Expert Systems With Applications 209 (2022) 118254

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Wafer map failure pattern recognition based on deep convolutional neural network

Shouhong Chen [a,b], Yuxuan Zhang [b], Xingna Hou [b], Yuling Shang [b], Ping Yang [a,*]

[a] Laboratory of Advanced Design, Manufacturing & Reliability for MEMS/NEMS/ODES, School of Mechanical Engineering, Jiangsu University, Zhenjiang 212013, China
[b] Guilin University of Electronic Technology, Guangxi Key Laboratory of Automatic Detecting Technology and Instruments, School of Electronic Engineering & Automation, Guilin 541004, China

ARTICLE INFO

Keywords:
Wafer map
Pattern recognition
Deep convolutional neural networks (DCNN)
Information entropy fusion

ABSTRACT

The objective of this paper is to propose a systematic failure pattern recognition for wafer map based on neural networks. A deep convolutional neural network (DCNN) model which includes the convolutional layer, batch normalization layer, Relu layer, maximum pooling layer, full connection layer, Softmax layer and classification layer is established for the problem of failure pattern recognition of wafer map. This model is an end-to-end 19-layer network, which can actively learn and automatically extract effective classification features. After grayscale and median filtering, the wafer map can be imported into the network for automatic failure classification without special feature extraction. On this basis, we build a dual-source DCNN structure by combining decision-level information entropy fusion. The verification results of the model in the actual wafer map database WM-811K show that the model exhibits good performance in identifying nine kinds of common failure patterns, and has advantages in identifying non-pattern wafer patterns without failure patterns. The dual-source DCNN structure has a better classification effect than the single-source DCNN structure, and the overall recognition accuracy of the dual-source DCNN structure reaches 98.34%.

1. Introduction

The wafer is named for its circular shape. A wafer is a basic silicon wafer for manufacturing a semiconductor element and can be divided into a plurality of chips. They are important fundamental components in integrated circuits (ICs) (Gupta, Ruiz, Fowler, & Mason, 2006; Yang & Qin, 2009; (Liu, Shi, Yang, & Yang, 2018). Increasingly complex chip designs make semiconductor manufacturing processes extremely complex and expensive (Chien, Chang, & Wang, 2014). Even with advanced precision equipment, highly automated production processes, fine monitoring, clean rooms and high-quality professionals, wafers are still prone to defects in semiconductor manufacturing. Wafer defects can result in reduced chip yields (Gupta et al., 2006). Therefore, post-analysis becomes a necessary means to improve wafer yield (García, Sánchez, Rodríguez-Picón, Méndez-González, & Ochoa-Domínguez, 2019).

The wafer map is a database system. The data of the corresponding wafer in the physical parameter test is recorded in each wafer map. We can use the wafer map to analyze and extract defect information in the wafer. Different defect cluster shapes appear in the wafer map (García et al., 2019), which represent different wafer defects, which are the failure patterns presented by the wafer map. It can be traced back to potential problems in the production process. For example, when the wafer map is shown as a random failure pattern, it may be caused by dust and chemical stains on the device. The scratch failure pattern may be a scratch in the manufacturing process. The edge ring failure mode may be caused by the etching problem or uneven temperature distribution during rapid annealing (Xie, Huang, Gu, & Cao, 2014). The wafer map shows that the center failure pattern may be caused by film deposition (Adly, Yoo, Muhaidat, & Al-Hammadi, 2014). The wafer map failure pattern contains manufacturing process information that helps engineers solve potential problems in the manufacturing process in a timely manner, making it easier for engineers to control the production process, stabilize and increase product yield, and reduce production costs Wu, Jang, & Chen, 2015). Effective recognition of wafer map failure patterns is valuable for semiconductor manufacturing.

Traditional wafer map defect pattern recognition is done manually. Experts in this field judge the defect types of the wafer map based on their professional knowledge (Chien et al., 2014), so as to find the root cause of wafer defects. However, this method is very complicated, time-

* Corresponding author.
E-mail addresses: cshgl@guet.edu.cn (S. Chen), hxngl@guet.edu.cn (X. Hou), syl@guet.edu.cn (Y. Shang), yangpingdm@ujs.edu.cn (P. Yang).

https://doi.org/10.1016/j.eswa.2022.118254
Received 29 April 2021; Received in revised form 3 July 2022; Accepted 20 July 2022
Available online 23 July 2022

consuming, and subjective (García et al., 2019; Xie et al., 2014), and the efficiency of identifying defects in the wafer map is not high. As a result, many scholars have studied various automated detection methods, which are faster and more efficient than traditional manual identification methods, and greatly reduce the cost of detection.

Wu et al. Wu et al., 2015) proposed a set of novel rotation and scale-invariant features to obtain a reduced representation of wafer maps, and used Support Vector Machine (SVM) to classify the wafer maps. Compared with region-based modeling methods, and spatial feature analysis methods, the biggest advantage of this method is that it can be used for large-scale data sets. Kang (2018) proposed a joint modeling method to perform classification and regression subtasks separately, and then integrate them into a single prediction model. This method can predict new defect types. Piao, Jin, Lee, and Byun (2018) proposed a decision tree ensemble learning-based wafer map failure pattern recognition method based on radon transform-based features. This method has the capability to aggregate and strengthen the discriminating power of different feature sets. Tan, Watada, Ibrahim, and Khalid (2015) proposes a hybrid classification method for Evolutionary Artificial Neural Networks (EANNs). This method aims to deal with the imbalanced data set problems in a semiconductor manufacturing operations. Liukkonen and Hiltunen (2018) proposed a modeling method combining Self-Organized Map (SOM) and K-means clustering. This method can trace back the individual wafers associated with each cluster or group, a further root causes analysis can be used to identify the primary reasons for bad quality after characterizing the systematically occurring quality patterns. Alawieh, Wang, and Li (2018) proposes a clustering method for identifying failure patterns at the wafer level and successfully extracted system failure patterns. The above-mentioned studies all use feature extraction methods to extract different features from different angles for classification. However, some feature information may be deviated or lost during this process.

With the development of computers, many scholars have applied machine learning methods to the defect pattern recognition and classification of wafer maps. Ishida, Nitta, Fukuda, and Kanazawa (2019) proposed a deep learning-based failure pattern recognition framework. The framework needs only wafer maps with and without target failure patterns to recognize, and ascertains the features of the target failure patterns automatically. Ma, Wang, Xie, Hong, and Mellmann (2021) proposed a construction of a machine learning model to achieve a fast and accurate wafer defect inspection process. Park, Jang, and Kim (2020) propose a class label reconstruction method for subdividing a defect class with various patterns into several groups, creating a new class for defect samples that cannot be categorized into known classes and detecting unknown defects.

