

Received February 11, 2022, accepted February 25, 2022, date of publication February 28, 2022, date of current version March 10, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3155513

# Supervised Multivariate Kernel Density Estimation for Enhanced Plasma Etching Endpoint Detection

JUNGYU CHOI<sup>1</sup>, BOBAE KIM<sup>1</sup>, SUNGBIN IM<sup>2</sup>, (Member, IEEE), AND GEONWOOK YOO<sup>2</sup>

<sup>1</sup>Department of Information and Communication Engineering, Soongsil University, Seoul 06978, South Korea

<sup>2</sup>School of Electronic Engineering, Soongsil University, Seoul 06978, South Korea

Corresponding author: Sungbin Im (sbi@ssu.ac.kr)

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) funded by the Korean Ministry of Science and ICT (MSIT) by the Development of AI-Based Four-Channel Optical Spectrum Analysis Platform for Real-Time Process Diagnostics of Semiconductor Etching Equipment under Grant 2020-0-02102.

**ABSTRACT** The advancement of semiconductor technology nodes requires precise control of their manufacturing process, including plasma etching, which is highly important in terms of the yield, cost, and device performance. Endpoint detection (EPD) is an imperative technique for controlling this process. Here, we propose a novel EPD scheme based on multivariate kernel density estimation (MKDE). The proposed approach is developed by extending the conventional unsupervised learning MKDE method to supervised learning. The performance of the proposed scheme is validated on randomly selected optical emission spectroscopy data collected from an industrial semiconductor manufacturing process. Because the proposed approach uses target values (labeling) of data, it demonstrates enhanced EPD performance compared to the conventional MKDE method, even without threshold presetting.

**INDEX TERMS** Multivariate kernel density estimation, plasma etching, endpoint detection, anomaly detection, supervised learning, semiconductor manufacturing.

## I. INTRODUCTION

Recently, with the improvement in computer functions and computational speed, artificial intelligence (AI)-based methods have been studied for detecting semiconductor plasma etching endpoints. This is because the existing etching endpoint detection (EPD) method depends on the experience of the engineers, and therefore, causes manufacturing process changes and errors. These changes and mistakes degrade the semiconductor manufacturing yield. The fabrication yield is a most critical indicator determining the success of the semiconductor manufacturing industry, and it is obtained as the ratio of the total number of wafers produced by the entire fabrication and evaluation process to the initial number of wafers. Therefore, various studies have been conducted to enhance the fabrication yield and address related problems in the semiconductor manufacturing process. AI has been adopted to improve the fabrication yield and conduct predictive maintenance of semiconductor equipment [1], [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Tong<sup>1</sup>.

Support vector machines [3] have been employed to classify low- and high-yield classes, and a polynomial neural network has been applied to model and control the plasma etching process [4]. Deep learning techniques have also been studied for virtual metrology to estimate quantities that are key to production but are expensive or unmeasurable, and to evaluate process quality [5].

In the semiconductor process, the plasma etching process is important, with the process nodes continuing to becoming small and shift toward new complex architectures [6], [7]. A vital part of the plasma etching process is the selective removal of thin layers of materials without damaging the layer beneath. Consequently, the plasma etching scheme for extreme ultraviolet patterning is challenging, and optimal EPD during the etching process is crucial for minimizing the process variations and errors.

Although various approaches have been proposed for EPD to control plasma etching, the most commonly used technique is monitoring the in situ optical emission spectra collected by optical emission spectroscopy (OES) during the process [8]–[11]. OES is an excellent technique for monitoring the

plasma emission intensity in plasma etching. However, the main problem is that large amounts of multidimensional spectroscopic data are collected owing to the measurement of 2,048 wavelength-specific intensity values over time. Therefore, research on dimensionality reduction of OES data using principal component analysis [12] and autoencoder neural networks as techniques for feature extraction has also been conducted [13]–[17]. In addition, a maximum separation clustering development study has been conducted to cluster variables with unique patterns and provide practical pattern expression using a few representative variables [18].

Multivariate analysis techniques have been used in the etching process to increase sensitivity [19], [20]. An investigation using K-means cluster analysis was attempted to improve the detection sensitivity of plasma etching endpoints [21]. K-means clustering is a nonhierarchical and centroid-based approach. It has been used to detect etching endpoints in real time and increase the sensitivity of optical emission signals. However, the K-means cluster analysis is based on the Euclidean distance, which has a limitation in that it does ignore the covariance of the cluster. Therefore, Gaussian mixture model-based plasma EPD has also been proposed for real-time monitoring [22]. A convolutional neural network (CNN) has been applied to detect endpoints using OES data [23]. Because OES data have specific patterns related to wavelengths, this study exploits the capability of a CNN model to recognize specific two-dimensional patterns, similar to image detection.

In this paper, we propose a novel plasma etching EPD approach based on multivariate kernel density estimation (MKDE). MKDE extends KDE to a multidimensional random vector, and KDE is a nonparametric statistical modeling method that generates a statistical model using only the given data, without using a parametric probability density function (PDF). KDE does not require a statistical moment or a specific PDF to estimate the probability distribution. It yields the kernel density function by combining the kernel functions provided by all feature vectors. Because KDE relies only on data, it is practical when the parametric PDF cannot represent the given data distribution.

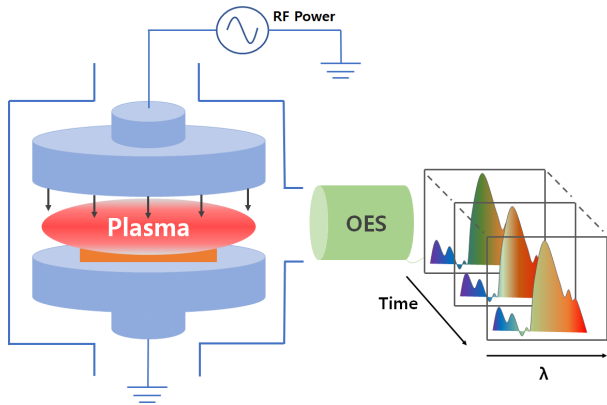
KDE is a technique specialized for classifying a small percentage of data by unsupervised learning and is commonly used for anomaly detection and data classification. Related applications include survey of motor fault detection and diagnosis by motor current characteristic analysis [24] and efficient automatic modulation classification for a modulated signal group [25].

Existing detection techniques rely extensively on unsupervised learning. However, the obtained results do not typically meet the criteria, and could be significantly improved by the supervision of some labeled features. Thus, important detection that conventional unsupervised learning cannot achieve is possible by the above approach [26]. Similarly, many algorithms improve clustering quality using labeled data pairs in unsupervised clustering. The addition of labeled

examples can improve the performance of anomaly detection. The detection rate of an algorithm can be improved and its false-positive rate can be reduced by further development of the model [27], [28]. In addition, there are infinite unlabeled data in real engineering problems, and in some cases, labels may be unavailable for all data. Therefore, learning by adding labeling to existing unsupervised learning is advantageous to integrate labeled and unlabeled data into the learning process. Specifically, exploiting data characteristics is an interesting and essential alternative for engineering classification problems [29].

