Engineering Failure Analysis 130 (2021) 105756

Contents lists available at ScienceDirect

Engineering Failure Analysis

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

AI classification of wafer map defect patterns by using dual-channel convolutional neural network

Shouhong Chen [a,b], Yuxuan Zhang [b], Mulan Yi [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] *Guangxi Key Laboratory of Automatic Detecting Technology and Instruments, School of Electronic Engineering & Automation, Guilin University of Electronic Technology, Guilin 541004, China*

ARTICLE INFO

Keywords:
Wafer map
Pattern recognition
Two-channel DCNN
Multi-source
ECOC-SVM

ABSTRACT

In semiconductor manufacturing, the presence of abnormal chips in wafer products will reduce the yield of wafer products. The identification of the wafer map can find related problems in the wafer production process, and post-analysis is a necessary means to improve the wafer yield. In this study, we proposed the method that image-based multi-source and two-channel convolutional neural network for wafer map classification. This method consists of the following three main steps. First, using two different deep convolution neural networks (DCNN) to form a two-channel DCNN feature extraction model and extracting multiple sets of advanced features from multi-source data. Second, processing the multi-group data features extracted from different channels to obtain two sets of new multi-source features. Third, the multi-source features in the two channels are further processed to become a new set of classification features. Then input the new features into the combined classification model of error correction code and support vector machine (ECOC-SVM) to classify the defect pattern of the wafer map. The data used in the experiment comes from data sets in actual production (WM-811K). The experimental results showed that the proposed method has a good effect on the defect pattern recognition of wafer maps, and the effectiveness of the proposed method is confirmed. There are a total of 33,256 wafer maps, and the overall classification accuracy of 6552 test sets has reached 96.4%.

1. Introduction

With the development of semiconductor technology, the design of integrated circuits is becoming more and more complicated. Semiconductor manufacturing is developing towards maximization of scale and miniaturization of process dimensions. Today's wafer production has a highly automated production process and high-precision production equipment. However, the manufacture of semiconductor wafers is complicated and time-consuming, involves many complicating chemical steps, and needs to monitor a large number of process parameters. Even if regulated by professionals, abnormal dies will inevitably appear in the wafer [1,2]. Yield rate refers to the qualified rate of each batch of products, which has always been attracting attention in the semiconductor manufacturing industry. The increase in yield means a reduction in production costs and an increase in profit margins, which are pivotal factors for IC companies to maintain their core competitiveness in the market. Abnormal problems in any step of the wafer manufacturing process

* Corresponding author.
E-mail address: yangpingdm@ujs.edu.cn (P. Yang).

https://doi.org/10.1016/j.engfailanal.2021.105756
Received 14 January 2021; Received in revised form 7 September 2021; Accepted 15 September 2021
Available online 20 September 2021

will cause defective dies on the wafer [3,4], such as film deposition problems, etching problems [5], uneven cleaning [6], uneven ultraviolet exposure, and scratches during material transportation [7]. These problems will reduce the yield of wafer products and increase production costs.

Wafer maps are the analysis and extraction of defect information in the wafer [8]. Abnormal dies in the wafer often exist in the form of clusters [9], showing a specific spatial pattern in the wafer map [10], which means there is a specific defect pattern in the wafer map. Identifying the defect patterns of these wafer maps can obtain information about abnormalities in the trace wafer production process [11], and provide clues for solving process failures in semiconductor manufacturing. For example, the Center type defect is usually caused by the change in the speed of the spinner during the chemical–mechanical flattening process or the uneven temperature during the oxidation process, while the Edge-Loc type defects are general caused by filming deposition problems. And the Edge-ring type defects are usually caused by etching or layer-to-layer alignment problems. Analyzing the defect mode of the wafer map can correspond to find out abnormal situations that may lead to low yield to maintain the quality level of the process. Therefore, post analysis has become a necessary means to improve wafer yield. The post analysis can accurately classify the defect patterns of the wafer map and quickly identify the root cause of the failures. With the rapid improvement of modern computing power, the application of automatic inspection methods based on machine learning to semiconductor production can improve efficiency and reduce costs. Therefore, there is no doubt use machine learning methods to establish a fast and effective automatic recognition and classification system. It is of great significance for efficiently identifying defect patterns of classified wafer maps, controlling potential problems in the wafer manufacturing process, and stabilizing and improving the yield.

At present, many experts and scholars have researched the classification of wafer map defect patterns. Among them, Yu et al. (2015) [12] proposed a method of using joint local and non-local linear discriminant analysis (JLNDA) to extract useful information from various features, which can identify new defect patterns. Piao et al. (2018) [13] proposed an integrated learning method of decision tree based on Radon feature for wafer pattern recognition. Fan et al. (2016) [11] proposed the OPTICS method and support vector machine (SVM) to deal with the multi-label classification problem of wafer maps. Saqlain et al. (2019) [6] proposed a voting integration method based on density, geometry, and Radon feature classification to identify defect patterns in classified wafer maps. These methods are all based on feature extraction to identify and classify the defect patterns of wafer maps. At this time, the classification results to a great degree depend on the effectiveness of the extracted features. However, in most cases, it is impossible to determine which kind of feature of the wafer map is the most representative and effective. In response to this problem, a study on the application of deep learning methods to the defect pattern recognition of the wafer map has emerged. Convolutional neural network (CNN) is a popular deep learning method [14,15], which can automatically extract meaningful features from the original input without manual feature extraction. Therefore, CNN is also used in many fields, for example, Ihor. K et al. (2018) [16] used the CNN method to study the rupture surface of the titanium alloy. Not only that, Ihor. K et al. (2019) [17] also proposed a CNN structure to automatically detect and calculate the quantitative parameters of ductile fracture dimples on digital images of fractures of different scales. In the field of wafer map research, Kyeong et al. (2018) [18] proposed to build a separate CNN model for each defect type and classify the wafer maps of mixed defect modes. Nakazawa et al. (2019) [19] established a CNN model using synthetic wafer maps, applied it to the classification of actual wafer maps, and used a convolutional auto-encoder to detect invisible wafer mapping patterns. Jin et al. (2019) [33] proposed a method for classifying wafer map defect patterns based on images combined with CNN and ECOC-SVM and obtained good classification results.

Although the ECOC-SVM classification result is improved compared to the direct classification of the classification layer of the CNN, the features extracted by the CNN are also richer than those extracted manually, but there is still a problem. It is that the features extracted by the CNN still have a certain degree of randomness, which can easily lead to the instability of the classification results. Therefore, this article combines multi-source data (three kinds of data) and uses a dual-channel DCNN to extract features from many aspects. After different combinations of these features, the ECOC-SVM [20,21] classification method is used to classify the combined features. The detailed steps of the method are as follows:

1. By using the two different DCNN models mentioned in this article to extract features from the original image data, and can obtain two sets of features of the same data. The artificial feature processing method of taking the maximum value, minimum value, average value, and direct splicing value of the two sets of data features of the dual-channel can obtain four new sets of features. Use ECOC-SVM classification for the four groups of features, and compare the effects of the four groups of feature classification and DCNN direct classification.
2. A DCNN model performs feature extraction on the three types of multi-source data to obtain three sets of data features. In the multi-source data, the grayscale of the original wafer map and two derived wafer map of the original wafer map (SA-DBSCAN wafer map and its intermediate image Auto-DBSCAN wafer map) proposed in the article [22] are used. The artificial feature processing method of taking the maximum value, minimum value, average value, and direct-splicing value of these three sets of data features can obtain four new sets of features. Eight groups of features can be obtained under two different DCNN models. Using ECOC-SVM to classify the eight groups of features and compare their classification effects, the classification results of the features after the average and direct splicing processing have obvious advantages compared with the maximum and minimum features.
3. Combine the features of dual-channel and multi-source data, and process the features of multi-source data acquired through two different DCNN models. In the two DCNN models, six sets of features can be obtained by taking the average of the three types of data features of the multi-source data and the result of direct splicing. Then take the maximum, minimum, average, and direct-splicing value features of the average and direct-splicing features of the three types of data to obtain eight sets of features. Finally, compare the effects of the eight groups of feature classification by using the ECOC- SVM classification.

