



A Defect Detection Method of Mixed Wafer Map Using Neighborhood Path Filtering Clustering Algorithm

Xingna Hou^{1,2} · Guanxiang Qin¹ · Ying Lu¹ · Mulan Yi¹ · Shouhong Chen¹

Received: 13 October 2023 / Accepted: 4 August 2024 / Published online: 2 October 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

As the wafer process becomes more complex, the probability of mixed-type defective wafer maps is constantly increasing. Therefore, it is necessary to perform effective filtering and denoising processing on the mixed-type defective wafer map to facilitate the identification. We propose a Neighborhood Path Filtering Clustering (NPFC) algorithm in this paper. For determining the number of clusters in the wafer map, we combine the clustering index of the Silhouette coefficient and the similarity index of Mahalanobis distance to propose a clustering similarity index SD. Then, calculating the compactness index between the clusters is judged whether the clusters are merged or not, to obtain the final clustering effect. The greatest advantage of this method is that it avoids the influence of algorithm parameter settings on the filtering effectiveness. Meanwhile, it can still effectively identify clusters of arbitrary shapes. The experimental results show that compared with median filtering, K-mean clustering and adjacency clustering, the NPFC algorithm achieves good results in the filtering and clustering effect.

Keywords Wafer map · Filtering · Denoising · NPFC · Clustering

1 Introduction

Due to advancements in semiconductor technology, the complexity of integrated circuit design has been steadily increasing. The semiconductor manufacturing industry is focused on achieving larger scales and smaller process dimensions. Present-day wafer production relies on highly automated processes and utilizes high-precision equipment. Nonetheless, the manufacturing of semiconductor wafers remains intricate and time-consuming, involving numerous complex chemical steps and requiring meticulous monitoring of various process parameters. Despite professional regulation, the occurrence of abnormal dies on the wafer is unavoidable

[1]. Therefore, improving the yield of semiconductor production is a crucial issue. The core step in the semiconductor manufacturing process is wafer manufacturing. The wafer manufacturing process requires a series of extremely complex operations such as oxidation, photolithography, etching, ion implantation, and metallization [2], which can easily cause wafer defects [3]. Therefore, after the wafer is manufactured, it is necessary to conduct an electrical test on each die on the wafer to mark whether there are defects in the wafer, and thus a wafer map [4] is generated. The wafer map obtained through testability inspection often forms a certain spatial pattern, which can reflect some problems in the production process. Realizing accurate classification of wafer maps can provide a theoretical basis for engineers to solve production problems, improve wafer production yield, and reduce production costs [5].

Wafer map defect pattern recognition and classification is an indispensable part of the semiconductor manufacturing process, because it can locate the cause of failures in the manufacturing process. In 2023, Jianhong Ma et al. [6] published a comprehensive review on wafer defect detection, which outlines the advancements in wafer defect pattern recognition in recent years. Naigong Yu et al. [7] developed an 8-layer CNN model to inspect wafer map defects

Responsible Editor: B. B. Bhattacharya.

✉ Shouhong Chen
cshgl@guet.edu.cn

¹ Guangxi Key Laboratory of Automatic Detecting Technology and Instruments, School of Electronic Engineering & Automation, Guilin University of Electronic Technology, Guilin 541004, China

² School of Architecture and Transportation Engineering, Guilin University of Electronic Technology, Guilin 541004, China

and a 13-layer model to classify defect patterns in 2019. In 2023, Chen et al. [8] proposed a deep convolutional neural network architecture incorporating improved convolutional block attention module (I-CBAM) for pattern recognition and classification of single-type wafer map defects, but the accuracy of Edge-Loc and Local still needs to be improved.

Some researchers have started using various methods for defect pattern recognition. For example: Li et al. [9] used semi-supervised learning for wafer map defect pattern recognition. This approach leverages unlabeled samples to improve recognition accuracy. Iljoo Jeong et al. [10] employed a geometric transformation-invariant convolutional neural network. Wang et al. [11] utilized rotation-invariant features Tziolas et al. [12] addressed the challenge of imbalanced datasets.

In addition to the specific spatial pattern of the wafer map itself, there is a lot of noise. The defect patterns in these wafer maps can be well identified by manual recognition [13, 14], but for machine learning methods, it is difficult to identify the real defect pattern types of the wafer maps. Therefore, it is necessary to filter the wafer map. Traditional filtering algorithms include spatial filtering, median filtering, mean filtering, etc. [14–16]. In 2022, Pourkaramdel et al. [17] proposed a new texture descriptor called complete local quartet pattern (CLQP), which is used to extract local texture features of fabric images and detect fabric defects in the training step. In 2023, Gulhan [18] proposed a new hybrid optical sensor based on deep learning to sense the micro-size defects on PCBs. These methods are extremely dependent on the setting of parameters. If the parameter threshold is set too small, the filtering will be incomplete, and if the parameter threshold is set too large, the defect pattern of the wafer map itself will be destroyed. Especially in the face of a large number of wafer maps, it is even more difficult to find a reasonable parameter that can be applied to all wafer maps. Therefore, such traditional filtering algorithms are not suitable for wafer maps.

(filter 算法) In addition to traditional filtering algorithms, many algorithms for filtering through clustering have also been derived [19]. For example, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is a typical algorithm based on density clustering, which can find clusters of arbitrary shapes and remove isolated noise points. The most important thing about the DBSCAN algorithm is to determine the two key parameters, the minimum neighborhood radius (Eps) and the minimum number of points (MinPts). Chen et al. [20] used the DBSCAN algorithm to filter the wafer map and chose to set the Eps parameter to $\sqrt{2}$ and the MinPts parameter to 3. The selection of parameters directly affects the accuracy of clustering. Like traditional filtering algorithms, the setting of this parameter is not suitable for large-scale wafer images with variable shapes. Facing the problem that the parameters are

difficult to determine, Chen et al. [21] proposed an adaptive DBSCAN clustering algorithm for wafer map data. The algorithm uses a comprehensive index of inter-cluster density and intra-cluster density to select optimal parameters and realizes adaptive filtering of large-scale wafer map data. However, this algorithm is aimed at single-type wafer maps. For mixed-type wafer maps, the number of clusters retained cannot be determined well.

However, as the wafer process becomes more complex and the wafer size becomes larger [22], the probability of mixed-type wafer maps is constantly increasing. If the size of the mixed-type wafer map and the single-type wafer map are the same in the classification and recognition process on the wafer map, the mixed-type wafer map is more susceptible to noise points than the single-type wafer map. Therefore, it is necessary to perform effective filtering and denoising processing on the mixed-type wafer map to facilitate the identification of the mixed-type wafer map. Kim et al. [23] proposed a connecting path filtering algorithm for 6 mixed-type wafer maps. The principle of this method is to find all possible connecting paths between two defect points. However the path length needs to exceed a certain set threshold, and the setting of this threshold is also related to the filtering effect. The same threshold setting is not necessarily applicable to all wafer maps. A. A. Ezzat et al. [24] proposed an adjacency clustering (AC) filtering algorithm based on graph theory for 12 mixed-type wafer maps. However, this method also requires manual setting of parameters, and then a large number of experiments on different parameters to select a suitable parameter, which is not suitable for large-scale wafer map data.

Among the above filtering algorithms for mixed-type wafer maps, the most prominent problem is still the selection of parameters. No matter what parameters are selected, they are not all applicable to different wafer maps. Therefore, this paper proposes a method for denoising mixed-type wafer maps—Neighborhood Path Filtering Clustering (NPFC), which is not affected by parameters. Based on the idea of clustering, this method uses the clustering index of the Silhouette coefficient and the similarity index of Mahalanobis distance to propose a cluster similarity index SD to determine the number of clusters in the wafer map. Then, by calculating the compactness index f between the clusters, it is judged whether the clusters are merged or not, to obtain the final clustering effect. The algorithm in this paper is mainly designed for the particularity that the wafer map is a matrix structure and consists of only three numeric values.

The rest of the paper is arranged as follows: The second section introduces in detail some concepts of clustering and related concepts redefined in this article and gives a detailed description of the specific steps of the basic principles of NPFC. The third section compares the method proposed in this paper with other experimental methods from the

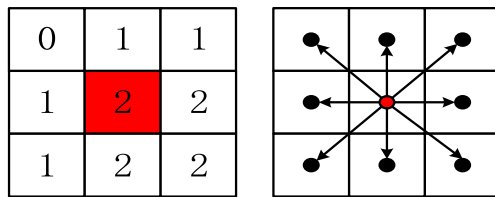


Fig. 1 Structure of the wafer map

perspectives of filtering and clustering. The fourth section summarizes the content of the full text and the outlook for future work.

