coatings

MDPI

Article

Machine Learning Prediction of Electron Density and Temperature from Optical Emission Spectroscopy in Nitrogen Plasma

Jun-Hyoung Park [1,2], Ji-Ho Cho [2], Jung-Sik Yoon [2] and Jung-Ho Song [1,2,*]

1 Fundamental Technology Research Division, Institute of Plasma Technology, Korea Institute of Fusion Energy, Gunsan-si 54004, Korea; pjh1126@kfe.re.kr
2 Plasma E. I. Convergence Research Center, Korea Institute of Fusion Energy, Gunsan-si 54004, Korea; jhcho@kfe.re.kr (J.-H.C.); jsyoon@kfe.re.kr (J.-S.Y.)
* Correspondence: jungho@kfe.re.kr

Abstract: We present a non-invasive approach for monitoring plasma parameters such as the electron temperature and density inside a radio-frequency (RF) plasma nitridation device using optical emission spectroscopy (OES) in conjunction with multivariate data analysis. Instead of relying on a theoretical model of the plasma emission to extract plasma parameters from the OES, an empirical correlation was established on the basis of simultaneous OES and other diagnostics. Additionally, we developed a machine learning (ML)-based virtual metrology model for real-time T_e and n_e monitoring in plasma nitridation processes using an in situ OES sensor. The results showed that the prediction accuracy of electron density was 97% and that of electron temperature was 90%. This method is especially useful in plasma processing because it provides in-situ and real-time analysis without disturbing the plasma or interfering with the process.

Keywords: optical emission spectroscopy; plasma nitridation; plasma parameters; machine learning; virtual metrology

check for updates

Citation: Park, J.-H.; Cho, J.-H.; Yoon, J.-S.; Song, J.-H. Machine Learning Prediction of Electron Density and Temperature from Optical Emission Spectroscopy in Nitrogen Plasma. *Coatings* **2021**, *11*, 1221. https://doi.org/10.3390/coatings11101221

Academic Editor: Alenka Vesel

Received: 30 July 2021
Accepted: 27 September 2021
Published: 6 October 2021

1. Introduction

Recently, plasma processing technology has played a crucial role in the surface modification of different materials, such as electronics, energy storage, automotive, health, or environmental applications [1,2]. Control of plasma processing methodologies can only occur by obtaining a thorough understanding of the physical and chemical properties of plasmas. However, most plasma processes that are currently used in the industry reflect an incomplete understanding of the reactive nature of plasma Thus, they are often non-predictive, and hence it is not possible to alter the manufacturing process without the risk of considerable product loss [3,4]. Basic plasma parameters, such as electron density, electron temperature, are useful information to understand the plasma process, to determine the proper recipe for plasma processes (etching, deposition, etc.), and also for the development of further fundamental understanding of the process. Thus, direct and quantitative real-time diagnostic sensors or techniques of industrial plasma processes generally pose a significant challenge [5].

Generally, there are two types of plasma diagnostic sensors, called invasive and non-invasive sensors. The invasive sensor is immersed directly into the plasma and can provide direct information on ion flux, electron density, temperature, and other properties in the early stage of the device's development. However, invasive sensors perturb the state of plasma and impossible to apply in mass production. Non-invasive sensors are limited in obtaining direct information of plasmas, but they are preferred in manufacturing because they do not perturb plasmas. Due to the abundant information that can be extracted from the data and the direct relationship of the data to the plasma process, optical emission spectroscopy (OES) is widely applied to IC fabrication as a non-invasive diagnostic sensor. OES provides useful information about plasma properties and can be used to identify specific excited species generated by plasmas. Although the richness of OES data is also a potential hindrance to effective interpretation and utility of the data, it allows for qualitative and quantitative analyses of the plasma and enables real-time and non-invasive monitoring of processes for various applications.

Previous research on OES measurements of plasma processes has largely focused on plasma monitoring, species identification, and determination of the electron temperature (T_e) and density (n_e). In order to use the OES method to determine the T_e and n_e, one usually applies the so-called line intensity ratio method, which requires a relative intensity calibrated spectroscopic system and a suitable physics-based model for excited species in the plasmas to be investigated, such as the coronal model or the collisional radiative model (CRM) [6–8]. The emphasis of these models is the identification of major production and depopulation processes under different plasma discharge conditions and for different kinds of excited species. The quality of the modelling results, which determines the accuracy of diagnostic results by the line intensity ratio method, depends on the existence and quality of the cross-section and reaction rate coefficient data for the collisional radiative processes. Due to the tremendous efforts in this area [9], the line intensity ratio method is expected to be further developed in the future. However, this physics-based model approach is still limited in industrial application to plasma processes that use mixture gases [3].

In recent years, many studies have demonstrated the effectiveness of machine learning (ML) for OES. The fundamental idea in ML is that, for many applications, training a computer algorithm for predicting or finding patterns in the behavior of a complex system by observing many input–output samples of its behavior can be significantly simpler than developing physics-based models. Many of the ideas underlying this data-driven approach to understanding complex systems have been known for years, but only recently has it become more practical to obtain and analyze the enormous quantities of data needed for the schemes to work. LeCun et al. [10] performed multivariate analysis of spectra and extract valuable predictive information from large datasets through data mining. Choi et al. [11] developed a machine learning-based virtual metrology (VM) model on film thickness in amorphous carbon layer deposition process using in situ OES sensor data. Additionally, Yue et al. [12], Han et al. [13], and Kim et al. [14] proposed principal component analysis (PCA) of OES spectrum data for endpoint detection of plasma etching processes in the semiconductor industry. Even though OES has the advantage of non-invasiveness, it provides a huge amount of information. Thus, it requires time consuming data analysis and feature extraction from data, based on expert domain knowledge. Thus, effective implicit feature extraction is of paramount importance, especially in semiconductor manufacturing VM. As a result, the analysis of the data is a big challenge.

Thus, in this study, we propose a non-invasive approach for monitoring plasma parameters such as electron temperature and density inside a radio-frequency (RF) plasma nitridation device using optical emission spectroscopy (OES) in conjunction with multivariate data analysis. Instead of relying on a theoretical model of the plasma emission to extract plasma parameters from the OES, an empirical correlation was established on the basis of simultaneous OES and other diagnostics. Additionally, we developed a machine learning (ML)-based virtual metrology model for real-time T_e and n_e monitoring in plasma nitridation process using an in-situ OES sensor.

