IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 23 June 2022, accepted 8 July 2022, date of publication 25 July 2022, date of current version 10 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3193783

RESEARCH ARTICLE

Virtual Metrology in Semiconductor Fabrication Foundry Using Deep Learning Neural Networks

TZE CHIANG TIN[1], SAW CHIN TAN[1], AND CHING KWANG LEE[2]
[1]Faculty of Computing and Informatics, Multimedia University, Cyberjaya 63100, Malaysia
[2]Faculty of Engineering, Multimedia University, Cyberjaya 63100, Malaysia

Corresponding author: Tze Chiang Tin (alvin.tin.i@gmail.com)

This work was supported by the Multimedia University, Cyberjaya Campus, Malaysia.

ABSTRACT Physical metrology inspections are crucial in semiconductor fabrication to ensure that wafers are fabricated within the production specification limits and to prevent faulty wafers from being shipped and installed in customer devices. However, it is not possible to examine every wafer, as such inspection would incur impractical costs on manpower, finances, and production cycle time (CT) of fabrication foundries (fabs). Virtual metrology (VM) presents an alternate approach to perform metrology inspection without incurring high costs using machine learning (ML) models. By leveraging historical equipment and process data, ML models can be calibrated to estimate the targeted metrology variables to estimate the quality of wafers, thereby performing virtual inspections on wafers. Recently, VM researchers have begun to introduce deep learning (DL) into VM research to examine its capability. Specifically, VM researchers experimented with a convolutional neural network (CNN). The targeted metrologies are those of plasma-based processes in both etching and chemical vapor deposition. The initial success has been reported by VM researchers. While various CNN-based VM models have been proposed for plasma-based fabrication processes, they have yet to be tested in the photolithography process. Motivated by the initial successes of CNN in plasma-based processes, this work is an initiative to experiment with CNN's performance in predicting the overlay errors of the photolithography process. Using data from a real fab, this study first establishes a baseline using the methodology of a prior study. The prediction results of the proposed CNN model are then compared with the baseline. The results showed that the CNN could further reduce prediction errors.

INDEX TERMS Virtual metrology, overlay, conjecture, regression, CNN.

I. INTRODUCTION

Semiconductor manufacturing is a highly stochastic and non-linear process [1]. It is one of the most complex processes in manufacturing industries with unique characteristics such as a complex and long series of sequential processing steps, process step dedication at designated tools, strict process window time frame, and dynamic product mix-run environment [2]. Wafers were fabricated through a series of process steps with a re-entrant loop for different fabrication processes. These steps are referred to as the process routes. After a wafer completes a particular fabrication process, metrology inspections of the wafer are required to ensure that the product design specifications are met. The inspection is performed by measuring the critical parameters of the wafer dies on the wafer surfaces to ensure that the parameters are within the tolerance limit. Certain metrology inspections of the physical properties of the wafer were also performed to ensure that there was no physical damage or the presence of particles that could jeopardize the functionality of the end devices. Therefore, metrology inspections are crucial for quality assurance of real fabs. Hence, various metrology inspection steps are performed in the process route to ensure that all quality checks are performed. As there are various types of metrology qualities to be examined throughout the fabrication cycle, this work focuses on the overlay quality of the wafer.

Overlay metrology inspection is a metrology inspection step in the photolithography process. The photolithography process remains one of the most crucial fabrication process steps because of the continuous demand for miniature

The associate editor coordinating the review of this manuscript and approving it for publication was Amin Zehtabian.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

devices [3]. The emphasis on the photolithography process is the minimization of the overlay error. As device miniature devices require feature size shrinkage and linewidth reduction in the fabricated integrated circuit (IC), it is crucial to ensure that the overlay errors are within the tolerance of the product specification for the end devices to operate appropriately. Hence, overlay metrology inspection is a critical metrology step in real fabs to meet the design requirements of modern electronic devices.

Although overlay metrology is crucial, these inspection steps are considered non-production-value steps because they do not contribute to the development of the end product. By contrast, frequent metrology inspections can reduce the throughput of fabs by incurring additional costs to the cycle time (CT). Performing inspections on every wafer is also impractical in terms of the cost of manpower and finances. Statistical sampling approaches have been conventionally employed in real fabs for metrology inspections. In a fab, wafers are usually transported in a cassette of 25 slots. Each unit of a transportation cassette is commonly called a wafer lot. Wafers stepping through the same process route are stored together in a single unit of wafer lot for ease of process time, arrangement, and transportation across the fab. Depending on the fabrication process, the wafers in the wafer lot will either enter the process chamber at the same time, or back-to-back. Hence, using the statistical sampling approach, the quality of the wafer is determined by performing sampling between wafer lots and between wafers in a wafer lot. Wafer lot sampling was carried out by placing intervals between the wafer lots to be measured. Once selected for measurement, a designated number of wafers in the wafer lot undergo metrology inspection at the physical station. The metrology quality results obtained are used to determine the metrology quality of the wafer lot and serve as an indicator of the stability of the fabrication process valid for a designated period of time. The statistical sampling approach significantly reduces the cost of performing a metrology inspection.

The conventional sampling approach allows metrology inspection to be carried out without incurring impractical costs and resources; it is weak in quality assurance. However, it is possible that the uninspected wafer lot contains wafers with faulty metrology quality owing to the intermittent instability of the fabrication equipment. This faultiness could not be detected through a conventional sampling approach, resulting in the escape of faulty wafers that ended up in customers' devices. To address the weakness of the statistical sampling approach, virtual metrology (VM) has emerged as either an alternate solution or a complementary solution to the statistical sampling strategies, to perform metrology inspection on uninspected wafers without the high cost of resources and additions to CT.

A VM is a conjecture model that predicts or estimates the target metrology variables of interest. Machine learning (ML) algorithms are typically leveraged for their predictive capabilities. The sources of data involved in building a VM generally consist of historical data from both metrology results and process conditions sampled by the sensors of the fabrication equipment [4], [5]. The success of ML in predicting the results of complex processes across research domains is clear. An area of notable success in ML is the advent of deep-learning (DL) models. Recently, VM researchers have begun experimenting with DL models to examine their prediction capability in metrologies of plasma-based etching and chemical vapor deposition (PECVD) processes. In contrast to previous works, this study aims to propose a convolutional neural network (CNN) based VM model to predict the overlay metrology quality of the photolithography process.

Therefore, the research questions of this work can be formulated as below:

1) How to apply CNN model to predict the overlay metrology values of wafers?

2) Does CNN model perform better than the comparison baseline?

The remainder of this paper is organized as follows. Section II introduces related VM studies on the photolithography process. Section III presents the study's research methodology. Section IV illustrates the experimental setup and analysis of the results obtained from the experiments. Finally, Section V concludes the paper.

II. RELATED WORKS

A. RESEARCH BACKGROUND

In [6], a VM that conjure the overlay errors of wafers was presented by the authors. The photolithography equipment studied in the authors' work consisted of two chucks. Hence, VM models were required for each of the two chucks. 37 equipment sensors were identified as the key sensors for collecting equipment data for the authors' work. The data collection period spanned 8 months. The data collected from chuck 1 contained 1612, while the data for chuck 2 contained 1563. Four summary statistics were used to derive the statistical measures for the feature set. These four statistical measures were minimum, maximum, mean, and variance. Using these four statistical measures, 148 process representatives are derived. Dimension-reduction schemes for both variable selection and variable extraction methods were used to perform feature selections. The results of the experiment showed that the kNN obtained the highest prediction accuracy. The authors then proposed a run-to-run (R2R) process control system utilizing exponentially weighted moving average (EMWA) with the addition of embedding the proposed VM into the control scheme. The proposed process control system was tested using Monte Carlo simulations. The evaluation obtained from a series of simulations demonstrated that the proposed R2R control system is capable of automatically tuning the process recipe settings to rectify the far-drifted overlay metrology measurement from the baseline values.

