

## Article

# Improved Plasma Etch Endpoint Detection Using Attention-Based Long Short-Term Memory Machine Learning

Ye Jin Kim <sup>1</sup>, Jung Ho Song <sup>2</sup>, Ki Hwan Cho <sup>2</sup>, Jong Hyeon Shin <sup>2</sup>, Jong Sik Kim <sup>2</sup> , Jung Sik Yoon <sup>2</sup> and Sang Jeon Hong <sup>1,\*</sup> 

<sup>1</sup> Department of Semiconductor Engineering, Myongji University, Yongin 17058, Republic of Korea; 80231133@mju.ac.kr

<sup>2</sup> Plasma E.I. Convergence Research Center, Korea Research Institute of Fusion Energy, Daejeon 34133, Republic of Korea; jungcho@kfe.re.kr (J.H.S.); khcho@kfe.re.kr (K.H.C.); shin0020@kfe.re.kr (J.H.S.); jongsik@kfe.re.kr (J.S.K.); jsyoon@kfe.re.kr (J.S.Y.)

\* Correspondence: samhong@mju.ac.kr

**Abstract:** Existing etch endpoint detection (EPD) methods, primarily based on single wavelengths, have limitations, such as low signal-to-noise ratios and the inability to consider the long-term dependencies of time series data. To address these issues, this study proposes a context of time series data using long short-term memory (LSTM), a kind of recurrent neural network (RNN). The proposed method is based on the time series data collected through optical emission spectroscopy (OES) data during the SiO<sub>2</sub> etching process. After training the LSTM model, the proposed method demonstrated the ability to detect the etch endpoint more accurately than existing methods by considering the entire time series. The LSTM model achieved an accuracy of 97.1% in a given condition, which shows that considering the flow and context of time series data can significantly reduce the false detection rate. To improve the performance of the proposed LSTM model, we created an attention-based LSTM model and confirmed that the model accuracy is 98.2%, and the performance is improved compared to that of the existing LSTM model.

**Keywords:** plasma etch; endpoint detection; machine learning



**Citation:** Kim, Y.J.; Song, J.H.; Cho, K.H.; Shin, J.H.; Kim, J.S.; Yoon, J.S.; Hong, S.J. Improved Plasma Etch Endpoint Detection Using Attention-Based Long Short-Term Memory Machine Learning.

*Electronics* **2024**, *13*, 3577. <https://doi.org/10.3390/electronics13173577>

Academic Editors: Claudio Turchetti and Laura Falaschetti

Received: 31 July 2024

Revised: 2 September 2024

Accepted: 5 September 2024

Published: 9 September 2024



**Copyright:** © 2024 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

Over the past decade, 3D-NAND flash memory technology has rapidly enhanced the bit storage density per unit area by increasing the number of vertically stacked gate layers. In recent years, the semiconductor industry has witnessed a significant increase in demand for high-capacity storage devices, such as 3D-NAND flash memory, driven by the rapid growth of various sectors, including cloud services and mobile devices. The latest generation of 3D-NAND features more than 200 layers of vertical gate stacks. To manufacture such a highly integrated 3D-NAND flash memory, high-aspect-ratio (HAR) etching technology is essential, enabling the precise etching of numerous layers, ranging in the several hundreds [1,2]. In these structures, accurately controlling the etching depth and precisely detecting the etching endpoint for each layer has become increasingly important. Consequently, reliable etch endpoint detection (EPD) techniques are crucial. EPD plays a vital role in determining the yield and quality of the etching process, and its significance continues to grow [3].

Accurate etch endpoint detection plays a crucial role in improving manufacturing process yield and reducing process failure rates. In the plasma etching process, the intensity of the wavelength corresponding to the thin-film layer to be etched is monitored, and the change in intensity provides a real-time view of the etching progress. The principle of EPD is that when the etched material is fully reacted, and no more by-products are produced, a change in the intensity of the corresponding wavelengths is detected. Various EPD techniques have been developed in recent years [4,5]. However, several limitations

still exist. Traditional EPD algorithms have primarily relied on signal changes at a single wavelength [6–8]. Such single-wavelength-based EPD methods are vulnerable to noise and signal fluctuations, which can lead to a low signal-to-noise ratio [9,10]. This may pose a challenge to the precision of endpoint detection.