Deep convolutional neural networks (DCNN) (Hinton & Salakhutdinov, 2006; Schmidhuber, 2015; Yang, Yang, Qiu, Wang, & Li, 2014; Krizhevsky et al., 2012) is an end-to-end learning model that can learn in enough training data to obtain highly representative hierarchical image features and establish a direct mapping from raw data to corresponding tags. One of the biggest advantages of DCNN over the above methods is that it does not require specific feature extraction. This reduces the time and workload of automatic detection, and also reduces the possibility of generating new features or losing original features. So, the characteristics of the original wafer map can be preserved to improve the accuracy of the recognition. Chen, Zhang, Yi, Shang, and Yang (2021) proposed a method of wafer pattern classification based on image multi-source dual channel convolutional neural network. Weighting different DCNN according to the entropy calculated by the posterior probability of different DCNN can improve the final classification effect. Chen, Liu, Li, and Wang (2018) proposed an adaptive weighted multi-classifier fusion decision algorithm for rail crack recognition. Based on the concept of calculating information entropy according to a posteriori probability in this article and the advantages of DCNN, this paper proposes a method for identifying and classifying failure patterns in wafer maps using a dual-source DCNN. After each DCNN outputs the probability value of the classification category, the method of information entropy fusion at decision level is adopted to improve the accuracy of wafer failure pattern recognition. The main content of the method in this paper is as follows:

(1) Preparation of the wafer map data set. This article mainly focuses on the classification of the eight types (Center, Donut, Edge-loc, Edge-ring, Local, Near-full, Random, and Scratch) of defect wafer maps in the actual database, and a non-defect (None) type wafer maps, a total of nine types of wafer maps.
(2) Wafer map data preprocessing. First, the original wafer map data is converted into grayscale map data. Using single-channel grayscale map data as the input data of the DCNN network can reduce the computational complexity of the network. Then the grayscale wafer map data is subjected to median filtering and denoising, and the discrete noise points in the wafer map are removed, so that the characteristics of the wafer map are more obvious, thereby improving the recognition accuracy. It is worth mentioning that before the wafer map data is input to the DCNN network, there is no need to perform feature separation or extraction on the wafer map.
(3) Establish DCNN classification model. This paper establishes a 19-layer DCNN model consisting of four convolutional layers, four normalization layers, four Relu layers, three maximum pooling layers, a fully connected layer, a Softmax layer, and a classification layer. The model can be directly used to classify the wafer map of nine failure types (Center, Donut, Edge-loc, Edge-ring, Local, Near-full, Random, None, and Scratch). The advantage of this DCNN model is that it has only 19 layers and a relatively deep DCNN network structure, which avoids the risk of overfitting and achieves a faster training speed. The training speed is under 27 min.
(4) Build a dual-source DCNN structure and combine the method of decision-level information entropy fusion to improve the final classification effect. Firstly, the same DCNN structure model is adopted, and the original wafer map and the preprocessed wafer map are used as the input of each DCNN respectively. Then each DCNN outputs a probability vector corresponding to the nine failure types of the wafer map, and the probability values of the two DCNNs are combined at the decision layer. Finally, the method of decision-level information entropy fusion is used to classify the failure patterns of the wafer map.

Through the above steps, the classification of the wafer map can be achieved, and the classification result reaches 98.34% on the WM-811K data set, which proves the effectiveness of the method in this paper. In this paper, the method of dual-source DCNN is better than the classification effect of single source DCNN. Using this DCNN in industrial inspection can improve the classification accuracy of wafer map failure pattern recognition. In this way, it can provide production personnel with information about potential problems in the production process more quickly and accurately, help them solve problems better, and improve wafer production yield.

The rest of this paper is organized as follows: the second part details the proposed method in this paper. The third part is the experimental results and comparison. And the fourth part is a summary of the article.

2. Proposed methods

2.1. Overview

We proposes a dual-source DCNN structure based on decision-level information entropy fusion. First, the original wafer map is converted into grayscale wafer map, and the preprocessed image of the wafer map is obtained by median filtering. Then, the original wafer map and the preprocessed wafer map are used as input to train two DCNN models,

and the probability scores of the predicted labels of the wafer map are obtained. Finally, the information entropy is used as the indicator of fusion to realize the classification of the wafer map. The overall framework of this paper is shown in Fig. 1.

2.2. Median filtering

Median filtering (Tang, Ni, Zhao, & Li, 2018) is a nonlinear signal processing technique based on the statistical theory of sorting, which can effectively suppress noise. Its basic principle is that the gray value of each pixel of the digital image is replaced by the median of the gray values of all the pixels in the neighborhood window of the point, so that the surrounding pixel values are close to the true value. This method eliminates isolated noise points and protects certain details of the image. The principle of median filtering is shown in Fig. 2.

Median filtering methods are commonly used for image de-noising, which is especially useful for speckle noise and salt and pepper noise. This method uses a window of a two-dimensional sliding template structure, and the value in the window is a pixel value, the pixels in the window are sorted according to the size of the pixel value to generate a monotonously rising (or falling) two-dimensional data sequence, and the median is used instead of the original value. The median filter formula can be expressed as follows:

$$P_j = med\{P_{i,j}, P_{i+1,j}, \ldots, P_{i+k-1,j}\} = P_{i+\frac{(k+1)}{2},j} \quad (i,j,k \in Z^+, \quad i = 1,2,\ldots,h-k+1, i \geqslant k) \tag{1}$$

$$P = med\{P_j, P_{j+1}, \ldots, P_{j+k-1}\} = P_{j+\frac{(k+1)}{2}} \quad (j,k \in Z^+, j = 1,2,\ldots,w-k+1, j \geqslant k) \tag{2}$$

In the function, the $\{P_{i,j}, P_{i+1,j}, \cdots, P_{i+k-1,j}\}$ sequence represents a sequence of k adjacent pixel values from the i-th row in the j-th column, which are arranged in ascending or descending order. h is the total number of column pixels of the image. The $\{P_j, P_{j+1}, \cdots, P_{j+k-1}\}$ sequence represents a sequence of k adjacent median pixels arranged in ascending or descending order, and w is the total number of rows of pixels of the image. And k is the size of the domain window, usually with an odd number.