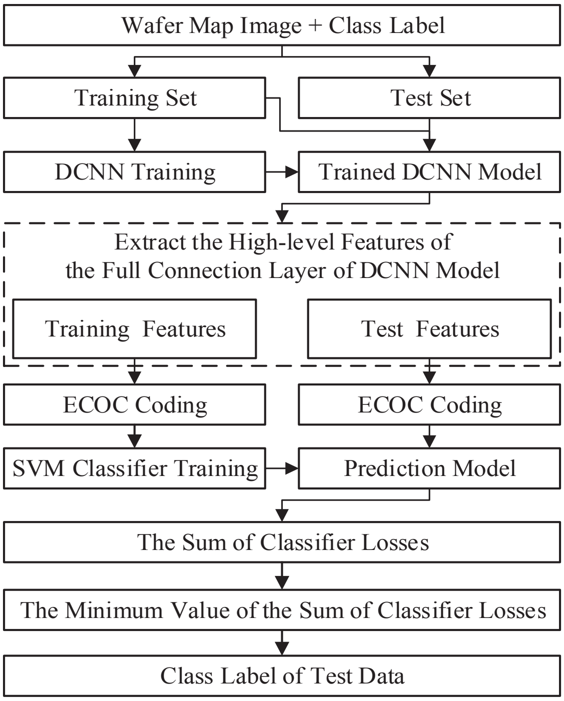


Fig. 1. DCNN + ECOC-SVM classification flowchart.

The rest of the paper is organized as follows. The Methods and Model section gives the framework of the method. The Experiments and Results section reports the experimental results of the proposed method. Finally, the work content of this paper is summarized in the Conclusion section.

2. Methods and model

2.1. DCNN + ECOC-SVm

With the development of computer technology, when the computer is fast enough, deep learning has higher performance in object recognition and has an advantage in large data set samples. It is widely used in the field of computer vision research. As one of the important network models in deep learning, DCNN [23,24] can actively learn advanced image features from sufficient training data and establish a direct mapping from original image data to corresponding labels. In addition, the DCNN network does not require specific feature engineering, can reduce interference, retain useful original image features as much as possible, improve recognition accuracy, and show good performance in image analysis. The DCNN model mainly includes an input layer, an output layer, and a hidden layer located between the two layers. Among them, the hidden layer is mainly composed of multiple convolutional layers, normalization layers, pooling layers, and fully connected layers with varying numbers. The fully connected layer can integrate multiple feature maps input by the upper network layer to obtain effective classification features to classify the input image. Therefore, high-level features of input data based on DCNN classification can be extracted from the fully connected layer. The Error Correction Output Coding Algorithm (ECOC) [25] has the feature of automatic error correction, which can reduce variance and bias errors and improve classification accuracy. ECOC is often used to solve multi-classification problems, which is solved by decomposing multi-classification problems into multiple two-class classification problems [23,26]. SVM is a two-class linear classifier based on statistical learning theory. The kernel function allows SVM to be used to solve nonlinear problems. The basic idea is to correctly divide the hyperplane of the positive and negative data in the training set and maximize the interval between the two data. ECOC-SVM is to use the SVM classifier as the basic classifier of ECOC to build a classification model.

Fig. 1 shows the DCNN + ECOC-SVM classification flowchart used in this article. Use the DCNN model to automatically extract the features of the wafer maps, and input the advanced features derived from the fully connected layer of the DCNN model into the ECOC-SVM model to train an 8-class wafer maps defect pattern classification model. The ECOC-SVM model compares the loss sum of the SVM classifier and selects the category with the smallest loss sum as the sample prediction value of the model.

The minimum loss sum in the ECOC-SVM model is defined as:

$$\widehat{k} = argmin_k \frac{\sum_{l=1}^{L} |m_{kl}| g(m_{kl}, s_l)}{\sum_{l=1}^{L} |m_{kl}|} \tag{1}$$

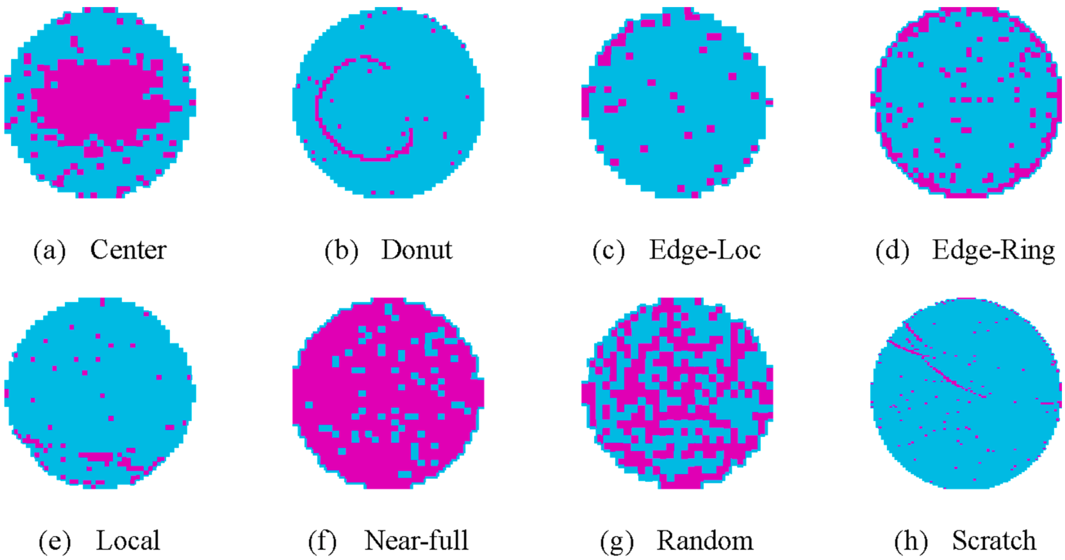


Fig. 2. Eight types of wafer map defects.

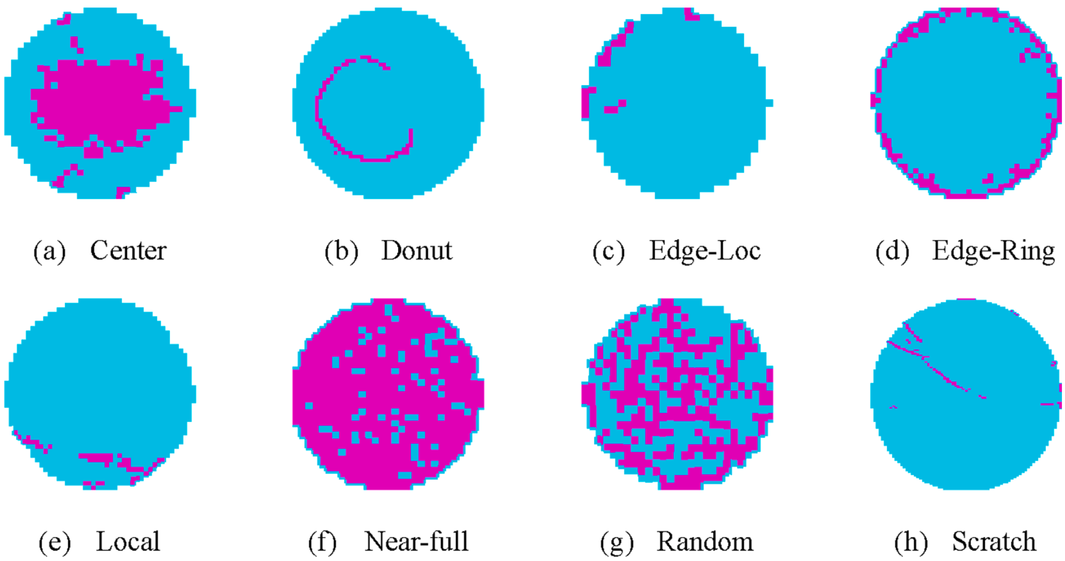


Fig. 3. Auto-DBSCAN wafer map.

where m_{kl} is the element in the k row and l column of $M_{3\times3}$, S_i is the score of L_i giving the predicted classification, and $g(m_{kl}, s_l)$ is the loss function.

2.2. Multi-source

This article uses the large-scale real wafer map data set (WM-811k) provided by Wu et al. [3]. The 8 defect pattern types of Center, Donut, Edge-loc, Edge-ring, Local, Near-full, Random, and Scratch are classified and studied in the data set. Fig. 2 shows the wafer map of the 8 types of defect modes in the WM-811K studied in this paper.

