRef]
27. Shukla, S.; Meghana, K.M.; Manjunatrh, C.R.; Santosh, N. Comparison of Wireless Network over Wired Network and its Type. *Int. J. Res. -Granthaalayah* **2017**, *5*, 14–20. [CrossRef]
28. Bailey, C.; Hutchison, K.; Wilders, M. Vacuum Systems for ALD. *Solid State Technol.* **2006**, *49*, 30–33.
29. Hur, M.; Lee, J.; Yoo, H.; Kang, W.; Song, Y.; Kim, D.; Lee, S. Reduction of byproduct particle size using low-pressure plasmas generated by a cylindrical-shaped electrode. *Vacuum* **2012**, *86*, 1834–1839. [CrossRef]
30. ISO 18436-2: 2014 Condition Monitoring and Diagnostics of Machines—Requirements for Qualification and Assessment of Personnel—Part 2: Vibration Condition Monitoring and Diagnostics. Available online: https://www.iso.org/standard/50447.html (accessed on 6 October 2021).
31. Liu, F.T.; Ting, K.M.; Zhou, Z.-H. Isolation Forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; pp. 413–422. [CrossRef]
32. Vishwakarma, M.; Purohit, R.; Harshlata, V.; Rajput, P. Vibration Analysis & Condition Monitoring for Rotating Machines: A Review. *Mater. Today Proc.* **2017**, *4*, 2659–2664. [CrossRef]
33. John, C.P. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *Adv. Large Margin Classif.* **1999**, *10*, 61–74. Available online: https://citeseerx.ist.psu.edu/viewdoc/versions?doi=10.1.1.41.1639 (accessed on 6 October 2021).

34. Chen, C.; Zhang, Q.; Yu, B.; Yu, Z.; Skillman-Lawrence, P.; Ma, Q.; Zhang, Y. Improving protein-protein interactions prediction accuracy using XGBoost feature selection and stacked ensemble classifier. *Comput. Biol. Med.* **2020**, *123*, 103899. [CrossRef] [PubMed]
35. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Adv. Neural Inf. Processing Syst.* **2017**, *30*, 3146–3154. Available online: https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html (accessed on 6 October 2021).
36. Zhang, W.; Zhao, X.; Li, Z. A Comprehensive Study of Smartphone-Based Indoor Activity Recognition via Xgboost. *IEEE Access* **2019**, *7*, 80027–80042. [CrossRef]
37. Lior, R. Ensemble-Based Classifiers. *Artif. Intell. Rev.* **2010**, *33*, 1–39. [CrossRef]
38. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]