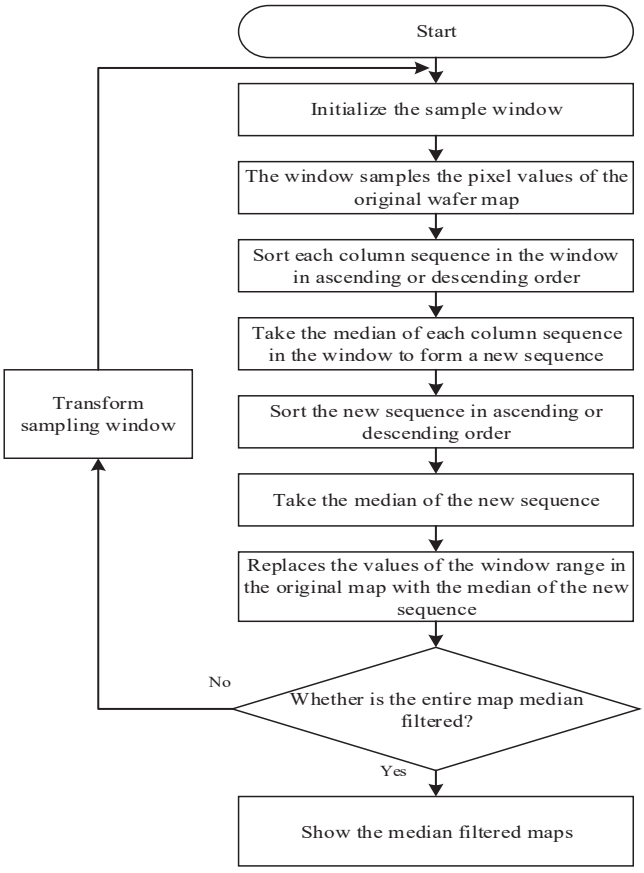


Fig. 2. The principle of median filtering.

2.3. Structure of DCNN

DCNN uses the matrix format to input feature data and uses weights to build a shared network structure, which is very effective in reducing the learning complexity of the network model. In this paper, we use the DCNN model to identify and classify wafer map failure patterns. It trains

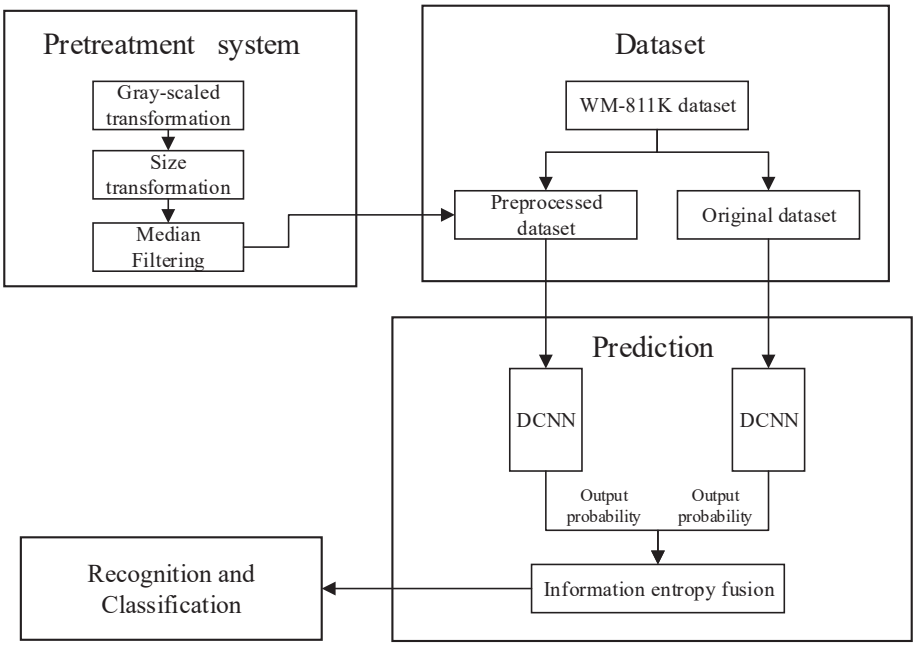


Fig. 1. The failure pattern recognition scheme.

the DCNN with training data and evaluates it with test data. The structure of the DCNN model proposed in this paper is shown in Fig. 3. This model consists of four convolutional layers, four normalization layers, four Relu layers, three max pooling layers, one fully connected layer, one Softmax layer, and one classification layer.

After the wafer map data is input to DCNN, the following operations will be performed:

(1) The input layer imports the wafer map data into the DCNN, and then convolution layer convolutes with the input wafer map data. The convolution layer has 75 convolution kernels, and different features of the input image are fully extracted to obtain 75 feature maps.
(2) The obtained feature maps are batch normalized, and the statistical distribution of the samples is standardized to reduce the difference between samples, stabilize the values, and accelerate the training and convergence speed of the model.
(3) In the Relu layer, a Relu nonlinear excitation function is added to the feature map to help the system express complex features. It uses the max pooling layer to reduce the dimension of the feature map, which reduces the computational complexity.
(4) The max pool size has a window size of 2×2, and its step size is 2. The size of the feature map obtained after the max pool layer is reduced by three-quarters.

After the above steps, the reduced feature map will enter the new convolutional layer, and then enter the batch normalization layer, the Relu layer, and the max pooling layer. However, there will be slight differences in the parameter settings each time feature map enter the layer. After repeating the above operation twice, enter the final convolutional layer. After passing through the normalization layer and the Relu layer again, the feature map enters the fully connected layer to integrate different types of local feature information. And input the feature size, so that the output feature size of the fully connected layer meets the input requirements of the Softmax layer. Finally enter the classification layer for classification. The model can be directly used to classify the wafer map of nine failure types (Center, Donut, Edge-loc, Edge-ring, Local, Near-full, Random, None, and Scratch). The advantage of this DCNN model is that it has only 19 layers and a relatively deep DCNN network structure, which avoids the risk of overfitting and achieves a faster training speed. After the wafer map enters the DCNN, the parameter settings and the change of the feature map are shown in the Table 1.

2.4. Decision-level information entropy fusion

The original wafer map contains all the original feature information, and the preprocessed wafer map can better highlight the main feature pattern of wafer map defects. Therefore, for different forms of images of the same wafer map, DCNN can obtain different features. By fusing the features of different forms of images, the accuracy of wafer map recognition and classification can be improved. There are two ways of feature fusion: feature layer fusion and decision layer fusion. We adopts the decision layer fusion method, as shown in Fig. 4.

The method of decision-level information entropy fusion at the decision layer first obtains the output probability of the corresponding category after the two DCNN networks are classified. Then, the output probabilities are fused by information entropy, and different weights are assigned to the output probabilities of different DCNN models by weighting. Finally, the results are comprehensively evaluated and the final classification results are obtained. The original wafer map and the

Table 1
Parameter settings of the network layers.