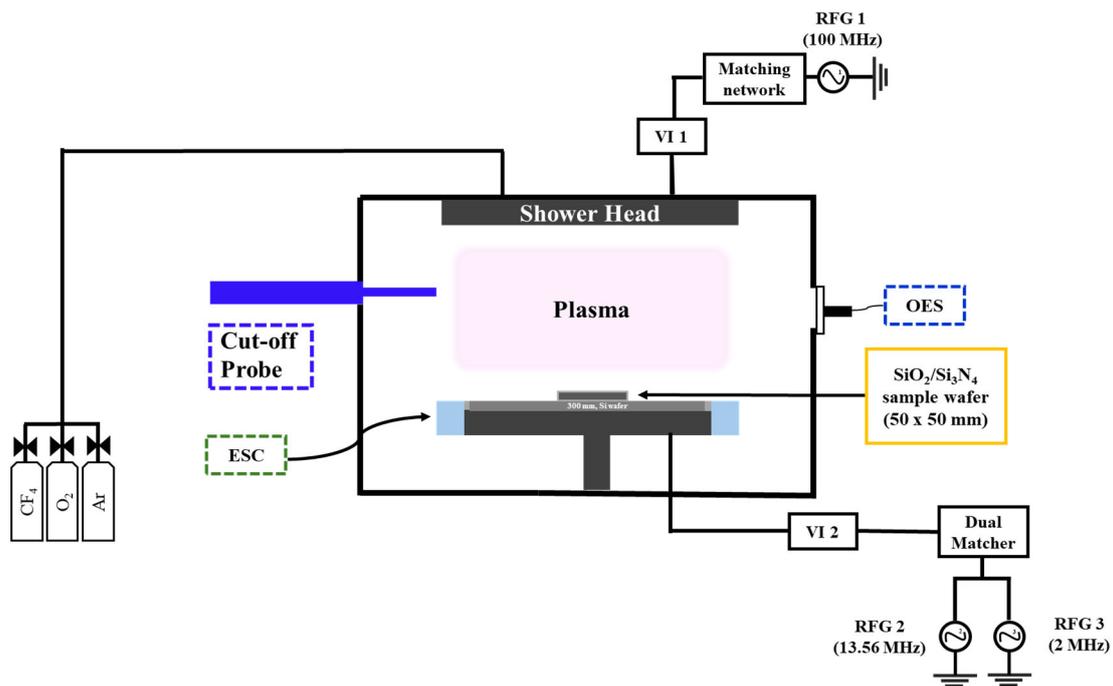
While recent studies have focused on improving EPD performance by employing multiple wavelengths [11,12], most EPD algorithms still rely on signal patterns at specific time points for decision-making. This reliance on instantaneous signal patterns makes it challenging to predict the EPD moment in advance or accurately detect the endpoint, as the algorithms mainly focus on the signal change at the exact EPD time. On the other hand, the optical emission spectroscopy (OES) time series data measured during the entire etching process contains information spanning from the beginning to the end of the etching procedure. However, existing EPD algorithms often struggle to fully account for the temporal dependencies embedded within these time series data, lacking consideration for the comprehensive flow and context of the data. As a result, there is a pressing need for data analysis and decision-making approaches that effectively incorporate these long-term temporal dependencies.

In recent years, efforts have been made to address these issues by leveraging deep learning techniques [13–16]. Kim et al. utilized a convolutional neural network (CNN) based on OES data to detect the etch endpoint [17], but it has a limitation in directly modeling the temporal dependencies of time series data, as the authors transformed the OES data into a two-dimensional format to be used as input for the CNN. Hwang et al. proposed a method for detecting anomalies in time series data from semiconductor manufacturing processes using a long short-term memory (LSTM) autoencoder [18]. This suggests that the LSTM can be effectively utilized for analyzing semiconductor process data, but there is still a lack of direct application to EPD. Existing EPD studies have primarily focused on endpoint detection itself, with a tendency to rely on signal patterns at specific time points. The LSTM is a type of recurrent neural network (RNN) designed to address the long-term dependency problem [19–22]. The LSTM can effectively learn long-term patterns in time series data through its cell state and multiple gates [23–25]. These characteristics of the LSTM make it suitable for analyzing time series data, such as the intensity values of wavelengths obtained through OES sensors.

In this study, we propose an EPD method that considers the entire flow and context of time series data using the LSTM. Through the LSTM, we can comprehensively analyze the change patterns of OES signals from the past to the present, considering the temporal dependencies and interactions between multiple wavelengths. The goal is to improve EPD performance by reflecting the context of the entire time series rather than simply looking at the signal at a specific time point. By doing so, we expect to increase accuracy and reduce the false detection rate.

## 2. Experiment

Experiments were conducted using an in-house modified commercial 300 mm capacitive coupled plasma (CCP) etch system operating at frequencies of 2 MHz, 13.56 MHz, and 100 MHz, as depicted in Figure 1. A commercial 300 mm high-end manufacturing etch chamber, in-house modified with triple-RF power systems for the purpose of plasma equipment intelligence, located in the Plasma E.I. Convergence Research Center in Korea Institute of Fusion Energy, was employed for the experiment. The triple-RF power systems consist of a 100 MHz VHF source power in the upper electrode and a combination of 13.56 MHz and 2 MHz LF bias power in the lower electrode. The upper electrode utilizes a very-high-frequency (VHF) power source to control the ion flux incident on the substrate, while the lower electrode employs a low-frequency (LF) power source to control the ion energy incident on the substrate. For the etching target, SiO<sub>2</sub>/Si<sub>3</sub>N<sub>4</sub>/Si coupon wafers (50 mm × 50 mm) were etched using a mixture of CF<sub>4</sub>, O<sub>2</sub>, and Ar gasses according to the recipe provided in Table 1.



**Figure 1.** A schematic description of the 300 mm CCP-type dielectric etcher, called the data acquisition oxide etch (DaO) system, located in the Korea Research Institute of Fusion Energy, Korea.

**Table 1.** Oxide etch recipe.

Pressure (mTorr)	RF Power (Watt)			Gas (sccm)			Time (s)
	2 MHz	13.56 MHz	100 MHz	CF <sub>4</sub>	O <sub>2</sub>	Ar	
20	1500	500	800	240	45	60	90

We acquired real-time OES data by connecting an OES sensor to the viewport sidewall of the etch chamber. Using the non-invasive OES sensor, we measured the wavelength and intensity of the light emitted from the plasma. The measurement range spanned from 200 nm to 1100 nm, and the wavelength information was obtained at 100 ms intervals. OES is commonly used to understand the energy state transitions of radicals, reactive ions, and atoms within a plasma chamber [26]. In this experiment, the emission wavelength intensities of radicals and reactive ions in the plasma were measured and identified.