This paper is organized as follows: in Section 2, the experimental settings and OES spectrum data characteristics are given. In Section 3, we introduce the multivariate data analysis approach for monitoring T_e and n_e during the plasma nitriding process using OES sensing data with other plasma parameters. The prediction results are presented in Section 4, and, finally, Section 5 summarizes the work with future research directions.

2. Experimental Setup and Plasma Characteristics

An inductively coupled plasma (ICP) system (Figure 1) for the plasma nitriding process (PN- ICP, homemade) was used to diagnose plasma properties during the plasma nitriding process. The process reactor for nitriding silicon oxide film was 450 mm in diameter, and an impedance-adjustable antenna was used. The antenna was mounted on top of the exterior of the reactor using ceramic insulation. The distance between the ceramic insulator and the substrate was 100 mm, and the substrate diameter was 300 mm. The radio frequency (RF) power supply used for plasma generation was Cito Plus (13.56 MHz, 1 kW) from Comet (Wünewil-Flamatt, Switzerland), and the impedance matching network used was AGS (13.56 MHz, 1 kW) from Comet. The process gas was N_2 (N_2, 99.999%), and the flow rate and pressure were controlled using a mass flow controller (MFC, Mass-Flo®, 1000 sccm, MKS, Andover, MA, USA). The base pressure of the reactor was 10^{-6} Torr, and this was maintained using a turbomolecular pump (STP-1303C, Edwards, Burgess Hill, UK) and a dry pump (GX100N, Edwards, Burgess Hill, UK).

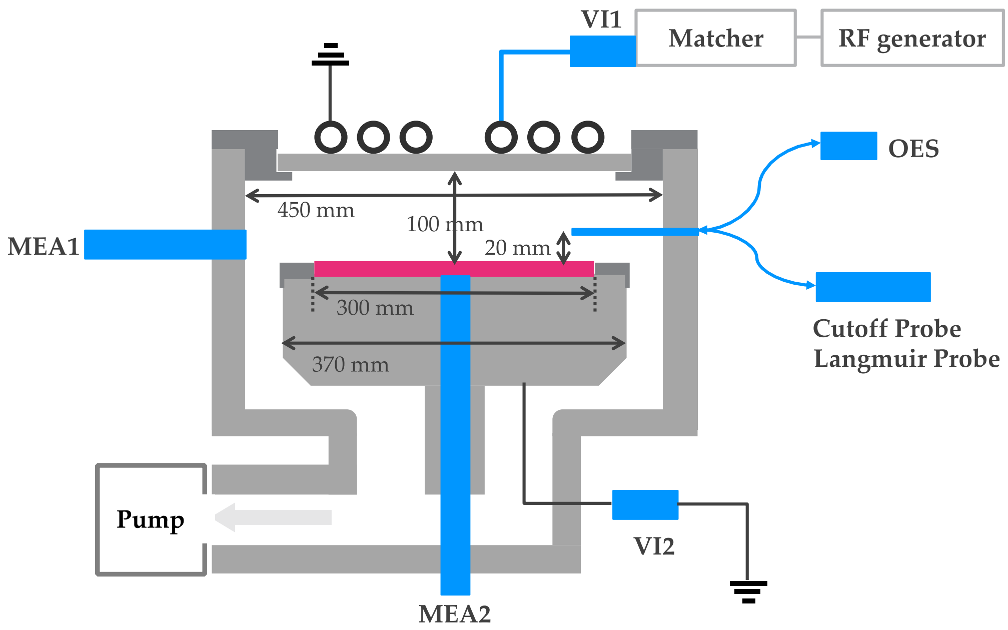

Figure 1. Schematic of the experimental device. VI probe (VI1, VI2), mass/energy analyzer (MEA1, MEA2), optical emission spectroscopy (OES), self-plasma OES (SP-OES), cutoff probe (CP), and Langmuir probe (LP).

To diagnose the plasma, the following diagnostic devices were installed in the PN-ICP reactor (Figure 1). A cutoff probe (CP, homemade) [15] and Langmuir probe (LP, homemade) were used to measure the plasma electron density (n_e) and other plasma properties such as electron temperature (T_e), floating potential (V_f), plasma potential (V_p), and ion saturation current (I_{isat}). Optical emission spectroscopy (OES, HR4000, Ocean optics, Dunedin, FL, USA) was used to measure the wavelength and intensity. The wavelength range of the OES was 200–1100 nm. To measure the RF power harmonic properties at the antenna, the VI probe (VI1, OCTIV POLY, Impedans, Dublin, Ireland) was installed, and the VI probe (VI2, OCTIV SUITE, Impedans, Dublin, Ireland) was installed on the substrate. The VI1 and VI2 measures the first to 15th harmonic components. The components measured in harmonics are the four values of voltage, current, phase, and harmonic phase, which are recorded as time-series data. For gas species analysis, a residual gas analyzer (RGA, Prisma Pro QMG 250, Pfeiffer Vacuum, Ablar, Germany) and self-plasma OES (SP-OES, AEGIS-7W, Nanotech, Yongin, Korea) were installed in the pump exhaust. Two mass/energy analyzers were installed at the bottom of the substrate and on the reactor wall to measure the mass and energy of the ionic species (MEA1: PSM, MEA2: EQP, Hiden, Warrington, UK).

Plasma parameters were measured at the RF powers of 400, 600, and 800 W, and the pressures of 10, 20, and 30 mTorr. The characteristics of PN-ICP plasmas are shown in Figure 2. The data analysis used 54 experimental datasets, and Figure 2 depicts the average value and standard deviation of each plasma parameter.