In [7], the authors proposed a novel detection method utilizing machine learning models to detect faulty wafers during the photolithography process. According to the authors, faulty wafers are identified based on their large metrology value deviations from baselines. In a real production

environment, the faultiness encountered varies according to the number of occurrences. For faultiness that has rare occurrences or those that are novel, it is not possible to detect them accurately using either binary or multiclass classifiers because samples to train these models would either be too small or have no prior samples at all. Hence, the author proposed the use of a novelty detection model to build a practical solution for a real production environment. The authors further addressed the weakness of the (SPC) system as a fault detection system. SPC has been conventionally used as a fault-detection system in real fabs. It employs statistical-based methods to detect process faults using data from a fault detection and classification (FDC) system. Data from the FDC system are usually termed FDC data. Owing to high-frequency data sampling at the equipment sensors, FDC data can be regarded as data from a direct recording of the process conditions. SPC detection methods in SPC were known to have inherent weaknesses. First, SPC monitors only a list of process parameters known to have a direct high impact on wafer quality. Second, SPC-based methods assume no interdependencies between the process variables that affect wafer quality when the opposite is true in actuality. Some SPC employ principal component analysis (PCA) to address this limitation but using PCA comes at the cost of losing crucial process information. Third, the SPC detection methods assume that the characteristics of the process data are both linear and unimodal. This assumption is not realistic because in the real world, the opposite is true. Finally, although FDC data are data from direct process condition monitoring, a wafer's quality cannot be directly inferred from the FDC data. The information pertaining to wafer qualities must be further derived from the FDC data. Hence, SPC-based methods are not suitable for use in faulty wafer detectors. A potential solution is to estimate the metrology values through VM, thereby determining the quality of the wafers. However, according to the authors, regression models are insensitive to sudden large deviations that are typically found in faulty wafers. In addition, owing to the novelty of the fault, similar samples may not exist in the training data to augment sensitivity. Hence, enhancement works are still necessary for VM to serve as faulty wafer detectors. In view of this need, the authors propose a novel fault detector method that utilizes machine learning models. Datasets of 2583 and 2509 wafer sensors, together with the measured metrology values, were collected from the two chucks of real photolithography equipment. Using statistical measurements, 148 features were derived for the feature selection set. Using dimension reduction schemes, the best process representatives among the 148 features were selected for the prediction task. To examine the capability of the proposed method, we first used cross-validation (CV) followed by moving windows (MW). The results showed that the one-class support vector machine (1-SVM) had the best prediction capability for CV, while both k-Means clustering and PCA methods had the same winning scores for MW.

In [8], the authors proposed a VM model with adaptive update capability to address the degrading prediction performance of the VM model over time. The authors pointed out that the prediction performance of a VM can degrade over time owing to external influences, such as the various maintenance activities performed on the fabrication equipment. Such activities could potentially change the baseline performance of the equipment and thereby introduce differences between the equipment representative data used to calibrate the VM initially and the latest representative data of the equipment. Hence, VM models must be recalibrated using latest equipment data. However, not all external disturbances in the equipment cause a significant shift in the performance baseline of the equipment. As such, recalibration may not be necessary after each external activity has been completed. Therefore, the authors propose a VM model that can perform adaptive updates. The proposed model employs an ensemble artificial neural network (ANN) algorithm. Using an ensemble ANN as the prediction algorithm, the reliability of a VM over time can be monitored by measuring the variation in its prediction accuracy. The proposed model was evaluated using data from two real photolithography equipment with six corresponding metrology variables for the fabrication process. The results from the experiments showed that the VM model was able to maintain its prediction accuracy through adaptive updates by incurring a lower recalibration cost than the comparison models. The authors also demonstrated the capability of the proposed solution as an anomaly process-event detector.

In [9], the authors experimented on the effectiveness of using transfer learning to deploy a reliable VM model in a real fab. The VM model, which is a data-driven model, is a virtual inspection that utilizes the conjecture model to estimate the values of metrology variables. These conjecture models are typically machine learning models, with ANN being one of the algorithms. To calibrate a conjecture model from scratch, sufficient historical observations of both equipment and metrology data are required so that the conjecture model can reach a steady modelling state internally before it can carry out the given prediction task. For new fabrication equipment introduced into a real fab, such a historical dataset would not be available. Collecting historical data usually spans months, leading to the slow deployment of VM models for new equipment in fabs. To overcome this challenge, the authors propose the use of transfer learning to build the required VM model. A strict requirement for transfer learning is that the learned source equipment must be similar to the targeted equipment. The authors experimented with two transfer learning strategies: model weight transfer and feature representation transfer. The experiment was carried out using real data from two photolithography processes to compare the performance of VM models built using transfer learning and those built using independent learning. The results obtained merited the use of a transfer learning strategy as a rapid deployment for a competent NN-based VM model in the event that independent learning from scratch is not possible.

In [10], the authors proposed a joint modelling approach to create faulty wafer detection for fabs. The presence of

various internal and external factors in real fabs inevitably causes faults in wafers. Faulty wafers must be thoroughly inspected to determine the required actions. The presence of faulty wafers could also indicate an unhealthy equipment state; thus, the equipment must also be examined. Owing to the limitations of SPC-based detection and statistical sampling methods, conjecture modelling through VM has been actively applied in real fabs in recent decades to detect faulty wafers in an effort to enhance yield. Typically, conjecture tasks can be categorized into two categories: classification and regression. The classification task is suited for categorical conjectures, whereas the regression task is suited for continuous values. In the author's view, these two predictive tasks can be jointly modelled to create a VM solution for fabs. Hence, the author proposed a joint modelling method that detects faulty wafers. The proposed model first conjures the metrology errors of the targeted metrology variables through a regression model. The conjured outputs then become inputs for a binary classification model. to determine the normality of the wafers. Datasets from two separate photolithography devices were collected from a real fab for a period of 7.5 months. Data for 2583 wafers were recorded in the first dataset, whereas the data for 2509 wafers were recorded in the second dataset. Each wafer dataset consisted of 102 process variables and four metrology values. The metrology values were made available in both numerics and binaries for each wafer. The faulty wafers present in the dataset were less than 1%, on average. The proposed model outperformed the comparison models in evaluation tests. The ANN model performed the best for both prediction tasks.

In [11], the authors proposed a lot-level modelling approach to address real fab conditions, where process conditions could not be sampled at high frequency owing to various limitations and events. From the literature review, [12] categorized the high data sampling frequency for the fabrication process as sampling frequencies of 1 Hertz (Hz) or higher. Hence, the low data sampling frequencies were 1 Hz and below. The FDC data frequency used in prior VM works to develop competent prediction models for real fabs are considered high sampling frequency data, as these data are sampled at a rate lower than 1 Hz. Therefore, to construct a competent VM model with low sampling frequency data, a different modelling paradigm may be required. The authors proposed a lot-level modelling approach that utilizes the information of other wafers in the same wafer lot as the wafer targeted for metrology quality prediction. Certain process and wafer qualification criteria are placed to ensure the reliability of such modelling, such as the selection of only the fabrication process with at least the process capability index C_{pk} greater than 1.33, and only wafers of mass production products with non-rework conditions. A joint-modeling approach that first performs novelty detection, followed by regression estimation of the overlay metrology values, the authors presented the experimental results in [13]. The experimental results obtained were significant enough for the authors to pursue further enhancement in future work.

B. DATA-DRIVEN MODELS

Machine learning models are data-driven models. These are non-parametric models with no fixed structures or parameters [14]. Neural networks (NN) and deep learning (DL) are non-parametric models. The advent of DL [15] and its reported successes [16] have drawn much attention from researchers. Researchers have experimented with DL in areas such as data reduction [17], natural language processing (NLP) [18], numerical digit recognition [19], and object detection [20]. These published results demonstrate the capability of DL to handle a given prediction task. The DL algorithm used in these studies was a convolutional neural network (CNN). In the area of VM studies, researchers have also begun exploring the potential use of CNN to identify latent features in the process data in order to augment the prediction accuracy of targeted metrology.