2 Related Work

A wafer map is a two-dimensional digital matrix structure that differs significantly from a typical image. As shown in the left side in Fig. 1, a matrix composed of three numbers zero, one, and two (zero means no grain, one means normal grains, two means defective grains), and the pixel value of a typical image is 0–255, with a relatively large range of values. Therefore, the traditional filtering method is not suitable for this kind of image structure with only three digits. If each wafer grain is represented by coordinates, each grid cell can be regarded as a point. As shown on the right side of Fig. 1, The data points in a wafer map are distributed in a relatively regular pattern, with the distance between the central point and its eight neighboring points being either 1 or $\sqrt{2}$.

Traditional clustering methods are typically designed for scattered data, so clustering filtering methods need to be adjusted to be applicable to wafer maps. In this paper, a filtering algorithm based on clustering principles is proposed to address the specificity of wafer map structures.

2.1 Algorithm Dependent Definition

This article defines some concepts related to points and clusters, with examples illustrated in in Fig. 2. Given a data set sample $X = \{x_1, x_2, \dots, x_n\}$, the specific concepts are as follows:

(1) Initial point: Before forming a cluster, the first selected point O is the initial point (the initial point is randomly selected), $O \in X$.

(2) Direct path connected: If there is a point P in the 3×3 neighborhood of the initial point O , then the point P , point R and point O are direct path connected.

(3) Path connected: If a point Q is directly connected to a connected point P , then point Q is path-connected to the initial point O .

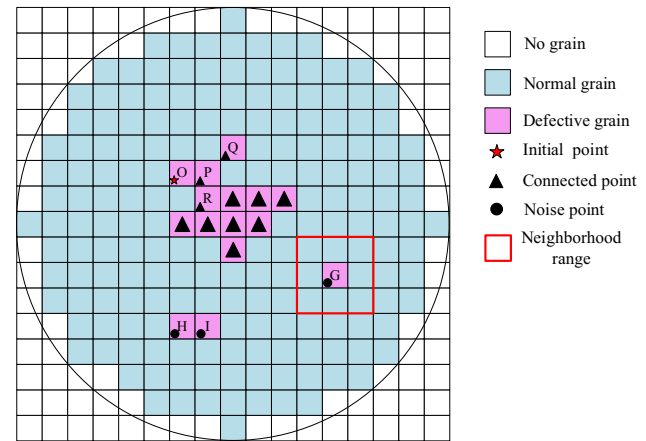


Fig. 2 Specific definition example

(4) Connected points: A point P that is directly connected to the initial point O is referred to as a connected point. In Fig. 2, all black triangles represent connected points.

(5) Cluster: A non-empty set where any point within the set is either direct path connected or path connected to an initial point is classified as a cluster. A cluster D should satisfy the following two conditions:

- ① $x_i \in D, x_j \in D \Rightarrow$ point x_i is path-connected with point x_j .
- ② $x_i \in D$, point x_j is path-connected with point $x_i \Rightarrow x_j \in D$.

As shown in Fig. 2, the defect point marked with a pentagon and the surrounding defect points marked with triangles form a cluster D .

(6) Core cluster: The collection of all clusters sorted according to the number of data samples from high to low is $\pi = (C_1, C_2, \dots, C_i, \dots, C_M)$, where $M > i > 2 > 1$, and select the Num clusters with the most data points from among them as the core cluster $(C_1, C_2, \dots, C_i, \dots, C_{Num})$, where $Num > i > 2 > 1$. Overall, $M \geq Num$.

(7) Noise points: A noise point is a defect point with no other defect points in its 3×3 neighborhood. Similarly, twin defect points that only have each other in their respective 3×3 neighborhoods are also considered noise point. For instance, point G , H and I in Fig. 2 are examples of noise points.

From the definitions above and as illustrated in Fig. 2, it can be seen that regardless of whether point O , P , Q , or any other point marked with a triangle is chosen as the initial point, a cluster D can always be formed. Therefore, compared to the DBSCAN clustering algorithm, the neighborhood path filtering algorithm does not require identifying core points of clusters or checking whether points directly reachable from or density-reachable to core points need to continue meeting the criteria for cluster expansion. This

approach reduces computational complexity while still effectively capturing clusters of arbitrary shapes.

2.2 Neighborhood Path Filtering Clustering

In response to the preprocessing challenges posed by mixed defect pattern wafer maps, this paper introduces the Neighborhood Path Filtering Clustering (NPFC) algorithm, which circumvents the parameter selection issues encountered in typical filtering methods and clustering-based filtering methods. The flowchart of the NPFC algorithm is depicted in Fig. 3.

(1) Form preliminary clusters by using the initial point as the core. Identify connected points within the smallest neighborhood of 3×3 around the initial point. Include points directly or path-connected within this area.

(2) Mark and remove noise points and clusters. Mark isolated defect points or twin defect points [22] not in any cluster as noise points, and remove noise points.

(3) Merge clusters. The Num clusters with the largest number are selected as the core clusters, and the remaining clusters are merged with the Num clusters as the core. Calculate the distance between the remaining clusters and the

core cluster, the closest and the farthest distance relationship between the core cluster and the central origin, and the closest and farthest distance relationship between the remaining clusters and the central origin. Through the above relationship, it is determined whether the cluster needs to be merged with the core cluster and which core cluster to merge with.

2.2.1 Preliminary Clustering of Wafer Map

The types of defects in the wafer map are composed of small clusters. The preliminary clustering aims to form multiple small clusters while eliminating isolated noise points and twin noise points. For clarity, the preliminary clustering steps of the wafer map are illustrated in Fig. 4. The specific clustering steps are as follows:

(1) First select the initial point. The defect point marked with a pentagon is designated as the initial point in Fig. 4 ②, and there are two defective points in the 3×3 neighborhood of the initial point.

(2) Divide the connected point and the initial point into the same cluster. As shown in Fig. 4 ③, two defect points found are marked with triangles, and the pentagram mark and the triangle mark are grouped into the same cluster.

(3) Find the point that is path connected with the initial point. Query the 3×3 neighborhood of the triangle marked defect point in Fig. 4 ③. Two defect points are found and marked as triangles, as shown in Fig. 4 ④.

(4) Repeat the process of step (3) until there are no unmarked defect points in the 3×3 neighborhood of the defect points to be queried, as shown in Fig. 4 ⑦, that is, a cluster is formed.

(5) Repeat steps 1) 2) 3) 4) until all the defect points on the wafer map are divided into clusters or noise points, as shown in Fig. 4 ⑧.

The defect types on a wafer map are all composed of small clusters. By employing the aforementioned clustering method, it is possible to capture clusters of any shape. This approach is versatile and applicable to various types of wafer maps.

2.2.2 Determine the Number of Clusters in the Wafer Map

Preliminary clustering divides the wafer map into many small clusters, but the overall defect type pattern of the wafer map is less than or equal to the number of preliminary clusters. Therefore, it is necessary to determine the total number of clusters representing the wafer map's overall defect pattern before proceeding with the merging of these small clusters. Through the observation of many wafer maps, the number of defect types in the wafer map is generally 1 to 4. To prevent the unobserved wafer map may have more defect types, this article considers the number of defect types to be 1 to 5. Therefore, the number of clusters is also in the range

