Mixed-type Defect Pattern Classifications -Takumi Maeda

Daisuke Takada, Sumika Arima
arimalab298@gmail.com
University of Tsukuba
1-1-1, Tennodai, Tuskuba, Ibaraki pref., 305-8573 Japan
Phone: +81 -298-535-558 Fax: +81 -298-535-558

Introduction

The classification problem (CP) for wo-dimensional defect patterns (DP) has been increasing at an accelerated pace since the disclosure of the wafer map (WM) open data WM-811K [1], and there have been many applications of deep learning such as Convolutional Neural Network (CNN) in recent years [2][3]. Developments in the research area have focused primarily on improving accuracy, and two main approaches have been taken: network sophistication and feature addition. Accuracy improvement through network sophistication is converging, while the DP superposition problem is once again being discussed with a context of features construction from WM [3]. The problem of DP superposition (SP) is inevitable in reality and is becoming increasingly important as products become more sophisticated.

The two main topics of this study are as follows:

1. A robust and accurate classification method that can deal with the SP problem of DPs, no matter how many SP are involved.
2. Identification of important causal variables of DPs to prevent and reduce defects, rather than just classifying DPs.

In this paper, we will present the proposed method and its performance, focusing mainly on 1.

Discussion

Latest study [3] has dealt with SP problems up to 4 classes, and it has been found that for SP classification problems, a significant loss of accuracy occurs in classes 3 or more. This is mainly due to the limitation of identifying WMs based on their 2-D information only in nature. On the other hand, systematic defects are mainly caused by physical, chemical, or mechanical cause. Therefore, it must be possible to improve the accuracy by analyzing candidate of causal variables (CVs) together with WM. It can be considered natural in principle. Therefore, in this study, we propose an image multimodal approach to analyze WM data of 2D matrix together with multiple source variables [4].

We present a two-step classification and extend WM-811K data (Fig. 1) to evaluate the SP DP classification problem. The first step is a preprocess to discriminate defective and non-defective wafers, and the second step is to classify the defective wafers by extending CNN (for instance, Nakazawa's CNN [2]) for multimodal analysis (Fig. 2). The underlying network is not limited to CNN or so as in [2][3], but can also be a transformer, which has been increasingly studied in recent years [4].

Data・Verification conditions

For extending the WM-811K to the multimodal data for evaluation, additional features are attached by cluster of Gausian Mixture Model (GMM) based on the wafers' defect rates (Fig.3). CVs are also integrated for SP DPs (Fig.4).
In this study, the missing rates of the additional features (i.e. CVs) were varied from 0, 0.2, 0.4, 0.6, 0.8, and 1.0 for numerical validation.

Numerical Result

Even with the 4-classes superposition, the accuracy remains higher than 0.85 for a missing ratio of 0.6 or less (Fig.5a). It outperforms previous studies (Table 1). In addition, looking at the individual classes, only two classes of SP have the recall of less than 0.8, indicating that a high level of accuracy is maintained in almost all classes. (Fig.5b)

Consideration

Although the accuracy is somewhat lower than that of the 2-class superposition, there is no significant difference between the 3- and 4-class SP and all achieve high accuracy up to a missing ratio of 0.6 (Fig.5a). That is one of the major differences of the proposed method from previous studies.
In particular, comparing the results of the four classes with the previous study[3], all of the methods proposed in the previous studies have low accuracy for ScratchCenterEdge-RingEdge-Loc (Table.1), however, our proposed method achieves 0.966 (even in missing ratio 0.4) all of them show high accuracy. This suggests that our proposed method is effective in improving classification accuracy.it is confirmed that any SP problem can be classified with high accuracy when the missing rate of causal features is less than 0.6 ($\leq$0.6) (Fig.5a). That means it can be applied to practical cases with high rate of missing values.

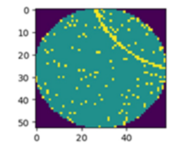
Conclusion

In this paper, we present results for up to 4 classes of superpositions, which is equivalent to the previous study, due to space limitations.

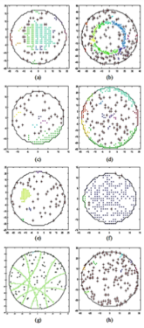
Reference

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

[2] Nakazawa, T. and Kulkarni, D. V. (2018). Wafer Map Defect Pattern Classification and Image Retrieval Using Convolutional Neural Network. IEEE Transaction on Semiconductor Manufacturing, 31(2). IEEE:309-314.
[3] Wang, R. and Chen, N. (2022). Detection and Recognition of Mixed-Type Defect Patterns in Wafer Bin Maps via Tensor Voting. IEEE Transaction on Semiconductor Manufacturing, 35(3). IEEE:485-493.
[4] Patent JP (Application No.) 2023-097287, S. Arima, H. Ito, H. Watanabe, D. Takada, T. Maeda, S. Hoshino (Applicant: University of Tsukuba), "Determination device, learning device and teacher data generation device" , June 13, 2023.
[5] Wei, Y. and Wang, H. (2022). Mixed-Type Wafer Defect Recognition With Multi-Scale Information Fusion Transformer. IEEE Transaction on Semiconductor Manufacturing, 35(3). IEEE:341-352.