C. CONVOLUTIONAL NEURAL NETWORK (CNN)

Similar to conventional NN, CNN also encompass self-optimizing units called neurons. However, CNN are fundamentally different from conventional NN because they are designed specifically to perform pattern recognition on images [21]. Thus, CNNs follow the assumption that the input is comprised of images. Each color in an image is modelled as a channel. For a typical RGB-colored image, the input contains three channels, whereas for a grayscale image, the input contains two channels.

The spatial dimensionality of an input refers to its width and height. In CNN, the neurons are organized into three dimensions to handle the spatial dimensionality of the input, as well as the depth of the input. Such neuron organization allows neurons within a layer to connect to a larger region of the layer preceding it. In traditional ANNs, only a shallow region connection can be achieved between layers.

In terms of architecture, the CNN comprises three types of layers. These three layers are the convolutional, pooling, and fully connected layers. Among these three layers, the convolutional and pooling layer are the two layers contribute to the uniqueness of the CNN. In the convolutional layer, the output neurons are only connected to their nearby input neurons, instead of to all the input neurons. When multiple convolutional layers are connected, the effectiveness of the feature extraction from a given image increases as each layer attempts to retrieve specific features that may be relevant to the given prediction problem [22], [23]. The main component of the convolutional layer was a learnable kernel. These kernels are usually small in spatial dimensionality but are spread along the entire depth of the input. When a given input enters a convolutional layer, the filters in the layer convolve across the data dimensions found in the input to generate 2D mappings of the input [21]. When the generated mappings arrive at the pooling layer, a pooling mechanism is used to optimize the number of features required to achieve effective learning for a given prediction problem [22]. Finally, in the fully connected layers, the prediction results are generated based on the activation functions of the neurons.

The success of CNN in virtual metrology has also been reported in recent studies. In [24], the authors proposed a deep learning model for the VM of real plasma etching equipment. The proposed model has a high performance resilience against process chamber condition variations. The authors leveraged optical emission spectroscopy (OES) data that contain large amounts of in-situ process chamber condition information to derive significant process performance indicators. The deep learning algorithm employed by the authors is a CNN. CNN have been proven capable of achieving significant performance in research problems that can be represented by 2D images. Although the 2D data structure OES data is similar to that of a 2D image, the authors pointed out that training a CNN using OES data is not as straightforward as training a CNN with 2D images. The reason for this complication is that treating OES data as an image will result in a significant loss of process information. Owing to these differences, the neural network configuration of the CNN must be altered before it can be trained effectively using OES data. The authors performed two configuration changes in the DL. The first configuration change was performed during the convolution calculations. This change was performed so that the DL could handle the time series aspect of the OES data. The second configuration was performed using the normalization method of DL. This change was necessary to prevent the loss of the signal intensity information in the OES data. With these two modifications, the authors termed the proposed modified DL model the OESNet. To evaluate the success of the modification, a comparison was made between the prediction accuracy of OESNet and that of the DL models from the ImageNet Large-scale Visual Recognition Challenge (ILSVRC) using real OES data. The comparison showed that OESNet had better generalization capabilities, prediction accuracies, and shorter prediction times than the comparison models. OESNet also has higher resilience against condition variations in the process chamber.

In [25], DL was applied as a feature extractor for VM. The authors pointed out that large-scale implementation of VM in real fabs is still not possible because the existing feature extraction methods encounter limitations when handling large and complex fabrication process data. An example of the data provided by the authors is OES data. When handling such data, automated feature extraction is more feasible than manual feature extraction in terms of time and scalability. However, existing automated feature extraction methods are still not sufficiently robust to extract critical information. Acknowledging the capability of CNN in handling highly complex and nonlinear data, the authors proposed the use of convolutional autoencoders (CA) as an automated feature extractor for real OES data. Such a proposal is novel, as the performance of CNN as a feature extractor is yet to be evaluated in VM research. The features extracted by the CNN were then used by a regression model to predict the etch rate of a real plasma etching process. The authors termed this type of VM DeepVM. Using the same approach, the authors also created a variety of DeepVM models for comparison using different types of autoencoders. Prediction accuracies were obtained between the DL and non-DL VM models in the experiment. The comparison results show that the DeepVM models outperformed the non-DL VM models. Despite this success, the authors remarked that such a performance can only be obtained when dealing with complex 2D data structures similar to images. The proposed method may not be able to achieve this performance in the context of a conventional tabular data structure.

In [26], the authors presented the use of a DL model to build a VM that predicts the electrical properties of a real CVD process. The authors addressed four research challenges in their work. First, there is a lack of research on multistage VM. Multistage studies are necessary because metrology inspection is usually performed only after a sequence of process steps. The cause of faultiness could be the previous steps instead of the immediate process step before the metrology inspection. Second, the time dimension has yet to be taken into account in existing studies of multi-stage VM for CVD. Third, feature extraction and prediction tasks are modelled explicitly and distinctively. Separate modelling may result in unoptimized end solutions. Fourth, existing feature extractions are performed over statistical measures derived from the raw data instead of directly inferred from the raw data. Although statistical measurements provide descriptive information on the fabrication process, they could also result in information loss, thereby reducing the prediction accuracy of the model. These four research challenges in VM led the authors to propose a multi-stage CNN model, termed CNN-GPR, which is the VM model for the CVD process. CNN has the advantage of implicitly performing both feature extraction and prediction tasks, thereby further optimizing VM solutions. However, as a DL algorithm, they are prone to overfitting. To avoid this issue, the authors employed a Gaussian process model (GPR) to provide a quantitative measurement of the prediction uncertainty of the CNN. Although GPR can provide a useful indicator to gauge the model's performance, the use of GPR imposed a limitation on the proposed model, whereby performance instability would occur on the model when the data have high dimensions. The authors also modified the learning algorithm of the CNN from the backpropagation learning algorithm to maximize the posterior density distribution. The proposed CNN-GPR model was evaluated using a real CVD process comprising four processing stages. The evaluation results showed that CNN-GPR performed better than the other models. The proposed CNN-GPR model can also provide a confidence level measurement for each prediction

III. METHODOLOGY

This work employs the methodology proposed in [11] to evaluate the performance of the CNN in predicting the overlay errors of wafers. The authors of [11] proposed a joint model of a novelty detection task and a regression task. Joint modeling first performs novelty detection to filter wafers that are predicted to be faulty from entering the regression task.

Wafers that are predicted to be faulty are routed to the overlay metrology station for thorough physical inspection. Wafers that are predicted to be normal will enter the regression task, thereby reducing prediction errors owing to noise introduced by faulty wafers. Although two prediction tasks are present in the methodology, this study focuses only on the prediction accuracy of the regression task. The baseline of this study was based on the results published in [13].

A. CNN's ARCHITECTURE

Although DL models d not follow a fixed approach to formulate their architecture, CNN tend to have a common architecture [21]. The two CNN architectures frequently found in the literature are depicted in Figure 1 and Figure 2. We denote these two architectures as Types A and B, respectively. In Type A architecture, the convolutional layers are first stacked, followed by repeated pooling layers, and finally, the fully connected layers. In Type B architecture, two convolutional layers are stacked, followed by a pooling layer. This structure was repeated, as required. According to [21], the Type B architecture is strongly recommended because stacking multiple convolutional layers increases the capability of the model to extract complex features from the input.