The derivative image of the original image obtained by a certain method can highlight some characteristics of the image. Using the multi-source data features composed of the original wafer maps and its derived wafer maps to classify the defect pattern of the wafer maps can make the classification features include more effective image features and improve the effect of image recognition and classification. The multi-source data used in this paper includes the original wafer map and the grayscale wafer map of its two derived images (Auto-DBSCAN wafer map and SA-DBSCAN wafer map) [20]. Fig. 2, Fig. 3, and Fig. 4 are the original wafer maps, Auto-DBSCAN wafer maps, and SA-DBSCAN wafer maps of the 8 types of defect modes used in this article.

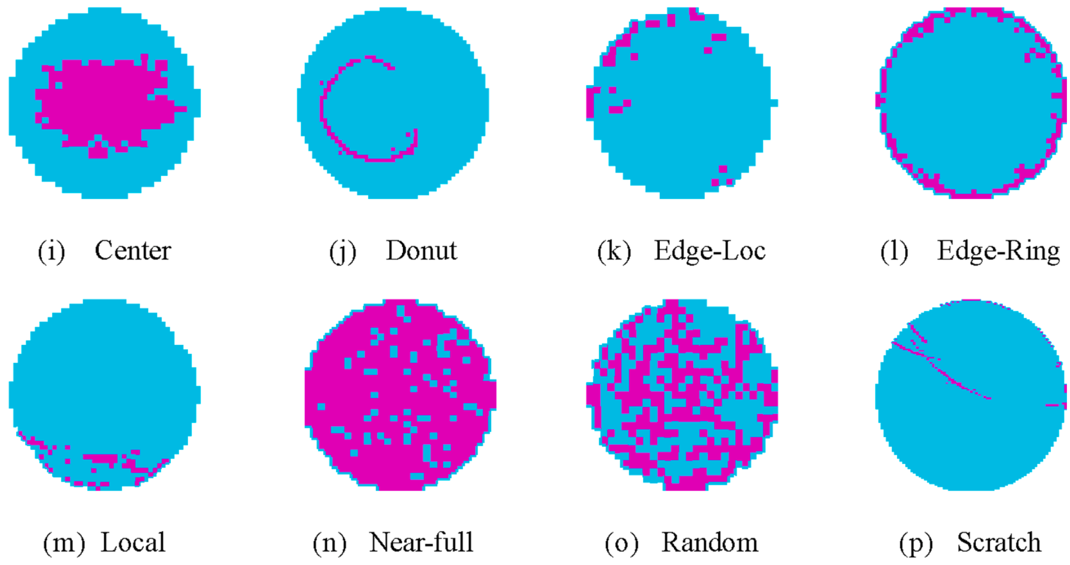


Fig. 4. SA-DBSCAN wafer map.

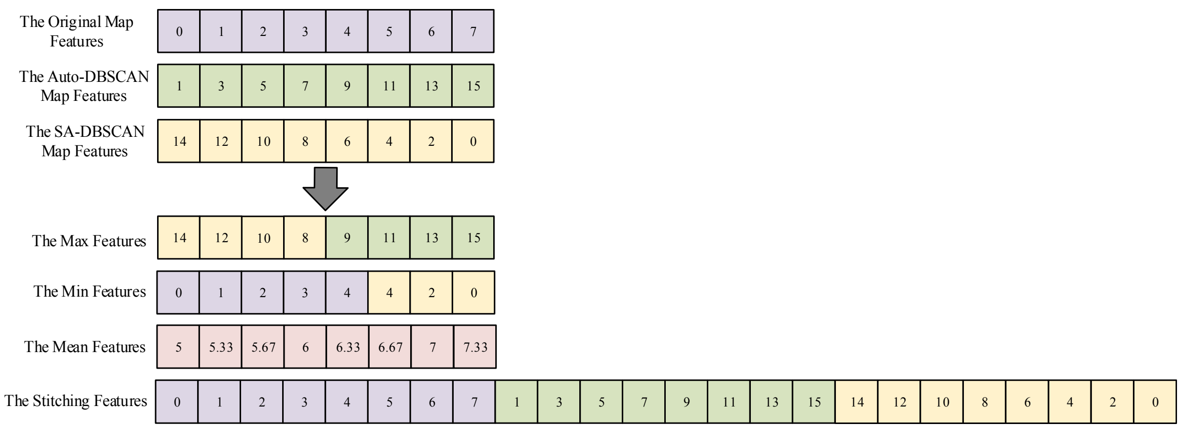


Fig. 5. Diagram of multi-source feature combination.

The SA-DBSCAN wafer maps are a derivative image of the original wafer map based on the adaptive DBSCAN method. The Auto-DBSCAN wafer maps are an intermediate image obtained in the process of obtaining the SA-DBSCAN wafer maps. Obtain these two kinds of wafer maps mainly through the following two steps:

1. Obtain the optimal DBSCAN parameter value pair (Eps, MinPts) of the wafer map through the KANN-DBSCAN algorithm proposed by Li Wenjie et al. [27], the mathematical expectation method, and the optimal parameter selection method set in the article [22]. Where Eps is the minimum neighborhood radius of the core point in the DBSCAN algorithm, and MinPts is the minimum number of clustering points contained in the Eps-neighbor of the core point in the DBSCAN algorithm. At this stage, the original wafer map is processed by the optimal parameter DBSCAN to obtain an Auto-DBSCAN map. The Auto-DBSCAN wafer map removes part of the isolated defect points to better highlight the defect point cluster pattern in the wafer map.
2. Use the method given in article [22] to perform secondary selection and retention of the clusters and noise points divided by DBSCAN to obtain the SA-DBSCAN wafer map containing the main spatial patterns of the wafer map defect mode. The SA-DBSCAN wafer map only retains clusters and scatter points with similar properties to the largest cluster in the wafer map, which can better highlight the main spatial patterns of the wafer map defect mode.

After the above processing, we get three different forms of wafer maps for each wafer map, namely original wafer map, Auto-DBSCAN wafer map and SA-DBSCAN wafer map. The reasons for choosing these three different forms of wafer map are as follows:

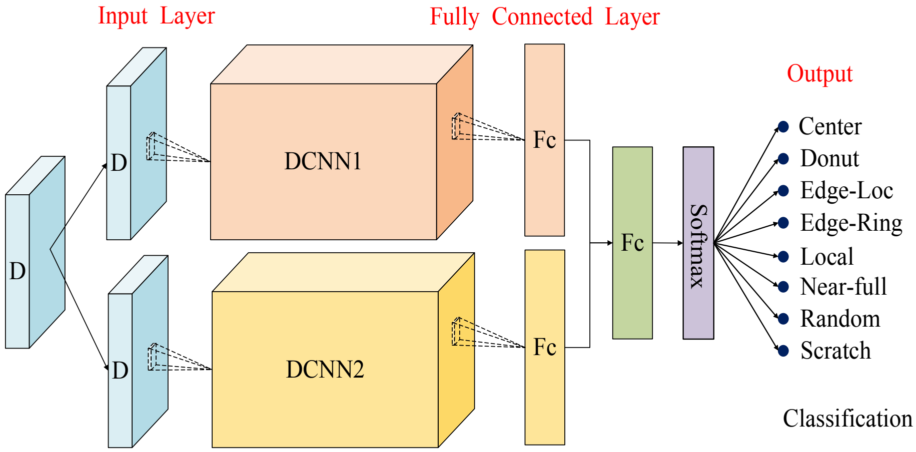


Fig. 6. The Framework of Dual-Fc-DCNN.