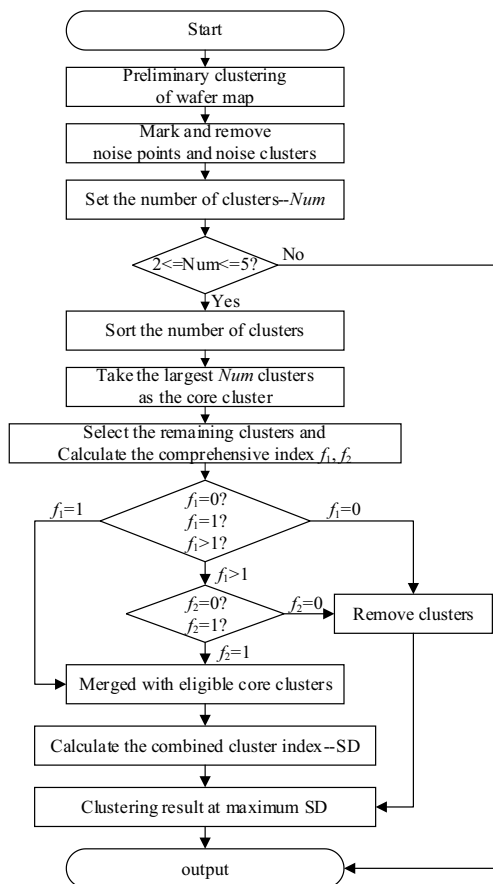


Fig. 3 Flowchart of NPFC algorithm

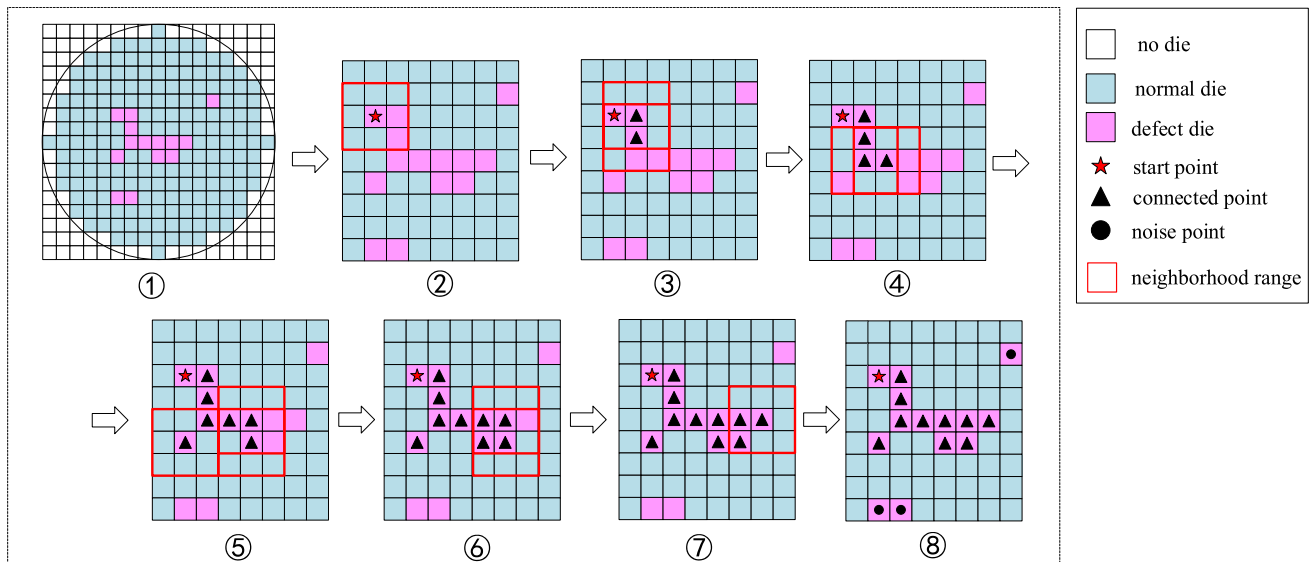


Fig. 4 Diagram of preliminary clustering

of 1 to 5. When discussing mixed-type wafer maps, the number of clusters ranges from 2 to 5. This paper proposes a new clustering similarity index (SD) to determine the number of clusters in wafer maps. It combines the Silhouette coefficient and the Mahalanobis distance.

(1) Silhouette coefficient.

The Silhouette coefficient [25] is a way to evaluate the clustering effect. Assuming that the data sample $X = \{x_1, x_2, \dots, x_n\}$ has been divided into number Num clusters by the algorithm. The silhouette coefficient for a sample point x_i in the k -th cluster is calculated as shown in Eq. (2-1).

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (1)$$

Among them, $a(i)$ is the average value of the distance from the sample point x_i and other points in the k -th cluster, which is used to indicate the degree of separation between clusters. $b(i)$ is the minimum value of the distance from the sample point x_i and other points in the k -th cluster.

(2) Mahalanobis distance.

Mahalanobis distance [26] is a metric used to evaluate the similarity between data points. For one-dimensional sample data, The specific calculation is shown in Eq. (2).

$$D_M(x) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)} \quad (2)$$

When evaluating the similarity between two images using the Mahalanobis distance, let the sample data of the first image be $X = \{x_1, x_2, \dots, x_n\}^T$, and the sample data

of the second image be $Y = \{y_1, y_2, \dots, y_n\}^T$. The specific calculation is shown in Eq. (3).

$$D_M(x, y) = \sqrt{(x - y)^T \Sigma^{-1} (x - y)} \quad (3)$$

In Eq. (3), Σ is the covariance matrix of x and y .

(3) Clustering similarity index SD.

The silhouette coefficient is a metric used to evaluate the clustering effectiveness of different algorithms based on the same raw data. However, when determining the number of clusters in this study, the comparison is made between the clustering effects before and after merging or removing clusters. Since these comparisons involve data with certain differences, relying solely on the silhouette coefficient could lead to inaccuracies. Therefore, this paper uses a combined metric, SD, which integrates both the silhouette coefficient and the Mahalanobis distance (used as a similarity measure), to select the appropriate number of clusters. The calculation method of SD is shown in the formula (4).

$$SD = \frac{1}{n} \sum_{i=1}^n S(i) + D_M(X, Y) \quad (4)$$

The larger the silhouette coefficient, the better the clustering effect. The larger the Mahalanobis distance, the greater the similarity between wafer maps. Therefore, the larger the SD, the better. The process for determining the number of clusters in wafer map clustering is shown in Fig. 5. The number of clusters, Num , ranges from 2 to 5, and the optimal number of clusters is the one where the SD is the largest.

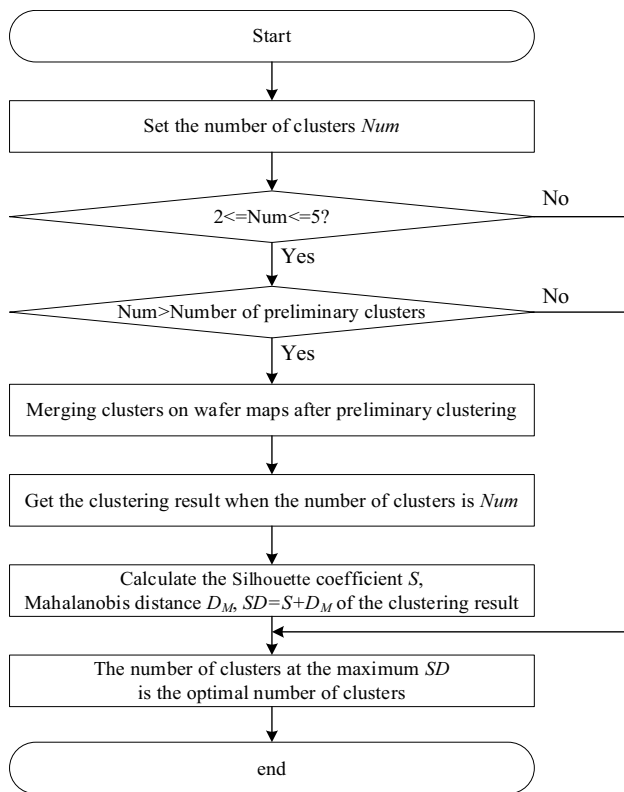


Fig. 5 Flowchart for determining the number of clusters

2.2.3 Remove or Merge Wafer Map Clusters

The initial clustering results in many small clusters. To obtain a clustering result with an appropriate number and reasonable division of data samples, further cluster merging and removal are required. Given the number of clusters, Num , the specific steps for merging or removing clusters are as follows:

Combine the clusters whose closest distance between clusters does not exceed $\sqrt{8}$ ($\sqrt{8}$ is the distance from the center of the entire square to the center of the diagonal square when the neighborhood is 5×5).

Find the largest Num clusters as core clusters and merge them with other clusters.

Merge or remove clusters. The rule are as follows:

Calculate the radius relationship between the remaining clusters and the core clusters. Take the center of the wafer map as the origin. Compute the distance from the origin to each defect point within each cluster. Use the maximum and minimum distances from the origin to determine the maximum and minimum radius distances for each cluster. Then, assess the relationship between the maximum and minimum radius distances between the remaining clusters and the core clusters. The schematic diagram of the radius relationship is shown in Fig. 6. The red markers indicate the core clusters, and the red

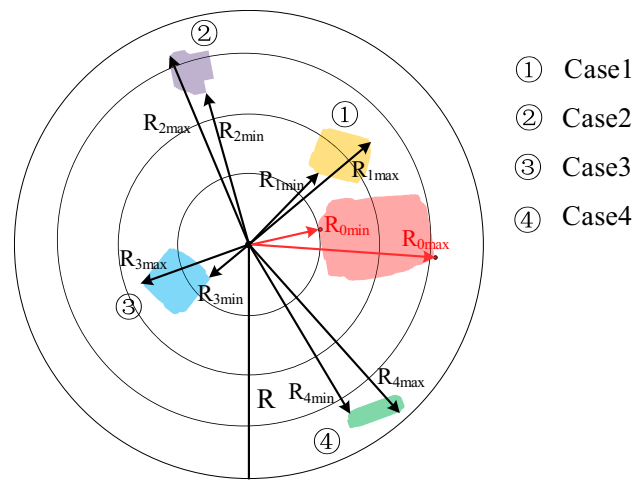


Fig. 6 Schematic diagram of radius distance relationship

arrows represent the maximum and minimum radius distances of the core clusters.