When etching SiO<sub>2</sub> using a CF<sub>4</sub>/O<sub>2</sub>/Ar plasma, the main chemical species generated through chemical and physical reactions are the reaction products in the surface reaction mechanism. The surface reactions can be simplified as SiO<sub>2</sub> + CF<sub>x</sub> → SiF<sub>y</sub> + CO<sub>z</sub> + ⋯ and Si<sub>3</sub>N<sub>4</sub> + CF<sub>x</sub> → SiF<sub>y</sub> + CO<sub>z</sub> + ⋯. The OES wavelength features selected in this process were the peaks related to the reactants F, O, and C, and the peaks related to the reaction products in the reaction surface mechanism, CO, CN, and SiF. The wavelengths used can be found in Table 2.

**Table 2.** Information about the wavelengths used for EPD.

Species	Wavelength (nm)
SiF	336, 440
CO	482, 560, 561
CN	386, 387
F	685, 677, 703
C <sub>2</sub>	516
O	777, 844

All coupon wafers were etched for a pre-determined fixed time. To ensure the complete etching of the SiO<sub>2</sub> layer, over-etching was performed. During this process, it was confirmed through ellipsometry that approximately 200 nm of the Si<sub>3</sub>N<sub>4</sub> layer was etched. This was carried out to verify that the SiO<sub>2</sub> layer was fully etched. Therefore, the optical emission signals collected through OES serve as time series data spanning the entire etching process. By using these data as input for the LSTM model, the model can learn long-term dependencies and detect the etch endpoint.

### 3. LSTM Method

The LSTM is a model designed to address the limitations of the RNN, specifically the problems of vanishing gradients and exploding gradients. The gradient information in the LSTM is more effective in maintaining and utilizing information over longer sequences compared to the RNN [27].

As shown in Figure 2, the LSTM consists of three gates: the input gate, forget gate, and output gate. Each gate has the following roles. The input gate determines how to update the internal state based on the current state and the previous hidden state [28]. It decides what information should be stored in the cell state. First, it determines what information to store from the current input and the previous hidden state through a sigmoid layer. Then, the hyperbolic tangent (tanh) layer generates a new candidate vector. The forget gate determines what information should be forgotten and to what extent, from the previous cell state, by passing it through a sigmoid layer. The output gate determines how much of the cell state should be reflected in the hidden state and decides the hidden state to be passed to the next time step of the LSTM [28].

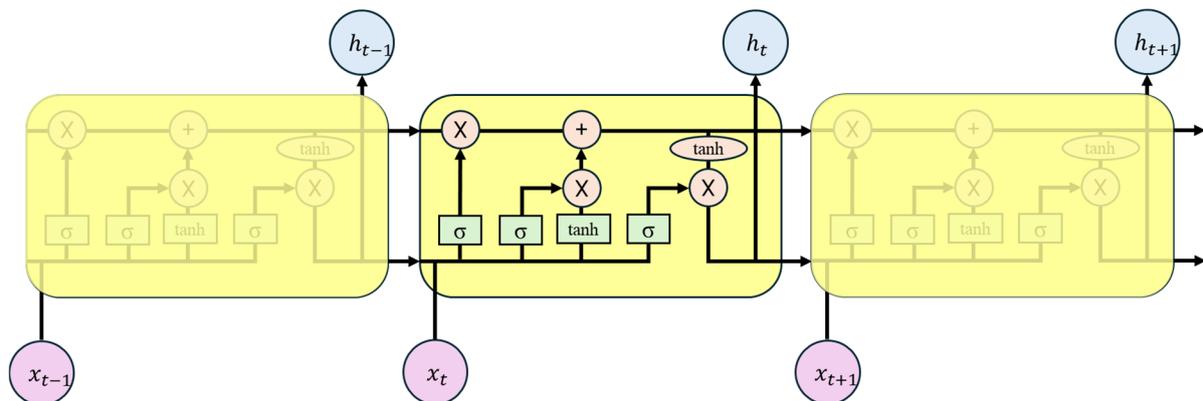


Figure 2. LSTM diagram [29].

The LSTM controls the cell state using three gates. The cell state contains all the core information, and the hidden state is processed whenever necessary to propagate information in a form that exposes only the required information for each time step. The equations for the LSTM are as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (2)$$

$$\vec{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \quad (3)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \vec{C}_t, \quad (4)$$

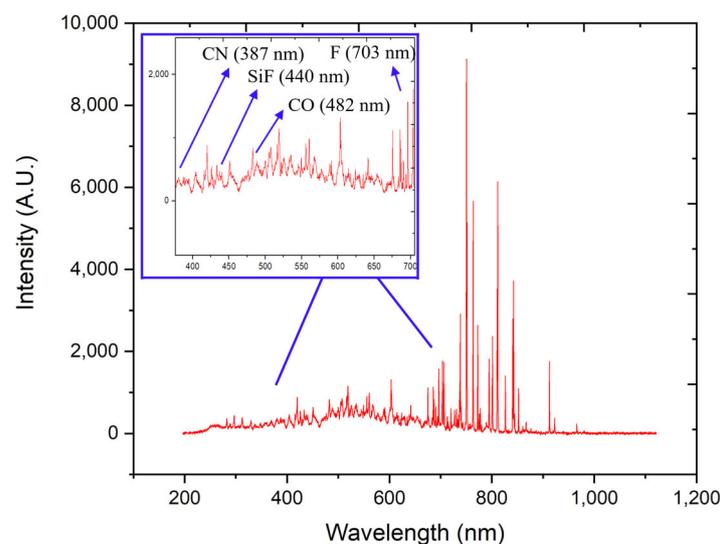
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad (5)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (6)$$