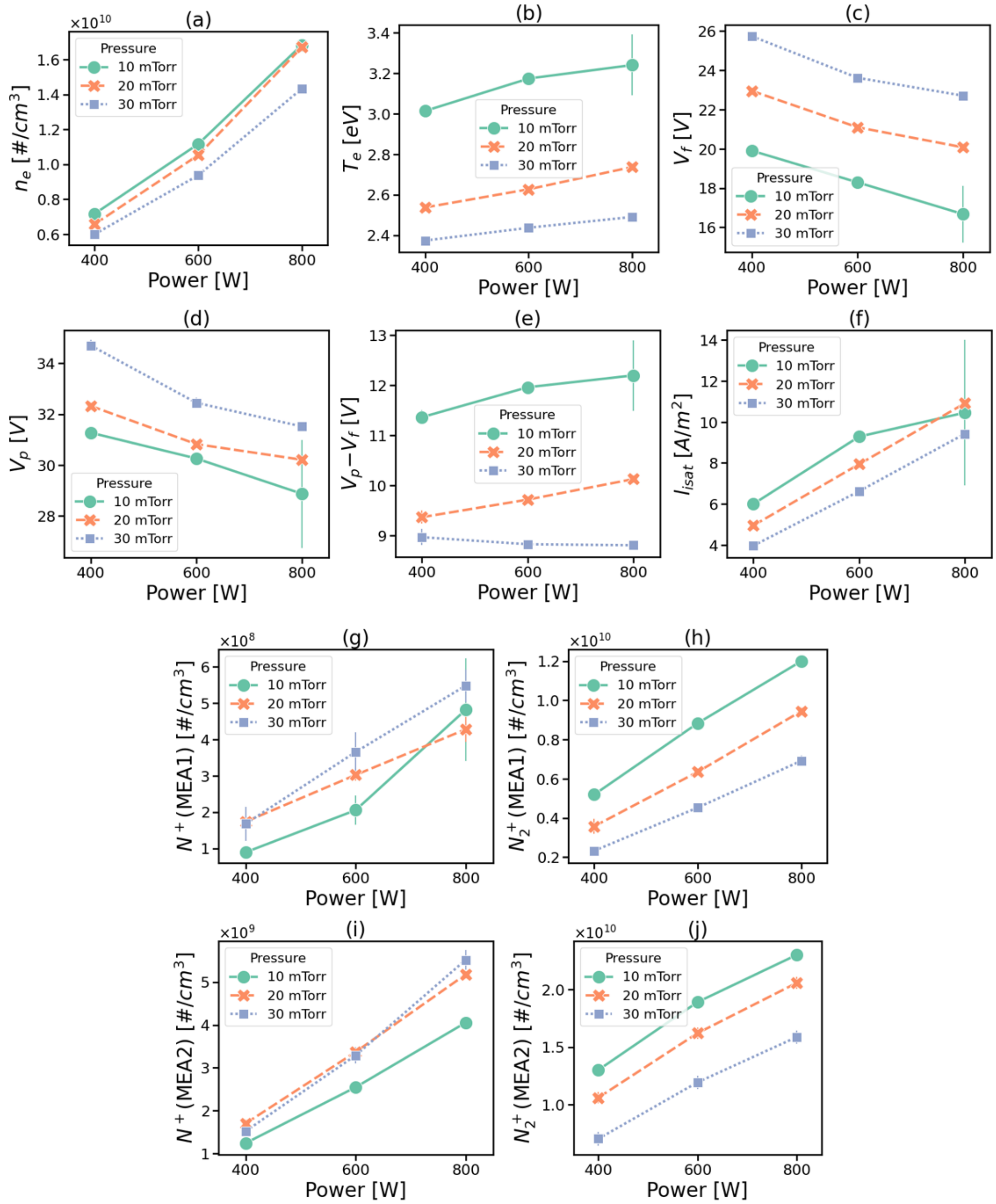

Figure 2. Plasma characteristics of PN-ICP system. (**a**) Electron density (n_e), (**b**) electron temperature (T_e), (**c**) floating potential (V_f), (**d**) plasma potential (V_p), (**e**) acceleration potential ($V_a = (V_p - V_f)$), (**f**) ion saturation current (I_{isat}), and N^+, N_2^+ ion number (#) density of MEA1(wall) (**g**,**h**) and MEA2 (substrate) (**i**,**j**). N_2 gas flow rate: 66 sccm (10 nTorr), 119 sccm (20 mTorr), and 174 sccm (30 mTorr).

The plasma electron density range of PN-ICP is 6×10^9 cm^{-3}–1.5×10^{10} cm^{-3}, which increases with increasing RF power and decreasing gas pressure. The plasma electron temperature range is 2–3.4 eV, which slightly increases with increasing RF power and decreasing gas pressure. Figure 2g–j represents the number density of the N^+ and N_2^+ ionic species measured with MEA1 (wall) and MEA2 (substrate). Here, the ion energy distributions (IEDs) of four ionic species (N^+, N_2^+, H_2O^+, and O_2^+) were measured, according to the experimental conditions, to obtain the relative ratio values.

The OES data were collected at 0.1 s intervals, and we detected a 200–1100 nm wavelength, divided by 3,648 wavelengths with 0.5 nm spectral resolution. To perform the wavelength calibration, a helium (He) and mercury (Hg) lamp were used [16]. The Atomic Spectral Database by the National Institute of Standards and Technology [17] was used as reference for the wavelength values of the atomic spectral line.

Figure 3 presents the optical emission spectrum of PN-ICP system operated in 10, 20, and 30 mTorr at the applied power of 400, 600, and 800 W. The main emission peaks correspond to several transition lines of atomic nitrogen, N (746 nm), and molecular nitrogen. The spectrum is characterized by the first positive system (FPS) (N_2, $B^3\Pi_g \rightarrow A^3\Sigma_u^+$ transitions), which is in the range of 478–1100 nm, and the second positive system (SPS) (N_2, $C^3\Pi_u \rightarrow B^3\Sigma_g^+$ transitions), which is in the range of 268–546 nm. The characteristic spectrum of the first negative system (FNS) (N_2^+, $B^2\Sigma_u^+ \rightarrow X^2\Sigma_g^+$ transitions) is shown in the 286–587 nm range. These wavelength classifications are given in [18].