Figure 6 illustrates the following four possible cases between the remaining clusters and the core clusters:

Case 1: The radius range of the cluster is entirely within the radius range of the core cluster, namely:

$$R_{1min} \geq R_{0min}, R_{1max} \leq R_{0max}$$

Case 2: The maximum radius distance of the cluster exceeds the maximum radius distance, namely:

$$R_{2min} \geq R_{0min}, R_{2max} > R_{0max}$$

③ Case 3: The minimum radius distance of the cluster exceeds the minimum radius distance of the core cluster, namely:

$$R_{3min} < R_{0min}, R_{3max} \leq R_{0max}$$

④ Case 4: The radius range of the cluster is completely outside the radius range of the core cluster, namely:

$$R_{4min} > R_{0min}, R_{4max} > R_{0max}$$

If it belongs to the ①②③ cases, set a parameter $\alpha = 1$. If it belongs to the ④ case, $\alpha = 0$, directly remove the cluster.

(2) The distance between the cluster and the core cluster.

Assuming that the initial set of clusters after preliminary clustering is $P_1 = \{C_1, C_2, \dots, C_N\}$, the core cluster set is $P_2 = \{C_1, \dots, C_{Num}\}$, then the rest of clusters set is $P_3 = \{C_{Num+1}, \dots, C_N\}$. Calculate respectively the nearest distance $nDist$, the farthest distance $fDist$, and the average distance $mDist$ between the sample points in the P_2 cluster and those in cluster P_3 , as shown in Eqs. (5, 6 and 7).

$$nDist = \min(Dist(X, Y)) \quad (5)$$

$$fDist = \max(Dist(X, Y)) \quad (6)$$

$$mDist = \text{mean}(Dist(X, Y)) \quad (7)$$

where $X = \{x_1, x_2, \dots, x_m\}$ represents the sample point of cluster C_i in the core cluster, $X \in C_i$, $C_i \in P_2$. Similarly, $Y = \{y_1, y_2, \dots, y_n\}$ represents the sample points in cluster C_j of the remaining clusters, $Y \in C_j$, $C_j \in P_3$. $Dist$ denotes the Euclidean distance, \min is the minimum value, \max denotes the maximum value, and mean denotes the average value.

(3) The compactness after merging clusters.

Assume $X = \{x_1, x_2, \dots, x_m\}$, where $X \in C_i$ and $C_i \in P_2$, $Y = \{y_1, y_2, \dots, y_n\}$, where $Y \in C_j$ and $C_j \in P_3$. After merging the data into $Z = \{x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n\}$, the compactness Den after merging clusters is given by Eq. (8) [27]. The larger compactness of Den is, the more suitable it is to merge the two clusters.

$$Den = \text{mean}(\text{mean}(Dist(Z, Z))) \quad (8)$$

(4) The closeness index.

The relationship between the radius and distance of composite clustering clusters and core clusters, as well as the compactness after merging two clustering clusters, establish the comprehensive indicator of closeness f as the final criterion for whether to merge. The expression of the closeness index f is shown in Eqs. (9 and 10).

$$f_1 = \alpha * \text{Count}(C_{\min(nDist)} \cap C_{\max(Den)} \cap C_{(mDist < R)}) \quad (9)$$

$$f_2 = \alpha * \text{Count}(C_{\min(nDist)} \cap C_{\min(fDist)} \cap C_{\max(Den)} \cap C_{(mDist < R)}) \quad (10)$$

where $C_{\min(nDist)}$ represents the core cluster with the minimum nearest distance to the current clustering cluster, $C_{\min(fDist)}$ represents the core cluster with the minimum furthest distance to the current clustering cluster, and $C_{\max(Den)}$ represents the core cluster with the maximum density after merging with the current clustering cluster, $C_{(mDist < R)}$ represents the intersection symbol, indicating that the average distance to the current clustering cluster is less than the wafer map radius R , \cap represents the intersection symbol, and $\text{Count}()$ means counting operations.

Regarding the decision of whether to merge or remove a specific cluster (other than the core cluster), and determining which core cluster to merge with, the following methods are used for assessment:

Judge f_1 . If for any core cluster, $f_1 = 0$, then remove that cluster.

If $f_1 = 1$, merge the cluster with the corresponding core cluster.

If $f_1 > 1$, $f_2 = 0$, remove the cluster.

If $f_1 > 1$, $f_2 = 1$, merge the cluster with the corresponding core cluster.

Apply the above judgments to each cluster until all remaining clusters in the wafer map are either merged or removed. The process of clustering and filtering on the wafer map using the above method is illustrated in Fig. 7. Initially, after preliminary clustering, a preliminary clustering image is obtained as shown in Fig. 7(b). Then, through determining the number of clusters, merging, and removing clusters, a clustered image is obtained as shown in Fig. 7(c). Finally, the representation is restored to the original image format, resulting in the filtered map as shown in Fig. 7(d).

3 Experimental Results and Analysis

The Neighborhood Path Filtering Clustering (NPFC) algorithm is a filtering method based on clustering concepts. To validate the effectiveness of the NPF algorithm proposed in this paper for preprocessing wafer maps with mixed defect patterns, this section compares its filtering and clustering performance with several existing methods.

3.1 Dataset

3.1.1 Data

The wafer map data used in this paper is 110 mixed-type wafer maps selected from the marked wafer maps in the WM-811 K database provided by Ming-Ju Wu et al. [28]. Among the 110 mixed-type wafer maps, there are 6 wafer maps labeled Center, 3 wafer maps labeled Donut, 30 wafer maps labeled Edge-Loc, 8 wafer maps labeled Edge-Ring, 60 wafer maps labeled Edge-Loc, and 3 wafer maps labeled Scratch. For ease of comparative experimental study, this paper uses the same 12 wafer maps with an article [24] as in reference (not duplicated with the 110 wafer maps selected above). Figure 8 shows the 12 mixed-type wafers.

3.2 Comparison Index

(1) Evaluation index of filtering effect.

Common filtering evaluation indexes are mainly as follows:

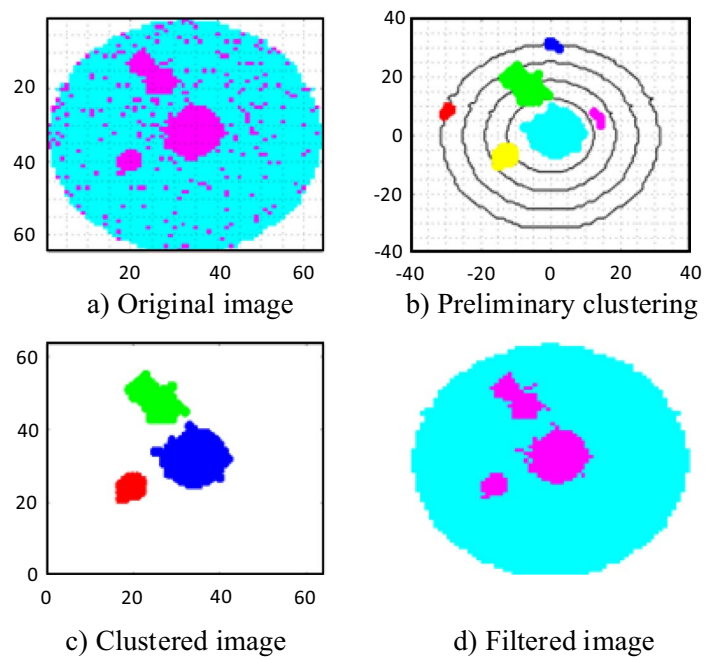
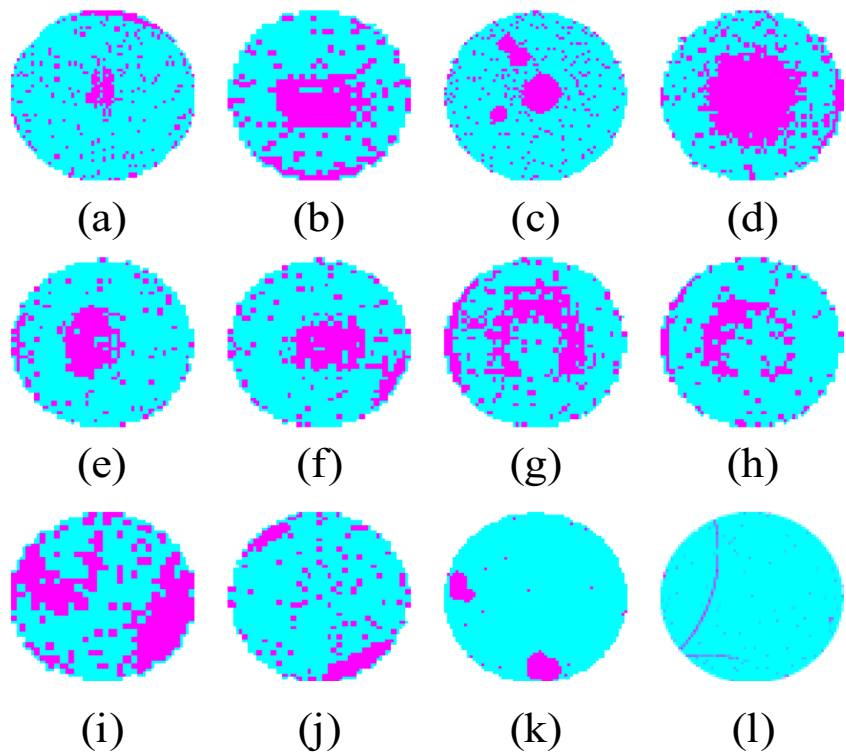
① Mean Absolute Error (MAE).