In these equations,  $f_t$ ,  $i_t$ , and  $o_t$  represent the forget, input, and output gates, respectively. The function  $\sigma$  denotes the sigmoid function. The matrices  $W_f$ ,  $W_i$ ,  $W_c$ , and  $W_o$  are the weight matrices for the corresponding gates, and the vectors  $b_f$ ,  $b_i$ ,  $b_c$ , and  $b_o$  are the bias vectors for the corresponding gates.  $h_{t-1}$  represents the hidden state of the previous time step,  $x_t$  is the input at the current time step,  $\vec{C}_t$  denotes the candidate cell state,  $C_{t-1}$  is the cell state from the previous time step, and  $C_t$  is the cell state at the current time step. These are all used to update and output the internal state of the LSTM.

#### A. Data preprocessing

In this study, we conducted etching experiments and acquired a total of 10 datasets. Each acquired dataset consisted of approximately 20,000 OES intensity values. The intensity values measured through OES can vary depending on the environment and sensor settings. Therefore, the preprocessing stage is essential to enhance the performance of the LSTM model. In this stage, the OES sensor data with different intensity values were preprocessed by normalizing between 0 and 1 using MinMaxScaler to adjust them to a consistent range. The algorithm proposed for EPD utilized the OES peak intensity data, which effectively reflect the characteristics of the etch endpoint, as shown in Figure 3, along with the aforementioned wavelengths. The preprocessed data were used as input data to detect the etch endpoint. The labeling results showed that the ratio of class 0 to class 1 was approximately 6:4, indicating a relatively balanced dataset. The dataset was divided into training and validation sets with an 8:2 ratio, which is a commonly used split ratio in machine learning.



**Figure 3.** Example of OES data with selected peaks, corresponding to chemical species in the plasma.

#### B. LSTM modeling and results

In this study, to detect the etch endpoint of  $\text{SiO}_2$ , we designed the LSTM modeling, as shown in Figure 4, considering the entire flow and context of the time series data. The modeled LSTM consists of three stacked layers, each performing a unique role. The layer configuration was determined through an iterative process, where we repeatedly evaluated performance to identify the optimal structure. The modeled LSTM consists of three stacked layers, each performing a unique role. The first layer contains 64 cells. The second layer includes 32 cells to remember the values from the first layer. In the third layer, a fully connected layer and the rectified linear unit (ReLU) activation function are used to accurately detect the etch endpoint based on the remembered state information. Finally, by using the sigmoid activation function, the output values are restricted between 0 and 1, enabling the identification of the situations before and after the etch endpoint. This configuration provides an optimal balance between performance and training speed.

during the model design process. The LSTM model structure and parameter values can be found in Table 3.

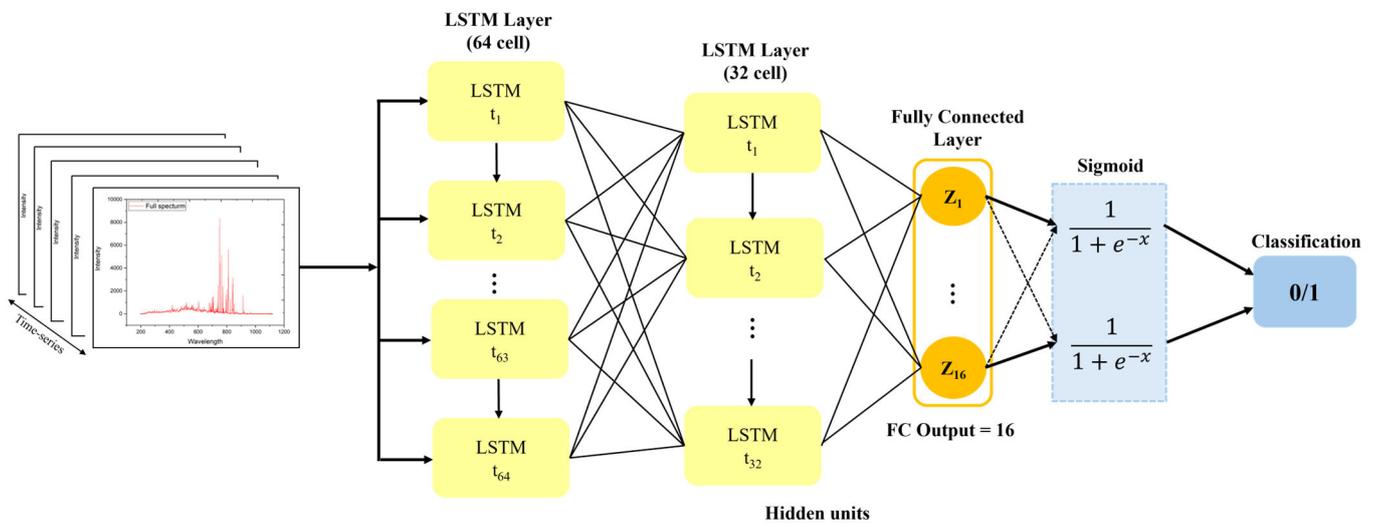


Figure 4. LSTM neural network schematic.

Table 3. LSTM model structure and parameter values.