To adapt the CNN from solving image classification to predicting the metrology values of targeted metrology variables, the CNN is modified in the following aspects [22]. First, the input to the CNN model contains the number of channels according to the input features in the historical data in contrast to the image colors. Second, the CNN model outputs are the predicted numerical values of the targeted metrology variables instead of the class labels in image classification problems. Third, the features extracted from the convolutional and pooling layers are the inter-dependency behavior between the input features derived from equipment sensor data, in contrast to the features extracted in image classification problems, such as the edges and shapes of an object.

Let X denote the set of historical data, x denote the statistical features of each wafer lot in X, i denotes the index of x, w denotes the index of each wafer lot observed, and W denotes the set of wafers that participated in the observation. The input to the CNN can be expressed as in Equation (1): The features denoted by x were derived using the same statistical descriptors presented in [11].

$$X = \begin{bmatrix} [x_i, x_{i+1}, x_{i+2}, x_{i+3}, \ldots x_I]_0 , \\ \ldots \\ [x_i, x_{i+1}, x_{i+2}, x_{i+3}, \ldots x_I]_w \end{bmatrix}, w \in W \quad (1)$$

The extracted interdependent behavior features are a combination of convolutional and pooling layers [22]. Let *Pooling* denote the pooling mechanism of the pooling layers; L denotes the depth of the CNN; l denotes the lth layer of the CNN, where $l \in L$. Let C_l denote the total number of convolutional filters in lth layer; X_l denotes the input to lth layer; O_l denotes the output of lth layer; w_l and b_l denote the weights and bias of neurons in lth layer, σ_l denotes the activation function of the lth layer, and j denotes the channel index considering the multiple convolutional filters in the convolutional layer [22]. The output of the first convolutional and pooling layers can be written as

$$o_l^j = Pooling\left(\sigma\left(w_l^j x_l^j + b_l^j\right)\right), j \in [1, C_l], l = 1 \quad (2)$$

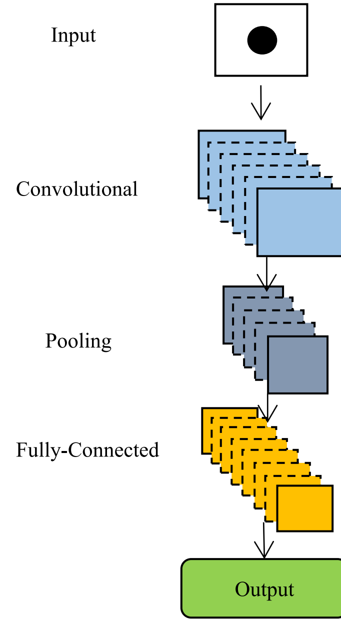

FIGURE 1. Common CNN architecture – Type A.

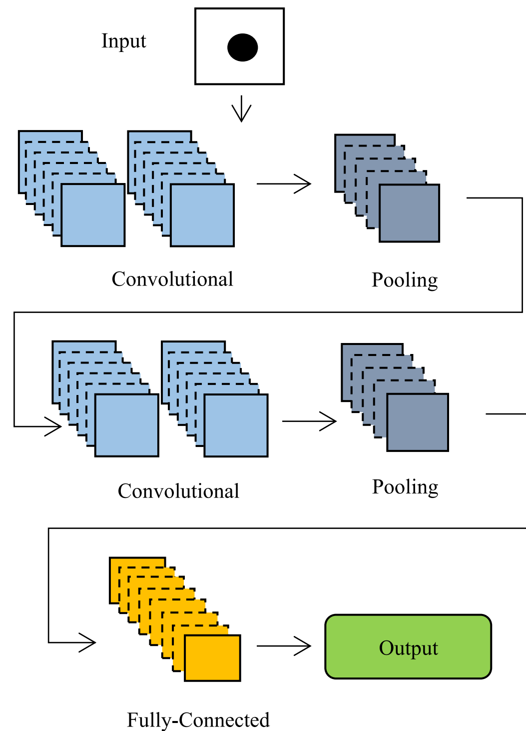

FIGURE 2. Common CNN architecture – Type B.

TABLE 1. Confusion matrix.

		Predicted	
		Normal	Faulty
Actual	Normal	TP	FP
	Faulty	FN	TN

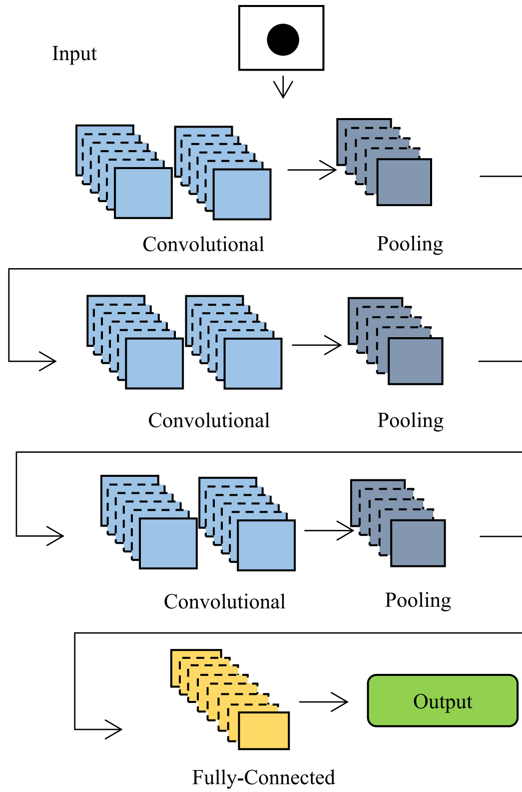

FIGURE 3. CNN architecture – Type C.

TABLE 2. The number of features selected for each of the target metrology variables using stepwise linear regression.

Target Variables, Y	Number of Features
1	23
2	21
3	27
4	21
5	19
6	17
7	24

The output in the lth convolutional and pooling layer, where $l \neq 1$, can be written as

$$o_l^j = Pooling\left(\sigma_l \sum_{r=1}^{C_{l-1}} \left(w_l^j x_l^r + b_l^j\right)\right), \quad j \in [1, C_l], l > 1 \tag{3}$$

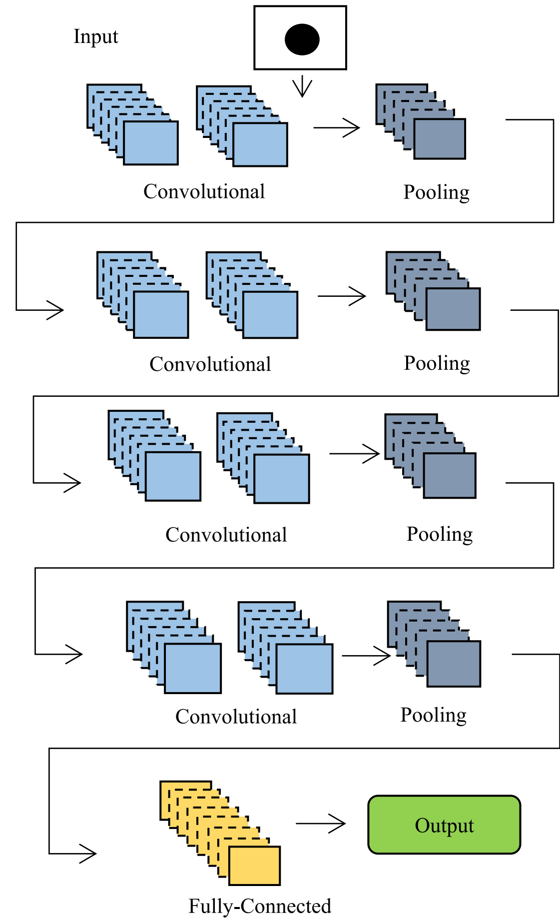

FIGURE 4. CNN architecture – Type D.