Owing to the characteristics of the semiconductor etching process, OES data are composed chiefly of signals generated during the etching process, which present stationary attributes over time. However, when the etching process approaches its endpoint, as exhibited by the rapid changes in the signal characteristics, it is stopped. Therefore, the number of signal samples near the etching endpoint is relatively very small, and observation of a signal near it can be considered as an anomaly. Therefore, MKDE can be employed to detect the etching endpoint. However, the application of KDE requires a certain threshold to determine whether an input feature belongs to the category of endpoint features. Advantageously, datasets obtained from the semiconductor manufacturing process typically contain endpoint information because real-world data generally retain the corresponding information from wafer processing. In this study, this provides a basis for detecting an etching endpoint by applying MKDE twice, without presetting the threshold. In the learning phase for MKDE, previously collected OES data can be used to develop two MKDE models, one each for nonendpoint and endpoint sections. In the learning phase using such data, the non-EPD section data for MKDE are employed for non-EPD, whereas the PDF for EPD is estimated using relatively fewer data labeled with EPD times. Therefore, the algorithm developed in this study belongs to the class of supervised learning. Specifically, the primary concept is to estimate the overall PDF by applying MKDE to a given OES dataset for non-EPD and relatively fewer OES samples near the EPD time to construct the second PDF at that moment. By comparing the magnitudes of these two PDFs, the probabilities of the features from the OES data are computed to determine whether the input feature is for EPD. The final determination is made when the decision for each input feature is repeated several times in a row, to obtain the final EPD feature. The advantage of the proposed approach is that advanced setting of the threshold is not required.

There are several contributions of this study. The PDFs of the non-EPD and EPD sections can be obtained during the semiconductor plasma etching process by extending the conventional MKDE method to a supervised learning version. The plasma etching endpoint can be detected without a threshold using the proposed MKDE, which has a simpler structure than a general deep-learning model. Moreover, it has the advantage of obtaining the probability for the etching endpoint by obtained the PDFs with a small amount of data.



**FIGURE 1.** Schematic of plasma etcher equipped with optical emission spectrometer and fluctuation of obtained multiwavelength OES spectra with time.

Finally, the proposed approach outperforms the conventional unsupervised MKDE.

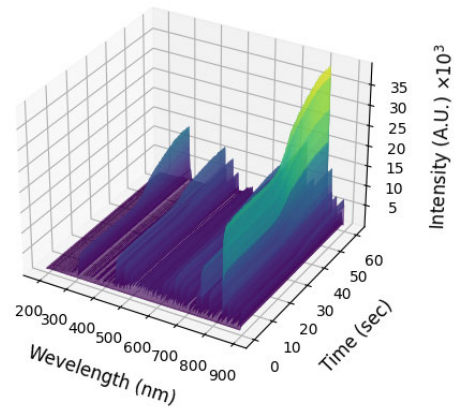
The remainder of this paper is organized as follows. Section II describes the OES data used in this study. Section III briefly explains the basic concept of MKDE, introduces the proposed approach, and describes the preprocessing, feature selection, and flag setting steps. Section IV discusses the experimental procedure and results, and Section V presents the conclusions.

**II. OES DATA**

In the plasma etching process, EPD is conducted by monitoring OES spectra collected during this process. A schematic of a plasma chamber equipped with an optical emission spectrometer and its multiwavelength OES data are shown in Fig. 1. The plasma etching equipment is an inductively coupled plasma reactive ion etching system with a radiofrequency (RF) power supply. The optical emission spectrometer, which is a fiber optical sensor system used to collect the plasma emission intensity, is fixed at the sidewall of the chamber using a quartz window viewport. Under low pressure, a reactive plasma containing perfluorocarbons is generated by RF power, bombards the wafer surface, and reacts with the targeted materials. Accordingly, the reactants and by-products of the etching cause fluctuations in the optical emission spectra at a particular time. The obtained OES data depend on the target materials. Furthermore, the size of the features to be etched degrades the signal-to-noise ratio of the OES data [30]. The OES measurement is conducted conveniently without intervention while providing reliable real-time information on the etching process.

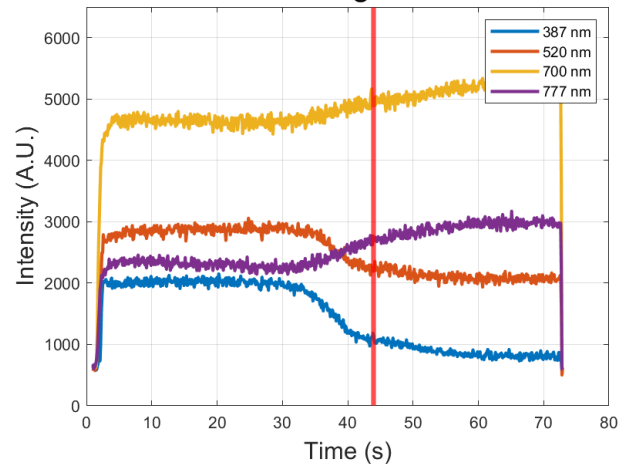
In general, OES data are vast and multidimensional, being functions of the wavelength, time, and intensity. However, high-resolution data are required to achieve the desired sensitivity and accuracy for EPD as the feature size decreases, because EPD is realized by monitoring the shift of the emission peak. Fig. 2 shows a three-dimensional plot of actual OES spectra used in this study. The spectra consist

**3D Plot of Observed OES Data**



**FIGURE 2.** Three-dimensional plot of OES spectrum samples.

**Selected Wavelengths of OES Data**

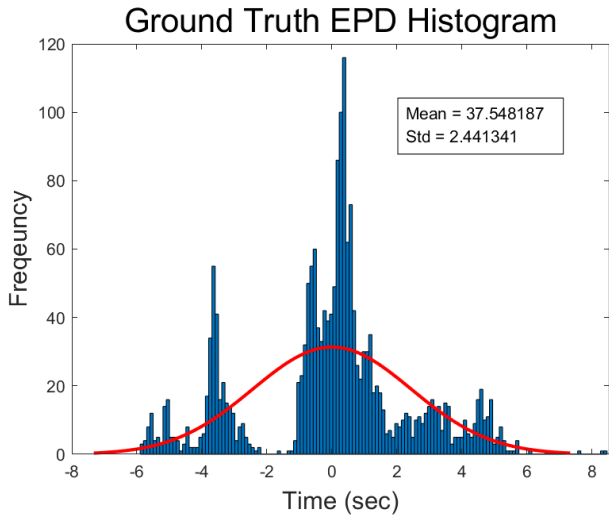


**FIGURE 3.** Intensity waveforms at wavelengths of 387 nm, 520 nm, 700 nm, and 777 nm, which are related to CN, CO, F, and SiF over time. Red vertical line denotes ground-truth EPD time.

of 2048 components ranging from 190.0 to 892.8 nm, and the sampling rate is 0.1 s for approximately 60 s.

In this study, the molecular species of CN, CO, F, and SiF are selected for breaking silicon dioxide covalent bonding using perfluorocarbon gases, thus forming volatile byproducts. The emission peak lines corresponding to 387 nm, 520 nm, 700 nm, and 777 nm of the OES signal are employed for EPD. It is well known that these wavelengths reflect well the characteristics of an etching endpoint section and that their intensity fluctuations show the changes in the chamber plasma during the etching process [31], [32]. Fig. 3 presents exemplary waveforms of the four wavelengths used as features.

The ground-truth EPD time is chosen using the sensor cluster manager toolbox software, SCM<sup>TM</sup> (Prime Solution Co., LTD), and process engineers. The EPD times are marked on the OES datasets by verifying the processed 1,911 wafers in



**FIGURE 4.** Ground-truth EPD time histogram with mean of 37.55 s and standard deviation of 2.44 s.

the semiconductor manufacturing process. The results show that the ground-truth EPD times are acceptable, even if they may have some errors. Fig. 4 shows a histogram of the ground-truth EPD times, with a mean of 37.55 s from the etching start time and a standard deviation of 2.44 s.

### III. PROPOSED APPROACH

#### A. MULTIVARIATE KERNEL DENSITY ESTIMATION

MKDE is a nonparametric statistical modeling method that estimates a PDF using only data, without defining any information on a specific PDF and correlation. MKDE is a technique with a very high degree of freedom in the expression of the distribution function because its shape is not fixed and can be flexibly expressed according to the data. It defines the same kernel function for all data components. When all kernel functions are summed according to the optimal bandwidth, it yields the joint PDF estimate [33]. Specifically, the multivariate kernel density estimator is a function of the estimated probability density of a random vector [34].