The smaller the value of MAE, the smaller the deviation from the original image and the better the image quality.

$$MAE = \frac{\sum_{i=1}^L \sum_{j=1}^W |G(i, j) - G'(i, j)|}{L \times W} \quad (11)$$

② Normalized Mean Square Error (NMSE).

NMSE is a measurement method based on energy normalization. The smaller the value of NMSE, the better the image quality.

Fig. 7 Process of NPFC**Fig. 8** Original wafer map

$$NMSE = \frac{\sum_{i=1}^L \sum_{j=1}^W |G(i,j) - G'(i,j)|^2}{\sum_{i=1}^L \sum_{j=1}^W |G(i,j)|^2} \quad (12)$$

③ Signal-to-Noise Ratio (SNR).

The larger the value of SNR, the better the image quality.

$$SNR = 10 \log_{10} \left[\frac{\sum_{i=1}^L \sum_{j=1}^W |G(i,j)|^2}{\sum_{i=1}^L \sum_{j=1}^W |G(i,j) - G'(i,j)|^2} \right] \quad (13)$$

Peak signal-to-noise ratio (PSNR).

The higher the value of PSNR, the better the image quality.

$$PSNR = 10 \log_{10} \left[\frac{255^2 \times L \times W}{\sum_{i=1}^L \sum_{j=1}^W |G(i,j) - G'(i,j)|^2} \right] \quad (14)$$

where $G(i,j)$ represents the pixel value of the original image, $G'(i,j)$ represents the pixel value of the filtered image, and L and W respectively represent the length and width of the wafer map.

(2) Evaluation index of the internal clustering effect.

Assuming that the clusters of the clustering results of the wafer map are divided into $C = (C_1, C_2, \dots, C_K)$, the definitions are as follows: the average distance between samples is $Avg(C)$, the longest distance between samples is $diam(C)$, the nearest distance between cluster samples from different cluster is $d_{min}(C_i, C_j)$ and the center distance between clusters is $d_{cen}(C_i, C_j)$, as shown in formulas (15, 16, 17, 18 and 19).

$$avg(C) = \frac{2}{|C|(|C| - 1)} \sum_{1 \leq i < j \leq |C|} dist(x_i, x_j) \quad (15)$$

$$diam(C) = \max_{1 \leq i < j \leq |C|} dist(x_i, x_j) \quad (16)$$

$$d_{min}(C_i, C_j) = \min_{x_i \in C_i, x_j \in C_j} dist(x_i, x_j) \quad (17)$$

$$d_{cen}(C_i, C_j) = dist(\mu_i, \mu_j) \quad (18)$$

where μ represents the center point of cluster C , $\mu = \frac{1}{|C|} \sum_{1 \leq i \leq |C|} x_i$.

Common metrics used for evaluating internal clustering performance (Vinh et al. 2009) includes the Davies-Bouldin Index (DBI), Dunn Index (DI), Calinski-Harabasz (CH) index, and Generalized Dunn Index (GDI). Respectively as shown in formulas (19, 20, 21 and 22).

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{j \neq i} \left(\frac{avg(C_i) + avg(C_j)}{d_{cen}(C_i, C_j)} \right) \quad (19)$$

$$DI = \min_{1 \leq i \leq K} \left\{ \min_{j \neq i} \frac{dmin(C_i, C_j)}{\max_{1 \leq l \leq K} diam(C_l)} \right\} \quad (20)$$

$$CH = \frac{r - K}{K - 1} \frac{\sum_{k=1}^K r_k \|G^k - G\|^2}{\sum_{k=1}^K \sum_i \|C^k - G^k\|^2} \quad (21)$$

$$GDI = \frac{\min_{k \neq k'} \frac{1}{r_k + r_{k'}} (\sum_i \|C_i^k - G^k\| + \sum_j \|C_j^{k'} - G^{k'}\|)}{\max_k \max_{i \neq j} \|C_i^k - C_j^k\|} \quad (22)$$

where G represents the center point of all coordinates, G^k represents the center point of all defect point coordinates in the k -th cluster, r represents the number of all defect points on the wafer map, and r_k represents the number of defect points in the k -th cluster. The smaller the value of DBI, the better the clustering effect. On the contrary, the larger the value of DI, CH, and GDI, the better the clustering effect.

(3) Evaluation index of the external clustering effect.

The Infinite Warped Mixture Model (iWMM) is a non-parametric Bayesian model that can find non-linearly separable clusters with complex shapes (non-Gaussian distribution). Through the warping process, complex shapes with non-Gaussian distributions can be modeled in the latent space using simpler Gaussian distributions, so that the model can be effectively inferred [23]. For a detailed introduction to iWMM, please refer to the paper [23, 24].

In addition to internal clustering indicators, In addition to internal clustering metrics, this paper also uses three external metrics of the iWMM model to compare with adjacency clustering [24]. The definitions of the three external indicators are as follows:

Assuming that the true label of the data set D is $L = \{L_1, L_2, \dots, L_n\}$, the predicted label is $\hat{L} = \{\hat{L}_1, \hat{L}_2, \dots, \hat{L}_n\}$, TP, FP, TN, and FN respectively represent the number of samples that meet the four conditions in Table 1, then $TP + FP + TN + FN$ is equal to the total number of samples N of the dataset D .

① Rand Index (RI)

$$RI = \frac{TP + TN}{TP + FN + FP + TN} \quad (23)$$

It can be seen from the above formula that the range of RI is $[0, 1]$. When the predicted label is completely inconsistent with the real label, $RI = 0$. When the predicted label is completely consistent with the real label, $RI = 1$. Therefore, the larger the value of RI, the more accurate the prediction.

Table 1 Four relationships between true labels and predicted labels

True label Predict Label	Positive	Negative
Positive	TP(True Positive)	FN(False Negative)
Negative	FP(False Positive)	TN(True Negative)

② Adjusted Rand Index (ARI)

$$ARI = \frac{N * (TP + FN) - [(TP + TN)(TP + FP) + (FP + FN)(TN + FN)]}{N^2 - [(TP + TN)(TP + FP) + (FP + FN)(TN + FN)]} \quad (24)$$

The larger the value of ARI, the more similar the predicted label and the real label, and the more accurate the prediction result.

③ Normalized mutual information (NMI)

$$NMI(L, \hat{L}) = \frac{I(L, \hat{L})}{H(L, \hat{L})} \quad (25)$$

$$I(L, \hat{L}) = \sum_{i=1}^L \sum_{j=1}^{\hat{L}} \frac{n_{ij}}{n} \log\left(\frac{n_{ij}/n}{(\sum_{j=1}^{\hat{L}} n_{ij})(\sum_{i=1}^L n_{ij})/n^2}\right) \quad (26)$$

$$H(L, \hat{L}) = \sum_{i=1}^L \sum_{j=1}^{\hat{L}} \frac{n_{ij}}{n} \log\left(\frac{n_{ij}/n}{(\sum_{i=1}^L n_{ij})/n}\right) \quad (27)$$

where n_{ij} represents the same number of samples that have the same true label belonging to cluster i and the predicted label belonging to cluster j , n represents the number of cluster samples in the cluster. The larger the NMI, the better the clustering effect.

3.3 Experimental Result and Comparison

3.3.1 Comparison of Filtering Effects

To validate the filtering effectiveness of the method proposed in this paper, a comparison was made with median filtering (MF) using four filtering evaluation metrics: MAE, NMSE, SNR, and PSNR. The filtering effects of NPFC and MF on the 12 wafer maps are shown in Fig. 9. The results for the 110 wafer maps are shown in Table 2, and the comparison of filtering metrics for the 12 wafer maps is presented in Table 3. In all tables of the paper, superior data is highlighted in bold, with “↑” indicating that a higher value is better for that metric and “↓” indicating that a lower value is better.