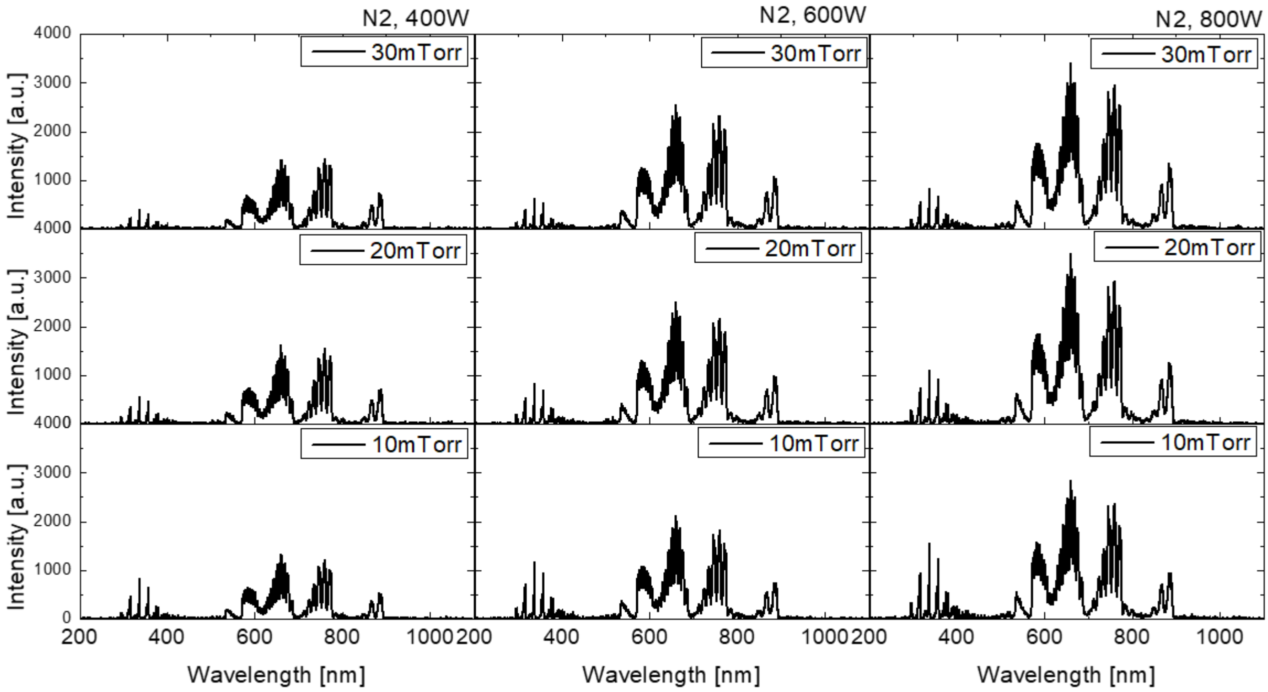

Figure 3. The OES spectrum measured for the exciting RF frequency of 13.56 MHz at 10, 20, 30 mTorr and 400, 600, 800 W.

The spectra are dominated by strong molecular features, which peak around 300–400 nm (SPS, FNS) and 500–800 nm (FPS). The FPS intensity increases with applied power and slightly increases with gas pressure. However, the SPS and FNS intensity slightly increases with applied power and decreases with gas pressure. In nitrogen plasmas, the atomic nitrogen species are mainly produced by electron-impact processes such as the dissociative collisions between electrons and nitrogen molecules or between electrons and nitrogen molecular ions (dissociative recombination) [8]. The most complete studies on the dissociative recombination of the nitrogen molecular cations can be found in Little et al. [19] and Abdoulanziz et al. [20]. Thus, the SPS and FNS depend on the electron temperature, and this trend is in good agreement with the electron temperature of the PN-ICP system (in Figure 2b).

3. Data Correlation Analysis

In this section, we introduce a multivariate data analysis approach for monitoring T_e and n_e during the plasma nitriding process using OES. Usually, assessing plasma parameters, such as T_e, and n_e, by OES requires complex theoretical modeling of the electronic states of the atoms and ions of the plasma. Additionally, in order to derive this theoretical spectrum which may be compared with the experiment, various scattering cross sections and reaction rate coefficients are required. To overcome the involved difficulties, we applied a data-driven approach which avoided the microscopic plasma model for extracting the plasma parameters from and optical emission spectrum. We measured the emission spectra simultaneously with the plasma parameters in the PN-ICP system and analyzed the correlation using principal component analysis (PCA).

This section is divided into subsections, providing a concise description of the plasma parameters correlations, their interpretation, and the PCA approaches.

3.1. Correlation between Diagnostic Data

The regression analysis by the coefficient of determination (R^2) method was used to find the correlation between the plasma parameter data (n_e, T_e, V_p, V_f, $V_a(= V_p - V_f)$, I_{isat}, N^+, and N_2^+), which were measured from the PN-ICP system. The data used for the regression analysis were obtained from 54 sets of measured data from different diagnostic system. Figure 4 shows the result of the correlation regression between the plasma parameters. The results with an R^2 value of 0.8 or higher out of a total 45 analysis results are summarized in Table 1.

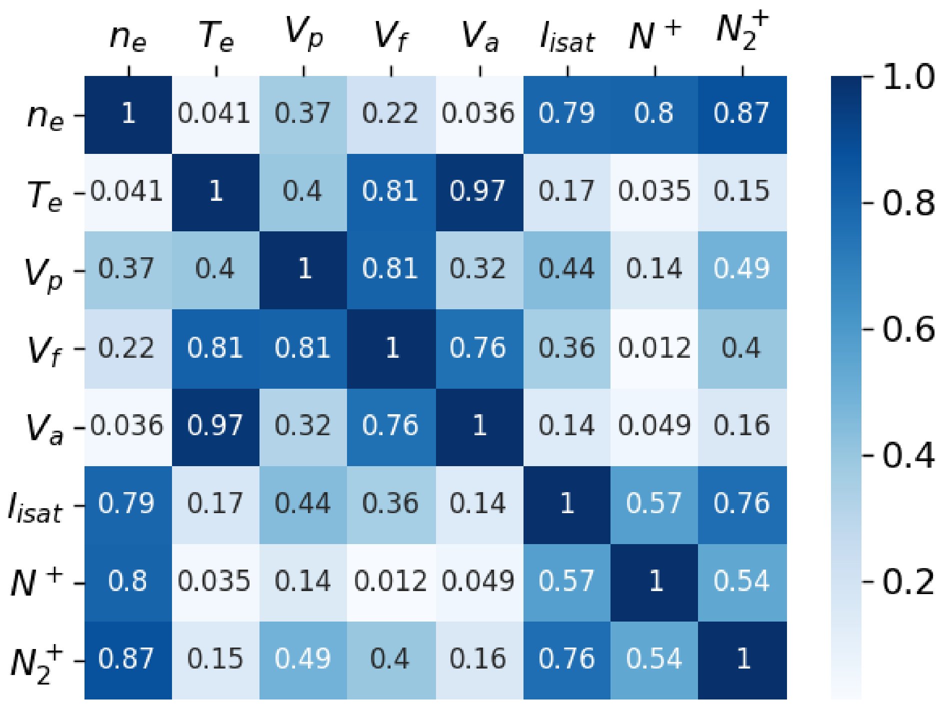

Figure 4. Regression results between the plasma parameters.