Layer	Pooling size	Stride	Feature map	Filter size
convolutional	1	1	64	5×5
batch normalization	\	\	62	\
Relu	\		62	\
max pooling	0	2	31	2×2
convolutional	1	1	29	5×5
batch normalization	\	\	29	\
Relu	\	\	29	\
max pooling	0	2	14	2×2
convolutional	1	1	12	5×5
batch normalization	\	\	12	\
Relu	\	\	12	\
max pooling	0	2	6	2×2
convolutional	1	1	6	3×3
batch normalization	\	\	6	\
Relu	\	\	6	\
Softmax	\	\	\	
classification	\	\		\

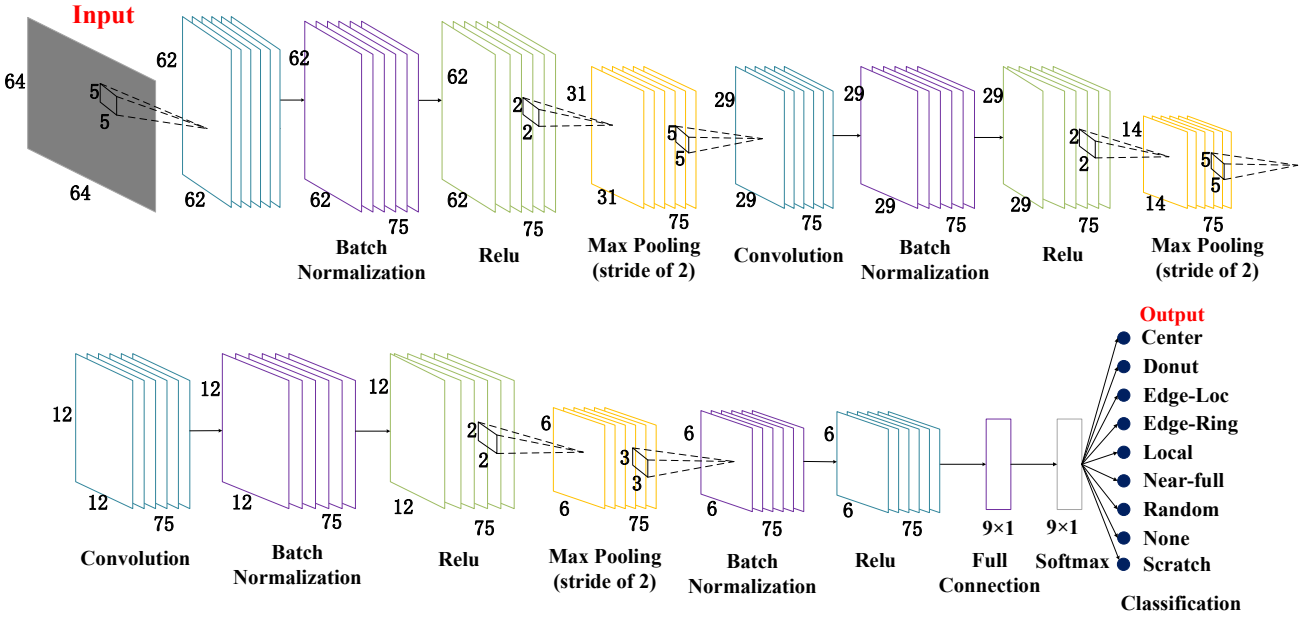


Fig. 3. The proposed DCNN model.

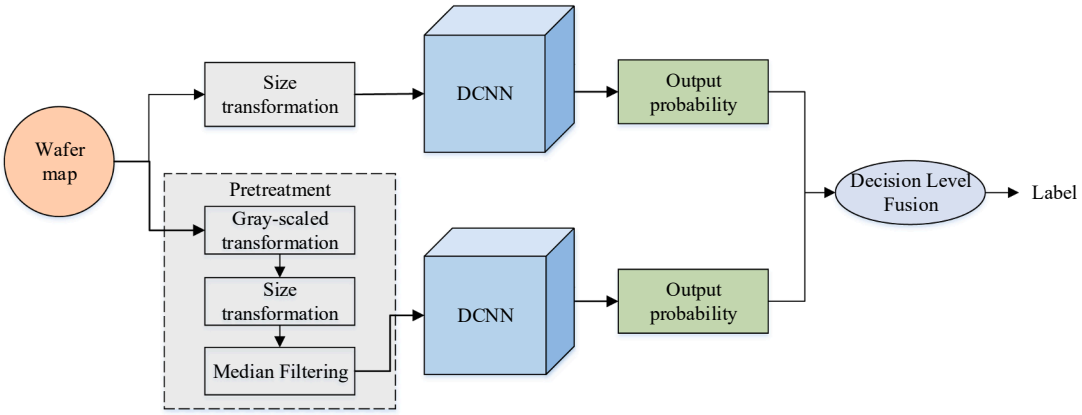


Fig. 4. Failure pattern recognition based on decision-level information entropy fusion.

preprocessed wafer map will respectively obtain a probability vector corresponding to the nine wafer map defect categories after passing through the DCNN network. After the original wafer map and the preprocessed wafer map pass through the DCNN network, a probability vector corresponding to the nine defect categories of the wafer map will be obtained respectively. If the maximum probability obtained by the two DCNN classifications is in the same category, the final classification result does not change, which is consistent with the DCNN classification result. If the maximum probability obtained by the two DCNN classifications is not in the same category, the decision fusion is performed, and the category is finally determined according to the weight distribution.

For the same wafer map, the original wafer map and the preprocessed wafer map are respectively input into the DCNN structure, and two different probability vectors can be obtained, which are combined into a probability matrix as shown in formula (3).

$$P(x) = \begin{bmatrix} P_{11}(x) & P_{12}(x) & \cdots & P_{1n}(x) \\ P_{21}(x) & P_{22}(x) & \cdots & P_{2n}(x) \end{bmatrix} \tag{3}$$

where $\times$ is the input sample, n is the failure category, corresponding to the probabilities of nine failure categories in the wafer map, and each row indicates the output probability prediction of each DCNN for the wafer map. The category corresponding to the maximum probability value in each row is the predicted classification result. If the maximum probability value is much higher than the remaining probability value, the classification result is more reliable. If the maximum probability value is not much different from other probability values, it is impossible to determine which category the classification result belongs to, and the respective uncertainty is large. The uncertainty of classification is expressed using the information entropy H(x) shown in Equation (4).

$$H_i(x) = -\sum_{j}^{n} P_{ij}(x)\log_2 P_{ij}(x), \mathrm{i} = 1, \quad 2 \tag{4}$$

where $P_{ij}(x)$ represents the probability of the i-th DCNN output about the j-th class. The smaller the information entropy, the more information is provided, and the more certainty the classification is. The output probability information entropy of the two DCNNs is calculated, and different weights are assigned to different DCNN. The weight calculation formula is shown in formula (5).

$$W_i = \frac{\exp(-H_i(x))}{\sum_{p=1}^{2}(-H_p(x))} \tag{5}$$

After the weight of each DCNN is obtained, it is multiplied by each row of the probability matrix in equation (3) to obtain a new probability output matrix $P'(x)$ as in equation (6).

$$P'(x) = \begin{bmatrix} w_1 P_{11}(x) & w_1 P_{12}(x) & \cdots & w_1 P_{1n}(x) \\ w_2 P_{21}(x) & w_2 P_{22}(x) & \cdots & w_2 P_{2n}(x) \end{bmatrix} \tag{6}$$

The weighted sum of $P'(x)$ by column can get the probability vector of n categories. The category corresponding to the maximum probability is the final classification result obtained after decision fusion, as shown in formula (7).

$$label(x) = \underset{j=1,2,\ldots,n}{\operatorname{argmax}} \left[\sum_{i=1}^{2} w_i P_{ij}(x) \right] \tag{7}$$

3. Experimental results and discussion

In this section, we identify and classify-nine common failure pattern types in the wafer map according to a database provided by Ming-Ju Wu et al. Also, we compare the proposed method with the existing method. The link of the uploaded code is: https://github.com/Shouhong-Chen1/wafer_map.