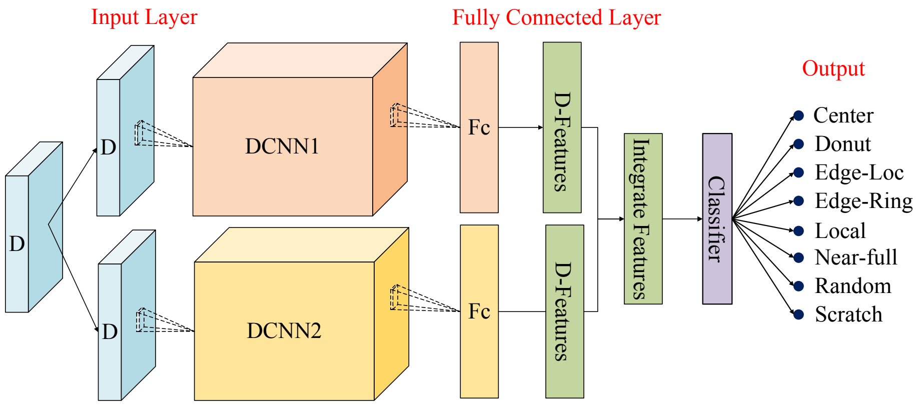


Fig. 7. The Framework of Dual-Int-DCNN.

(1) The original wafer map can show a complete wafer defect pattern.
(2) The Auto-DBSCAN wafer map is the image after removing the noise points. At this time, the features are more prominent than the original wafer map. Although there may be some noise points, it also retains the defect features more completely.
(3) The SA-DBSCAN wafer map may have some loss of defect features, but it can also highlight the main feature defect patterns in the wafer map.

By acquiring features from the three different forms of images of the wafer map, the complete characteristics of the wafer map can be obtained from more aspects.

Fig. 5 is a schematic diagram of the processing method proposed in this paper for multiple sets of different source data features of the wafer map. Processing methods include taking the maximum, minimum, and average values of multiple sets of different source data features, and directly splicing multiple sets of different source data features. This paper uses the DCNN model to extract features from various data sources in the multi-source data and integrates the three corresponding features of the sample to obtain a set of multi-source features. Then use ECOC-SVM classifier to study the effect of these multi-source feature recognition and classification.

If the DCNN model is used to extract features from the grayscale image of the original wafer maps, Auto-DBSCAN wafer maps, and SA-DBSCAN wafer maps, one sample can obtain three sets of 8-dimensional features (the Original Map Features, the Auto-DBSCAN Map Features, and the SA-DBSCAN Map Features) from different data sources as shown in Fig. 5. Take the maximum values, minimum values, average values, or direct-splicing values of the three groups of features at the corresponding positions to obtain a new set of multi-source features (the max features, the min features, the mean features, or the stitching features). When the maximum value, minimum value, or average value is taken, the integrated multi-source feature is still an 8-dimensional feature, which is the same as the dimension of a single sample feature. However, when the direct splicing value is taken, the new multi-source feature is a 24-dimensional feature composed of the original wafer map features, the Auto-DBSCAN wafer map features, and the SA-DBSCAN wafer map features from the front to back. The internal order of the features of the original wafer map, Auto-DBSCAN wafer map, and SA-DBSCAB wafer map remains unchanged, and the multi-source feature dimension at this time is three times the feature dimension of a single sample.

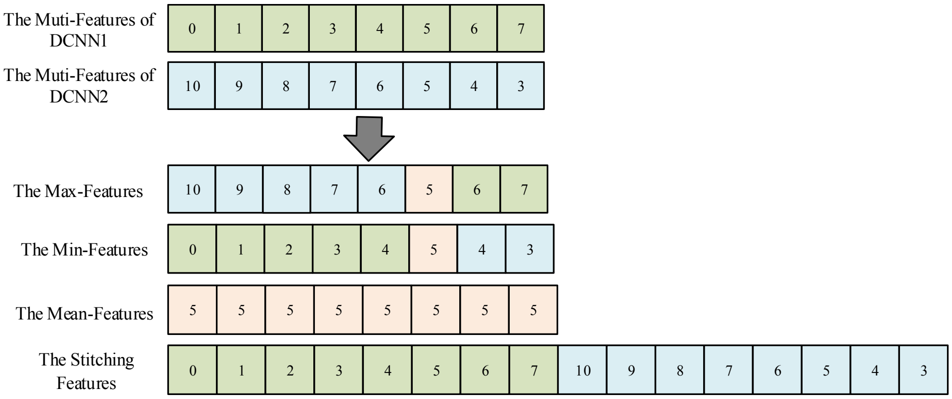


Fig. 8. Diagram of the Integrate Features of the Dual-Int-DCNN Model.

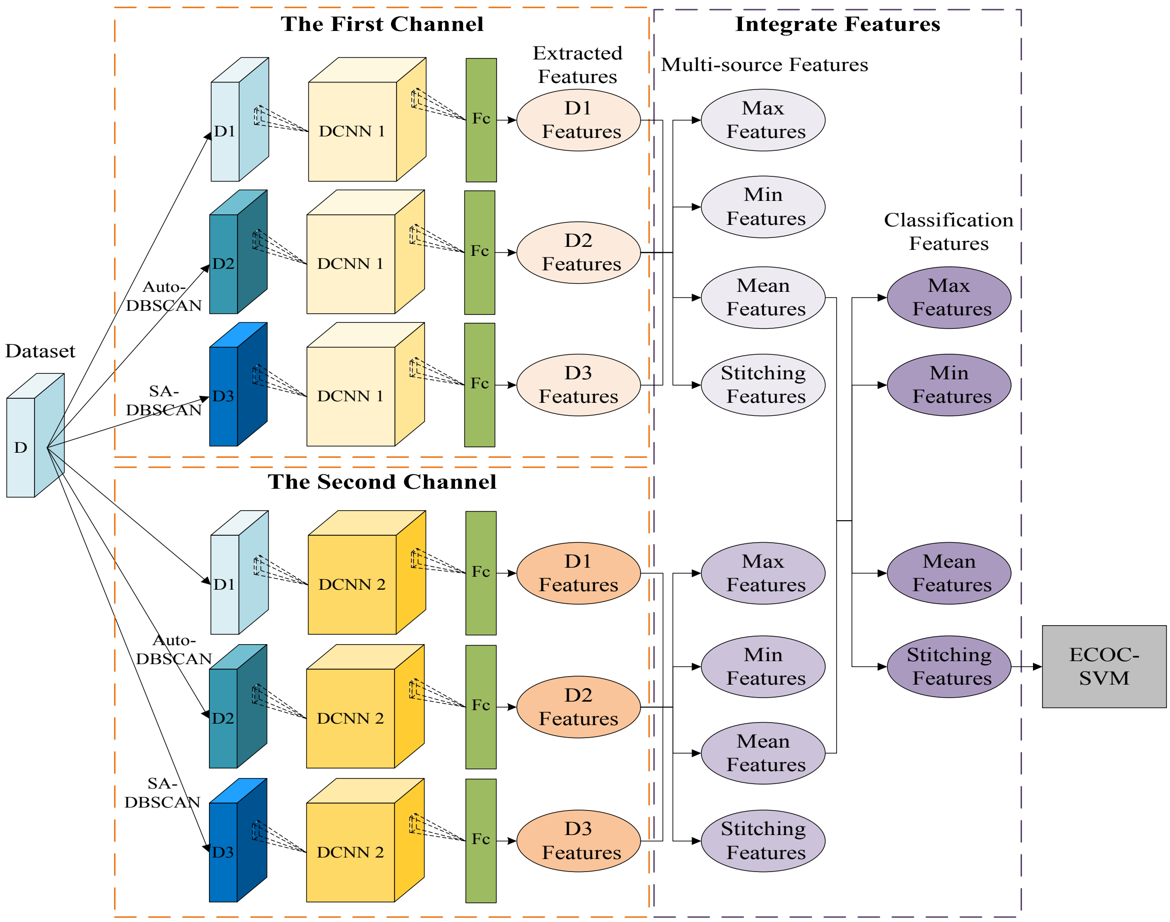
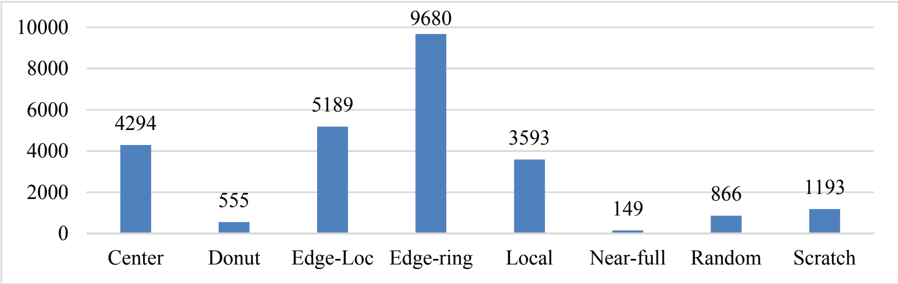


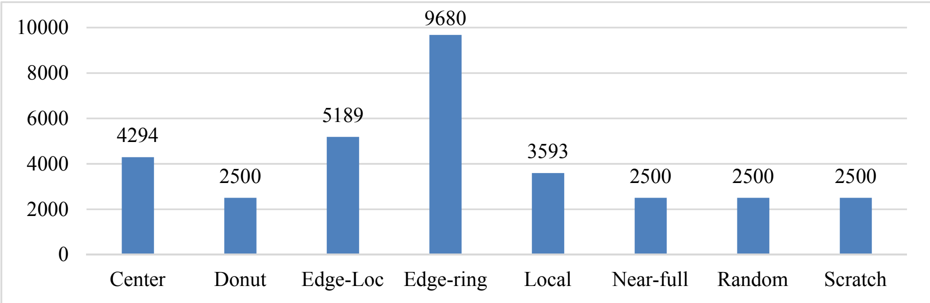
Fig. 9. Overall experimental Framework.