Let  $\vec{x} = (x_1, x_2, \dots, x_d)^T$  be a  $d$ -dimensional random vector with density  $f$ , and let  $\vec{y}_i = (y_{i1}, y_{i2}, \dots, y_{id})^T$  denote the independent random samples extracted from  $f$ . The typical form of the multivariate kernel density estimator for a real vector of  $\vec{x}$  is

$$\hat{f}_{\mathbf{H}}(\vec{x}) = \frac{1}{n} \sum_{i=1}^n K_{\mathbf{H}}(\vec{x} - \vec{y}_i), \quad (1)$$

where  $n$  is the number of random samples,  $K_{\mathbf{H}}$  is a scaled kernel function with kernel function  $K$  and defined by  $K_{\mathbf{H}}(\vec{z}) = |\mathbf{H}|^{-\frac{1}{2}} K(\mathbf{H}^{-\frac{1}{2}} \vec{z})$ , and  $\mathbf{H}$  is a  $d \times d$  bandwidth matrix, which is positive definite and symmetric [35].

There are several options for the bandwidth matrix [36]–[38]. However, in this study, Silverman’s rule [38], which is commonly applied for MKDE, is adopted. Note that this rule approximates the optimal bandwidth in terms of the

mean integrated squared error for the Gaussian random variables. Let the bandwidth matrix,  $\mathbf{H}$ , be diagonal as follows:

$$\mathbf{H} = \text{diagonal}(h_1, h_2, \dots, h_d). \quad (2)$$

The component,  $h_\ell$ , for  $\ell = 1, 2, \dots, d$  is expressed as

$$h_\ell = \sigma_\ell \left\{ \frac{4}{(d+2)n} \right\}^{1/(d+4)} \quad (3)$$

where  $\sigma_\ell$  is the standard deviation of the  $\ell$ -th component of the feature vectors,  $\vec{y}_i$ , for  $i = 1, 2, \dots, n$ . Note that in practice, the sample estimator substitutes  $\sigma_\ell$ . Equation ((3)) is used in practice, even if most of the data are not Gaussian.

The kernel function,  $K$ , is non-negative, real-valued, and integrable and satisfies the following prerequisites:  $\int_{-\infty}^{+\infty} K(u) du = 1$  and  $K(-u) = K(u)$  for any  $u$ . Many types of kernels  $K(\cdot)$  can be found in the relevant literature, and common symmetric kernels are Gaussian, uniform, Epanechnikov, triangular, biweight, and triweight types [39], [40]. The shape of  $K(\cdot)$  has little influence on the estimator shape [33], [40], [41]. This is crucial because the smoothing parameter,  $\mathbf{H}$ , determines the degree of smoothing. When  $\mathbf{H}$  is extremely small, the estimator shows insignificant details. An extremely large  $\mathbf{H}$  causes oversmoothing of the information in the samples, which in effect may hide some of the essential characteristics. Therefore, a compromise is required.

In general, the semiconductor manufacturing process has the following unique characteristics. Although most of the data in the etching process of a wafer are in the non-EPD state, the etching endpoint appears for a relatively short time after the etching process is completed. In addition, one etching endpoint should necessarily result in a successfully etched wafer by the above process. For example, when EPD occurs in the 300th OES sample on a wafer, 299 features emerge from the etching process and one etching endpoint. Compared with the features during the etching process, the features near the etching endpoint can be regarded as anomalies, with a relatively low occurrence frequency. Therefore, in this study, the etching endpoint of the semiconductor manufacturing process is detected in a supervised manner by applying MKDE, the existing unsupervised method, twice using the above aspects.

Existing KDE-based anomaly detection requires a threshold value. Finding the optimal threshold needs a tuning process based on known data. Furthermore, because the bandwidth of the kernel has a significant influence on KDE performance, the optimal bandwidth is selected by trial and error using available data.

#### B. DEVELOPMENT OF PROPOSED APPROACH

MKDE is known as unsupervised learning, which requires no labels for training data and creates a PDF using the entire dataset. Therefore, it produces a low probability for a sample with a small proportion of data, which is detected as an anomaly. However, in practice, OES data from the semiconductor etching process are composed of non-endpoint and endpoint sections. The non-endpoint section is from the beginning of the process before the endpoint, whereas

the endpoint section is located immediately after the non-endpoint section. In general, the endpoint section contains only a few samples, and thus, is much shorter than the non-endpoint section. Therefore, the OES data obtained from the practical manufacturing process contain ground truths, which are actual EPD time values. Thus, these EPD times can be exploited in experiments. In this approach, two kernel densities are estimated. The first kernel density is assessed using the data for the non-endpoint section, whereas the second one is obtained using the features of the endpoint section.

Fig. 5 illustrates a flowchart of the MKDE-based EPD algorithm developed for the semiconductor etching process. For the learning phase, the training data are utilized for estimating the two kernel densities for the non-endpoint and endpoint sections, which are denoted by “KDE (Non-EPD)” and “KDE (EPD)” in the figure, respectively. When the algorithm operation starts, the moving average filter is applied in the preprocessing stage to reduce the noise before extracting the features of the OES data considering the measurement noise contamination. Subsequently, only the wavelengths that are very sensitive to the change in the etching process are selected and used as features, as mentioned in the previous section. Following this, the feature,  $\vec{x}$ , of each time sample is applied to Models 1 and 2 for obtaining the probabilities,  $f_1(\vec{x})$  and  $f_2(\vec{x})$ . Probabilities  $f_2(\vec{x})$  and  $f_1(\vec{x})$  represent the probabilities of a feature being and not being an endpoint, respectively. A flag is set for each time sample. When  $f_1(\vec{x}) \geq f_2(\vec{x})$ , the flag of a time sample is set as zero; otherwise, it is set as one. The final decision for the etching endpoint is made when the number of flags exceeds the preset threshold, which is determined from experiments in terms of the optimal performance.

The MKDE-based approach employed in this study avoids these two difficulties. The existing MKDE requires threshold value setting, whereas the proposed scheme does not need to set the threshold value by applying the second MKDE. Because the conventional MKDE is a non-parametric density estimation scheme, the MKDE-based approach employed in this study can overcome the problem of model assumptions. This approach exploits both labelled and unlabelled samples. It also has lower complexity compared to CNNs [39].

### C. PREPROCESSING

In this study, the proposed algorithm uses the wavelengths of 387 nm, 520 nm, 700 nm, and 777 nm for EPD. As previously mentioned, these wavelengths reflect well the characteristics of the etching endpoint. Example waveforms of the wavelengths used as features are shown in Fig. 3. The intensity values observed by OES may vary depending on the sensor setting and environment. Therefore, preprocessing is essential to complement the performance of the KDE model. In this step, moving average filtering and normalization to a specific reference sample on the time axis are performed to mitigate the measurement noise and remove the characteristics dependent on the value of the waveform, respectively.

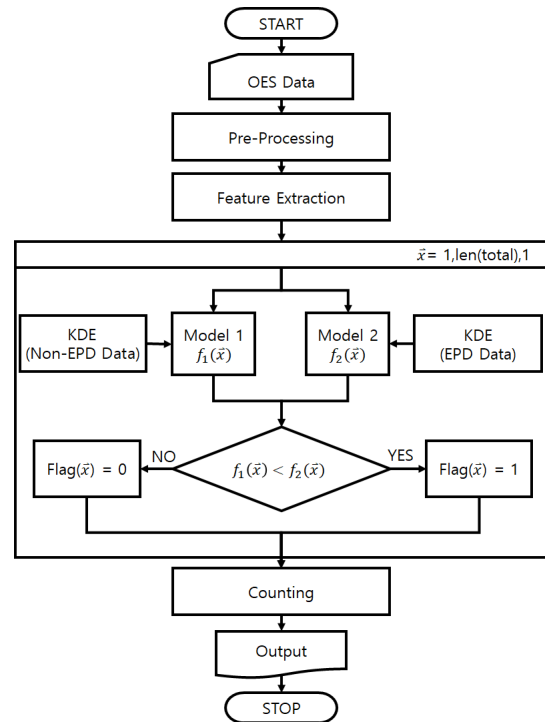


FIGURE 5. Flowchart of MKDE-based etching endpoint detection.