The features learned through the convolutional and pooling layers were then concatenated through a flattening process before passing to the fully connected layers for prediction. Let *Flatten* denote the flattening process, which can be expressed as

$$o_l^{Flatten} = Flatten\left(o_L^1, o_L^2, o_L^3, \ldots, o_L^k\right), k = C_L \tag{4}$$

Finally, the model's prediction is the output from the fully connected layers and can be written as

$$\tilde{y} = w_F o_L^{Flatten} + b_F \tag{5}$$

The activation function present in each layer of the neural network has two purposes. The first is to transform the output of each layer into a manageable and scaled data range, and the second is to enable the neural network to mimic a very complex nonlinear function through a combination of activation functions from various layers [22]. The common activation functions used in the context of forecasting for CNN are hypertangent (tanh), rectified linear unit (ReLU), and variants of ReLU, called leaky rectified linear unit (ReLU). Equations (6), (7), and (8) define the activation functions.

$$\tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \tag{6}$$

TABLE 3. Prediction performance of the classification models.

Target Variable, Y	Best Algorithm	Calibration Period	TPR (%)	FPR (%)	TPR – FPR (%)
1	KNN	3	68.15	32.49	35.66
2	KNN	5	69.43	47.10	22.33
3	1-SVM	3	74.05	40.45	33.60
4	1-SVM	4	70.19	43.44	26.75
5	1-SVM	5	71.73	55.10	16.63
6	KNN	3	69.13	37.66	31.47
7	1-SVM	5	73.55	36.17	37.38
Average			70.89	41.77	29.12

TABLE 4. Comparison of classification results with prior works.

Research Work	TPR (%)	FPR (%)	TPR – FPR (%)
[3]	73.49	45.85	27.64
[39]	71.34	47.49	23.86
This work	70.89	41.77	29.12

$$\mathrm{ReLU}(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{7}$$

$$\mathrm{Leaky\ ReLU}(x) = \begin{cases} x, & x > 0 \\ 0.01x, & x \leq 0 \end{cases} \tag{8}$$

B. CNN PARAMETERS TUNING AND MODEL SELECTION

The two critical factors that affect the performance of a CNN are 1) the hyperparameters of the convolutional layer and the pooling layer, and 2) the depth of the CNN. The hyperparameters of the convolutional layer are the size of the convolutional filter and learnable kernel, while the hyperparameters of the pooling layer are the pooling size and pooling method [22]. No general rules can be applied directly to hyperparameter selection. Well-known examples can be found in. Two well-known examples are AlexNet [16] and LeNet [27]. The depth of the CNN refers to the number of convolutional and pooling layers used. The recommendation from the literature suggests that the depth should not be too large or too small [21], [22], so that the CNN is capable of learning complex relations and simultaneously maintaining model convergence. Therefore, increasing the depth of the CNN should only be applied when hyperparameter tuning does not increase the prediction accuracy, and the increment of the depth of the CNN should stop when the prediction accuracy does not improve further and yet experiences difficulty in model convergence.

C. FEATURES SELECTION

Feature selection is a crucial step for identifying features that can potentially augment the prediction accuracy of conjecture models. Studies [6] and [13] have identified that stepwise feature selection has the feature selection method that provides the best predictors to the prediction models. To maintain consistency with the methodology of [13], this study employs a stepwise feature selection method as feature selection method.

D. PERFORMANCE METRICS

For the classification task, two measurement metrics were used to measure the performance of the classification model. These two metrics are the True Positive Rate (TPR), and the False Positive Rate (FPR). In the context of this study, a true event is the event when a faulty wafer is found. Hence, TPR would refer to the rate that the model successfully identified a faulty wafer, while the FPR would refer to rate that the model identified a normal wafer as faulty. These two terms can be further illustrated using a confusion matrix, as shown in Table 1.

For the regression task, two measurement metrics were used, as well, to measure the performance of the regression model. These two metrics are mean squared error (MSE) and mean absolute specification error (MASE). During model training, the MSE was used to measure the prediction error of the model. The combination of both hyperparameters and the depth of the CNN that obtains the lowest mean squared error (MSE) during its training is selected as the prediction model for the VM. Let n represent the number of predictions performed, y represent the observed measurement value, $\tilde{y}$ denote the predicted measurement value, j denote the subscript of each prediction, and MSE can be written as in Equation (9).

$$MSE = \frac{1}{n}\sum_{j=1}^{n}(y_j - \tilde{y}_j)^2 \tag{9}$$

The next metric employed was the mean absolute scaled error (MASE). MASE was used by both [6] and [13] as one of the metrics to evaluate the performance of the regression models because it allows the prediction performance of a VM model to be measured based on the error tolerance of targeted metrology. The error tolerance of a metrology quality is the permitted metrology error boundary, which does not result in significant defects in a wafer die. (Kang et al. [6]). Let e represent the error tolerance of a targeted metrology and let MASE be calculated using the formulation in Equation (10).

$$MASE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{e}\right| \times 100\% \tag{10}$$

TABLE 5. CNN structures experimented and averaged MSE obtained with 8 convolutional filters.

Type	CNN Structure	Window Sizes			
		2	3	4	5
B	{8 conv + 8 conv + max pool} → {8 conv + 8 conv + max pool} → {Fully Connected}	0.19	0.14	0.12	0.11
C	{8 conv + 8 conv + max pool} → {8 conv + 8 conv + max pool} → {8 conv + 8 conv + max pool} → {Fully Connected}	0.17	0.16	0.11	0.10
D	{8 conv + 8 conv + max pool} → {8 conv + 8 conv + max pool} → {8 conv + 8 conv + max pool} → {8 conv + 8 conv + max pool} → {Fully Connected}	0.20	0.17	0.15	0.12

TABLE 6. CNN structures experimented and averaged MSE obtained with 8 and 16 convolutional filters.

Type	CNN Structure	Window Sizes			
		2	3	4	5
B	{8 conv + 8 conv + max pool} → {16 conv + 16 conv + max pool} → {Fully Connected}	0.19	0.17	0.14	0.10
C	{8 conv + 8 conv + max pool} → {16 conv + 16 conv + max pool} → {16 conv + 16 conv + max pool} → {Fully Connected}	0.17	0.14	0.10	0.09
D	{8 conv + 8 conv + max pool} → {16 conv + 16 conv + max pool} → {16 conv + 16 conv + max pool} → {16 conv + 16 conv + max pool} → {Fully Connected}	0.25	0.18	0.12	0.11

TABLE 7. CNN structures experimented and averaged MSE obtained with 16 and 32 convolutional filters.

Type	CNN Structure	Window Sizes			
		2	3	4	5
B	{16 conv + 16 conv + max pool} → {32 conv + 32 conv + max pool} → {Fully Connected}	0.23	0.18	0.16	0.12
C	{16 conv + 16 conv + max pool} → {32 conv + 32 conv + max pool} → {32 conv + 32 conv + max pool} → {Fully Connected}	0.18	0.13	0.09	**0.08**
D	{16 conv + 16 conv + max pool} → {32 conv + 32 conv + max pool} → {32 conv + 32 conv + max pool} → {32 conv + 32 conv + max pool} → {Fully Connected}	0.20	0.19	0.15	0.13

IV. EXPERIMENTAL RESULTS

This section presents the experimental design, followed by a presentation of the results obtained for the classification and regression tasks. Finally, the results were analyzed and discussed based on the experimental results obtained.

A. EXPERIMENTAL DESIGN

The data used in this study were acquired from a real fab. This study applied the same setting as in [13], which targets a list of photolithography processes of stepper-type photolithography equipment with $C_{pk} > 1.33$. Data for 2000 wafers over a period of seven months, with January 2019 being the starting month. Each wafer dataset consisted of 18 equipment sensor observation values and overlay errors measured for seven overlay variables. After the data preprocessing steps to remove incomplete data, data for 1900 wafers were available to conduct the experiment.