Layer (Type)	Output Shape	Parameter No.
lstm (LSTM)	(None, 10, 64)	18,688
lstm_1 (LSTM)	(None, 32)	12,416
dense (Dense)	(None, 16)	528
dense_1 (Dense)	(None, 1)	17

The LSTM model training process is illustrated in Figure 5. During the training of the LSTM model, the wavelength intensity values from Table 2 were bundled in 0.5 s intervals and input into the model. This allows the LSTM to determine the occurrence of EPD based on the intensity changes over a 0.5 s period, distinguishing between the situations before and after the endpoint. The parameters set during the training process were as follows: the learning rate was 0.001, the number of training iterations was 200, and the batch size was 32. To reduce the training time, the learning process was set to stop if the same loss is obtained for 10 consecutive iterations. With these conditions, the model achieved an accuracy of 97.1% upon completion of the training. To evaluate the generalization ability of the trained model, we performed a 10-fold cross-validation and found that the LSTM model was able to maintain a stable predictive performance, with an average accuracy of 97.1%, even on data not used for training.

As shown in Figure 6, the performance of the model was evaluated using a 10-fold cross-validation and the loss and accuracy graphs of the model. The loss graph reveals that the training loss decreases rapidly in the initial stages but gradually becomes more gradual. This indicates that the model is being optimized through learning. The validation loss also exhibits a similar trend, and no signs of overfitting were observed. As shown in Table 4, when evaluating the accuracy using test data, the model achieved a 97.1% accuracy in identifying EPD, with a false detection rate of approximately 1%. Consequently, the LSTM model demonstrated high accuracy by effectively learning the temporal dependencies of the time series data.

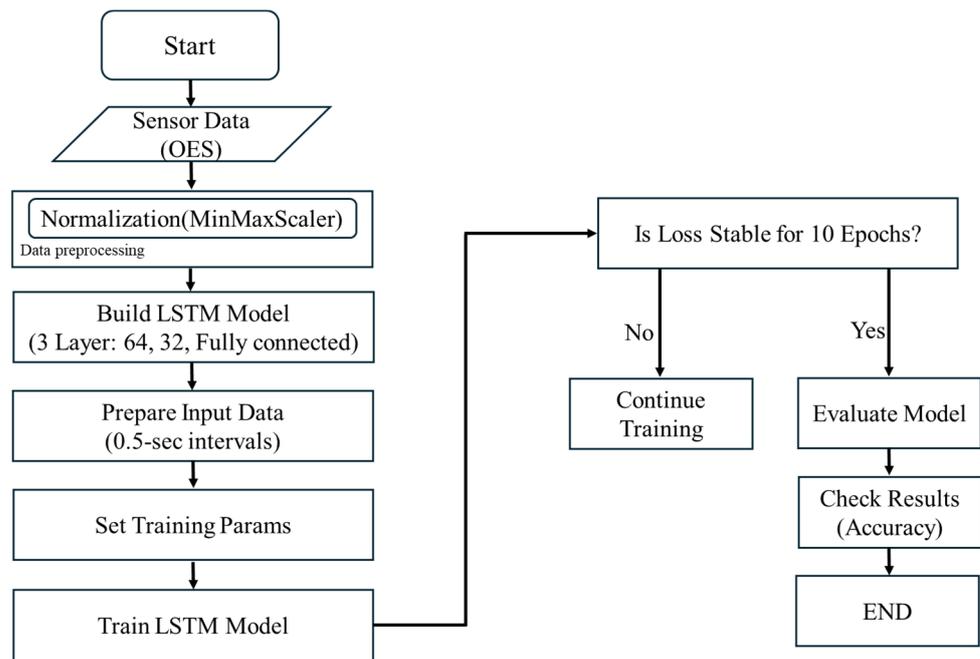


Figure 5. Flowchart of the training process.

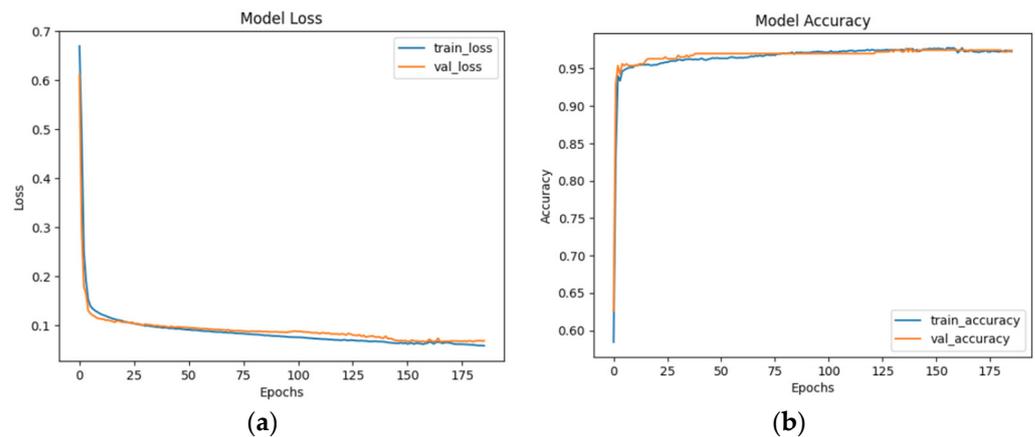


Figure 6. Modeling performance evaluation: (a) loss graph of LSTM-based model; (b) accuracy of LSTM-based model.

Table 4. LSTM model confusion matrix metrics.