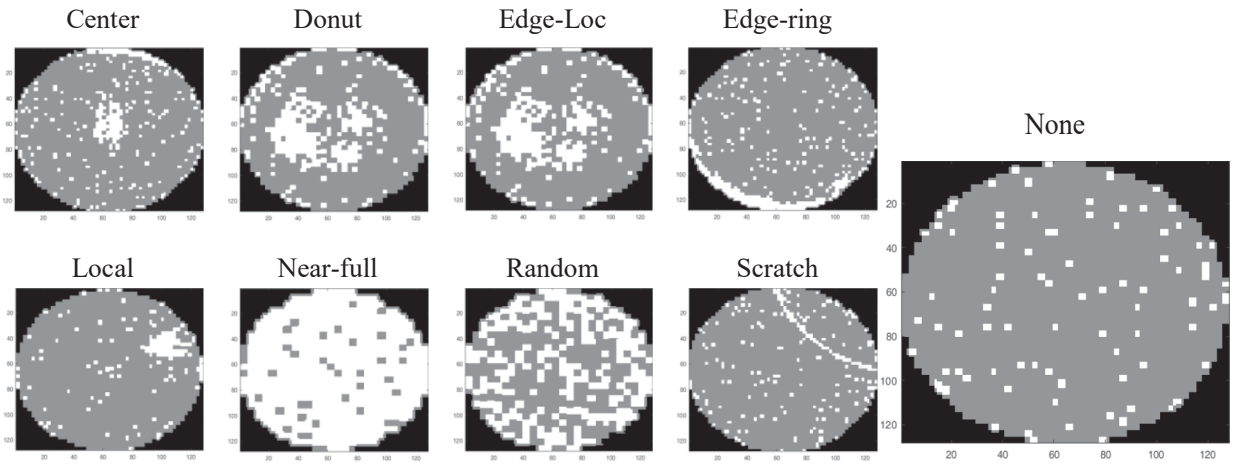
3.1. The database

The database used in this paper is the WM-811K database provided by Wu et al. (2015). The database includes 811,457 wafer maps collected from actual production, of which approximately 20% of the wafer maps have been annotated by experts in the field. The annotated wafer map includes eight types of wafer maps with common failure patterns (Center, Donut, Edge-loc, Edge-ring, Local, Near-full, Random, and Scratch) and one type of non-pattern (None) wafer map. The grayscale patterns of the 9 types wafer maps are shown in Fig. 5. Since all data in the WM-811K database comes from the actual production process, the data in the WM-811K database is unevenly distributed. The data distribution in the data set is shown in Fig. 6.

In this paper, we perform experiments using 53925 annotation wafer maps in the WM-811K training set. The distribution of various data in the experimental data set is shown in Fig. 7. The experiment randomly divides the 53925 annotation wafer maps into 80% and 20% to form a training data set and a test data set, respectively.

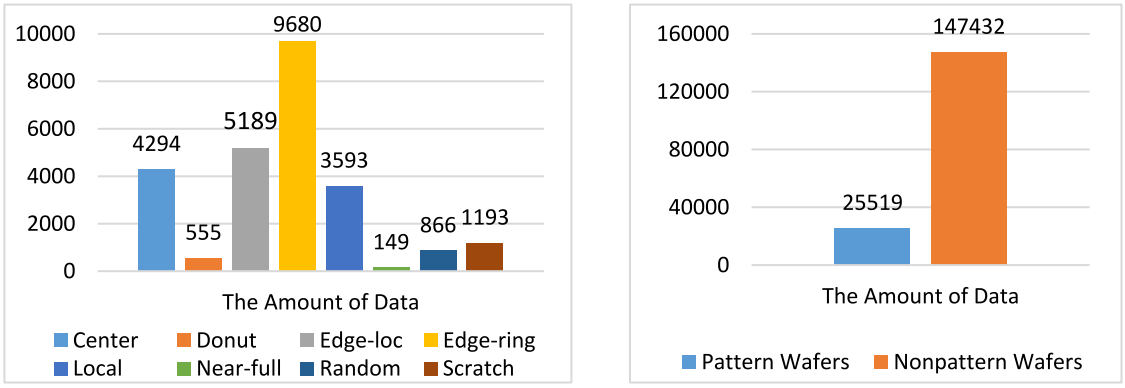
3.2. Pretreatment

The pretreatment process of this article is mainly composed of three parts. First, the color wafer map is converted into a grayscale wafer map, and the 3-channel operation is converted into a single-channel operation, thereby reducing the complexity of subsequent calculations. The second, the grayscale wafer map is converted to a uniform size of 64 $\times$ 64. Finally, the median filtering method is used to enhance the wafer map features without adding other features. The pre-processed wafer



(a) Eight common failure pattern types in the wafer map (b) Nonpattern wafer map

Fig. 5. Wafer map type.



(a) Distribution of wafer map data in various failure patterns in the data set (eight types)

(b) Distribution of wafer map data with failure pattern and nonpattern in the data set

Fig. 6. Data distribution in the WM-811K data set.

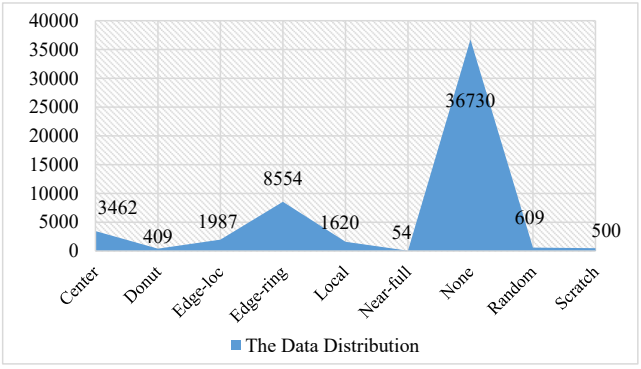


Fig. 7. Distribution of various types of data in experimental data sets.

map is shown in Fig. 8.

3.3. Learning rate and evaluation index

In the process of training the neural network, the learning rate has very important impact on the recognition accuracy and efficiency of the network. A higher learning rate can accelerate the convergence process of the neural network to some extent, but this may cause the gradient to drop too fast and miss the optimal value while the lower learning rate can prevent the network under-fitting to a certain extent, but it will reduce the speed of the neural network convergence process, reduce the network efficiency, and cannot guarantee the high accuracy of the map recognition. The learning rate affects the accuracy and efficiency of DCNN for map recognition. Therefore, choosing the right learning rate is very important.