Table 1. Regression results between the plasma parameters for an $R^2 > 0.8$.

Ranking	Parameters	R^2 Score
1	T_e vs. $V_a(= V_p - V_f)$	0.97
2	n_e vs. N_2^+	0.87
3	T_e vs. V_f	0.81
4	V_f vs. V_p	0.80
5	n_e vs. N^+	0.80

The V_a value is defined as the acceleration potential, and it is the difference between the plasma potential and floating potential $\left(V_p - V_f\right)$. A PN-ICP system in non-bias substrates, usually the energy of the particles incident on the substrates was determined by V_a. T_e is strongly correlated with V_a. Figure 5 shows the strongly correlated T_e and V_a in nine sets of data for experiment numbers 30–38. However, the R^2 for T_e vs. V_f was 0.81, but T_e vs. V_p was 0.4; the correlation was low.

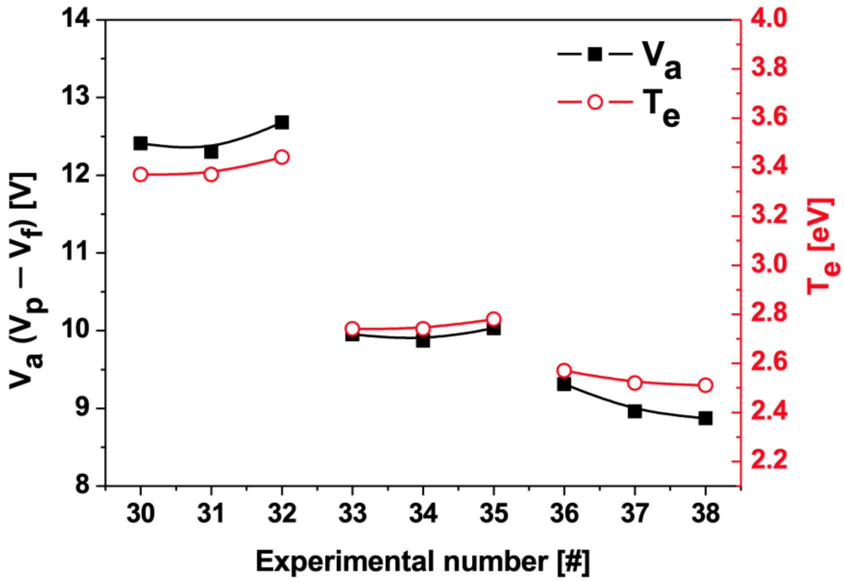

Figure 5. Electron temperature and acceleration potential for different operation condition in the PN-ICP system.

In general, the ion density variation depended on the n_e, and as a result, the I_{isat} measured on the LP also changed. Thus, n_e and I_{isat} are expected to be closely related, with a regression analysis showing an R^2 of 0.79. The n_e value was also strongly correlated with the N^+, and N_2^+ densities measured in MEA, and the R^2 was ≥ 0.8. The results of the correlation between the plasma parameters indicate that V_a was closely related to T_e, and n_e was highly correlated to the ionic species N^+, and N_2^+.

3.2. Principal Component Analysis (PCA) for OES Data vs. Plasma Parameters

The correlation between the plasma parameters and OES was also analyzed. Instead of using a physics-based model for the analysis of individual lines, here, we used all the spectral information in terms of relative intensities available in a spectrum of a given operational condition by applying the PCA method [21]. The PCA method can simplify the complex correlations between several measured lines and the resulting plasma parameters by reducing the dimensions of the datasets. Thus, PCA has been suggested as a method for analyzing large amounts of OES data [14,22]. PCA transforms the input variables to a set of orthogonal variables known as principal components (PCs), which are linear combinations of the original variables. Since the covariance matrix of the original input matrix X is symmetric, it can be decomposed into $X^T X = V \Omega V^T$, where the diagonal elements of the Ω are the eigenvalues and the columns of V are the eigenvectors of $X^T X$, the coefficients of the original variable are the eigen vectors V. The PCA transformation is thus

$$Z = V^T \left(X - \overline{X}\right) \tag{1}$$

where $\overline{X}$ is the vector of average values of each variable X, and Z is the coordinate in the transformed space. Once the PCA is complete, the reduced PC variables can be used as the input matrix for modeling.

In this study, the raw OES data were three-dimensional (3-D), where the three dimensions were operation condition, wavelength, and time. Plasma parameters were measured at the RF powers of 400, 600, and 800 W, and the pressures of 10, 20, and 30 mTorr, meaning a total of nine experimental runs were conducted, and each experiment lasted 30 s while the PN-ICP system was monitored by the OES system. Emission intensity was recorded every 0.1 s, and the collected data were stored in a local computer as ASCII files. The OES system collected 3600 data points in the wavelength range of 200–1100 nm. When plasma was stabilized under each experimental condition, the time dimension was reduced because the plasma state was the same during operation. Thus, the OES data had a two-dimensional matrix structure. Before the PCA, we performed a standardization process. The main purpose of the standardization was to adjust the range of variables so that each variable contributed equally to the PCA. Here we used open source (scikit-learn PCA) from Python, and the overall accuracy was around 80% [23].

The plasma parameters used in this analysis were n_e, T_e, V_p, V_f, $V_a (= V_p - V_f)$, I_{isat}, N^+, and N_2^+, which were measured from the PN-ICP system. The results with an R^2 value compared with OES data are given in Table 2. The OES data highly correlated with the n_e and N^+; however, the correlation with T_e, V_p, V_f, and V_a was low. This result implied the possibility of inferring n_e through the OES data.

Table 2. Regression results between the OES data and nine plasma parameters measured in the PN-ICP system.

	n_e	T_e	V_f	V_p	V_a	N^+ (MEA1)	N_2^+ (MEA1)	N^+ (MEA2)	N_2^+ (MEA2)
R^2 Score	0.92	0.05	0.20	0.41	0.02	0.90	0.62	0.96	0.65

Here, we further investigated the relationship between OES data and n_e, T_e. The regression analysis was performed using Pearson's correlation coefficient on the entire wavelength of OES data for more accurate correlation analysis. Pearson's correlation coefficient is a test that measures the statistical relationship, or association, between two continuous variables. It is known as the best method of measuring the association between variables of interest because it is based on the method of covariance. It gives information about magnitude of the association, or correlation, as well as the direction of the relationship [24].