First, to mitigate the effect of measurement noise, a moving average filter of ten samples (1 s) is applied to the data. Note that KDE estimates the probability density based on the given dataset, whereas the etching endpoint is detected by the change in the intensity value, not by the simple intensity value. If the intensity value varies, the etching endpoint cannot be correctly detected by observing only the value. Therefore, it is necessary to normalize  $x(k)$  and use the normalized version,  $x_n(k)$ , as a feature component. The normalized version,  $x_n(k)$ , at time  $k_s$  is given by  $x_n(k) = x(k)/x(k_s)$ ,  $k = 1, 2, \dots, n$ , where  $k_s$  is the start time when the OES equipment normally operates and  $n$  is the length of the OES data. The moving average filter output waveforms shown in Fig. 3 are presented in Fig. 6, and the normalized waveforms are shown in Fig. 7.

### D. FLAG SETTING

For each feature vector  $\vec{x}$ , the two KDE models generate the non-EPD and EPD probabilities of  $f_1(\vec{x})$  and  $f_2(\vec{x})$ , respectively. When  $f_1(\vec{x}) > f_2(\vec{x})$ , the flag is set as zero, implying the feature is EPD; otherwise, it is one, implying the feature is non-EPD. Finally, when the number of repetitions exceeds the preset number, the EPD feature is determined. This preset number is also optimized in terms of performance. Note that a feature with a flag of one is not the etching endpoint matching the ground truth; however, it is nearer to an endpoint section than to a non-endpoint section. Consequently, the final EPD feature is determined based on the number of flag repetitions.

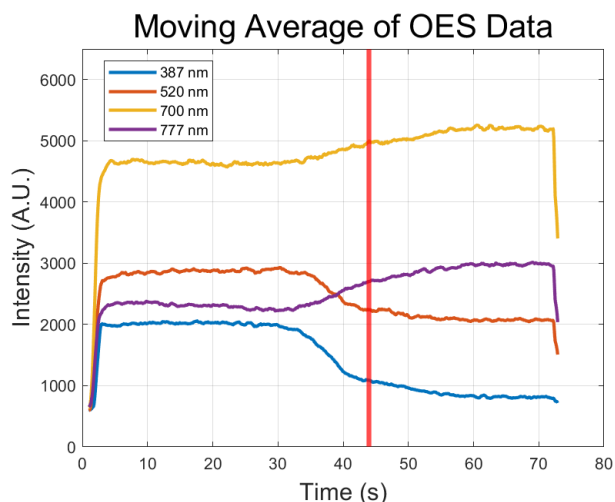


FIGURE 6. Waveform of moving average filter with ten samples averaging for four wavelengths.

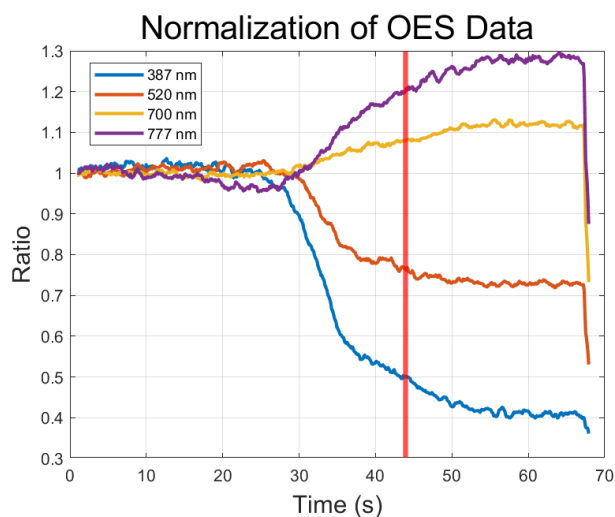


FIGURE 7. Normalized waveforms of moving average filter output with  $k_s = 50$ .

Fig. 8 shows an example of the result of the entire etching process of an arbitrary wafer. Probabilities  $f_1$  (red) and  $f_2$  (blue) presented in Fig. 8 are obtained using the non-EPD and EPD models for features that occur in 0.1-s units during the entire etching process, respectively. For each feature, the flag is set by evaluating the relative relationship between  $f_1$  and  $f_2$ . Fig. 8 (b) presents an enlarged image of Fig. 8 from the 200th to the 400th sample. The black vertical line is the ground-truth EPD, and the pink dotted line is the EPD estimated by the proposed method. As we will present subsequently, the estimated EPD position is where one flag is continuously repeated 16 times. In this case, the difference between the ground truth and the estimate is five samples.

#### IV. EXPERIMENT

This section presents the performance of the proposed approach using OES data, as previously explained. First,

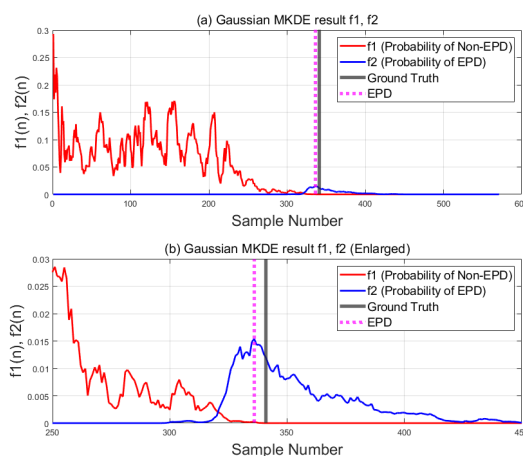


FIGURE 8. Examples of non-EPD and EPD probabilities  $f_1$  and  $f_2$ , according to sample number when using Gaussian kernel. Note that bottom figure is enlarged image from 200th to 400th sample.

the computational environment and dataset to be used are described. The evaluation metrics used in the experiment for performance investigation are briefly introduced. First, the performance of the proposed approach is compared with that of the conventional MKDE scheme for a Gaussian kernel. Subsequently, the kernel among Gaussian, Epanechnikov, triangle, and box kernels that provides the best performance is investigated. Furthermore, the effect of bandwidth variation is demonstrated in terms of the evaluation metrics. Finally, the statistical distribution of the prediction error is presented. Note that the experiment results in this section are obtained by averaging over ten trials.

The proposed and conventional models are developed using MATLAB 2020a, and the computing environment is a 32-core 3.69-GHz CPU, 64-GB RAM, and an RTX 2080 Super GPU.

#### A. DETAILS OF DATASET

OES dataset obtained from 1911 wafers is used in this experiment. A total of 1300 wafers are randomly selected to create the non-EPD and EPD MKDE models. The total number of features used to construct the non-EPD MKDE model is 280,323. Concurrently, for the EPD MKDE model, 1300 feature vectors are selected at the EPD time, and two additional feature vectors before the EPD time and two after the EPD time are taken for each wafer. Thus, five features are chosen from a wafer, and the total number of features for the EPD KDE model is 6500. Note that the EPD time obtained during the etching process is an experimentally determined ground truth, which may have errors. Tests are performed using the data obtained from the remaining 611 wafers. The results presented in the following subsections are obtained by experimenting with this process ten times.