We use MATLAB to implement the proposed algorithm. We train the proposed DCNN model with a learning rate of 0.5, 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005, 0.0001 and 0.00005, respectively. The number of iterations is 10 times. The training situation of DCNN with different learning rates is compared from the perspective of training time and accuracy. The expression of accuracy is shown in Equation (8), and the training situation of DCNN is shown in Fig. 9.

$$Accuracy = \frac{true\ \ positivess + true\ \ \ negatives}{dataset} \tag{8}$$

As shown in Fig. 9, the accuracy and training time under different learning rates reflect the convergence of the DCNN model. The training

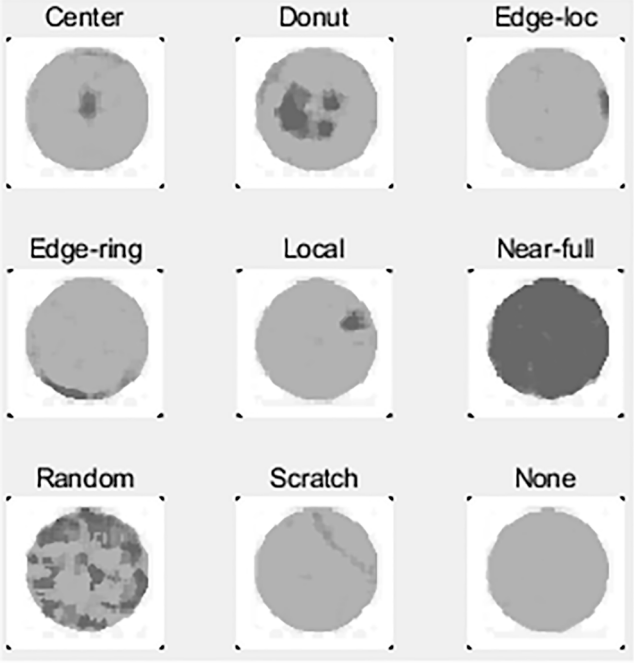


Fig. 8. Pre-processed wafer map.

time at all learning rates is below 27 min. The training speed is under 27 min. When the learning rate is 0.5, the training precision of the DCNN is very stable, and the convergence effect is poor as the number of iterations increases. Moreover, the experimental results show that DCNN cannot correctly identify various failure patterns in the training network when the learning rate is 0.5. At this point, DCNN classifies all map data as non-pattern map. This indicates that the learning rate of 0.5 is too high, so that the proposed DCNN cannot be effectively trained.

When the learning ratios are 0.0001 and 0.0005, from Fig. 8 we can see that in both cases, the training set accuracy of the DCNN is close to 100%. And the DCNN has a test accuracy of 98.2% when learning rate is 0.0005, with high test accuracy and relatively short time required. Therefore, in this paper, we choose the learning rate of 0.0005 to train the proposed DCNN.

In order to verify the validity and universality of the proposed DCNN for wafer map identification, we perform a tenfold cross-validation test on DCNN. It randomly selects 80% of each type of data in the data set as training data, and the remaining 20% data as test data. To evaluate the results, we use the Accuracy, Precision, Recall and F1-Score. The formula is shown in (8)–(11).

$$Precision = \frac{true\ positives}{true\ positives + false\ positives} \tag{9}$$

$$Recall = \frac{true\ positives}{true\ positives + false\ negatives} \tag{10}$$

$$F1 - Score = \frac{2 \times precision \times recall}{precision + recall} \tag{11}$$

3.4. Experiment and result analysis

The DCNN structure in Fig. 3 is used to conduct a tenfold cross-validation experiment on the original wafer maps and the pre-processed wafer maps. Then, by combining the two trained models with different inputs, the structure shown in Fig. 4 is used for decision fusion of wafer map features. The experimental results are as follows.

3.4.1. Comparison of overall accuracy from ten experiments

The tenfold cross-validation experiment was carried out using the above three different methods, and the accuracy of each experiment is shown in Fig. 10.

In Fig. 10, the classification accuracy of using the original wafer map as the input of DCNN can reach 98.20%. From the comparison of ten experiments, it can be seen that if the preprocessed wafer map is used as the input of DCNN, a higher classification accuracy can be obtained than the original wafer map. On this basis, by combining two different input DCNN models and adopting the method based on decision-level information entropy fusion can achieve higher accuracy, with the highest accuracy reaching 98.61%.

3.4.2. Comparison of evaluation index from different failure types from wafer maps

The tenfold cross-validation experiment was carried out using the above three different methods, and the average of the Precision, Recall and F1-Score of ten experiments for each failure type of the wafer map were calculated. The experimental results are shown in Table 2, Table 3 and Table 4.

From the precision index of ten experiments for nine failure types is shown in Table 2, it can be seen that the precision of the method using decision-level information entropy fusion is better than the other two methods in the types of Center, Edge-Loc, Edge-Ring, Local, None, and Scratch. Especially in the Scratch type has a big improvement. The precision of DCNN training classification using preprocessed wafer maps is slightly higher on the types of Donut, Near-full and Random, but not much different on Random types. The overall precision average is better using the method of decision-level information entropy fusion.

From the recall index of ten experiments for nine failure types in Table 3, it can be seen that the recall of the method using decision-level information entropy fusion is better than the other two methods in the types of Center, Edge-Loc, Edge-Ring, Random, and None. Especially in the Random type, the recall rate reaches 100%. The recall of DCNN training classification using preprocessed wafer maps is slightly higher in the types of Local and Scratch. Overall, the recall of the method using