Seven descriptive statistical measures described in [13] were used to derive the statistical process features for the data of each of the 18 equipment sensors. Using the same statistical measures, 126 (18×7) measures were derived for the feature selection set. Owing to the data privacy requirement of a real fab, the real names of the targeted metrology variables

TABLE 8. CNN structures experimented and averaged MSE obtained with 32 and 64 convolutional filters.

Type	CNN Structure	Window Sizes			
		2	3	4	5
B	{32 conv + 32 conv + max pool} → {64 conv + 64 conv + max pool} → {Fully Connected}	0.23	0.20	0.18	0.13
C	{32 conv + 32 conv + max pool} → {64 conv + 64 conv + max pool} → {64 conv + 64 conv + max pool} → {Fully Connected}	0.17	0.14	0.10	0.09
D	{32 conv + 32 conv + max pool} → {64 conv + 64 conv + max pool} → {64 conv + 64 conv + max pool} → {64 conv + 64 conv + max pool} → {Fully Connected}	0.24	0.21	0.17	0.14

TABLE 9. The optimum window size and prediction algorithm for the target overlay variables without joint modelling.

Target Variable, Y	Window Size	Algorithm	MSE	MASE
1	5	Elastic Net	0.06	10.23
2	5	Elastic Net	0.06	10.07
3	5	Elastic Net	0.07	10.27
4	5	Elastic Net	0.06	10.16
5	5	CNN	0.05	9.78
6	5	CNN	0.05	9.81
7	5	CNN	0.05	9.95

TABLE 10. Comparison of MSE and MASE with prior works using non-joint modelling of this work.

Research Work	Averaged MSE	Averaged MASE
[14]	0.042	5.65
[39]	0.060	10.59
This work	0.057	10.04

will not be revealed in this study. To address these targeted overlay variables, the variable acronyms Y1, Y2, ..., Y7 will be used to indicate the targeted metrology variables.

Similar to the work by [6] and [13], the data used to calibrate the prediction models were partitioned into moving window sizes of two, three, four, and five months. The window size of the evaluation set was set to one month. The first data point of the window is the first date of the starting month, whereas the end data point of the training data is the last date of the ending month. The evaluation data will be data from subsequent months.

B. CLASSIFICATION TASK

The work in [13] employed a joint prediction method to create a smart sampling system. The first prediction task is a classification class using novelty detection methods to identify faulty wafers. To compare the VM models for the regression task, it was also necessary to implement the classification task in this work. Table 2 lists the number of features selected using the same feature selection method as in [13].

Using the selected features, novelty detection models were created and evaluated using k-nearest neighbors (kNN) and 1-SVM. Table 3 lists the prediction performance of these two models. Table 4 compares the results obtained in this study with those of previous studies.

C. REGRESSION TASK

Following the parameter settings in [16], [21], [22], and [23], the learnable kernels were set to a size of three for each convolutional layer, the pooling size was set to two for each pooling layer, and max pooling was selected as the pooling method. For the CNN architecture, we experimented with three architectures. The first architecture was **Type B** which is presented in Section III. **Type B** has two convolutional layer depths. We propose the next two architectures for the experiments, which are denoted as **Type C** and **Type D** architectures.

The **Type C** architecture varies from **Type B** architecture by extending the convolutional depth to three. The **Type D** architecture varies from the Type B architecture by extending the convolutional depth to four. Figure 3 depicts **Type C** architecture while Figure 4 **Type D**.

D. RESULTS ANALYSIS AND DISCUSSIONS

The VM is an alternative to expensive physical metrology inspections in real fabs. By using historical processes and metrology information, prediction models can be constructed to gauge the metrology quality of the targeted wafers. As a data-driven model, the quality of data used to calibrate prediction models is crucial. One such quality is the data sampling rate. Previous studies have mainly focused on realizing high-competency VM models using equipment data sampled

TABLE 11. The optimum window size and prediction algorithm for the target overlay variables with joint modelling.

Target Variable, Y	Window Size	Algorithm	MSE	MASE
1	5	Elastic Net	0.05	8.03
2	5	Elastic Net	0.05	8.07
3	5	CNN	0.06	8.76
4	5	CNN	0.06	9.06
5	5	CNN	0.05	8.58
6	5	CNN	0.04	8.65
7	5	CNN	0.05	8.75

at a high frequency. The common ones are the FDC data. However, in the real world, unforeseen and unavoidable situations render such data temporarily unavailable. To continuously sustain the product line, the authors of [11] proposed an alternate modelling approach when only low-frequency equipment data were available. The proposed system was published in their following work in [13].

In [11] and [13], the authors have yet to explore deep-learning-based models. The reason given by the authors was that the aim of the work was to first evaluate the efficacy of the proposed modelling approach. Hence, it was necessary to evaluate using the same models as in prior works that have achieved the best performance. Hence, this study intends to examine the capability of deep learning models in view of this research opportunity. VM researchers began experimenting with the capability of CNN in 2018.

The scope of these research works revolved around OES data for plasma-based fabrication processes. As there is still a research contribution of CNN in the photolithography process, this work aims to deliver an initial research contribution by examining the prediction capability of CNN in the overlay quality of the photolithography process based on the VM modelling of [11]. The results showed that with the CNN model, the prediction error of the regression task can be further reduced compared with the results obtained in [13].

Tables 5–8 show the average MSE obtained for all seven metrology variables for each CNN architecture using these convolutional filters. From these tables, the Type C architecture, using a mix of 16 and 32 convolutional filters, obtained the lowest MSE during the training phase. Hence, the architecture and parameter settings were used to construct the prediction model proposed in this work. The comparison models employed are those using the k-nearest neighbor and elastic net. By performing a similar examination as in [13], the models were first evaluated without joint modelling, followed by evaluation with joint modelling. For non-joint modelling, the optimum window size and the best-performing algorithm for each of the overlay variables are listed in Table 9. A comparison of the MSE and MASE with prior works is presented in Table 10. The evaluation results for the joint modelling are listed in Tables 11 and 12, respectively. Figure 5 shows the prediction performances of kNN, Elastic Net, and CNN with joint modelling

From the results obtained, the non-joint modelling method did not yield a significant improvement. The results obtained

TABLE 12. Comparison of MSE and MASE with prior works using joint modelling of this work.

Research Work	Averaged MSE	Averaged MASE
[6]	0.042	5.65
[13]	0.055	9.59
This work	0.051	8.56

in this work were slightly lower than those in [13], with an improvement of 0.003 for the average MSE and 0.55 for the average MASE. With joint modelling, the improvement was more significant. Compared to the results from [39], the average MSE was lower by 0.004, and the average MASE was lower by 1.03. Observing the best algorithms for targeted overlay variables, a CNN was selected for most of the overlay variables when joint modelling was used. The filtering of faulty wafers from the regression task gave the CNN an upper hand in its performance over an elastic net. Although the prediction results of the regression task were improved using CNN compared to the FDC-based approach in [6], the performance of the non-FDC-based VM models still requires improvement. The contributing source to the higher prediction error could be the high mixture of process recipes present in the test set. In real fabs, a single photolithography layer comprises various individual photolithography processes. These process recipes are grouped under the same layer because they aim to accomplish the same fabrication process, but on different products. Owing to the various complexities in circuit design across end products, the properties of the chemical, mechanical, and electrical interactions could also vary across process recipes, resulting in various degrees of differences. As such, there are specific differences in the process characteristics of each of the individual process recipes in the same layer of the photolithography process. Hence, when the prediction is performed at the layer level, variances may occur in the internal modelling of the prediction models with regard to the relationship between the process data and targeted metrology. In the work done by [6], the impact of the recipe mix run was not reported. This shows that with the fundamental difference in non-FDC-based modelling, additional techniques are needed in this work compared to [6] to reduce the prediction error in order to achieve a performance similar to that of the baseline.