#### B. PERFORMANCE MEASURE

In the experiments, performance evaluation of the proposed approach is in terms of the accuracy, receiver operating

characteristic (ROC) curve, and  $F_1$ -score. Accuracy is an indicator of the similarity between predicted data and actual data. However, an imbalanced distribution of label values, such as in this study, can distort the performance evaluation. An ROC curve is commonly known to be effective for imbalanced datasets. However, to construct an ROC curve, the threshold value needs to be varied and the classifier outputs should be real-valued, and not labels such as 0 and 1. Therefore, the ROC curve is inapplicable to the proposed approach, which produces a label depending on the comparison result of the outputs of the two models. In this experiment, the ROC curve is applied only to the conventional MKDE scheme to determine the threshold for the optimal performance. The  $F_1$ -score is also known to be effective for imbalanced datasets and is bounded between 0 and 1. An  $F_1$ -score close to 1 represents good performance, whereas a zero  $F_1$ -score fails in classification [42].

For binary classification, accuracy can be defined as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

where  $TP$ ,  $TN$ ,  $FP$ , and  $FN$  stand for “true positive,” “true negative,” “false positive,” and “false negatives,” respectively. The ROC curve shows the variation in the true positive rate (TPR) with the false positive rate (FPR,  $1 - TNR$ , where TNR denotes true negative rate). The TPR is occasionally called *Recall* and represents the number of predicted positives among all real positive cases. The TNR represents the level at which the actual negative value should be accurately predicted. The TPR and the FPR are expressed as follows:

$$TPR = \frac{TP}{TP + FN} = Recall \quad (5)$$

and

$$FPR = 1 - \frac{TN}{FP + TN} = \frac{FP}{FP + TN} \quad (6)$$

The FPR should also be considered in the detection performance. An ROC curve can rate the goodness-of-fit of a detection approach, reflecting the FPR [43], [44]. An ROC curve presents the model performance in a two-dimensional form. Hence, an intuitive decision is difficult based on performance comparison. Therefore, a simple approach for showing performance is necessary. The area under the ROC curve (AUC) is a number between zero and one representing the unit square area under the ROC curve, and its computation is simple [45].

The  $F_1$ -score is the harmonic mean of both the *Precision* and  $TPR$  (*Recall*), which is given by

$$F_1 = \frac{2}{\frac{1}{Precision} + \frac{1}{TPR}} = 2 \times \frac{Precision \times TPR}{Precision + TPR}, \quad (7)$$

where  $Precision = \frac{TP}{TP+FP}$ . *Precision* indicates the number of positives for all positive predictions. Note that the  $F_1$ -score accounts for both the false positives and false negatives by the *Precision* and the  $TPR$  (*Recall*). To obtain a higher  $F_1$ -score, both the *Precision* and  $TPR$  need to be increased in terms of the harmonic mean.

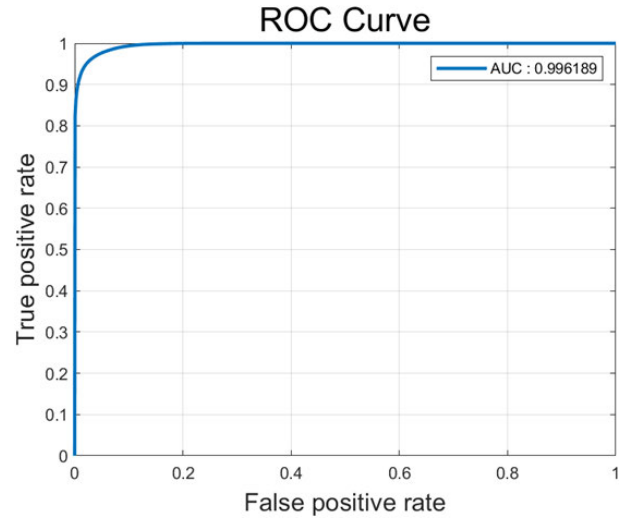


FIGURE 9. ROC curve of conventional MKDE approach with Gaussian kernel.

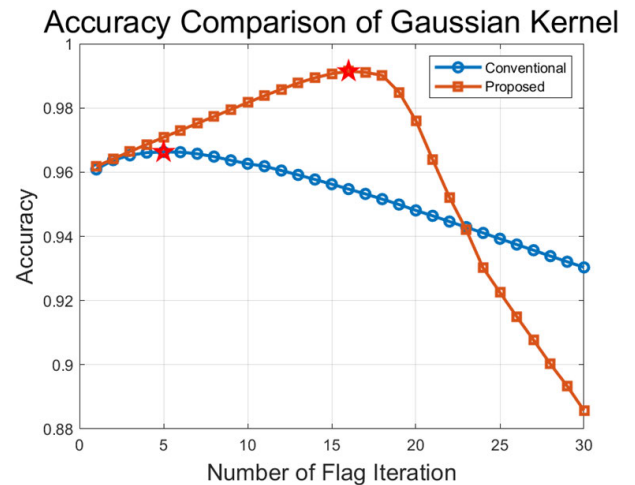


FIGURE 10. Accuracy curves of conventional and proposed MKDE approaches with Gaussian kernel versus number of flag repetitions. Optimal threshold value is applied to conventional approach curve.

### C. PERFORMANCE COMPARISON OF CONVENTIONAL AND PROPOSED APPROACHES

The performance of the proposed approach is evaluated and compared with that of the conventional MKDE approach under the same conditions employing the most commonly used Gaussian kernel. Bandwidth  $H$  is computed using Silverman’s rule (3). For the conventional MKDE approach, the optimal threshold is estimated using the ROC curve shown in Fig. 9, whose AUC is 0.9962. Note that the proposed approach does not require any threshold value because it compares the two outputs from the EPD and non-EPD models for binary classification, as mentioned in the previous section. The optimal operating point on the ROC curve is obtained as  $FPR = 0.0285$  and  $TPR = 0.9595$  when the threshold is 0.9997. Using this threshold, the accuracy curves of the conventional and proposed schemes versus the number of flag

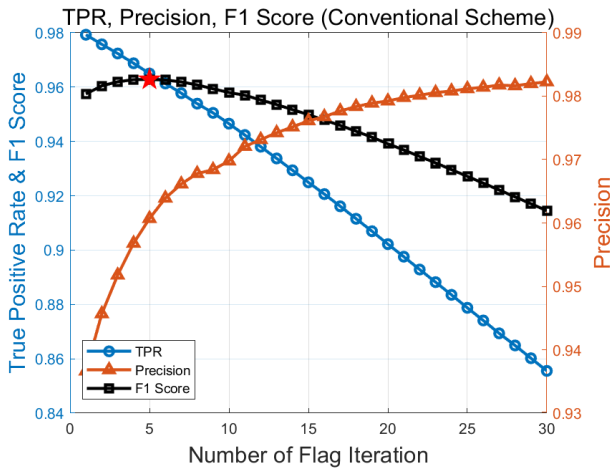


FIGURE 11. TPR, precision, and  $F_1$ -score curves of conventional approach versus number of flag repetitions.

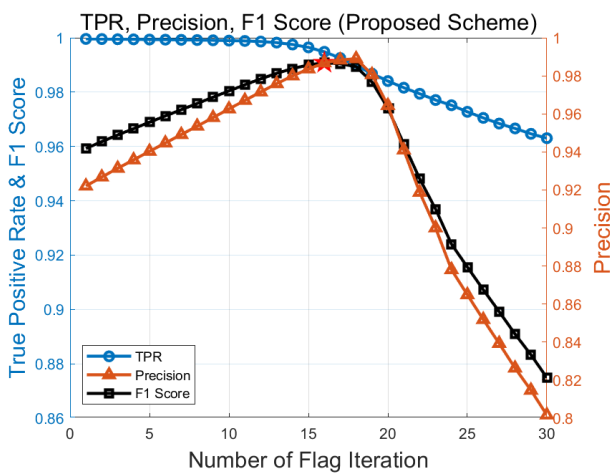


FIGURE 12. TPR, precision, and  $F_1$ -score curves of proposed approach with Gaussian kernel versus the number of flag repetitions.

iterations are displayed in Fig. 10. The conventional method achieves a maximum accuracy of 0.9663 at five iterations, whereas the proposed method realized a maximum accuracy of 0.9914 at 16 iterations. The TPR, FPR, and  $F_1$ -score curves are shown in Figs. 11 and 12 for the conventional and proposed methods, respectively. The conventional scheme achieves a maximum  $F_1$ -score of 0.9628, whereas the proposed scheme realizes a maximum of 0.9908. In terms of the accuracy and the  $F_1$ -score, the proposed approach is superior to the conventional approach, which is based on supervised learning.