From the visual comparison in Fig. 9, it can be observed that compared to median filtering, NPFC method's wafer

Table 2 Comparison of the filtering effect of 110 wafer maps

method	Average value(↓)		Number of superior values(↑)	
	MF	NPFC	MF	NPFC
MAE	0.0327	0.0238	11	99
NMSE	0.0270	0.0179	11	99
SNR	16.3432	19.6420	11	99
PSNR	64.1825	67.2271	11	99

Fig. 9 Comparison of median filtering and NPFC

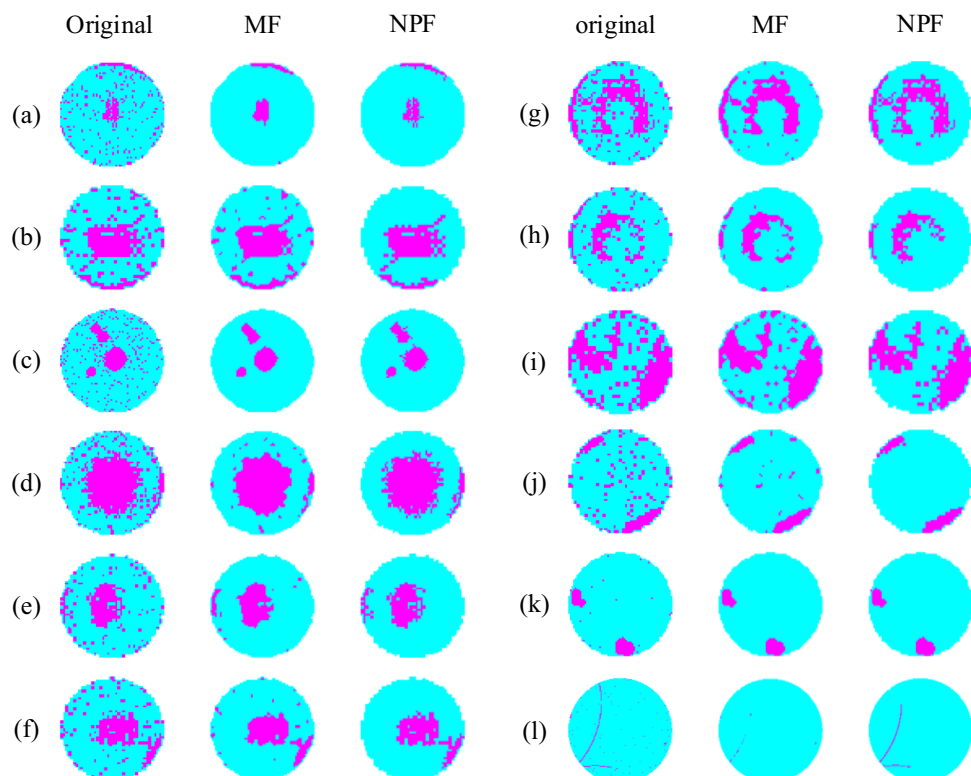


Table 3 Comparison of the filtering effect of 12 wafer maps

	MAE(↓)		NMSE(↓)		SNR(↑)		PSNR(↑)	
	MF	NPFC	MF	NPFC	MF	NPFC	MF	NPFC
(a)	0.0679	0.0608	0.0609	0.0545	12.16	12.63	59.81	60.29
(b)	0.0598	0.0688	0.0395	0.0455	14.03	13.42	60.36	59.75
(c)	0.0608	0.0544	0.0530	0.0475	12.76	13.24	60.29	60.77
(d)	0.0728	0.0532	0.0419	0.0306	13.78	15.14	59.51	60.87
(e)	0.0603	0.0415	0.0502	0.0340	12.99	14.69	60.26	61.95
(f)	0.0603	0.0554	0.0484	0.0428	13.15	13.69	60.16	60.69
(g)	0.0635	0.0422	0.0439	0.0290	13.57	15.38	60.07	61.87
(h)	0.0591	0.0522	0.0495	0.0427	13.05	13.69	60.31	60.95
(i)	0.0476	0.0527	0.0266	0.0294	15.75	15.31	61.35	60.91
(j)	0.0681	0.0681	0.0604	0.0591	12.19	12.28	59.71	59.80
(k)	0.0059	0.0039	0.0063	0.0042	22.00	23.76	70.45	72.21
(l)	0.0128	0.0071	0.0154	0.0085	18.13	20.71	67.06	69.65

images (a) to (k) have removed more noise points while retaining the main spatial patterns intact. In contrast, the wafer image (l), after median filtering, shows that the basic spatial patterns have been disrupted, whereas the NPFC method retains them completely. According to the comparison results of the four evaluation metrics—MAE, NMSE, SNR, and PSNR—shown in Tables 2 and 3, the NPFC algorithm scores higher in most of these indicators than the median filtering method. In the comparison of 110 wafer images, the NPFC method has a superior number of 99 in all four indicators, which is significantly higher than the median filtering. Additionally, the average values are better with the NPFC than with the median filtering. Therefore, the NPFC algorithm proposed in this paper is superior to the median filtering method.

3.3.2 Comparison of Clustering Effects

(1) Comparison of SD index for different numbers of clusters.

When determining the number of clusters *Num*, the SD index is used to determine the number of clusters. The larger the value of the SD index, the better the cluster effects. Therefore, the number of clusters corresponding to the maximum SD value is chosen as the final number of clusters. The SD indicators for different numbers of clusters for the wafer images are shown in Table 4 (clustering stops when *Num* exceeds the actual number of core clusters that the wafer images can form, hence some cluster numbers do not have data).

From Table 4, it can be seen that only the wafer image (c) has a final cluster number of 3, while the other images have a cluster number of 2. Upon comparison, these results align with the actual number of clusters that should be present in the wafer images. Similarly, for the remaining 110 wafer images, the determined number of clusters also meets the

Table 4 Comparison of SD index for different numbers of clusters

	Number of cluster Num			
	2	3	4	5
(a)	2.3994	\	\	\
(b)	2.2649	\	\	\
(c)	2.4018	2.4597	\	\
(d)	2.3014	\	\	\
(e)	2.3748	\	\	\
(f)	2.2566	\	\	\
(g)	2.2545	\	\	\
(h)	2.2011	2.1231	2.0859	\
(i)	2.5110	2.4622	\	\
(j)	2.6256	2.5443	2.5885	2.5719
(k)	2.7577	\	\	\
(l)	1.8269	1.6912	1.7372	1.7886

expected requirements. Therefore, the method used in this paper to determine the number of clusters is feasible.

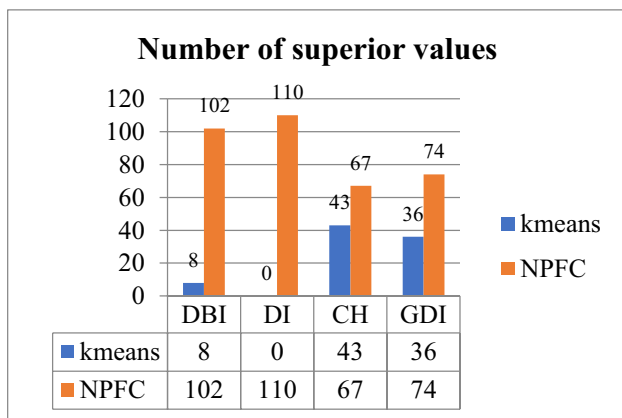
(2) Comparison of the NPFC clustering and K-means clustering.

The NPFC method is a filtering technique based on the clustering process. Therefore, this paper compares the clustering process of the NPFC method with the traditional K-means clustering method using four clustering indicators: DBI, DI, CH, and GDI. The clustering results for 12 wafer images using different clustering methods are shown in Table 5, while the clustering results for 110 wafer images are illustrated in Fig. 10.

From the comparison of the four clustering indicators for the 12 wafer images in Table 5, it is evident that the NPFC algorithm has an absolute advantage over the K-means algorithm in terms of the DBI, DI, and GDI indicators. For the CH indicator, both algorithms perform similarly. Additionally, from the comparison of the optimal clustering

Table 5 Comparison of four indexes of different clustering methods for 12 wafer maps

	DBI(↓)		DI(↑)		CH(↑)		GDI(↑)	
	kmeans	NPFC	kmeans	NPFC	kmeans	NPFC	kmeans	NPFC
(a)	0.3671	0.2857	14.5602	18.0278	322.49	407.82	0.1707	0.1926
(b)	0.9136	0.5303	1.0000	3.0000	661.41	471.97	0.2269	0.2974
(c)	0.4669	0.4142	6.3246	4.1231	278.85	294.27	0.2245	0.3121
(d)	0.8591	0.4622	1.0000	8.6023	941.30	371.92	0.2021	0.2758
(e)	0.4808	0.3891	9.4868	9.4868	344.92	330.16	0.2803	0.3653
(f)	0.6830	0.3947	1.0000	3.0000	603.01	531.26	0.1843	0.2677
(g)	0.6668	0.5799	1.0000	2.2361	845.76	537.65	0.2021	0.3093
(h)	0.9224	0.4494	1.0000	9.4868	378.02	211.20	0.2094	0.4642
(i)	0.5573	0.4226	1.0000	8.5440	1677.13	1791.71	0.2225	0.2781
(j)	0.4131	0.1386	1.0000	48.7545	631.11	970.36	0.2142	0.3075
(k)	0.1205	0.1205	30.4138	30.4138	940.71	940.71	0.3035	0.3035
(l)	0.5225	0.5177	1.0000	6.4031	217.70	120.26	0.3709	0.2195

**Fig. 10** Comparison of the optimal clustering results for the 110 wafer images

indicators for the 110 wafer images shown in Fig. 10, it can be observed that the neighborhood path clustering outperforms the K-means clustering. Therefore, the neighborhood path filtering algorithm proposed in this paper is superior.