As a result, Table 3 shows the Pearson correlation coefficient (R) values of 20 highly correlated wavelengths out of a total wavelength range (200–1100 nm). Since the OES spectrum in the PN-ICP system was dominated by strong molecular features (in Figure 3), 20 highly related wavelengths with plasma parameters were divided into two groups. One was in the range of 295.55–399.59 nm where SPS and FNS were mixed. The other group was in the range of 582.71–770.74 nm where FPS appeared. Here, only the SPS and FNS mixed wavelength range had higher relation with T_e, and it showed a higher relation with N_2^+ ion number density, which was measured by MEA1 and MEA2. The FPS wavelength range had a low relation with T_e, but a higher relation with N^+ ion number density, which was measured by MEA1 and MEA2. As discussed in Section 2 (in Figure 3), these results show the characteristics of the PN-ICP spectrum, where the FPS intensity increases with applied power and slightly increases with gas pressure, and the SPS and FNS intensity slightly increases with applied power and decreases with gas pressure. Thus, this shows that SPS and FNS mixed ranges are mainly correlated with the electron temperature, and the FPS range is correlated with electron density. This trend is in good agreement with electron temperature and density of the PN-ICP system (Figure 2a,b), showing the possibility of monitoring plasma parameters by utilizing specific wavelengths of OES spectrum.

Table 3. Pearson correlation coefficient R value between OES intensity at 20 wavelengths and plasma parameters.

Wavelength (nm)	n_e	T_e	V_f	V_p	V_a	N^+ (MEA1)	N_2^+ (MEA1)	N^+ (MEA2)	N_2^+ (MEA2)
295.55	0.92	0.70	−0.84	−0.86	0.64	0.62	1.00	0.66	0.98
315.98	0.87	0.78	−0.90	−0.88	0.73	0.53	1.00	0.57	0.98
337.15	0.82	0.84	−0.93	−0.88	0.79	0.46	0.99	0.49	0.96
357.49	0.85	0.82	−0.91	−0.88	0.76	0.50	0.99	0.54	0.97
370.93	0.89	0.75	−0.88	−0.87	0.70	0.57	1.00	0.61	0.98
375.40	0.86	0.79	−0.90	−0.88	0.74	0.52	0.99	0.55	0.98
380.14	0.84	0.82	−0.92	−0.88	0.77	0.48	0.99	0.51	0.97
389.34	0.96	0.60	−0.77	−0.82	0.54	0.71	0.98	0.76	0.97
394.07	0.89	0.76	−0.88	−0.87	0.71	0.56	1.00	0.60	0.98
399.59	0.85	0.80	−0.91	−0.88	0.75	0.50	0.99	0.54	0.97
582.71	0.93	0.10	−0.36	−0.56	0.02	0.90	0.72	0.99	0.75
589.39	0.93	0.08	−0.34	−0.55	0.00	0.90	0.70	0.99	0.74
645.18	0.91	0.03	−0.30	−0.52	−0.05	0.90	0.67	0.99	0.70
652.83	0.91	0.04	−0.30	−0.52	−0.04	0.90	0.67	0.99	0.70
660.73	0.91	0.05	−0.31	−0.52	−0.04	0.90	0.68	0.99	0.71
668.87	0.88	−0.02	−0.24	−0.47	−0.10	0.91	0.62	1.00	0.66
675.23	0.87	−0.04	−0.22	−0.46	−0.12	0.91	0.61	0.99	0.64
746.09	0.88	−0.03	−0.24	−0.46	−0.11	0.91	0.62	1.00	0.65
760.69	0.85	−0.09	−0.18	−0.42	−0.17	0.91	0.57	0.99	0.61
770.74	0.80	−0.18	−0.09	−0.34	−0.26	0.90	0.49	0.98	0.53

4. Machine Learning Prediction Method

In the previous section, in order to develop a predictive VM model to predict the plasma electron density and temperature, we performed a correlation analysis for the data characterization to better understand the PN-ICP system. Data characterization defines the relation between the operation condition data, plasma parameters, and OES spectrum data. Next, in this section, we provide a concise description of the machine learning prediction method for OES as a real-time VM plasma monitoring sensor.

Electron density and temperature measured by CP and LP were used as a target data for machine learning prediction. Since OES intensity has a linear correlation with RF power and gas pressure, a multi-linear regression method (MLR) [25] was used that can express the functional dependence relationship between input data and predictive data in a mathematical form. When considering the relationship between the k independent variables $(X_1, X_2, X_3, \cdots, X_k)$ and the dependent variable Y, a regression equation is established: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \epsilon_i$, where $\beta_0, \beta_1, \beta_2, \cdots, \beta_k$ are the regression coefficients to be estimated; $i = 1, 2, \cdots, n$ (n is the sample size); ϵ_i is the random error. In this study, the dependent variable was T_e and n_e; the independent variable was the intensity of 20 selected wavelengths, as shown in Table 3.

The input variable $X_1, X_2, X_3, \cdots, X_k$ selection was important for correct MLR development. There were three selection methods: forward selection, backward elimination, and stepwise selection [24,25]. The forward selection is a variable selection method that begins with a model that contains no variables. Then, it starts adding the most significant variables one after the other until a pre-specified stopping rule is reached or until all the variables under consideration are included in the model. The backward elimination method is the opposite of the forward selection method. The backward elimination method is a method of creating a model with all variables, and then deleting variables that do not contribute to model performance. The stepwise selection method is similar to the forward selection method, but the process of removing unimportant variables by examining the importance is added.

The importance is determined by the p-value of the variable [26]. The p-value is defined as the probability that a result (observed result) will be more extreme than the result obtained when the null hypothesis is true. It is usually based on a p-value less than

0.05 or 0.01. In this statistic, if the null hypothesis is true, it means that the probability that the sample actually corresponds to the statistic is less than 5% (or 1%). In other words, it can be said that the statistic has a much higher probability that the alternative hypothesis will be true with a 95% probability.

In this paper, the null hypothesis was that the selected wavelength was not related to electron temperature and electron density. Thus, the alternative hypothesis was that the wavelength chosen was related to electron temperature and electron density. Table 4 shows the most optimized wavelengths for predicting electron density and electron temperature with MLR as a result of selecting three variable selection methods.

Table 4. The optimized wavelengths for predicting electron density and electron temperature in the MLR method.