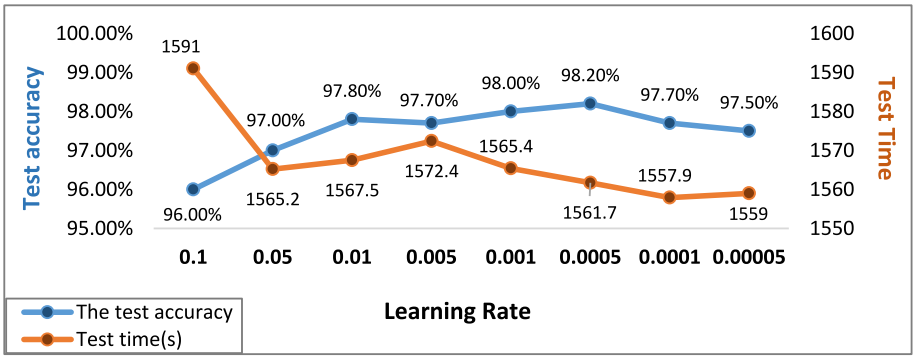
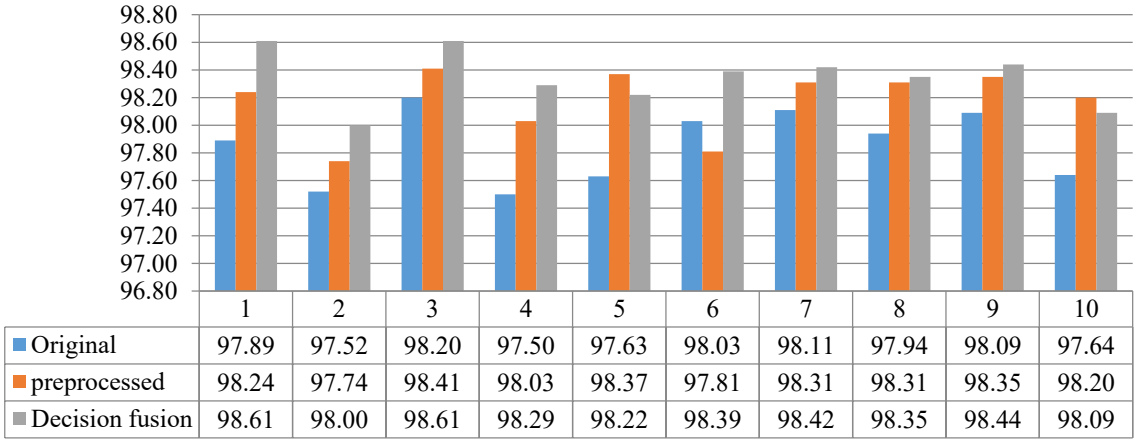


Fig. 9. Test of DCNN model at different learning rates.



	1	2	3	4	5	6	7	8	9	10
Original	97.89	97.52	98.20	97.50	97.63	98.03	98.11	97.94	98.09	97.64
preprocessed	98.24	97.74	98.41	98.03	98.37	97.81	98.31	98.31	98.35	98.20
Decision fusion	98.61	98.00	98.61	98.29	98.22	98.39	98.42	98.35	98.44	98.09

Fig. 10. The accuracy of tenfold cross-validation experiment (%).

Table 2
The Precision of each failure type in 10 cross-validations (%).

	Original	Preprocessed	decision-level information entropy fusion
Center	97.48	98.68	**98.85**
Dounut	92.10	**94.81**	94.02
Edge-Loc	88.42	91.65	**92.22**
Edge-Ring	99.29	99.44	**99.57**
Local	86.66	87.44	**89.80**
Near-full	83.29	**94.64**	93.57
Random	99.00	**99.02**	98.99
None	93.33	96.10	**96.40**
Scratch	71.00	69.69	**76.31**

Table 3
The Recall of each failure type in 10 cross-validations (%).

	Original	Preprocessed	decision-level information entropy fusion
Center	98.59	98.82	**98.96**
Dounut	93.63	90.94	**93.88**
Edge-Loc	87.87	88.92	**89.88**
Edge-Ring	99.10	98.98	**99.32**
Local	78.70	**81.98**	81.91
Near-full	**94.33**	90.33	92.33
Random	99.66	99.91	**100**
None	92.75	94.75	**95.23**
Scratch	50.40	**52.80**	50.40

Table 4
The F1-Score of each failure type in 10 cross-validations (%).

	Original	Preprocessed	decision-level information entropy fusion
Center	98.02	98.74	**98.90**
Dounut	92.61	92.71	**93.91**
Edge-Loc	88.12	90.16	**90.99**
Edge-Ring	99.19	99.21	**99.44**
Local	82.29	84.46	**85.53**
Near-full	87.29	91.68	**92.13**
Random	99.33	99.46	**99.49**
None	92.85	95.37	**95.80**
Scratch	58.24	59.48	**60.17**

decision-level information entropy fusion is better than the other two methods.

From the F1-Score index of ten experiments for nine failure types in Table 4, it can be seen that the method with decision-level information entropy fusion is better than the other two methods on all nine failure types of wafer maps. Therefore, the method using decision-level information entropy fusion proposed in this paper is effective.

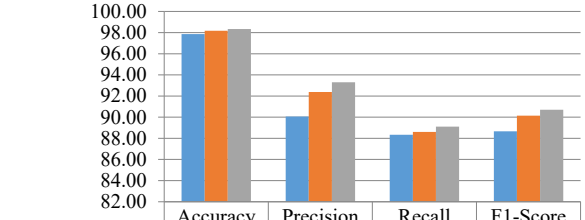
3.4.3. *Comparison from the average of each classification evaluation index*

Calculate the average value of the tenfold crossover experiment and the average value of each failure type of the three wafer map identification and classification methods proposed in this paper, and the overall Accuracy, Precision, Recall, and F1-Score can be obtained as shown in Fig. 10.

From the results in Fig. 11, it can be seen that the overall accuracy of DCNN has reached a high level, and the method using decision-level information entropy fusion is better than using the original wafer map and the preprocessed wafer map for DCNN training classification. Although the overall indicators have not improved much, it can be seen from Table 2, Table 3 and Table 4 that various indicators of types with less data such as Edge-Loc, Scratch and other types have been greatly improved.

3.4.4. *Comparison with other literatures*

Compare the method proposed in this paper with other methods, as shown in Table 5. Compared with other methods, this method identifies nine common failure patterns in the wafer map, which has achieved good results in identifying and classifying most failure patterns, and has high recognition accuracy. In addition, the proposed method has higher recognition accuracy for non-pattern wafer maps and higher overall recognition accuracy. Although the classification accuracy rate in Scratch is low, in general, the recognition and classification effect of the method in this paper is better.



	Accuracy	Precision	Recall	F1-Score
Original	97.85	90.06	88.34	88.66
Preprocessing	98.18	92.39	88.60	90.14
Decision fusion	98.34	93.30	89.10	90.71

Fig. 11. The average results of each classification (%).

Table 5
Comparison of the methods proposed with other methods (%).

Pattern Types	The Paper's method	(Wu et al., 2015)	(Piao et al., 2018)	(Park et al., 2020)
Average	98.34	94.63	90.50	91.20

4. Conclusions

We proposed a method for identifying failure modes of wafer maps based on dual-source deep convolutional neural networks (DCNN). First, a DCNN structure is proposed, which can automatically perform feature selection on input images and realize the recognition and classification of the wafer map. The original image can represent all the feature information of the wafer map, while the grayscaled and median filtered image can better highlight the main feature information of the wafer map. In order to obtain a high accuracy recognition and classification effect, two different forms of wafer maps are input into the DCNN training model respectively, and an output probability value is obtained. Finally, through the method of decision-level information entropy fusion of decision layer, the final classification result is obtained. The experimental results show that the dual-source DCNN structure combined with the decision-level information entropy fusion of the decision layer proposed in this paper has a better recognition and classification effect than the single-source DCNN. The experimental data in this article comes from a real database (WM-811K). The classification accuracy of the method proposed in this paper on this data set has reached 98.34%, which is higher than the classification accuracy of other papers.