**D. KERNEL TYPE**

The performance of the proposed approach is also investigated according to the kernel type: Gaussian, Epanechnikov, triangle, and box kernels. Fig. 13 shows the four accuracy curves of the proposed approach versus the number of flag repetitions obtained using the Gaussian, Epanechnikov, triangular, and box kernels. The Gaussian kernel is the best

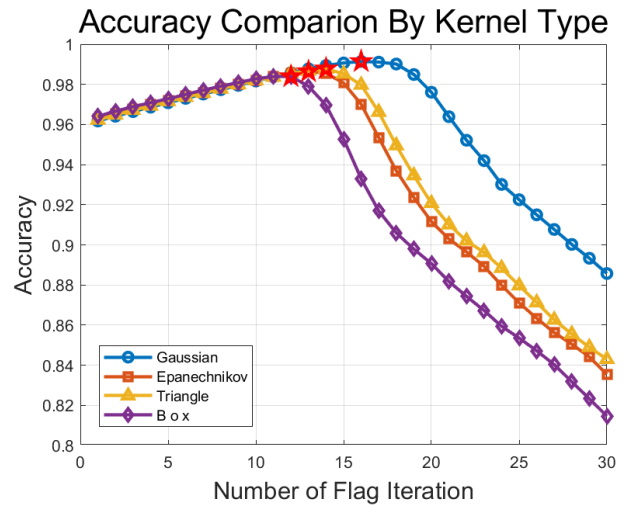


FIGURE 13. Accuracy curves of proposed approach with Gaussian, Epanechnikov, triangle, and box kernels versus number of flag repetitions.

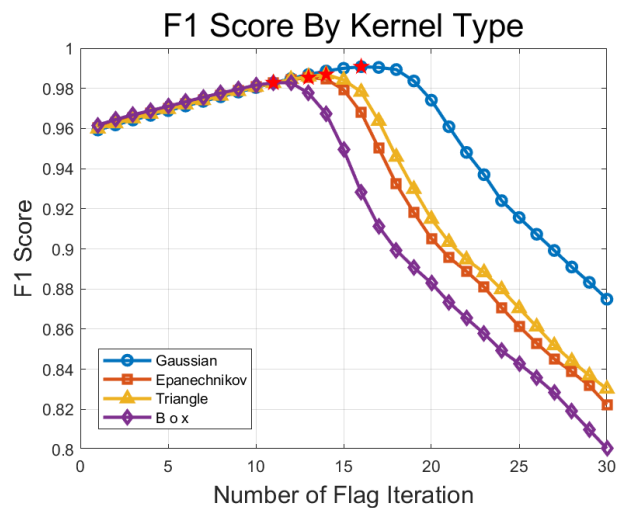


FIGURE 14.  $F_1$ -score curves of proposed approach with Gaussian, Epanechnikov, triangle, and box kernels versus number of flag repetitions.

kernel, followed by the triangle kernel, and the Epanechnikov and box kernels are the worst ones. The  $F_1$ scores of the proposed approach obtained using the four kernels are shown in Fig. 14, in which results similar to those in terms of accuracy are observed. The performance details are summarized in Table 1. As mentioned earlier, the proposed scheme with the Gaussian kernel demonstrates the best performance in terms of both the accuracy and  $F_1$ -score.

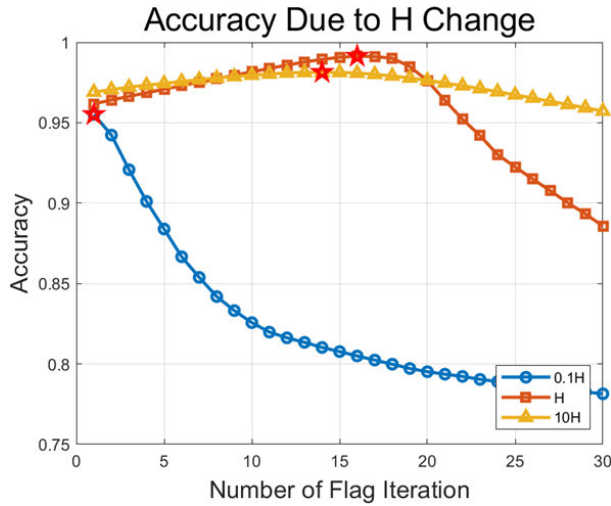
**E. BANDWIDTH H**

Subsequently, the performance according to the bandwidth is investigated. The bandwidth,  $H$ , considered here is computed using Silverman’s rule, and the detection performance is examined when the bandwidth is changed by 0.1 and 10 times. Fig. 15 shows the change in the accuracy

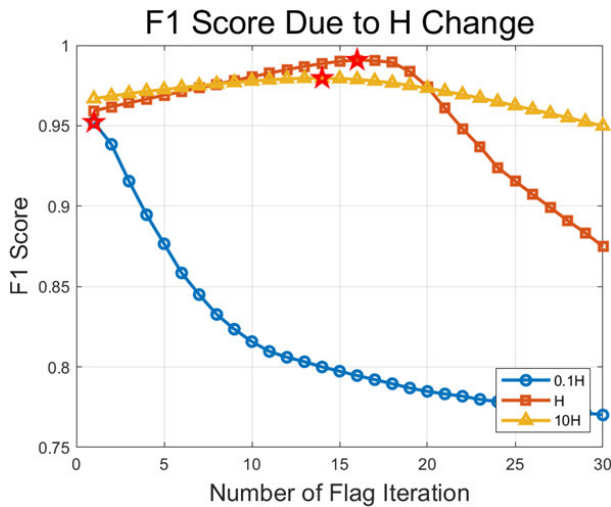


**TABLE 1.** Summary of accuracy and  $F_1$ -score of proposed approach. Note that last row corresponds to those of conventional scheme.

| Kernel Type       | Accuracy (No. of flags) | $F_1$ -score (No. of flags) |
|-------------------|-------------------------|-----------------------------|
| Gaussian          | 0.991424584 (16)        | 0.990804344 (16)            |
| Epanechnikov      | 0.986367966 (13)        | 0.985424941 (13)            |
| Triangle          | 0.987783129 (14)        | 0.986931227 (14)            |
| Box               | 0.98392009 (12)         | 0.982817472 (11)            |
| Conventional MKDE | 0.966296756 (5)         | 0.962840343 (5)             |

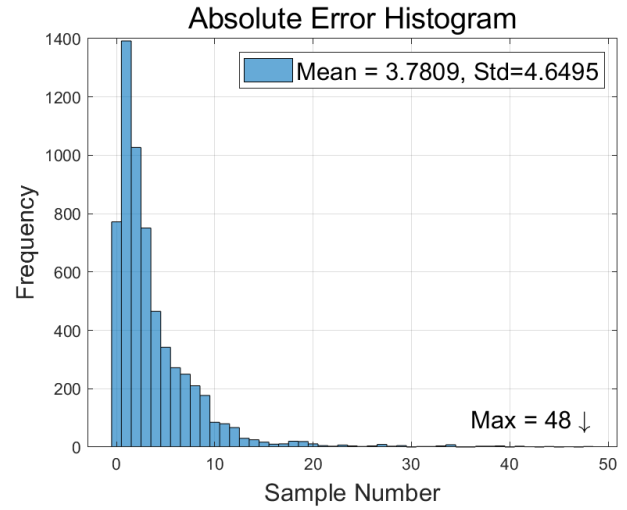


**FIGURE 15.** Accuracy curves of proposed approach with Gaussian kernel versus number of flag repetitions for three bandwidths: 0.1H, H, and 10H.