	Precision	Recall	F1 Score
Without EPD	0.95	0.99	0.97
With EPD	0.99	0.96	0.97
Accuracy			0.97

The accuracy of the proposed LSTM model was evaluated by comparing the predicted results with the actual results. To compare the performance of the LSTM-based model, support vector machine (SVM), Random Forest, and Adaboost models, which are models that do not consider temporal dependencies, were used as etch endpoint detectors. The SVM classifier demonstrated an accuracy of 93.6%, while RandomForest showed an accuracy of 94.8%. AdaBoost exhibited an accuracy of 92%. Consequently, the proposed LSTM model outperformed the other three models, confirming that a model considering temporal dependencies is superior. These model accuracy results are presented in Table 5.

**Table 5.** Machine learning model accuracy.

Model	Accuracy (%)
SVM	93.6
Random Forest	94.8
AdaBoost	92

On the other hand, deep learning models such as the Bidirectional LSTM (BiLSTM) and vanilla RNN were used to compare the performance of the LSTM model. BiLSTM showed an accuracy of 93.1%, while vanilla RNN demonstrated an accuracy of 96.4%. As a result, the proposed LSTM model achieved the highest performance, and these model accuracy results are presented in Table 6.

**Table 6.** Deep learning model accuracy.

Model	Accuracy (%)
BiLSTM	93.1
vanilla RNN	96.4
LSTM	97.1
Attention-based LSTM	98.2

To improve the LSTM model, an attention layer was added to the proposed LSTM model. This was carried out to more effectively capture the intensity change characteristics of the OES data as the etching process progresses, thereby enhancing the performance of the EPD model. The acquired OES data are characterized by a gradual change in intensity as the etching process progresses, with different intensity widths at certain critical points where the intensity changes in unit time. To effectively capture these change features, we used an attention mechanism.

The attention mechanism can focus on points in a sequence of OES data that have different intensity changes. It emphasizes the important parts of the sequence by assigning higher weights to points where the change in intensity represents a change from the previous state. During the learning process of the model, it evaluates the importance of each element of the input sequence and helps to ensure that important information is not lost [30]. The attention-based LSTM model was designed as shown in Figure 7. It consists of a basic LSTM layer that processes the input data, followed by an attention layer that evaluates the importance of each sequence. Finally, a fully connected layer is added to receive the output from the attention layer and generate the final predicted values. This ultimately produces an output that identifies the situations before and after the etch endpoint. The structure and parameter values of the attention-based LSTM model can be found in Table 7.

**Table 7.** Attention-based LSTM model structure and parameter values.

Layer (Type)	Output Shape	Parameter No.
lstm (LSTM)	(None, 10, 64)	18,688
lstm_1 (LSTM)	(None, 10, 32)	12,416
Attention (Attention)	(None, 32)	1088
dense (Dense)	(None, 16)	528
dense_1 (Dense)	(None, 1)	17

The training process for the attention-based LSTM model is the same as that of the LSTM model, with the only difference being the addition of the attention layer. When the training of the attention-based LSTM model was completed, it achieved an accuracy of approximately 98.2%. The performance of the model was evaluated using 10-fold cross-validation and the loss and accuracy graphs of the model. The results showed that, similar

to the LSTM model, the validation loss exhibited a similar trend, and no signs of overfitting were observed.

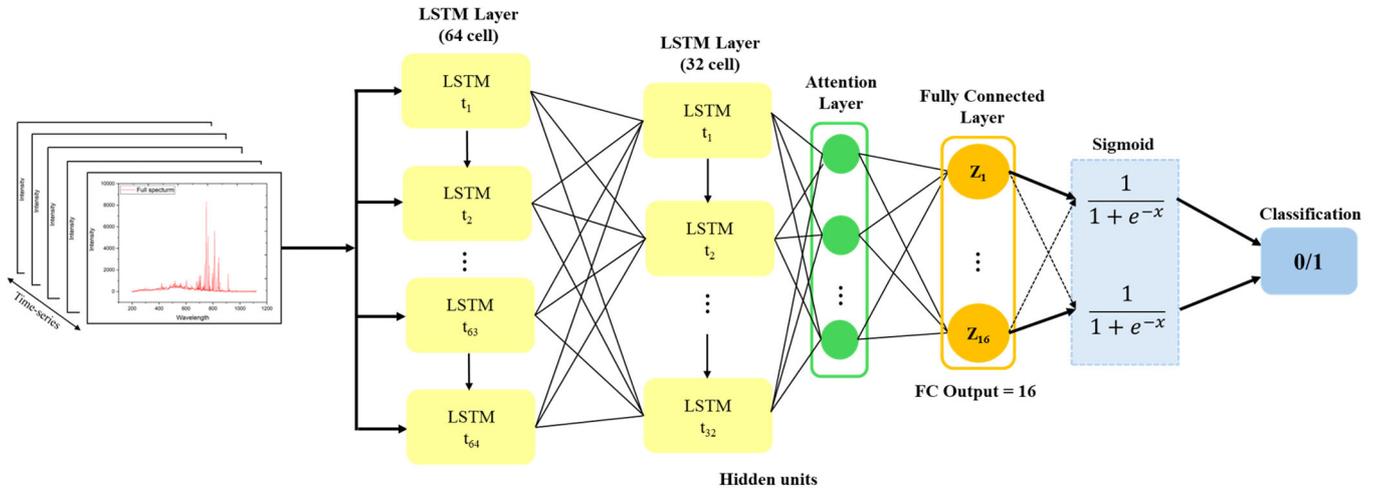


Figure 7. Attention-based LSTM neural network schematic.