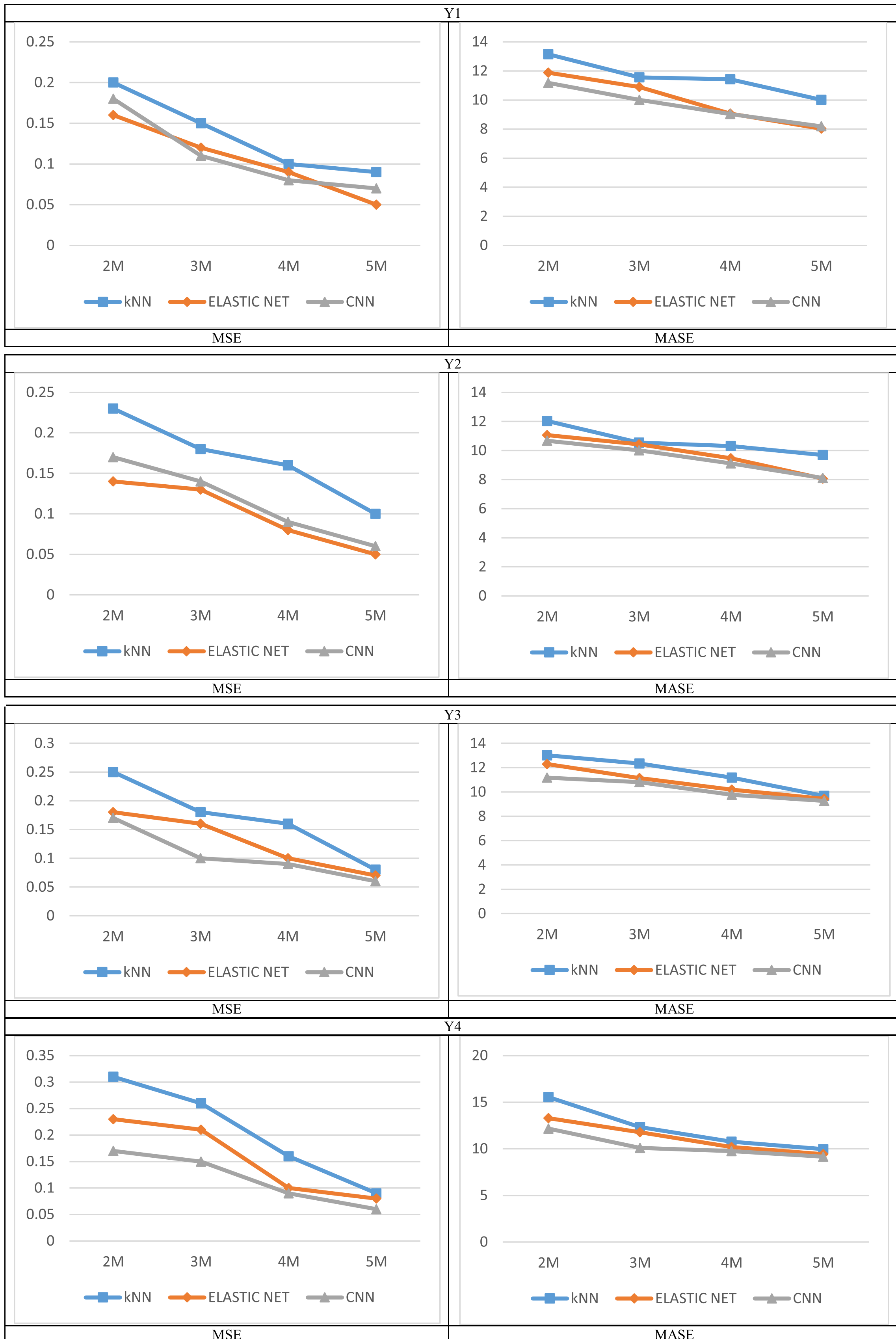

FIGURE 5. Prediction performances of kNN, Elastic Net, and CNN with joint modelling.

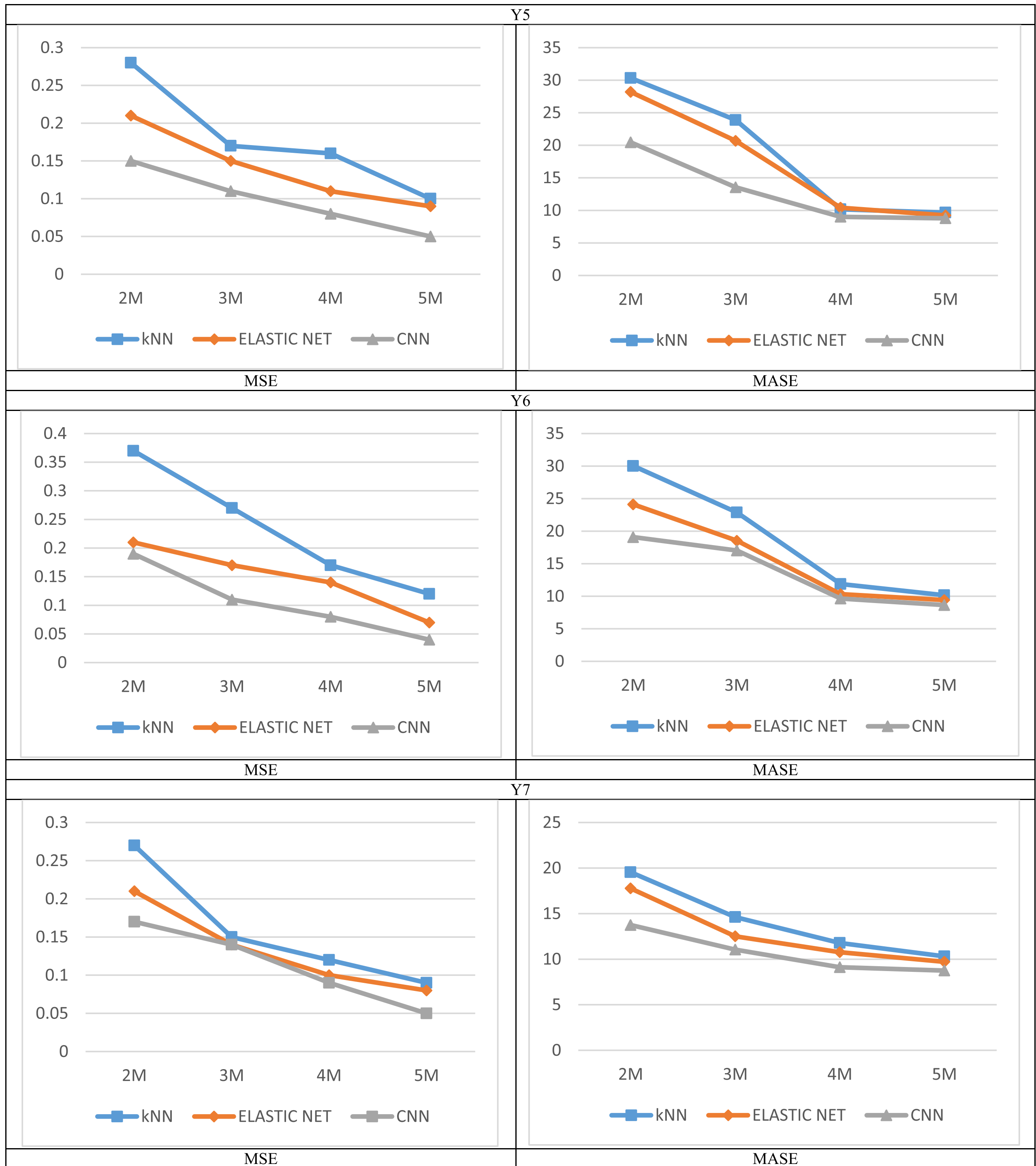

FIGURE 5. *(Continued.)* **Prediction performances of kNN, Elastic Net, and CNN with joint modelling.**

V. CONCLUSION

In this study, a VM utilizing a CNN model to predict the overlay errors of the photolithography process was presented. The aim was to examine whether the CNN model could deliver better prediction results using the methodology presented by the authors of [11] and [13]. Using data from a real fab, the experimental results showed that the CNN performed better than the comparison models. However, there remains a gap to bridge when compared with the work in [14]. The source of distortion in the prediction performance could be the effects of varying the process characteristics of each individual process recipe in a single photolithography layer. Hence, the first future work following this work is to carry out data-driven analysis to accurately form clusters of process

recipes according to the similarity of the process recipes' process behavior. Next, nonlinear feature extraction methods, such as automatic feature extraction using deep learning, will be explored to examine if this method can further extract process features that sequential-based feature extraction may not be possible. These two future works will lead to further uncovering the missing factors that can reduce prediction errors.