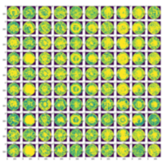


(a) Example of defect pattern of the wafer ([1])



	Class	# of wafers
(a)	Center	4294
(b)	Donut	555
(c)	Edge-Loc	5189
(d)	Edge-Ring	9680
(e)	Loc	3593
(f)	Near-full	149
(g)	Random	866
(h)	Scrach	1193
	None	147431
	Total	172950

(b) Classes of defect pattern ([6])



(c) Example of superposition defect pattern

Fig. 1. Data of Semiconductor Defect Patterns - WM-811K and its application.

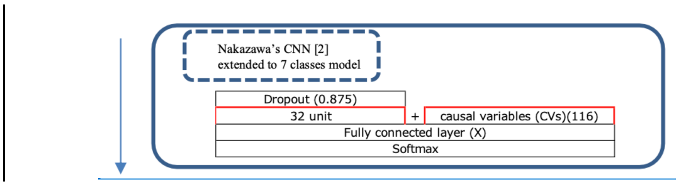
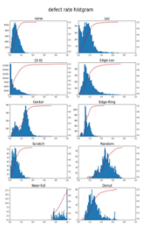
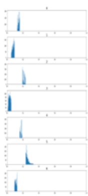


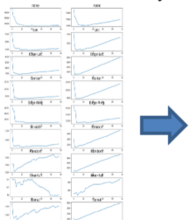
Fig. 2. Models proposed for multimodal analyses.



(a) Distribution of Defect rate by class



(b) Clustering result (e.g. "Loc" class)



DP class	# of clusters
Loc	9
Edge-Loc	3
Center	3
Edge-Ring	14
Scratch	3
Random	3

Left: AIC, Right: BIC

(c) The number of clusters determined by AIC/BIC criteria.

DP class	Cluster No.	Range of defect rate	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11
Edge-Loc	2	(0.36, 1.00]	1	1	1	1	1	1	1	1	1	1	1
Edge-Loc	1	(0.18, 0.36]	1	1	1	1	1	1	1	1	0	0	0
Edge-Loc	0	(0, 0.18]	1	1	1	1	1	0	0	0	0	0	0

(d) Example of "causal variables" added to each pair of a DP class and a cluster

Fig. 3. Clustering by DP class and causal variables.

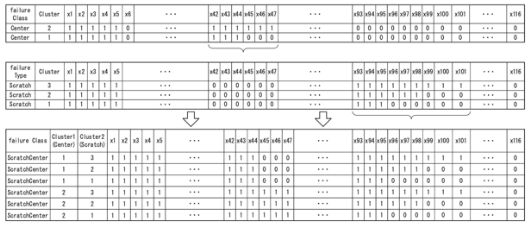


Fig. 4 Integrating CVs for SP DPs

Table. 1. Recall of DP SP (Scratch(s), Etch-Ring(r), Center(c), Edge-Loc(z))

Methods	Normal	S	Z	R	C	SZ	SR	SC	ZR	ZC	RC	SZR	SZC	SRC	ZRC	SZRC	Overall
TV+DecisionTree	**100**	**100**	**100**	99	99	92	89	**100**	**100**	**100**	**100**	**93**	**93**	90	**100**	**88**	**96.4**
TV+CNN	**100**	**100**	97	99	**100**	85	93	**100**	92	**100**	**100**	76	90	94	**100**	82	94.3
TV+2CNN	**100**	**100**	**100**	98	99	91	91	**100**	**100**	**100**	**100**	89	90	**99**	99	87	**96.4**
TV+CNN(s)	**100**	91	95	**100**	99	77	82	94	90	94	**100**	69	77	89	91	67	88.4
TV+2CNN(s)	**100**	91	**100**	99	99	88	89	93	**100**	**100**	**100**	90	85	98	**100**	83	94.7
HC+CNN	**100**	99	**100**	**100**	**100**	64	46	40	84	63	91	63	19	46	54	25	67.1
DBSCAN+CNN	**100**	80	**100**	52	**100**	63	57	83	64	**100**	50	59	51	48	61	47	70.5
BGM+CNN	**100**	**100**	44	79	98	**96**	74	98	88	88	96	72	88	65	94	41	82.6
Takeshi_CNN	99	97	95	98	99	84	**95**	92	89	96	99	76	81	84	90	76	90.6
Kiryong_CNN	99	99	**100**	95	99	84	93	95	98	**100**	97	73	84	96	93	79	92.8
Deformable_CNN	**100**	**100**	**100**	**100**	95	60	85	95	95	95	**100**	85	80	85	90	55	88.8