2.3. *Dual-channel*

Two different DCNN models are used to obtain different features of the image from different angles with different “eyes” so that the classification features contain more effective information and improve the recognition and classification effect of the wafer map. There are two ways to integrate the features of two DCNN models.



(a) Wafer map distribution of 8 types of defect modes in labeled WM-811K



(b) Wafer map distribution of 8 types of defect modes of Used dataset

Fig. 10. Data distribution of wafer map.

1. Use the new fully connected layer to connect the fully connected layers of the two DCNN models to integrate the features in the two DCNN models to form a new DCNN model and then classify, as shown in Fig. 6 for the Dual-Fc-DCNN model.
2. In this paper, the dual-channel features are extracted from two single-channel fully connected layers, and then the two sets of features are manually integrated to obtain new classification features and input into the classifier for classification. The Dual-Int-DCNN model shown in Fig. 7 uses this feature integration method to integrate the features in the two channels and input the ECOC-SVM classification model for classification. Fig. 8 is a diagram of integrating the features extracted by the dual-channel DCNN model to obtain classification features. Processing methods include taking the maximum, minimum, and average values of multiple sets of different source data features, and directly splicing multiple sets of different source data features.

If you use a dual-channel model for feature extraction on wafer map data, you can obtain two sets of 8-dimensional features from different data sources (the Multi-source Features of DCNN1 and the Multi-source Features of DCNN2) as shown in Fig. 8. For the two groups of features, take the maximum value, the minimum value, the average value, and direct splicing value of the feature to obtain a classification feature (the max features, the min features, the mean features, the stitching features). When the maximum value, minimum value, or average value is taken, the integrated multi-source feature is still an 8-dimensional feature, which is the same as the dimension of a single sample feature. However, when directly splicing the feature, the new multi-source features are 16-dimensional features combined in the order of the Multi-source Features of DCNN1 and the Multi-source Features of DCNN2. The internal order of the Multi-source Features of DCNN1 and the Multi-source Features of DCNN2 remains unchanged, and the classification feature dimension at this time is twice the feature dimension of a single sample.

2.4. Framework

Fig. 9 shows the main framework of the method proposed in this article. The main idea as follow:

1. The multi-source data of the wafer map is used to train the dual-channel DCNN model.
2. Extract three sets of DCNN features from the fully connected layer of the dual-channel DCNN after training. The three sets of extracted features of each channel are integrated by averaging to obtain a set of multi-source features, and a total of two sets of multi-source features are obtained for two channels.
3. The two sets of multi-source features are spliced to obtain a set of classification features, and the classification features and class labels are fed back to the ECOC-SVM classification model for verification.

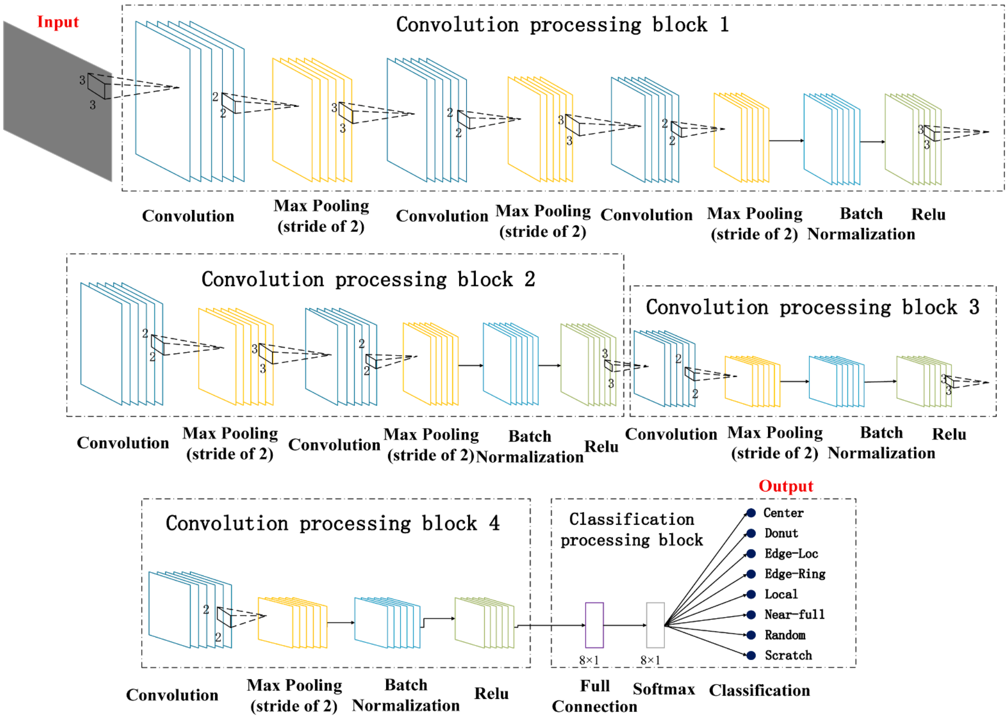


Fig. 11. Network structure of DCNN1.

3. Experiments and results

3.1. Dataset

The WM-811K data set [3] consists of 811,457 wafer maps collected from actual production, of which about 20% of the wafer maps have been annotated the defect types by experts in the field. Annotated wafer maps include wafer maps of eight common defect modes and None type wafer maps. The wafer map forms a specific pattern in the space, and each different pattern can be traced back to different problems in the production process, which is crucial for improving the production yield of wafers. In order to analyze the problems in production more clearly, the wafer map defect modes are divided into eight different types. The wafer maps of the eight defect modes are shown in Fig. 2, and the distribution of the wafer maps of the 8 defect modes in the WM811K data set is shown in Fig. 10(a). In the WM811K data set, the amount of data of Donut, Near-full, Random, and Scratch types is small. For machine learning algorithms, insufficient learning samples will lead to insufficient training and poor classification results. Therefore, this article uses the method of image rotation to expand Donut, Near-full, Random, and Scratch type data. Fig. 10(b) shows the distribution of wafer maps of 8 types of defect modes in the expanded data set. The original wafer map of the training data set used in this article are from 80% of the data of each data type in the expanded data set, a total of 26,704 wafer maps. And the original wafer map of the test data set are from the remaining 20% of the data in the expanded data set, totaling 6552.

All the work of recognition, classification, modeling and verification, sample data feature extraction, and machine learning algorithm processing are all carried out in the same environment. Use Windows7 to build a defect pattern recognition and classification model of wafer map and perform data processing. The workstation used is Dell Precision Tower 7910, the processor is configured with dual Intel® XeonTM processors E5-2698 v3 (16C, 2.3 GHz, Turbo, HT, 40 M, 135 W), the memory is 512 GB (16x32GB) 2133 MHz DDR4 LRDIMM ECC1, The graphics card is Nvidia Quadro K5200 8 GB high-end 3D graphics card. The modeling and experimentation of wafer pattern defect pattern recognition classification model and the processing of sample data by machine learning algorithm are completed under Matlab2017b.