As shown in Table 8, the attention-based LSTM model improved the etch endpoint detection performance by approximately 1% and reduced the false detection rate (F1 score) by approximately 1% compared to the LSTM model.

Table 8. Attention-based LSTM model confusion matrix metrics.

	Precision	Recall	F1 Score
Without EPD	0.98	0.98	0.98
With EPD	0.98	0.98	0.98
Accuracy			0.98

#### 4. Conclusions

In this study, we proposed an LSTM-based method for detecting the endpoint in etching processes. Unlike traditional approaches that rely on signal patterns at specific time points, this method demonstrated its effectiveness by considering all the information from the beginning to the end of the etching process contained in the entire OES time series data, and the LSTM model achieved a high accuracy of 97.1%. The performance of the model was further improved by applying the attention mechanism to the LSTM model. The attention mechanism assigns higher weights to points in the time series data where the intensity change differs from the previous state, allowing the model to focus on the critical information for etch endpoint detection. This resulted in an approximately 1% performance improvement compared to the LSTM model, with the attention-based LSTM model ultimately achieving a high accuracy of 98.2%. The proposed attention-based LSTM model overcomes the limitations of the existing EPD algorithms and addresses the issues of traditional methods that primarily focus on signal changes only at the precise EPD moment. By utilizing the temporal dependencies inherent in the entire OES time series data, the proposed model not only accurately identifies the EPD moment but also captures the precursory symptoms of EPD occurrence, enhancing the reliability of the system. In addition to improving endpoint detection accuracy, the proposed method also reduced the false detection rate, which can negatively impact yield. This improvement can contribute to enhanced EPD accuracy in high-aspect-ratio etching processes. Furthermore, the proposed LSTM-based method can be applied not only to the etching process but also to various stages of the semiconductor manufacturing process, such as cleaning processes, process monitoring, and anomaly detection. This can be utilized in various fields where time

series data analysis plays a crucial role and is expected to contribute to the technological advancement of semiconductor manufacturing processes.

**Author Contributions:** Conceptualization, S.J.H., Y.J.K. and J.H.S. (Jong Hyeon Shin); methodology, S.J.H., Y.J.K. and K.H.C.; software, Y.J.K. and J.H.S. (Jong Hyeon Shin); validation, J.S.K. and J.H.S. (Jung Ho Song); formal analysis, Y.J.K. and S.J.H.; investigation, Y.J.K., J.H.S. (Jong Hyeon Shin) and S.J.H.; data curation, Y.J.K. and J.H.S. (Jong Hyeon Shin); writing—original draft preparation, Y.J.K. and S.J.H.; writing—review and editing, J.H.S. (Jung Ho Song) and S.J.H.; visualization, Y.J.K. and J.H.S. (Jung Ho Song); supervision, S.J.H. and J.S.Y.; project administration, S.J.H. and J.S.Y.; funding acquisition, S.J.H. and J.S.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** Please add: This research was funded by the National Research Council of Science & Technology, via Plasma E.I. Conversion Research Center with grant number 1711121944, CRC20014-000.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author due to the restrictions from the associated semiconductor equipment industry.

**Acknowledgments:** The authors are grateful to J.B. Kim and the technical staff at KFE for their superior support on the high-end semiconductor process equipment for the experiment.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the writing of the manuscript.

## References

1. Kihara, Y.; Tomura, M.; Sakamoto, W.; Honda, M.; Kojima, M. Beyond 10 Mm Depth Ultra-High Speed Etch Process with 84% Lower Carbon Footprint for Memory Channel Hole of 3D NAND Flash over 400 Layers. In Proceedings of the 2023 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), Kyoto, Japan, 11–16 June 2023. [CrossRef]
2. Shim, S.I.; Jang, J.; Song, J. Trends and Future Challenges of 3D NAND Flash Memory. In Proceedings of the 2023 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), Monterey, CA, USA, 21–24 May 2023. [CrossRef]
3. Zakour, S.B.; Taleb, H. Endpoint in Plasma Etch Process Using New Modified W-Multivariate Charts and Windowed Regression. *J. Ind. Eng. Int.* **2017**, *13*, 307–322. [CrossRef]
4. Kim, C.S.; Lee, H.J.; Roh, H.R.; Park, T.; Lee, Y.; Han, J.; Kwon, S.; Lee, C.; Sun, J.; Yoon, K.; et al. Improvement of Plasma Etching Endpoint Detection with Data-Driven Wavelength Selection and Gaussian Mixture Model. *IEEE Trans. Semicond. Manuf.* **2023**, *36*, 389–397. [CrossRef]
5. Choi, J.; Kim, B.; Im, S.; Yoo, G. Supervised Multivariate Kernel Density Estimation for Enhanced Plasma Etching Endpoint Detection. *IEEE Access* **2022**, *10*, 25580–25590. [CrossRef]
6. Han, Y.S.; Shin, S.H.; Park, Y.K.; Hong, S.J.; Han, S.S. Endpoint Detection in Plasma Etching Using FFT and SVM. *ECS Trans.* **2013**, *52*, 907. [CrossRef]
7. Roland, J.P.; Marcoux, P.J.; Ray, G.W.; Rankin, G.H. Endpoint Detection in Plasma Etching. *J. Vac. Sci. Technol. A* **1985**, *3*, 631–636. [CrossRef]
8. Jeon, S.-I.; Kim, S.-G.; Hong, S.; Han, S.-S. Endpoint Detection of SiO<sub>2</sub> Plasma Etching Using Expanded Hidden Markov Model. In *Advances in Neural Networks (ISNN 2010)*; Zhang, L., Lu, B.-L., Kwok, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 464–471. [CrossRef]
9. Lee, S.; Choi, H.; Kim, J.; Chae, H. Spectral Clustering Algorithm for Real-Time Endpoint Detection of Silicon Nitride Plasma Etching. *Plasma Process. Polym.* **2023**, *20*, e2200238. [CrossRef]
10. Goodlin, B.; Boning, D.S.; Sawin, H.H. Quantitative Analysis and Comparison of Endpoint Detection Based on Multiple Wavelength Analysis. In Proceedings of the 201st ECS Meeting, Philadelphia, PA, USA, 12–17 May 2002; Available online: <https://mtlsites.mit.edu/researchgroups/Metrology/PAPERS/ECS02-Endpoint-Goodlin.pdf> (accessed on 10 July 2024).
11. Lee, S.; Jang, H.; Kim, Y.; Kim, S.J.; Chae, H. Sensitivity Enhancement of SiO<sub>2</sub> Plasma Etching Endpoint Detection Using Modified Gaussian Mixture Model. *IEEE Trans. Semicond. Manuf.* **2020**, *33*, 252–257. [CrossRef]
12. Chakroun, I.; Ashby, T.J.; Das, S.; Halder, S.; Wuyts, R.; Verachttert, W. Using Unsupervised Machine Learning for Plasma Etching Endpoint Detection. In Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods ICPRAM, Valletta, Malta, 22–24 February 2020; pp. 273–279. [CrossRef]
13. Kambara, M.; Kawaguchi, S.; Lee, H.J.; Ikuse, K.; Hamaguchi, S.; Ohmori, T.; Ishikawa, K. Science-Based, Data-Driven Developments in Plasma Processing for Material Synthesis and Device-Integration Technologies. *Jpn. J. Appl. Phys.* **2022**, *62*, SA0803. [CrossRef]
14. Buda, T.S.; Caglayan, B.; Assem, H. DeepAD: A Generic Framework Based on Deep Learning for Time Series Anomaly Detection. In *Advances in Knowledge Discovery and Data Mining*; Phung, D., Tseng, V.S., Webb, G.I., Ho, B., Ganji, M., Rashidi, L., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 577–588. [CrossRef]