**FIGURE 16.**  $F_1$ -score curves of proposed approach with Gaussian kernel versus number of flag repetitions for three bandwidths: 0.1H, H, and 10H.

versus the flag repetition for three bandwidths: 0.1H, H, and 10H. When 10H is used, there is no overall change in accuracy, whereas when 0.1H is used, the accuracy decreases sharply as the number of repetitions increases. This phenomenon seems to increase the error by classifying the feature vectors with similar components as those with different components by increasing the selectivity using a small bandwidth. The  $F_1$ -score curves for the three bandwidths are



**FIGURE 17.** Histogram of absolute errors between predicted and ground truth EPD times in terms of number of samples.

shown in Fig. 16. Furthermore, the best results of accuracy and  $F_1$ -score for the three bandwidths are summarized in Table 2. The number in parentheses is the number of repetitions with the best performance. As observed from the table, the bandwidth value calculated by Silverman’s rule shows the best performance in terms of both the accuracy and  $F_1$ -score.

**TABLE 2.** Comparison of accuracy and  $F_1$ -score of proposed scheme for three bandwidths of 0.1H, H, and 10H.

| Bandwidth | Accuracy (No. of flags) | $F_1$ -score (No. of flags) |
|-----------|-------------------------|-----------------------------|
| 0.1H      | 0.955173859 (1)         | 0.952166087 (1)             |
| H         | 0.991424584 (16)        | 0.990804344 (16)            |
| 10H       | 0.981392611 (14)        | 0.979548058 (14)            |

**F. PREDICTION ERRORS**

In this subsection, the error distribution between the predicted and ground truth EPD times using the proposed approach with the Gaussian kernel is presented. Fig. 17 shows a histogram of the absolute error values, where the x-axis represents the sample number and the y-axis the number of samples. In this graph, the mean is 3.7809 and the standard deviation is 4.6495. As the sampling interval is 0.1 s, the average error time is 0.3781 s.

**V. CONCLUSION**

In this paper, we propose a modified MKDE technique for detecting the endpoint of the plasma etching process, which typically relies on experienced engineers. To obtain a better performance, we extended the existing unsupervised learning MKDE to a supervised learning approach that divides the training data into non-endpoint and etch endpoint sections to construct PDFs in both cases. The performance of the proposed approach was evaluated by the measures of average accuracy and the  $F_1$ -score on randomly selected OES data.

The results showed that the proposed method is superior to the conventional MKDE scheme in terms of both the accuracy and  $F_1$ -score. This demonstrates that semiconductor etching EPD using the proposed method can more clearly distinguish non-endpoint and etching endpoint sections than the conventional MKDE.

Furthermore, the commonly used Gaussian kernel performs better than other three conventional kernels that were considered. The bandwidth effect was also investigated using three different bandwidths. This experiment demonstrated that the performance of the proposed method is highly dependent on the bandwidth and that the bandwidth is more critical than the kernel shape.

The characteristics of the proposed approach are summarized as follows: 1) The conventional MKDE method requires a threshold, whereas the proposed method does not. 2) Because the proposed method uses target values (labeling) of data, it belongs to supervised learning schemes and performs better than the conventional MKDE, which relies on unsupervised learning. 3) Many machine-learning schemes generally require a long model training time, whereas MKDE has a relatively short training time. This is also true for the proposed method.

In the future, we will further analyze the differences in the data for different chambers and combinations used for plasma etching and supplement the accuracy and  $F_1$ -score performance by optimization. In addition, we will continue studies to improve the performance by establishing additional feature extraction methods and preprocessing schemes by data analysis by various approaches.

## ACKNOWLEDGMENT

The authors would like to thank Prime Solution Company Ltd., for providing the experimental datasets.

## REFERENCES

- [1] K. Nakata, R. Orihara, Y. Mizuoka, and K. Takagi, "A comprehensive big-data-based monitoring system for yield enhancement in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 4, pp. 339–344, Nov. 2017.
- [2] B.-S. Kang, J.-H. Lee, C.-K. Shin, S.-J. Yu, and S.-C. Park, "Hybrid machine learning system for integrated yield management in semiconductor manufacturing," *Expert Syst. Appl.*, vol. 15, no. 2, pp. 123–132, Aug. 1998.
- [3] R. Baly and H. Hajj, "Wafer classification using support vector machines," *IEEE Trans. Semicond. Manuf.*, vol. 25, no. 3, pp. 373–383, Aug. 2012.
- [4] B. Kim, D. W. Kim, and G. T. Park, "Prediction of plasma etching using a polynomial neural network," *IEEE Trans. Plasma Sci.*, vol. 31, no. 6, pp. 1330–1336, Dec. 2003.
- [5] M. Terzi, C. Masiero, A. Beghi, M. Maggipinto, and G. A. Susto, "Deep learning for virtual metrology: Modeling with optical emission spectroscopy data," in *Proc. IEEE 3rd Int. Forum Res. Technol. Soc. Ind. (RTSI)*, Sep. 2017, pp. 1–6.
- [6] S. J. Pearton and D. P. Norton, "Dry etching of electronic oxides, polymers, and semiconductors," *Plasma Processes Polym.*, vol. 2, no. 1, pp. 16–37, Jan. 2005.
- [7] V. M. Donnelly and A. Kornblit, "Plasma etching: Yesterday, today, and tomorrow," *J. Vac. Sci. Technol. A, Vac. Surf. Films*, vol. 31, no. 5, Aug. 2013, Art. no. 050825.
- [8] M. Kanoh, M. Yamage, and H. Takada, "End-point detection of reactive ion etching by plasma impedance monitoring," *Jpn. J. Appl. Phys.*, vol. 40, no. 3R, p. 1457, 2001.
- [9] R. Chen, H. Huang, C. J. Spanos, and M. Gatto, "Plasma etch modeling using optical emission spectroscopy," *J. Vac. Sci. Technol. A, Vac. Surf. Films*, vol. 14, no. 3, pp. 1901–1906, May 1996.
- [10] A. J. M. Mackus, S. B. S. Heil, E. Langereis, H. C. M. Knoops, M. C. M. van de Sanden, and W. M. M. Kessels, "Optical emission spectroscopy as a tool for studying, optimizing, and monitoring plasma-assisted atomic layer deposition processes," *J. Vac. Sci. Technol. A, Vac. Surf. Films*, vol. 28, no. 1, pp. 77–87, Jan. 2010.
- [11] H. H. Yue, S. J. Qin, J. Wiseman, and A. Toprac, "Plasma etching endpoint detection using multiple wavelengths for small open-area wafers," *J. Vac. Sci. Technol. A, Vac. Surf. Films*, vol. 19, no. 1, pp. 66–75, Jan. 2001.
- [12] C. J. Pugh, "End point detection in reactive ion etching," Ph.D. dissertation, Univ. College London, London, U.K., 2013.
- [13] S. J. Hong, G. S. May, and D.-C. Park, "Neural network modeling of reactive ion etching using optical emission spectroscopy data," *IEEE Trans. Semicond. Manuf.*, vol. 16, no. 4, pp. 598–608, Nov. 2003.
- [14] H. H. Yue, S. J. Qin, R. J. Markle, C. Nauert, and M. Gatto, "Fault detection of plasma etchers using optical emission spectra," *IEEE Trans. Semicond. Manuf.*, vol. 13, no. 3, pp. 374–385, Aug. 2000.
- [15] C. Lee, S. W. Choi, and I.-B. Lee, "Sensor fault identification based on time-lagged PCA in dynamic processes," *Chemometric Intell. Lab. Syst.*, vol. 70, no. 2, pp. 165–178, Feb. 2004.
- [16] K. Han, K. J. Park, H. Chae, and E. S. Yoon, "Multi-way principal component analysis for the endpoint detection of the metal etch process using the whole optical emission spectra," *Korean J. Chem. Eng.*, vol. 25, no. 1, pp. 13–18, Jan. 2008.
- [17] D. A. White, B. E. Goodlin, A. E. Gower, D. S. Boning, H. Chen, H. H. Sawin, and T. J. Dalton, "Low open-area endpoint detection using a PCA-based  $T^2$  statistic and Q statistic on optical emission spectroscopy measurements," *IEEE Trans. Semicond. Manuf.*, vol. 13, no. 2, pp. 193–207, May 2000.
- [18] B. Flynn and S. McLoone, "Max separation clustering for feature extraction from optical emission spectroscopy data," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 4, pp. 480–488, Nov. 2011.
- [19] Q. P. He and J. Wang, "Fault detection using the K-nearest neighbor rule for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 4, pp. 345–354, Nov. 2007.
- [20] T. Hirai and M. Kano, "Adaptive virtual metrology design for semiconductor dry etching process through locally weighted partial least squares," *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 2, pp. 137–144, May 2015.
- [21] H. Jang, H. Lee, H. Lee, C.-K. Kim, and H. Chae, "Sensitivity enhancement of dielectric plasma etching endpoint detection by optical emission spectra with modified K-means cluster analysis," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 1, pp. 17–22, Feb. 2017.
- [22] S. Lee, H. Jang, Y. Kim, S. J. Kim, and H. Chae, "Sensitivity enhancement of SiO<sub>2</sub> plasma etching endpoint detection using modified Gaussian mixture model," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 2, pp. 252–257, May 2020.
- [23] B. Kim, S. Im, and G. Yoo, "Performance evaluation of CNN-based endpoint detection using *in-situ* plasma etching data," *Electronics*, vol. 10, no. 1, p. 49, Dec. 2020.
- [24] A. Giantomassi, F. Ferracuti, S. Iarlori, G. Ippoliti, and S. Longhi, "Electric motor fault detection and diagnosis by kernel density estimation and Kullback–Leibler divergence based on stator current measurements," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1770–1780, Nov. 2014.
- [25] H. Abuella and M. K. Ozdemir, "Automatic modulation classification based on kernel density estimation," *Can. J. Electr. Comput. Eng.*, vol. 39, no. 3, pp. 203–209, 2016.
- [26] J. Gao, H. Cheng, and P.-N. Tan, "Semi-supervised outlier detection," in *Proc. ACM Symp. Appl. Comput. (SAC)*, 2006, pp. 635–636.
- [27] S. Basu, M. Bilenko, and R. J. Mooney, "A probabilistic framework for semi-supervised clustering," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2004, pp. 59–68.
- [28] J. Gao, H. Cheng, and P.-N. Tan, "A novel framework for incorporating labeled examples into anomaly detection," in *Proc. SIAM Int. Conf. Data Mining*. Philadelphia, PA, USA: SIAM, 2006, pp. 594–598.
- [29] M. Zhao, Z. Tian, and T. W. S. Chow, "Fault diagnosis on wireless sensor network using the neighborhood kernel density estimation," *Neural Comput. Appl.*, vol. 31, no. 8, pp. 4019–4030, Aug. 2019.
- [30] H. E. Litvak, "End point control via optical emission spectroscopy," *J. Vac. Sci. Technol. B, Microelectron.*, vol. 14, no. 1, pp. 516–520, 1996.