(3) Comparison of NPFC algorithm and Adjacent Clustering algorithm.

The Adjacent Clustering (AC) [18] is a graph-theoretic clustering method, and it is also a method designed for mixed types of wafer maps. In this paper, we compare the clustering performance of the NPFC method with the Adjacent Clustering method using the clustering metrics and 12 wafer images referenced in [22]. The comparison results are shown in Table 6 and 7.

From the comparison of the clustering metrics CH and GDI in Table 6, it is clear that the clustering performance of the method proposed in this paper is significantly better than that of the Adjacent Clustering (AC) method. Furthermore, as shown in Table 7, for the RI, ARI, and NMI indicators, the NPFC method presented in this paper

demonstrates a decisive advantage over the AC method, with higher scores in all three metrics. Overall, the clustering effectiveness of the NPFC algorithm is superior to that of the Adjacent Clustering method.

The effectiveness of clustering mixed defect pattern wafer maps using the iWMM method is shown in Fig. 11. From the visual results in Fig. 11, this study also demonstrates good clustering performance.

3.3.3 Algorithm Complexity Analysis

For two-dimensional data, the basic time complexity of the traditional DBSCAN algorithm is $O(n^2)$ [29]. The Adjacent Clustering (AC) algorithm, which is based on graph theory, also has a time complexity of $O(n^2)$. This is because it involves calculations for both the separation function and the deviation function for any two different defect points [24]. In contrast, the NPFC algorithm only requires repeating the process of merging and removing clusters four times when determining the number of clusters. Therefore, the time complexity of the Neighborhood Path Filtering algorithm is $O(4n)$, that is, $O(n)$.

The space complexity of the traditional DBSCAN algorithm is $O(n^2)$. The Adjacent Clustering (AC) algorithm requires storing the relationships and weights between defect points, as well as a binary label for each defect point, resulting in a space complexity of $O(n^2)$. In contrast, the Neighborhood Path Filtering algorithm only needs to store the relevant metric data for four different cluster numbers during the selection process. Consequently, the space complexity of the Neighborhood Path Filtering algorithm is $O(4n)$, that is, $O(n)$. Based on the complexity analysis, the proposed method in this paper is more efficient than both the traditional DBSCAN algorithm and the AC clustering algorithm.

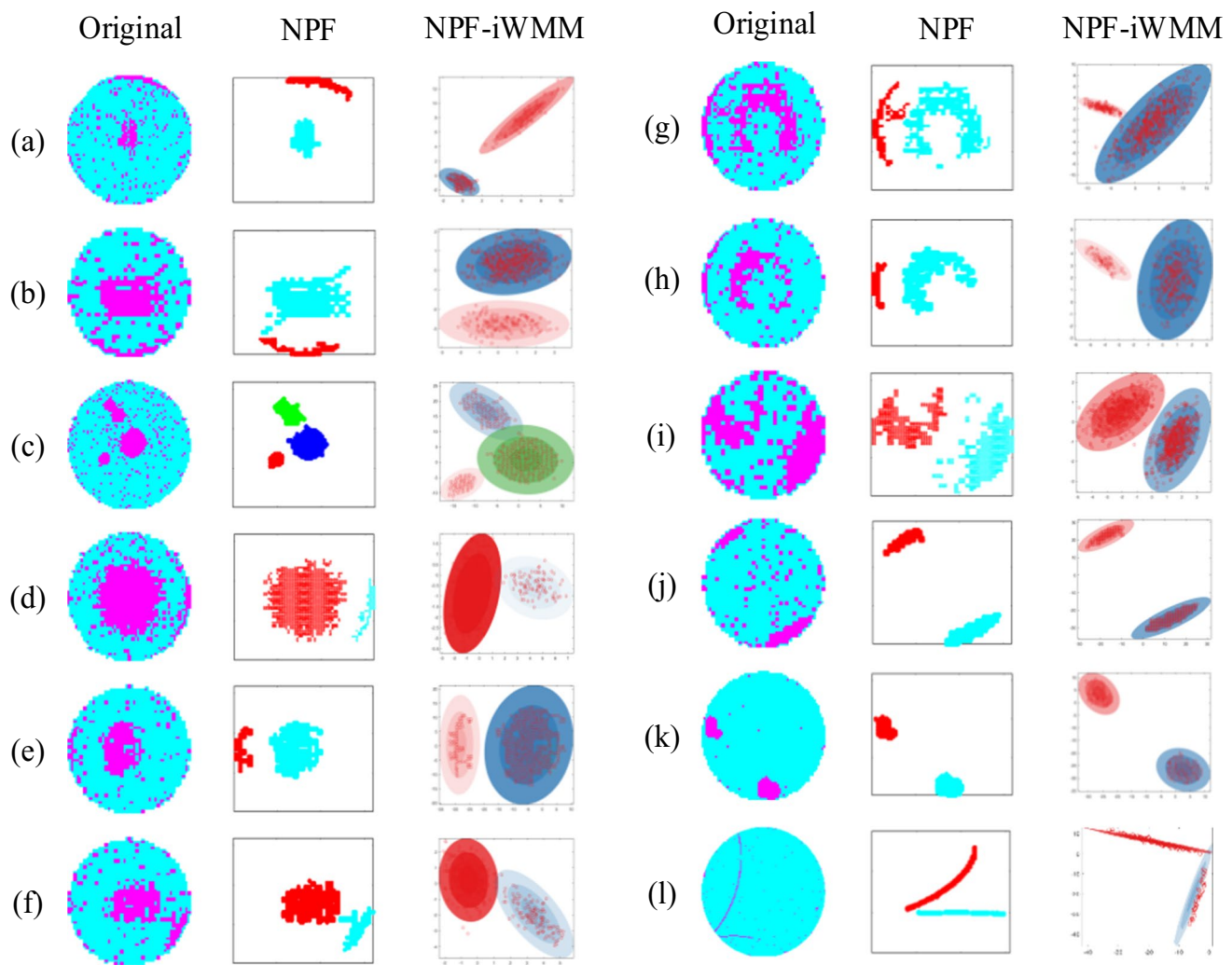


Fig. 11 iWMM result of wafer map

Table 6 Comparison of CH and GDI index of different clustering methods

	CH(↑)		GDI(↑)	
	AC	NPFC	AC	NPFC
(a)	703.0000	407.8214	0.1950	0.1926
(b)	218.0000	471.9661	0.2850	0.2974
(c)	682.3000	294.2727	0.2280	0.3121
(d)	254.8000	371.9185	0.3020	0.2758
(e)	222.8000	330.1629	0.2760	0.3653
(f)	304.4000	531.2631	0.2810	0.2677
(g)	148.4000	537.6503	0.2600	0.3093
(h)	240.8000	211.1955	0.3200	0.4642
(i)	593.4000	1791.7148	0.2230	0.2781
(j)	1565.8000	970.3568	0.2150	0.3075
(k)	8266.1000	940.7053	0.3000	0.3035
(l)	54.7000	120.2559	0.1840	0.2195

4 Conclusion

This chapter proposes a Neighborhood Path Filtering Clustering (NPFC) algorithm based on clustering concepts for wafer maps with mixed defect patterns. The algorithm extends clustering within the neighborhood range of defect points on the wafer map. It determines the number of clusters by using a comprehensive metric combining the silhouette coefficient and the Mahalanobis distance. Additionally, it uses indicators designed from compactness, distances between clusters, and the radius relationships between clusters and the center of the wafer map to decide whether clusters should be merged. The NPFC algorithm does not require the confirmation of defect points as core points or the assessment of whether points are density-reachable or density-connected to core points, as in the DBSCAN algorithm. Despite this, it can still effectively identify clusters of arbitrary shapes. The paper compares the filtering effect of

Table 7 Comparison of RI, AMI, and NMI of different clustering methods

	RI(↑)		ARI(↑)		NMI(↑)	
	AC	NPFC	AC	NPFC	AC	NPFC
(a)	0.9850	1.0000	0.9700	1.0000	0.9160	0.9566
(b)	0.9310	1.0000	0.8510	1.0000	0.7990	0.8927
(c)	0.9800	0.9929	0.9610	0.9894	0.9060	0.9226
(d)	0.9430	1.0000	0.8690	1.0000	0.8410	1.0000
(e)	0.9670	0.9977	0.9320	0.9955	0.8760	0.6749
(f)	0.9560	0.9684	0.9080	0.9369	0.8450	0.9379
(g)	0.9390	0.9502	0.8770	0.9010	0.8120	0.6751
(h)	0.9230	1.0000	0.8300	1.0000	0.7900	0.6596
(i)	0.8560	0.9996	0.6610	0.9991	0.6860	0.9841
(j)	0.9850	1.0000	0.9690	1.0000	0.9170	1.0000
(k)	0.9900	1.0000	0.9790	1.0000	0.9360	1.0000
(l)	0.9850	1.0000	0.9680	1.0000	0.9360	1.0000

12 mixed-type wafer maps that is better than median filtering. This paper compares the proposed method with existing methods in terms of filtering performance, clustering performance, and infinite Gaussian mixture warping effects. Based on the visual results and the quantitative comparison of various metrics, the results consistently indicate that the proposed method is superior.