15. Selmy, H.A.; Mohamed, H.K.; Medhat, W. A Predictive Analytics Framework for Sensor Data Using Time Series and Deep Learning Techniques. *Neural Comput. Appl.* **2024**, *36*, 6119–6132. [[CrossRef](#)]
16. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
17. Kim, B.; Im, S.; Yoo, G. Performance Evaluation of CNN-Based End-Point Detection Using in-Situ Plasma Etching Data. *Electronics* **2020**, *10*, 49. [[CrossRef](#)]
18. Hwang, R.; Park, S.; Bin, Y.; Hwang, H.J. Anomaly Detection in Time Series Data and Its Application to Semiconductor Manufacturing. *IEEE Access* **2023**, *11*, 130483–130490. [[CrossRef](#)]
19. Sherstinsky, A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *Phys. Nonlinear Phenom.* **2020**, *404*, 132306. [[CrossRef](#)]
20. Shiri, F.M.; Perumal, T.; Mustapha, N.; Mohamed, R. A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU. *arXiv* **2023**, arXiv:2305.17473. [[CrossRef](#)]
21. Liu, Y.; Ni, D.; Wang, Z. A Fault-Tolerant Soft Sensor Algorithm Based on Long Short-Term Memory Network for Uneven Batch Process. *Processes* **2024**, *12*, 495. [[CrossRef](#)]
22. Lee, K.B.; Kim, C.O. Recurrent Feature-Incorporated Convolutional Neural Network for Virtual Metrology of the Chemical Mechanical Planarization Process. *J. Intell. Manuf.* **2020**, *31*, 73–86. [[CrossRef](#)]
23. Abbasimehr, H.; Paki, R. Improving Time Series Forecasting Using LSTM and Attention Models. *J. Ambient Intell. Humaniz. Comput.* **2022**, *13*, 673–691. [[CrossRef](#)]
24. Hu, J.; Wang, X.; Zhang, Y.; Zhang, D.; Zhang, M.; Xue, J. Time Series Prediction Method Based on Variant LSTM Recurrent Neural Network. *Neural Process. Lett.* **2020**, *52*, 1485–1500. [[CrossRef](#)]
25. Shahi, T.B.; Shrestha, A.; Neupane, A.; Guo, W. Stock Price Forecasting with Deep Learning: A Comparative Study. *Mathematics* **2020**, *8*, 1441. [[CrossRef](#)]
26. 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](#)]
27. Wang, X.; Zhao, Y.; Pourpanah, F. Recent Advances in Deep Learning. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 747–750. [[CrossRef](#)]
28. Lu, Y.; Salem, F.M. Simplified Gating in Long Short-Term Memory (LSTM) Recurrent Neural Networks. In Proceedings of the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), Boston, MA, USA, 6–9 August 2017; pp. 1601–1604. [[CrossRef](#)]
29. Fang, W.; Chen, Y.; Xue, Q. Survey on Research of RNN-Based Spatio-Temporal Sequence Prediction Algorithms. *J. Big Data* **2021**, *3*, 97–110. [[CrossRef](#)]
30. Kardakis, S.; Perikos, I.; Grivokostopoulou, F.; Hatzilygeroudis, I. Examining Attention Mechanisms in Deep Learning Models for Sentiment Analysis. *Appl. Sci.* **2021**, *11*, 3883. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.