Table. 2. Definition for SP classes (Example)

(a) 2-class SP

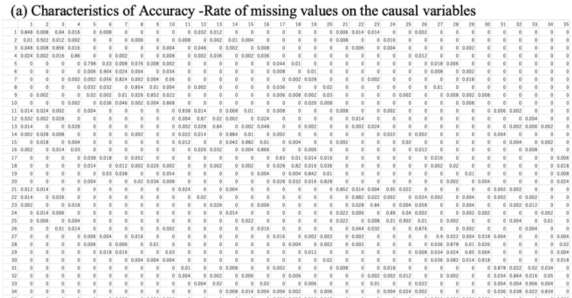
Class No.	Com bination of classes	Scratch	Center	Edge-Ring	Edge-Loc	Loc	Donut	Random
1	Loc					1		
2	Edge-Loc				1			
3	Center		1					
4	Edge-Ring			1				
5	Scratch	1						
6	Random							1
7	Donut						1	
8	Scratch+Center	1	1					
9	Scratch+Edge-Ring	1		1				
10	Scratch+Edge-Loc	1			1			
11	Scratch+Loc	1				1		
12	Scratch+Donut	1					1	
13	Scratch+Random	1						1
14	Center+Edge-Ring		1	1				
15	Center+Edge-Loc		1		1			
16	Center+Loc		1			1		
17	Center+Donut		1				1	
18	Center+Random		1					1
19	Edge-Ring+Edge-Loc			1	1			
20	Edge-Ring+Loc			1		1		
21	Edge-Ring+Donut			1			1	
22	Edge-Ring+Random			1				1
23	Edge-Loc+Loc				1	1		
24	Edge-Loc+Donut				1		1	
25	Edge-Loc+Random				1			1
26	Loc+Donut					1	1	
27	Loc+Random					1		1
28	Donut+Random						1	1

(b)4-class SP

Class No.	Combination of classes
1	ScratchCenterEdge-RingEdge-Loc
2	ScratchCenterEdge-RingLoc
3	ScratchCenterEdge-RingDonut
4	ScratchCenterEdge-RingRandom
5	ScratchCenterEdge-LocLoc
6	ScratchCenterEdge-LocDonut
7	ScratchCenterEdge-LocRandom
8	ScratchCenterLocDonut
9	ScratchCenterLocRandom
10	ScratchCenterDonutRandom
11	ScratchEdge-RingEdge-LocLoc
12	ScratchEdge-RingEdge-LocDonut
13	ScratchEdge-RingEdge-LocRandom
14	ScratchEdge-RingLocDonut
15	ScratchEdge-RingLocRandom
16	ScratchEdge-RingDonutRandom
17	ScratchEdge-LocLocDonut
18	ScratchEdge-LocLocRandom
19	ScratchEdge-LocDonutRandom
20	ScratchLocDonutRandom
21	CenterEdge-RingEdge-LocLoc
22	CenterEdge-RingEdge-LocDonut
23	CenterEdge-RingEdge-LocRandom
24	CenterEdge-RingLocDonut
25	CenterEdge-RingLocRandom
26	CenterEdge-RingDonutRandom
27	CenterEdge-LocLocDonut
28	CenterEdge-LocLocRandom
29	CenterEdge-LocDonutRandom
30	CenterLocDonutRandom
31	Edge-RingEdge-LocLocDonut
32	Edge-RingEdge-LocLocRandom
33	Edge-RingEdge-LocDonutRandom
34	Edge-RingLocDonutRandom
35	Edge-LocLocDonutRandom

	The rate of the missing value on the causal variables					
	1.0	0.8	0.6	0.4	0.2	0.0
Average (only single class)	0.8741	0.9640	0.9930	0.9975	0.9987	1.0000
Average (2-class SP with single)	0.6939	0.9033	0.9636	0.9623	0.9967	1.0000
Average (3-class SP)		0.6807	0.8310	0.9680	0.9945	1.0000
Average (4-class SP)		0.4600	0.8578	0.9586	0.9945	1.0000

(a) Characteristics of Accuracy -Rate of missing values on the causal variables



(b) Confusion matrix (the diagonal line is the Recall) (Rate of missing value=60 %)

Fig. 5. Numerical results of 4-class superpositions (all pairs).