Wavelength (nm)	n_e		T_e	
	Use	p-Value	Use	p-Value
295.55	Yes	0.000	Yes	0.000
315.98	Yes	0.000	Yes	0.000
337.15	No	0.141	Yes	0.000
357.49	Yes	0.003	Yes	0.000
370.93	Yes	0.000	Yes	0.000
375.40	Yes	0.000	Yes	0.000
380.14	No	0.227	Yes	0.000
389.34	Yes	0.000	Yes	0.000
394.07	Yes	0.001	Yes	0.000
399.59	No	0.238	Yes	0.035
582.71	Yes	0.000	-	-
589.39	Yes	0.000	-	-
645.18	Yes	0.000	-	-
652.83	Yes	0.000	-	-
660.73	Yes	0.000	-	-
668.87	Yes	0.002	-	-
675.23	Yes	0.000	-	-
746.09	No	0.886	-	-
760.69	Yes	0.000	-	-
770.74	No	0.117	-	-

Figure 6 shows the results of comparing the electron temperature and density predicted by the MLR method using the selected wavelength with the electron temperature and density measured by the Langmuir probe as described in Section 2, which were measured at the RF powers of 400, 600, and 800 W and the pressures of 10, 20, and 30 mTorr. As can be seen in Figure 5, although 15 wavelengths were selected to predict the electron density, and 5 wavelengths were selected to predict the electron temperature, the MLR prediction results are good agreement with the measured data; the prediction accuracy of electron density was 97%, and the electron temperature was 90%. This shows that the electron density and electron temperature can be predicted by selecting meaningful regression coefficients from the wavelength measured by OES.

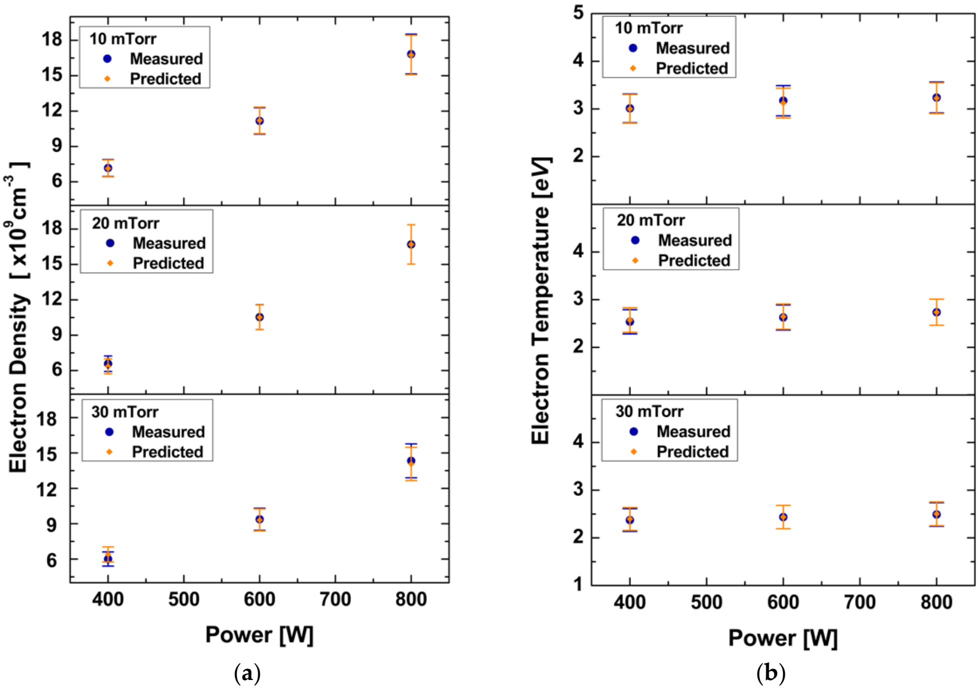

Figure 6. Comparison between the MLR method (predicted) and Langmuir probe data (measured). (**a**) is plasma electron density, (**b**) is plasma electron temperature.

5. Summary and Discussions

A method for the non-invasive extraction of plasma parameters from an OES using a correlation based on a diagnostic data set of simultaneously measured OES and other diagnostics has been presented and validated. Instead of relying on a theoretical model of the plasma emission to extract plasma parameters from the OES, an empirical correlation was established on the basis of simultaneous OES and other diagnostics; various diagnostic tools were constructed; and plasma parameters were measured and analyzed. The correlation between the measured data was investigated using regression analysis. The correlation analysis between the plasma parameters and OES spectrum indicated that 20 wavelengths out of a total wavelength range (200–1100 nm) were significantly closely related to the plasma parameters, including the ionic species density. This result indicates the possibility of inferring the plasma parameters and ionic species density from the OES data.

Additionally, we developed a machine learning-based virtual metrology model for real-time T_e and n_e monitoring in plasma nitridation process using an in situ OES sensor. Electron density and temperature measured by CP and LP were used as a target data for machine learning prediction. Since OES intensity has a linear correlation with RF power and gas pressure, a multi-linear regression method (MLR) was used that can express the functional dependence relationship between input data and predictive data in a mathematical form. The result show that the prediction accuracy of electron density was 97% and that of electron temperature was 90%. This method is potentially powerful in real-time control and monitoring because it uses non-invasive and in situ sensor readings for prediction while the wafer is processing.

Author Contributions: Conceptualization, J.-H.P., J.-H.C., J.-S.Y. and J.-H.S.; methodology, J.-H.P., J.-H.C., J.-S.Y. and J.-H.S.; software, J.-H.P., J.-H.C., J.-S.Y. and J.-H.S.; validation, J.-H.P., J.-H.C., J.-S.Y. and J.-H.S.; formal analysis, J.-H.P., J.-H.C., J.-S.Y. and J.-H.S.; investigation, J.-H.P., J.-H.C., J.-S.Y. and J.-H.S.; resources, J.-H.P., J.-H.C., J.-S.Y. and J.-H.S.; data curation, J.-H.P., J.-H.C., J.-S.Y. and J.-H.S.; writing—original draft preparation J.-H.P., J.-H.C., J.-S.Y. and J.-H.S.; writing—review and editing, J.-H.C.; visualization, J.-H.P., J.-H.C., J.-S.Y. and J.-H.S.; supervision, J.-H.P., J.-H.C., J.-S.Y. and J.-H.S.; project administration, J.-H.P., J.-H.C., J.-S.Y. and J.-H.S.; funding acquisition, J.-H.P., J.-H.C., J.-S.Y.

and J.-H.S. All the authors contributed to writing, reviewing, and editing the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Council of Science & Technology (NST) grant provided by the Korean government (MSIT) (No. CRC-20-01-NFRI) and R&D Program of "Plasma Big-Data ICT Convergence Technology Research Project (code No. 1711124799)" through the Korea Institute of Fusion Energy (KFE) funded by the Government funds, Republic of Korea.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are contained within the article.