3.2. Classification of integrated multi-source-DCNN features

The convolutional layer is the network layer for feature extraction in the DCNN model. The convolutional layer is based on the feature extraction method of local perception field and weight sharing, which can greatly reduce the computational complexity and quickly extract effective high-level features of the input data [24]. The batch normalization process in the batch normalization layer can stabilize the value, reduce the difference of different samples, and improve the generalization ability of the network. It is one of the important techniques for training modern neural networks [28]. The pooling layer is an important network layer for dimensionality reduction. The processing performed by the pooling layer is to replace the single point result in the feature map with the feature

Table 1
Classification results of the combined feature of multi-source data sets (%).

Methods	Accuracy	Precision	Recall	F1Score
Single-Source-DCNN1-Features + ECOC-SVM	95.7	95.7	95.5	95.6
Max-Multi-Source-DCNN1-Features + ECOC-SVM	52.6	74.4	50.4	51.9
Min-Multi-Source-DCNN1-Features + ECOC-SVM	37.1	37.0	29.2	25.9
Stitching-Multi-Source-DCNN1-Features + ECOC-SVM	95.8	95.7	95.6	95.7
Mean-Multi-Source-DCNN1-Features + ECOC-SVM	96.2	96.2	95.9	96.0

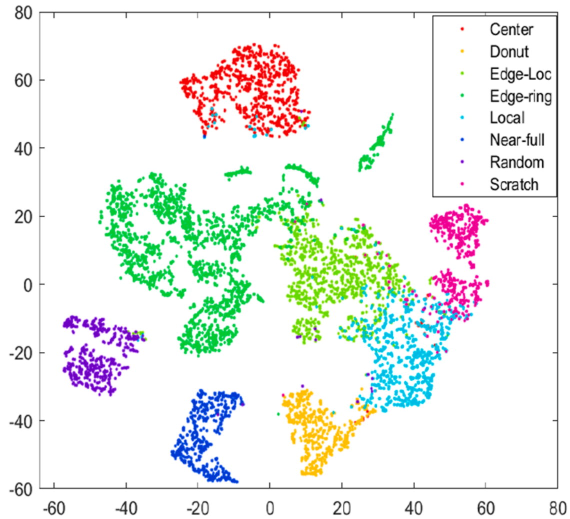


Fig. 12. Distribution of Single-Source DCNN Features.

statistics of its neighboring regions, simplify the feature map, and reduce the dimensionality and computational complexity of the feature map. The fully connected layer [29] is usually located in the last few layers of the neural network, and each neuron in the fully connected layer is interconnected with all the neurons in the upper layer. The first fully connected layer in the DCNN model will unfold and integrate the feature map output by the upper layer into a vector form. The Rectified linear unit (Relu) is a commonly used neuron activation function in the DCNN model, which can maintain the convergence rate of the model in a stable state.

Based on the above basic principles, this paper proposes the DCNN1 model architecture as shown in Fig. 11. The architecture of DCNN1 includes an input layer, seven convolutional layers, seven maximum pooling layers, four normalization layers, four Relu layers, a fully connected layer, a Softmax layer, and a classification layer.

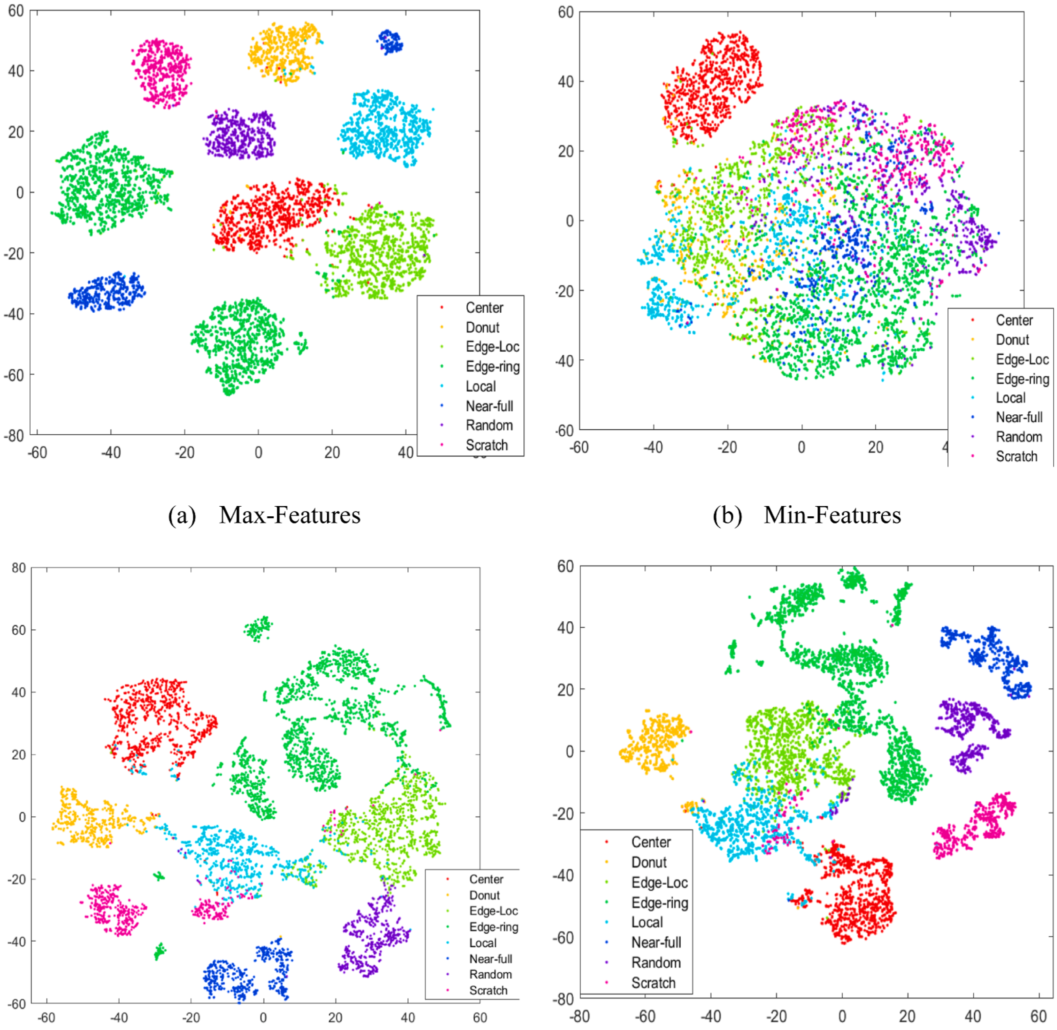
In this part, a separate data set and a multi-source data set are used for DCNN-ECOC-SVM classification and comparison. This part uses the DCNN1 model to extract features from multi-source input data to obtain multiple sets of high-level wafer map features, and use ECOC-SVM to classify the features after multiple sets of features. The classification accuracy, precision, recall, and F1score are shown in Table 1. This article finally adopts the Mean-Multi-Source-DCNN-Features + ECOC-SVM method.

It can be seen from Table 1 that when comparing the results of multiple classifications and taking the average of the features of the multi-source data, the classification effect of the multi-source data is better than other classification effects. However, when the maximum and minimum features of multi-source features are taken, the classification effect is the worst. As shown in Fig. 12 and Fig. 13, the T-distributed Stochastic Neighbor Embedding (T-SNE) [30,31] method is used to further analyze the classification results given in Table 1.

From Table 1, Fig. 12, and Fig. 13, we can see that when averaging multi-source features, the classification result is better than the classification results using other features. The classification between various types of data is the clearest, with less data that is easily confused, and the classification accuracy rate is high. When processing the maximum and minimum values of multi-source data, the distance between some types of data is relatively large. In addition, the maximum and minimum features are extremely extreme for multi-source data, and cannot accurately describe the features of the wafer map, so the accuracy is not high and the classification effect is not good.

3.3. Classification of integrated Dual-DCNN features

Based on the dual-channel principle, this paper proposes the second model DCNN2. As shown in Fig. 14, the structure of DCNN2 includes one input layer, seven convolutional layers, seven Maximum Pooling layers, five Normalization layers, five Relu layers, one Fully Connected layer, one Softmax layer, and a Classification layer.



(a) Max-Features (b) Min-Features

(c) Mean-Features (d) Stitching-Features

Fig. 13. Distribution of Multi-Source DCNN1 Features.