- [31] I. Chakroun, T. Ashby, S. Das, S. Halder, R. Wuyts, and W. Verachtert, "Using unsupervised machine learning for plasma etching endpoint detection," in *Proc. 9th Int. Conf. Pattern Recognit. Appl. Methods*, 2020, pp. 273–279.
- [32] H. Jang, H.-S. Lee, H. Lee, and H. Chae, "Non-invasive plasma monitoring tools and multivariate analysis techniques for sensitivity improvement," *Appl. Sci. Conv. Technol.*, vol. 23, pp. 328–339, Nov. 2014.
- [33] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Evanston, IL, USA: Routledge, 2018.
- [34] T. Duong, "KS: Kernel density estimation and kernel discriminant analysis for multivariate data in R," *J. Stat. Softw.*, vol. 21, no. 7, pp. 1–16, 2007.
- [35] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. Hoboken, NJ, USA: Wiley, 2015.
- [36] N. Zougab, S. Adjabi, and C. C. Kokonenjji, "Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation," *Comput. Statist. Data Anal.*, vol. 75, pp. 28–38, Jul. 2014.
- [37] Y.-J. Kang and Y. Noh, "Comparison study of kernel density estimation according to various bandwidth selectors," *J. Comput. Struct. Eng. Inst. Korea*, vol. 32, no. 3, pp. 173–181, Jun. 2019.
- [38] X. Zhang, M. L. King, and R. J. Hyndman, "A Bayesian approach to bandwidth selection for multivariate kernel density estimation," *Comput. Statist. Data Anal.*, vol. 50, no. 11, pp. 3009–3031, Jul. 2006.
- [39] A. Gramacki, *Nonparametric Kernel Density Estimation and Its Computational Aspects*. Cham, Switzerland: Springer, 2018.
- [40] S. Węglarczyk, "Kernel density estimation and its application," in *Proc. ITM Web Conf.*, vol. 23, 2018, p. 37.
- [41] P. Hall and J. S. Marron, "On the amount of noise inherent in bandwidth selection for a kernel density estimator," *Ann. Statist.*, vol. 15, no. 1, pp. 163–181, Mar. 1987. [Online]. Available: <http://www.jstor.org/stable/2241075>
- [42] P. Branco, L. Torgo, and R. P. Ribeiro, "A survey of predictive modeling on imbalanced domains," *ACM Comput. Surv.*, vol. 49, no. 2, pp. 1–50, Nov. 2016, doi: [10.1145/2907070](https://doi.org/10.1145/2907070).
- [43] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.
- [44] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Dec. 2005.
- [45] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, vol. 398. Hoboken, NJ, USA: Wiley, 2013.



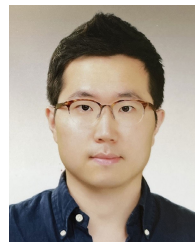
**JUNGYU CHOI** received the B.S. and M.S. degrees in electronic engineering from Soongsil University, in 2020 and 2022, respectively. He is currently pursuing the Ph.D degree with the Graduate School. His research interests include digital signal processing, artificial intelligence, anomaly detection, and smart factories.



**BOBAE KIM** received the B.S. degree in computer engineering from Academic Credit Bank System, in 2017, and the M.S. degree in information security from Soongsil University, Seoul, South Korea, in 2019, where she is currently pursuing the Ph.D. degree in information and communication engineering. Her research interests include machine-learning and signal processing.



**SUNGBIN IM** (Member, IEEE) received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, South Korea, in 1986 and 1988, respectively, and the Ph.D. degree in electrical and computer engineering from The University of Texas at Austin, in 1994. Since 1995, he has been with the School of Electronic Engineering, Soongsil University, Seoul. His research interests include the innovative application of digital nonlinear and statistical signal processing in the analysis, interpretation, and modeling of random data characterizing nonlinear physical phenomena and systems.



**GEONWOOK YOO** received the B.S. degree in electrical engineering from Korea University, Seoul, South Korea, in 2005, the M.Eng. degree in biomedical engineering from Cornell University, and the Ph.D. degree in electrical engineering from the University of Michigan, Ann Arbor, in 2011. He was a Senior Engineer at Samsung Electronics, from 2010 to 2014, and a Senior Researcher with the Korea Electronics Technology Institute, from 2014 to 2016. In 2016, he joined Soongsil University, Seoul, as an Assistant Professor with the School of Electronic Engineering. His research interests include III-N and ultrawide bandgap semiconductor-based devices, fabrication processes, and heterogeneous integration.

• • •