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

References

1. Vandenabeele, C.R.; Lucas, S. Technological challenges and progress in nanomaterials plasma surface modification—A review. *Mater. Sci. Eng. R Rep.* **2020**, *139*, 100521. [CrossRef]
2. Kanarik, K.J.; Lill, T.; Hudson, E.A.; Sriraman, S.; Tan, S.; Marks, J.M.; Vahedi, V.; Gottscho, R.A. Overview of Atomic Layer Etching in the Semiconductor Industry. *J. Vac. Sci. Technol. A Vac. Surf. Film.* **2015**, *33*, 020802. [CrossRef]
3. Samukawa, S.; Hori, M.; Rauf, S.; Tachibana, K.; Bruggeman, P.; Kroesen, G.; Whitehead, J.C.; Murphy, A.B.; Gutsol, A.F.; Starikovskaia, S.; et al. The 2012 Plasma Roadmap. *J. Phys. D Appl. Phys.* **2012**, *45*, 253001. [CrossRef]
4. Bruggeman, P.; Verreycken, T.; Gonzalez, M.A.; Walsh, J.L.; Kong, M.G.; Leys, C.; Schram, D.C. Optical emission spectroscopy as a diagnostic for plasmas in liquids: Opportunities and pitfalls. *J. Phys. D Appl. Phys.* **2010**, *43*, 124005. [CrossRef]
5. Bogaerts, A.; Tu, X.; Whitehead, J.C.; Centi, G.; Lefferts, L.; Guaitella, O.; Jury, F.A.; Kim, H.H.; Murphy, A.B.; Schneider, W.F. The 2020 plasma catalysis roadmap. *J. Phys. D Appl. Phys.* **2020**, *53*, 443001. [CrossRef]
6. Zhu, X.M.; Pu, Y.K. Using OES to determine electron temperature and density in low-pressure nitrogen and argon plasmas. *Plasma Sources Sci. Technol.* **2008**, *17*, 024002. [CrossRef]
7. Guerra, V.; Sá, P.A.; Loureiro, J. Kinetic modeling of low-pressure nitrogen discharges and post-discharges. *Eur. Phys. J. Appl. Phys.* **2004**, *28*, 125. [CrossRef]
8. Shakhatov, V.A.; Lebedev, Y.A. Kinetic excitation of $N_2(A^3\Sigma^+u, va)$, $N_2(C^3\Pi u, vC)$ and $N_2(B^3\Pi g, vB)$ in nitrogen discharge plasma as studied by means of emission spectroscopy and computer simulation. *High Energy Chem.* **2008**, *42*, 107. [CrossRef]
9. Itagaki, N.; Iwata, S.; Muta, K.; Yonesu, A.; Kawakami, S.; Ishii, N.; Karai, Y. Electron-temperature dependence of nitrogen dissociation in 915 MHz ECR plasma. *Thin Solid Film.* **2003**, *435*, 259–263. [CrossRef]
10. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [CrossRef] [PubMed]
11. Choi, J.E.; Hong, S.J. Machine learning-based virtual metrology on film thickness in amorphous carbon layer deposition process. *Meas. Sens.* **2021**, *16*, 100046. [CrossRef]
12. Yue, H.H.; Qin, S.J.; Wiseman, J.; Toprac, A. Plasma etching endpoint detection using multiple wavelengths for small open-area wafers. *J. Vac. Sci. Technol. A* **2001**, *19*, 66. [CrossRef]
13. Han, K.; Yoon, E.S.; Lee, J.; Chae, H.; Han, K.H.; Park, K.J. Real-time end-point detection using modified principal component analysis for small open area SiO_2 plasma etching. *Ind. Eng. Chem. Res.* **2008**, *47*, 3907. [CrossRef]
14. Kim, B.B.; Im, S.; Yoo, G. Performance Evaluation of CNN-Based End-Point Detection Using In-Situ Plasma Etching Data. *Electronics* **2021**, *10*, 49. [CrossRef]
15. Kim, J.; Seong, D.; Chung, K. Plasma frequency measurements for absolute plasma density by means of wave cutoff method. *Appl. Phys. Lett.* **2003**, *83*, 4725. [CrossRef]
16. National Physical Laboratory. *Certifcate of Calibration: Tungsten Ribbon Lamp, Gas Filled No. P300C*; Middlesex University: London, UK, 2010.
17. National Institute of Standards and Technology. NIST Atomic Spectra Database. September 2010. Available online: http://nist.gov/pml/data/asd.cfm (accessed on 1 July 2021).
18. Lofthus, A.; Krupenie, P.H. The spectrum of molecular nitrogen. *J. Phys. Chem. Ref. Data* **1977**, *6*, 113. [CrossRef]
19. Abdi, H.; Williams, L.J. Principal component analysis. *WIREs Comput. Stat.* **2010**, *2*, 433. [CrossRef]
20. Wang, C.Y.; Hsu, C.C. Development and testing of an efficient data acquisition platform for machine learning of optical emission spectroscopy of plasma in aqueous solution. *Plasma Sources Sci. Technol.* **2019**, *28*, 105013. [CrossRef]
21. Scikit-Learn. Available online: https://scikit-learn.org/stable/index.html (accessed on 1 July 2021).
22. Schober, P. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth. Analg.* **2018**, *126*, 1763. [CrossRef] [PubMed]
23. Freedman, D.A. *Statistical Models: Theory and Practice*; Cambridge University: Cambridge, UK, 2009; p. 26.
24. Koller, D.; Sahami, M. Toward Optimal Feature Selection. In Proceedings of the Thirteenth International Conference on International Conference on Machine Learning, Bari, Italy, 3–6 July 1996.

25. Susto, G.A.; Beghi, A. Least Angle Regression for Semiconductor Manufacturing Modeling. In Proceedings of the 2012 IEEE International Conference on Control Applications, Dubrovnik, Croatia, 3–5 October 2012; pp. 658–663.
26. Ioannidis, J.P. The proposal to lower P value thresholds to. 005. *JAMA* **2018**, *319*, 1429–1430. [CrossRef] [PubMed]