Table 2 shows the comparison results of DCNN + ECOC-SVM classification based on the Dual-Fc-DCNN model and the Dual-Int-DCNN model. The final method used in this article is the Stitching Dual-Int-DCNN-Features + ECOC-SVM.

From Table 2, we can compare the classification results of multiple DCNN + ECOC-SVM based on the Dual-Fc-DCNN model and the Dual-Int-DCNN model. When taking the stitching value of dual-channel features, the classification effect of multi-source data is better than that of other methods.

When establishing the basic models DCNN1 and DCNN2, this article compares the classification accuracy and training time of the two models with other classic CNN models. As shown in Table 3, under the same conditions and the same data, the classification accuracy of the two networks proposed in this paper is higher than that of LeNet, and the training time is greatly reduced. Compared with the network proposed in literature [33], the accuracy of DCNN1 is relatively low, and the accuracy of DCNN2 has been improved, but the training time of the two networks proposed in this paper is lower than that of the network proposed in literature [33]. In general, the training time required for other classic networks is 1.8 to 4.3 times the training time required for the network mentioned in this article.

3.4. Classification of integrated multi-source-dual-DCNN features

Combining the advantages of multi-source data and dual-channel DCNN model, this paper proposes an image-based defect pattern recognition method for wafer maps. The multi-source data composed of the original wafer map and its derived wafer map and the different feature extraction methods of the dual-channel DCNN model are used to obtain richer wafer map features. And select suitable feature data to form the classification feature and input it into the classification model to achieve the purpose of improving the

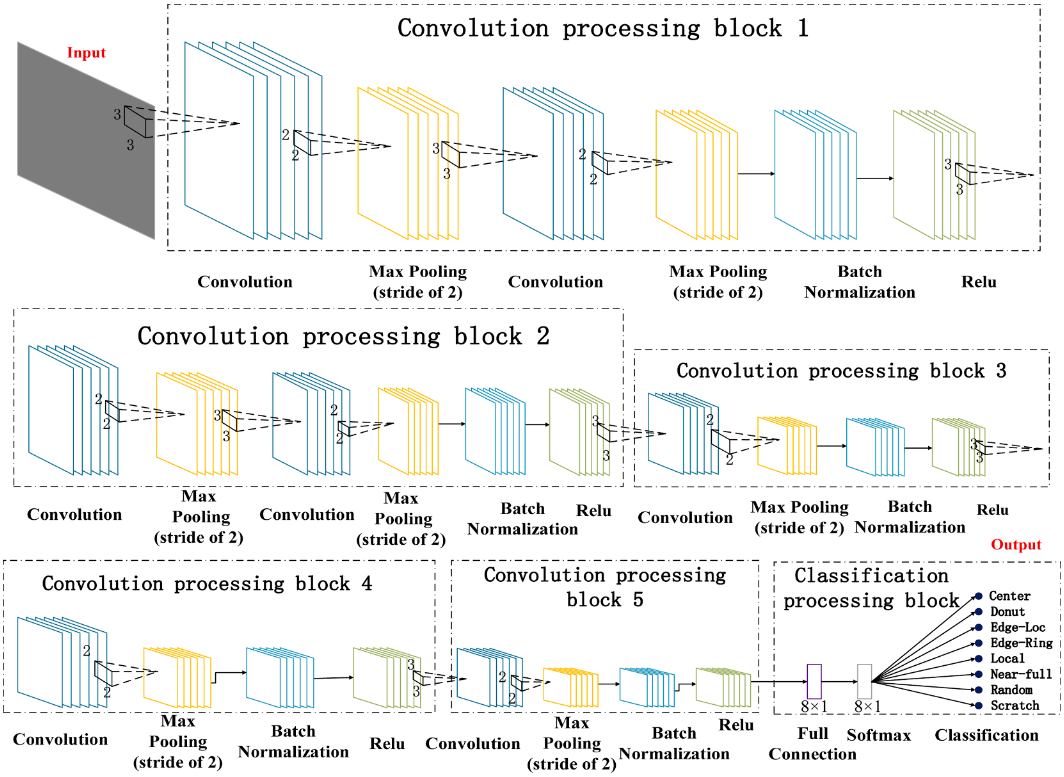


Fig. 14. Network structure of DCNN2.

Table 2
Comparison of DCNN + ECOC-SVM results of two Dual-DCNN models (%).

Methods	Accuracy	Precision	Recall	F1Score
Dual-Fc-DCNN-Features + ECOC-SVM	94.9	94.7	94.7	94.7
Max Dual-Int-DCNN-Features + ECOC-SVM	95.8	95.6	95.6	95.6
Min Dual-Int-DCNN-Features + ECOC-SVM	95.9	95.7	95.6	95.6
Mean Dual-Int-DCNN-Features + ECOC-SVM	95.9	95.8	95.7	95.7
Stitching Dual-Int-DCNN-Features + ECOC-SVM	96.1	95.9	95.9	95.9

Table 3
Comparison of accuracy and training time of different CNN structures.

	DCNN1	DCNN2	LeNet	Literature [33]
Accuracy(%)	93.2	93.6	81.7	93.3
Training time(s)	1295.87	1316.59	5473.63	2409.12

Table 4
The results of feature extraction from multi-source dual-DCNN models.

Method	Accuracy	Precision	Recall	F1Score
Literature [32]	93.25	\	\	\
Literature [33]	95.3	95.2	95.2	95.2
Dual-Fc-DCNN-Features + ECOC-SVM	94.9	94.7	94.7	94.7
Max Mean-Multi-Source-Dual-Int-DCNN-Features + ECOC-SVM	96.4	96.4	96.3	96.3
Min Mean-Multi-Source-Dual-Int-DCNN-Features + ECOC-SVM	96.3	96.3	96.2	96.2
Mean Mean-Multi-Source-Dual-Int-DCNN-Features + ECOC-SVM	96.4	96.4	96.3	96.3
Stitching Mean-Multi-Source-Dual-Int-DCNN-Features + ECOC-SVM	96.4	96.4	96.2	96.3

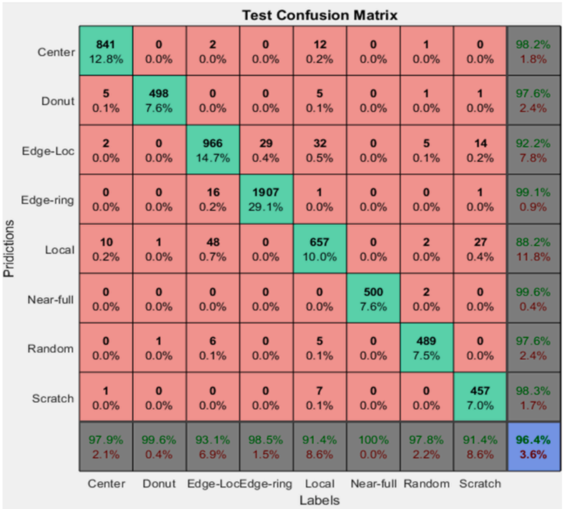


Fig. 15. Confusion matrix of multi-source data's mean value features.

classification accuracy. The proposed method first extracts multiple sets of data set features from the fully connected layers of every single channel in the dual-channel DCNN model and integrates multiple sets of data features obtained in a single channel into a set of multi-source features. At this time, two sets of multi-source features are obtained. Then the two sets of multi-source features are further integrated to obtain a set of features and input into ECOC-SVM, and the recognition and classification effects are studied. Table 4 compares the accuracy, precision, recall, and F1scores of the classification method proposed in this article and other literature classification methods. Among them, the method proposed in this study is Stitching Mean-Multi-Source-Dual-DCNN-Features + ECOC-SVM.

From the rough comparison of the classification results in Table 4 with the literature [32] and literature [33], the classification effect of the method proposed in this paper is better. The classification accuracy, precision, recall and F1score are 96.4%, 96.4%, 96.2% and 96.3%, respectively. Among them, the classification effect of DCNNs-And-Features-ECOC-SVM is the best in Table 3. Fig. 15 shows the classification confusion matrix of the DCNNs-Mean-And-Features-ECOC-SVM wafer map defect pattern recognition classification method proposed in this paper.