REFERENCES

[1] J. D. Schwartz and D. E. Rivera, "A system identification approach to PDE modeling of a semiconductor manufacturing process," *IFAC Proc. Volumes*, vol. 42, no. 10, pp. 964–969, 2009.
[2] K. Rozen and N. M. Byrne, "Using simulation to improve semiconductor factory cycle time by segregation of preventive maintenance activities," in *Proc. Winter Simul. Conf. (WSC)*, Washington, DC, USA, Dec. 2016, pp. 2676–2684.
[3] C.-F. Chien and Y.-J. Chen, "Manufacturing intelligence and smart production for industry 3.5 and empirical study of decision-based virtual metrology for controlling overlay errors," in *Proc. Int. Symp. VLSI Design, Automat. Test (VLSI-DAT)*, Taiwan, Apr. 2016, pp. 1–4.
[4] Q. P. He and J. Wang, "Large-scale semiconductor process fault detection using a fast pattern recognition-based method," *IEEE Trans. Semicond. Manuf.*, vol. 23, no. 2, pp. 194–200, May 2010.
[5] G. Verdier and F. Ariane, "Adaptive Mahalanobis distance and K-nearest neighbor rule for fault detection in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 1, pp. 59–68, Feb. 2011.
[6] P. Kang, D. Kim, H.-J. Lee, S. Doh, and S. Cho, "Virtual metrology for run-to-run control in semiconductor manufacturing," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2508–2522, Mar. 2011.
[7] D. Kim, P. Kang, S. Cho, H.-J. Lee, and S. Doh, "Machine learning-based novelty detection for faulty wafer detection in semiconductor manufacturing," *Expert Syst. Appl.*, vol. 39, no. 4, pp. 4075–4083, Mar. 2012.
[8] S. Kang and P. Kang, "An intelligent virtual metrology system with adaptive update for semiconductor manufacturing," *J. Process Control*, vol. 52, pp. 66–74, Apr. 2017.
[9] S. Kang, "On effectiveness of transfer learning approach for neural network-based virtual metrology modeling," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 1, pp. 149–155, Feb. 2018.
[10] S. Kang, "Joint modeling of classification and regression for improving faulty wafer detection in semiconductor manufacturing," *J. Intell. Manuf.*, vol. 31, no. 2, pp. 319–326, Feb. 2020.
[11] T. C. Tin, S. C. Tan, H. Yong, J. O. H. Kim, E. K. Y. Teo, C. K. Lee, P. Than, A. P. S. Tan, and S. C. Phang, "A realizable overlay virtual metrology system in semiconductor manufacturing: Proposal, challenges and future perspective," *IEEE Access*, vol. 9, pp. 65418–65439, 2021, doi: 10.1109/ACCESS.2021.3076193.
[12] A. U. Haq and D. Djurdjanovic, "Virtual metrology concept for predicting defect levels in semiconductor manufacturing," in *Proc. 49th CIRP Conf. Manuf. Syst.*, vol. 57, Jan. 2016, p. 580.
[13] T. C. Tin, S. C. Tan, H. Yong, J. O. H. Kim, E. K. Y. Teo, J. C. Y. Wong, C. K. Lee, P. Than, A. P. S. Tan, and S. C. Phang, "The implementation of a smart sampling scheme C_2O utilizing virtual metrology in semiconductor manufacturing," *IEEE Access*, vol. 9, pp. 114255–114266, 2021, doi: 10.1109/ACCESS.2021.3103235.
[14] J. A. Ramírez-Hernández, J. Crabtree, X. Yao, E. Fernandez, M. C. Fu, M. Janakiram, S. I. Marcus, M. O'Connor, and N. Patel, "Optimal preventive maintenance scheduling in semiconductor manufacturing systems: Software tool and simulation case studies," *IEEE Trans. Semicond. Manuf.*, vol. 23, no. 3, pp. 477–489, Aug. 2010.
[15] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, Jan. 2009.
[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1–9.
[17] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
[18] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proc. ICML*, 2008, pp. 160–167.
[19] I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, and V. Shet, "Multi-digit number recognition from street view imagery using deep convolutional neural networks," 2013, *arXiv:1312.6082*.
[20] B. Huval, A. Coates, and A. Ng, "Deep learning for class-generic object detection," 2013, *arXiv:1312.6885*.
[21] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015, *arXiv:1511.08458*.
[22] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.
[23] D. Stutz, "Understanding convolutional neural networks," in *Seminar Report, Fakultät für Mathematik, Informatik und Naturwissenschaften Lehr-und Forschungsgebiet Informatik VIII Computer Vision*. 2014. [Online]. Available: https://www.semanticscholar.org/paper/Understanding-Convolutional-Neural-Networks-Stutz/02180e8704cdffb76974da6250127fdfe3dc78e9
[24] T. Tsutsui and T. Matsuzawa, "Virtual metrology model robustness against chamber condition variation using deep learning," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 4, pp. 428–433, Nov. 2019.
[25] M. Maggipinto, A. Beghi, S. McLoone, and G. A. Susto, "DeepVM: A deep learning-based approach with automatic feature extraction for 2D input data virtual metrology," *J. Process Control*, vol. 84, pp. 24–34, Dec. 2019.
[26] X. Wu, J. Chen, L. Xie, L. L. T. Chan, and C.-I. Chen, "Development of convolutional neural network based Gaussian process regression to construct a novel probabilistic virtual metrology in multi-stage semiconductor processes," *Control Eng. Pract.*, vol. 96, Mar. 2020, Art. no. 104262.
[27] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 1989.

TZE CHIANG TIN received the M.Sc. degree in computer science from the University of Malaysia, Sarawak, in 2018. He is currently pursuing the Ph.D. degree with the Faculty of Computing and Informatics, Multimedia University, Malaysia.

SAW CHIN TAN received the M.Sc. degree in information technology from Coventry University, U.K., and the Ph.D. degree in information technology from Multimedia University, Malaysia, in 2008. Since 2002, she has been a Lecturer with the Faculty of Computing and Informatics. In 2008, she became a Senior Lecturer at the Faculty of Computing and Informatics, Multimedia University. Her research interests include software defined networking, optical communication, and ant colony optimization.

CHING KWANG LEE received the B.S. degree from the School of Information Science and the M.Sc. and Ph.D. degrees from the University of Kent, Canterbury, U.K., in 1982 and 1987, respectively. He has been a Chartered Engineer, since 1991. He was a Research Fellow in microwave antennas with a major in frequency-selective surfaces (FSS) at the University of Kent, from 1988 to 1990. From October 1990 to July 1991, he was a Research Scientist with the Electro-Optic Group, Division of Radio Physics, Commonwealth Scientific Industrial Research Organization (CSIRO), Australia. He was a Faculty Member of the School of Electronic Engineering, Nanyang Technological University, Singapore, from July 1991 to July 2010. He has been with FOE, Multimedia University, since November 2010.