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10836-024-06132-8>.

Funding The research work reported in this paper is supported by the National Natural Science Foundation of China (No. 61661013), Guangxi Natural Science Foundation (No. 2018GXNSFBA281175), Guangxi Key Laboratory of Automation Test and Instrumentation (No. YQ22102), Innovation Project of GUET Graduate Education (No. 2020YCX095, No. 2022YCX162, No. 2024YCX123), the Basic Ability Enhancement Program for Young and Middle-aged Teachers of Guangxi (No. 2022KY0197).

Data Availability The data that has been used is confidential.

Declarations

Competing Interests 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.

References

- 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. *J Intell Manuf* 25:961–972. <https://doi.org/10.1007/s10845-013-0791-5>
- Cheon S, Lee H, Kim CO, Lee SH (2019) Convolutional Neural Network for Wafer Surface Defect Classification and the Detection of Unknown Defect Class. *IEEE Trans Semicond Manuf* 32:163–170. <https://doi.org/10.1109/TSM.2019.2902657>
- Jeong Y-S, Kim S-J, Jeong MK (2008) Automatic Identification of Defect Patterns in Semiconductor Wafer Maps Using Spatial Correlogram and Dynamic Time Warping. *IEEE Trans Semicond Manuf* 21:625–637. <https://doi.org/10.1109/TSM.2008.2005375>
- Shim J, Kang S, Cho S (2020) Active Learning of Convolutional Neural Network for Cost-Effective Wafer Map Pattern Classification. *IEEE Trans Semicond Manuf* 33:258–266. <https://doi.org/10.1109/TSM.2020.2974867>
- Saqlain M, Jargalsaikhan B, Lee JY (2019) A Voting Ensemble Classifier for Wafer Map Defect Patterns Identification in Semiconductor Manufacturing. *IEEE Trans Semicond Manuf* 32:171–182. <https://doi.org/10.1109/TSM.2019.2904306>
- Ma J, Zhang T, Yang C et al (2023) Review of Wafer Surface Defect Detection Methods. *Electronics* 12:1787. <https://doi.org/10.3390/electronics12081787>
- Yu N, Xu Q, Wang H (2019) Wafer Defect Pattern Recognition and Analysis Based on Convolutional Neural Network. *IEEE Trans Semicond Manuf* 32:566–573. <https://doi.org/10.1109/TSM.2019.2937793>
- Chen S, Liu M, Hou X et al (2023) Wafer map defect pattern detection method based on improved attention mechanism. *Expert Syst Appl* 230:120544. <https://doi.org/10.1016/j.eswa.2023.120544>
- Li KS-M, Jiang X-H, Chen LL-Y et al (2022) Wafer Defect Pattern Labeling and Recognition Using Semi-Supervised Learning. *IEEE Trans Semicond Manuf* 35:291–299. <https://doi.org/10.1109/TSM.2022.3159246>
- Jeong I, Lee SY, Park K et al (2023) Wafer map failure pattern classification using geometric transformation-invariant convolutional neural network. *Sci Rep* 13:8127. <https://doi.org/10.1038/s41598-023-34147-2>
- Wang R, Chen N (2019) Wafer Map Defect Pattern Recognition Using Rotation-Invariant Features. *IEEE Trans Semicond Manuf* 32:596–604. <https://doi.org/10.1109/TSM.2019.2944181>
- Tziolas T, Theodosiou T, Papageorgiou K, et al (2022) Wafer Map Defect Pattern Recognition using Imbalanced Datasets. In: 2022 13th International Conference on Information, Intelligence, Systems and Applications (IISA). IEEE, Corfu, 1–8
- Coules HE, Orrock PJ, Seow CE (2019) Reflectance Transformation Imaging as a tool for engineering failure analysis. *Eng*

- Fail Anal 105:1006–1017. <https://doi.org/10.1016/j.engfailanal.2019.07.037>
14. Mohanty S, Hazra S, Paul S (2020) Intelligent prediction of engine failure through computational image analysis of wear particle. *Eng Fail Anal* 116:104731. <https://doi.org/10.1016/j.engfailanal.2020.104731>
 15. Wang C-H, Kuo W, Bensmail H (2006) Detection and classification of defect patterns on semiconductor wafers. *IIE Trans* 38:1059–1068. <https://doi.org/10.1080/07408170600733236>
 16. Wang C-H, Wang S-J, Lee W-D (2006) Automatic identification of spatial defect patterns for semiconductor manufacturing. *Int J Prod Res* 44:5169–5185. <https://doi.org/10.1080/02772240600610822>
 17. Pourkaramdel Z, Fekri-Ershad S, Nanni L (2022) Fabric defect detection based on completed local quartet patterns and majority decision algorithm. *Expert Syst Appl* 198:116827. <https://doi.org/10.1016/j.eswa.2022.116827>
 18. Ustabas Kaya G (2023) Development of hybrid optical sensor based on deep learning to detect and classify the micro-size defects in printed circuit board. *Measurement* 206:112247. <https://doi.org/10.1016/j.measurement.2022.112247>
 19. Chang C-W, Chao T-M, Horng J-T et al (2012) Development Pattern Recognition Model for the Classification of Circuit Probe Wafer Maps on Semiconductors. *IEEE Trans Compon, Packag Manufact Technol* 2:2089–2097. <https://doi.org/10.1109/TCPMT.2012.2215327>
 20. Chen S, Zhang Y, Yi M et al (2021) AI classification of wafer map defect patterns by using dual-channel convolutional neural network. *Eng Fail Anal* 130:105756. <https://doi.org/10.1016/j.engfailanal.2021.105756>
 21. Chen S, Yi M, Zhang Y et al (2021) A self-adaptive DBSCAN-based method for wafer bin map defect pattern classification. *Microelectron Reliab* 123:114183. <https://doi.org/10.1016/j.microrel.2021.114183>
 22. Tello G, Al-Jarrah OY, Yoo PD et al (2018) Deep-Structured Machine Learning Model for the Recognition of Mixed-Defect Patterns in Semiconductor Fabrication Processes. *IEEE Trans Semicond Manufact* 31:315–322. <https://doi.org/10.1109/TSM.2018.2825482>
 23. Kim J, Lee Y, Kim H (2018) Detection and clustering of mixed-type defect patterns in wafer bin maps. *IIE Transactions* 50:99–111. <https://doi.org/10.1080/24725854.2017.1386337>
 24. Ezzat AA, Liu S, Hochbaum DS, Ding Y (2021) A Graph-Theoretic Approach for Spatial Filtering and Its Impact on Mixed-Type Spatial Pattern Recognition in Wafer Bin Maps. *IEEE Trans Semicond Manufact* 34:194–206. <https://doi.org/10.1109/TSM.2021.3062943>
 25. Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
 26. Mahalanobis PC (2018) ON THE GENERALIZED DISTANCE IN STATISTICS. *Sankhyā. Indian J Stat* 80:S1–S7. <https://www.jstor.org/stable/48723335>
 27. Hu L, Zhong C (2019) An Internal Validity Index Based on Density-Involved Distance. *IEEE Access* 7:40038–40051. <https://doi.org/10.1109/ACCESS.2019.2906949>
 28. Ming-Ju Wu, Jang J-SR, Chen J-L (2015) Wafer Map Failure Pattern Recognition and Similarity Ranking for Large-Scale Data Sets. *IEEE Trans Semicond Manufact* 28:1–12. <https://doi.org/10.1109/TSM.2014.2364237>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Xingna Hou is a Ph.D. student from Guilin University of Electronic Technology. She is currently an associate professor with the School of Architecture and Transportation Engineering, Guilin University of Electronic Technology, China. Her research interests include computer-aided testing, defect recognition of wafer map, and deep learning.

Guanxiang Qin is a graduate student from Guilin University of Electronic Technology. He graduated from Guilin University of Electronic Technology. His research interests include computer-aided testing, defect recognition of wafer map, and deep learning.

Ying Lu is a graduate student from Guilin University of Electronic Technology. Her research interests include computer-aided testing, defect recognition of wafer map, and deep learning.

Mulan Yi received her master's degree from Guilin University of Electronic Technology in 2022. Her research interests are about defect recognition of wafer maps.

Shouhong Chen received his Ph.D. degree in Mechanical and Electronic Engineering from Jiangsu University, Zhenjiang, Jiangsu, China, in 2021. He is currently a Professor at the School of Electronic Engineering and Automation, Guilin University of Electronic Technology, China. His current research interests include computer-aided testing, defect recognition, and deep learning.