Applying the method proposed in this paper to industrial inspection can improve the classification accuracy of wafer map failure pattern recognition. This can provide production personnel with information on potential problems in the production process more quickly and accurately, helping them to better solve problems and improve wafer production yield.

CRediT authorship contribution statement

Shouhong Chen: Methodology, Data curation. **Yuxuan Zhang:** Conceptualization, Formal analysis. **Xingna Hou:** Validation. **Yuling Shang:** Formal analysis. **Ping Yang:** Idea, PI of the project. All authors analyzed data and take part in discussion on conclusion.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

The research work reported in this paper is supported by the National Natural Science Foundation of China (Nos. 51465013, 51975262, 61661013), Guangxi Natural Science Foundation (No. 2015GXNSFDA139003), Guangxi Key Laboratory of Automation Test and Instrumentation (No. YQ22102), Innovation Project of Guet Graduate Education (No. 2019YCXS085) the High Performance Computing Platform of Jiangsu University during the course of this work.

References

Adly, F., Yoo, P. D., Muhaidat, S., & Al-Hammadi, Y. (2014). Machine-Learning-Based Identification of Defect Patterns in Semiconductor Wafer Maps: An Overview and Proposal. *IEEE International Parallel & Distributed Processing Symposium Workshops, 2014*, 420–429. https://doi.org/10.1109/IPDPSW.2014.54

Alawieh, M. B., Wang, F., & Li, X. (2018). Identifying Wafer-Level Systematic Failure Patterns via Unsupervised Learning. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 37*, 832–844. https://doi.org/10.1109/TCAD.2017.2729469

Chien, C. F., Chang, K. H., & Wang, W. C. (2014). An empirical study of design-of-experiment data mining for yield-loss diagnosis for semiconductor manufacturing. *Journal of Intelligent Manufacturing, 2014*(25), 961–972. https://doi.org/10.1007/s10845-013-0791-5

Gupta, J. N. D., Ruiz, R., Fowler, J. W., & Mason, S. J. (2006). Operational planning and control of semiconductor wafer production. *Production Planning & Control, 17*, 639–647. https://doi.org/10.1080/09537280600900733

García, V., Sánchez, J. S., Rodríguez-Picón, L. A., Méndez-González, L. C., & Ochoa-Domínguez, H. J. (2019). Using regression models for predicting the product quality in a tubing extrusion process. *Journal of Intelligent Manufacturing, 30*, 2535–2544. https://doi.org/10.1007/s10845-018-1418-7

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science, 313*, 504–507. https://doi.org/10.1126/science.1127647

Ishida, T., Nitta, I., Fukuda, D., & Kanazawa, Y. (2019). Deep Learning-Based Wafer-Map Failure Pattern Recognition Framework. In *20Th International Symposium on Quality Electronic Design (ISQED)* (pp. 291–297).

Kang, S. (2018). Joint modeling of classification and regression for improving faulty wafer detection in semiconductor manufacturing. *Journal of Intelligent Manufacturing, 31*. https://doi.org/10.1007/s10845-018-1447-2

Krizhevsky A., Sutskever I., &Hinton G.E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems, Curran Associates Inc., 25, 1097-1105.

Liu, H., Shi, S., Yang, P., & Yang, J. (2018). An Improved Genetic Algorithm Approach on Mechanism Kinematic Structure Enumeration with Intelligent Manufacturing. *Journal of Intelligent & Robotic Systems, 2018*(89), 343–350. https://doi.org/10.1007/s10846-017-0564-z

Liukkonen, M., & Hiltunen, Y. (2018). Recognition of Systematic Spatial Patterns in Silicon Wafers Based on SOM and K-means. *IFAC-Papers OnLine, 51*, 439–444. https://doi.org/10.1016/j.ifacol.2018.03.075

Ma, Y., Wang, F., Xie, Q., Hong, L., & Mellmann, J. (2021). Machine learning based wafer defect detection. *IEEE Transactions on Semiconductor Manufacturing, 34*, 161–167.

Park, S., Jang, J., & Kim, C. O. (2020). Discriminative feature learning and cluster-based defect label reconstruction for reducing uncertainty in wafer bin map labels. *Journal of Intelligent Manufacturing, 32*, 251–263. https://doi.org/10.1007/s10845-020-01571-4

Piao, M., Jin, C. H., Lee, J. Y., & Byun, J. Y. (2018). Decision Tree Ensemble-Based Wafer Map Failure Pattern Recognition Based on Radon Transform-Based Features. *IEEE Transactions on Semiconductor Manufacturing, 31*, 250–257.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Chen, S., Zhang, Y., Yi, M., Shang, Y., & Yang, P. (2021). AI classification of wafer map defect patterns by using dual-channel convolutional neural network. *Engineering Failure Analysis, 130*, 1–14. https://doi.org/10.1016/j.engfailanal.2021.105756

Chen, W., Liu, W., Li, K., & Wang, P. (2018). Rail crack recognition based on adaptive weighting multi-classifier fusion decision. *Measurement, 123*, 102–114. https://doi.org/10.1016/j.measurement.2018.03.059

Tang, H., Ni, R., Zhao, Y., & Li, X. (2018). Median filtering detection of small-size image based on CNN. *Journal of Visual Communication and Image Representation, 51*, 162–168. https://doi.org/10.1016/j.jvcir.2018.01.011

Tan, S. C., Watada, J., Ibrahim, Z., & Khalid, M. (2015). Evolutionary Fuzzy ARTMAP Neural Networks for Classification of Semiconductor Defects. *IEEE Transactions on Neural Networks and Learning Systems, 26*, 933–950. https://doi.org/10.1109/TNNLS.2014.2329097

Wu, M. J., Jang, J. R., & Chen, J. L. (2015). Wafer Map Failure Pattern Recognition and Similarity Ranking for Large-Scale Data Sets. *IEEE Transactions on Semiconductor Manufacturing, 28*, 1–12. https://doi.org/10.1109/TSM.2014.2364237

Xie, L., Huang, R., Gu, N., & Cao, Z. (2014). A novel defect detection and identification method in optical Inspection. *Neural Computer & Applications, 24*(7-8), 1953–1962. https://doi.org/10.1007/s00521-013-1442-7

Yang, P., & Qin, X. (2009). A hybrid optimization approach for chip placement of multi-chip module packaging. *Microelectronics Journal, 40*, 1235–1243. https://doi.org/10.1016/j.mejo.2009.05.002

Yang, P., Yang, H., Qiu, W., Wang, S., & Li, C. (2014). Optimal approach on net routing for VLSI physical design based on Tabu-ant colonies modeling. *Applied Soft Computing, 21*, 376–381. https://doi.org/10.1016/j.asoc.2014.03.033