4. Conclusion

In this study, we combined multi-source data, dual-channel DCNN model, and ECOC-SVM to propose an image-based defect pattern classification method for wafer maps. The method mainly includes four steps of using the DCNN model to extract multi-source data features, multiple source data feature combination, dual-model feature integration, and ECOC-SVM classification. In the feature extraction step, the DCNN model is used to extract high-level features from multiple source data. In the multi-source feature combination step, the averaging method is used to synthesize multiple sets of data features into a set of multi-source features. In the dual-model feature integration step, the two sets of multi-source features extracted from the two DCNN models are spliced to synthesize a set of classification features. In the classification step, the ECOC-SVM model is trained using the classification features to classify the wafer map defect pattern. The experimental results on the test set divided in the WM811K data set shows that the proposed classification method achieves better results compared with other methods, and the classification accuracy rate reaches 96.4%.

Although this article uses a combination of a dual-channel model and a multi-source model, intuitively, the model consumes six times the time of a single training network. However, the model proposed in this paper can train six basic models at the same time in parallel, and the time required can still be approximately equal to the time of a single training network.

In the future, features can be obtained from more angles, and more processing methods can be used to integrate features. In this study, only the original wafer map of the dataset and its two derived wafer maps are used, a total of three data sources and two DCNN models, and the processing methods of the two features integration only consider the maximum value, minimum value, average value, and direct splicing value. However, in order to improve the classification accuracy, other derivative data, and other DCNN models, a combination of more sets of data sources and more DCNN models, and other feature integration methods can also be considered.

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.

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, 2018GXNSFBA281175), Guangxi Key Laboratory of Automation Test and Instrumentation (No. YQ15109, Q17103), Innovation Project of GUET Graduate Education (No. 2019YCXS085, 2020YCXS095).

References

[1] C.-F. Chien, K.-H. Chang, W.-C. Wang, An empirical study of design-of-experiment data mining for yield-loss diagnosis for semiconductor manufacturing, J. Intell. Manuf. 25 (5) (2014) 961–972.
[2] C.-H. Wang, W. Kuo, H. Bensmail, Detection and classification of defect patterns on semiconductor wafers, IIE Trans. 38 (12) (2006) 1059–1068.
[3] L. Xie, R. Huang, N. Gu, Z. Cao, A novel defect detection and identification method in optical inspection, Neural Comput. Appl. 24 (7-8) (2014) 1953–1962.
[4] S. Kang, Joint modeling of classification and regression for improving faulty wafer detection in semiconductor manufacturing, J. Intell. Manuf. 31 (2) (2020) 319–326.
[5] C.H. Jin, H.J. Na, M. Piao, G. Pok, K.H. Ryu, A novel DBSCAN-Based defect pattern detection and classification framework for wafer bin map, IEEE Trans. Semicond. Manuf. 32 (3) (2019) 286–292.
[6] Y.S. Jeong, S.J. Kim, M.K. Jeong, Automatic identification of defect patterns in semiconductor wafer maps using spatial correlogram and dynamic time warping, IEEE Trans. Semicond. Manuf. 21 (4) (2008) 625–637.
[7] M.Y. Fan, Q. Wang, B.V.D. Waal, Wafer defect patterns recognition based on OPTICS and multi-label classification. In: Proceedings of 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC 2016), 2016, pp. 4.
[8] M.J. Wu, J.S.R. Jang, J.L. Chen, Wafer map failure pattern recognition and similarity ranking for large-scale data sets, IEEE Trans. Semicond. Manuf. 28 (1) (2015) 1–12.
[9] M.H. Hansen, V.N. Nair, D.J. Friedman, Monitoring wafer map data from integrated circuit fabrication processes for spatially clustered defects, Technometrics 39 (3) (1997) 241–253.
[10] A. Drozda-Freeman, et al., The application and use of an automated spatial pattern recognition (SPR) system in the identification and solving of yield issues in semiconductor manufacturing, in: Proc. IEEE/SEMI Adv. Semicond. Manuf. Conf., 2007, pp. 302–305.
[11] M. Saqlain, B. Jargalsaikhan, J.Y. Lee, voting ensemble classifier for wafer map defect patterns identification in semiconductor manufacturing, IEEE Trans. Semicond. Manuf. 32 (2) (2019) 171–182.
[12] J. Yu, X. Lu, Wafer map defect detection and recognition using joint local and nonlocal linear discriminant analysis, IEEE Trans. Semicond. Manuf. 29 (1) (2016) 33–43.
[13] M. Piao, C.H. Jin, J.Y. Lee, J.-Y. Byun, Decision tree ensemble-based wafer map failure pattern recognition based on radon transform-based features, IEEE Trans. Semicond. Manuf. 31 (2) (2018) 250–257.
[14] O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, in: Proceedings of the BMVC, vol. 1, 2015, pp. 6.
[15] X. Liu, Z. Deng, Y. Yang, Recent progress in semantic image segmentation, Artif. Intell. Rev. 52 (2) (2019) 1089–1106.
[16] I. Konovalenko, P. Maruschak, O. Prentkovskis, R. Junevičius, Investigation of the rupture surface of the titanium alloy using convolutional neural networks, Materials 11 (2018) 2467, https://doi.org/10.3390/ma11122467.
[17] I. Konovalenko, P. Maruschak, J. Brezinová, J. Brezina, Morphological characteristics of dimples of ductile fracture of VT23M titanium alloy and identification of dimples on fractograms of different scale, Materials 12 (2019) 2051, https://doi.org/10.3390/ma12132051.
[18] K. Kyeong, H. Kim, Classification of mixed-type defect patterns in wafer bin maps using convolutional neural networks, IEEE Trans. Semicond. Manuf. 31 (3) (2018) 395–402.
[19] T. Nakazawa, D.V. Kulkarni, Anomaly detection and segmentation for wafer defect patterns using deep convolutional encoder–decoder neural network architectures in semiconductor manufacturing, IEEE Trans. Semicond. Manuf. 32 (2) (2019) 250–256.
[20] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297.
[21] Y.C. Tang, Deep learning using linear support vector machines, in: International Conference on Machine Learning, 2013.
[22] S.H. Chen, M.L. Yi, Y.X. Zhang, et al., Research on wafer map preprocessing based on optimized DBSCAN clustering algorithm, J. Control Decis. (2020) 1–9.
[23] M. Ferguson, R. Ak, Y.-T.-T. Lee, K.H. Law, Detection and segmentation of manufacturing defects with convolutional neural networks and transfer learning, Smart Sustain. Manuf. Syst. 2 (1) (2018) 210–237.
[24] Q.R. Zhang, M. Zhang, T.H. Chen, Z.F. Sun, et al., Recent advances in convolutional neural networks acceleration, Neurocomputing 323 (2019) 37–51.
[25] T.G. Dietterich, G. Bakiri, Solving multiclass learning problems via error-correcting output codes, J. Artif. Intell. Res. 2 (1995) 263–286.
[26] F. Al-shargie, T.B. Tang, N. Badruddin, M. Kiguchi, Towards multilevel mental stress assessment using SVM with ECOC: an EEG approach, Med. Biol. Eng. Compu. 56 (1) (2018) 125–136.
[27] T. Yuan, W. Kuo, S.J. Bae, Detection of spatial defect patterns generated in semiconductor fabrication processes, IEEE Trans. Semicond. Manuf. 24 (3) (2011) 392–403.
[28] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, Comput. Sci. (2015).
[29] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT press, Cambridge, 2016.
[30] M.V.D. Laurens, G. Hinton, Visualizing data using T-sne, J. Mach. Learn. Res. 9 (2008) 2579–2605.
[31] G. Crescenzio, C. Vito, A wafer bin map "Relaxed" clustering algorithm for improving semiconductor production yield, Open Comput. Sci. 10 (1) (2020) 231–245.
[32] Naigong Yu, Qiao Xu, Honglu Wang, Wafer defect pattern recognition and analysis based on convolutional neural network, IEEE Trans. Semicond. Manuf. 32 (4) (2019) 566–573.
[33] Cheng Hao Jin, Hyun-Jin Kim, Yongjun Piao, Meijing Li, Minghao Piao, Wafer map defect pattern classification based on convolutional neural network features and error-correcting output codes, J. Intell. Manuf. 31 (8) (2020) 1861–1875.