A GENERALIZED SEMICONDUCTOR WAFER DEFECT CLASSIFIER

Priyanshu Kumar Rai[1*], **Pratik Pal**[1*], **Akshay Agarwal**[2]
[1]Department of Electrical Engineering and Computer Science
[2]Department of Data Science and Engineering
Indian Institute of Science Education and Research, Bhopal
{priyanshu20, pratik20, akagarwal}@iiserb.ac.in

ABSTRACT

Silicon-based integrated circuits (ICs) and electronic devices are used in every possible electronic device, including high-performance computers fabricated from silicon wafers. Hence, ensuring the quality and reliability of silicon components is of utmost importance. This work focuses on developing and implementing computer vision and deep learning algorithms to detect defects in semiconductor manufacturing ICs, contributing to higher yields and reduced production costs.

1 INTRODUCTION

Semiconductor defects occur during the fabrication process of silicon wafers, and accurately classifying these defects is crucial for fabrication engineers. Moreover, manual classification is laborious and challenging, as shown in Figure 1. Recently, several research works have explored the potential of convolutional neural networks (CNNs) Ishida et al. (2019); Shinde et al. (2022) for semiconductor wafer defect classification using scanning electron microscopes (SEMs). Further, Piao et al. (2018); Saqlain et al. (2019) proposed ensemble classifiers based on improved defect detection performance. One significant drawback of the datasets capturing silicon wafer defects is that a few defects are more frequent as compared to others and hence lead to biased learning. The above limitation is visible in the performance of the existing algorithms, which fail to handle minority classes and, hence, lack generalizability. The proposed research addresses these issues by training noise augmentation to improve the generalizability of the feature space. Further, we have developed a custom CNN model by considering the computational cost of edge devices and comparing it with the existing state-of-the-art (SOTA) methods to demonstrate the efficacy of the proposed algorithm.

2 PROPOSED ALGORITHM

The proposed algorithm consists of two steps: (i) In the first step, we applied the autoencoder to generate the novel samples to upsample the minority classes and applied the noise layer to ensure that the generated samples are distinctive. This augmentation process yields a total of $39,023$ data points, ensuring a balanced distribution of classes. (ii) In the second stage, we have proposed a novel CNN to perform wafer defect classification. The parameters for training and the model architecture of both the autoencoder and the CNN are reported in Appendix A and Table 2, respectively.

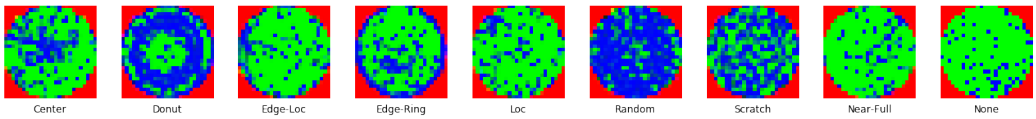

Figure 1: Sample examples of semiconductor wafer failure types in the WM-811k dataset.

3 DATASET

In this research, we have performed a 9 class classification to classify silicon wafers into defect and non-defect classes. For that, a benchmark dataset namely WM-811K Wu et al. (2014) containing

[*]These authors contributed equally to this work

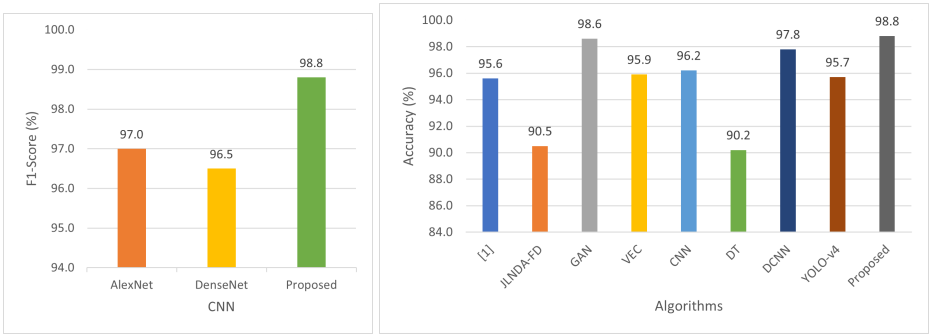

Figure 2: Comparison of the proposed wafer defect detector with the benchmark CNNs and SOTAs.

Table 1: Confusion matrix of the proposed algorithm reflecting the potential for detecting different defect classes.

True ↓ Predicted →	Center	Donut	Edge-Loc	Edge-Ring	Loc	Random	Scratch	Near-Full	None
Center	**936**	0	0	0	2	0	1	0	2
Donut	0	**923**	0	0	0	0	0	0	0
Edge-Loc	0	0	**1058**	8	2	0	4	0	3
Edge-Ring	0	0	2	**924**	0	0	0	0	2
Loc	8	0	11	0	**1009**	0	1	3	16
Random	0	0	0	0	0	**911**	2	0	0
Scratch	0	0	0	0	2	0	**912**	0	0
Near-Full	0	0	0	0	1	0	0	**599**	11
None	1	0	1	0	9	0	0	6	**730**

$811,457$ wafer maps of eight defect classes (Center: 4294; Donut: 555; Edge-Loc: 5189; Edge-Ring: 9688; Loc: 3593; Random: 866; Scratch: 1193; Near-Full: 149) collected from $46,393$ lots in real-world fabrication has been used. No defect class referred to as 'none' contains 13489 images.

4 Experimental Results and Analysis

To perform the extensive set of comparisons, in addition to our proposed custom CNN, we have trained the AlexNet Krizhevsky et al. (2012) and the DenseNet121 Huang et al. (2017) models for wafer defect classification. Further, we compare the performance of the proposed algorithm with several existing SOTAs, such as [1] Baly & Hajj (2012), Fisher-discriminant-based Joint Local and Nonlocal Linear Discriminant Analysis (JLNDA-FD) Yu & Lu (2016), Generative Adversarial Network (GAN) Ji & Lee (2020), Voting Ensemble Classifier (VEC) Saqlain et al. (2019), CNN Cheon et al. (2019), decision tree (DT) Piao et al. (2018), Deep CNN (DCNN) Chien et al. (2020), and YOLO-v4 Shinde et al. (2022).

The results of the proposed and existing SOTA CNNs, such as AlexNet and DenseNet, are reported using accuracy and the F-1 score to avoid any possible bias. The comparative results of the proposed algorithm with CNNs in terms of F-1 score are showcased in Figure 2 (left). The comparison reflects the exciting fact that the proposed algorithm surpasses both shallow (AlexNet) and deeper (DenseNet) models. *Similarly, as shown in Figure 2 (right), the proposed cost-effective algorithm shows its effectiveness by surpassing each of the existing SOTA algorithms by a significant margin*. Figure 1 visually showcases that it is hard to manually detect defects due to low inter-class and high intra-class variation; however, the confusion matrix reported in Table 1 shows that the proposed algorithm is effective and unbiased in detecting different defect classes, including none.

5 Conclusion

Once the product is developed, identification of any fault can lead to not only the waste of human resources but also monetary losses; therefore, it is critical to detect the wafer defects in the early stages. Low inter-class variations, manual classification, and the ineffectiveness of the existing algorithms demand an effective and unbiased algorithm. In this research, we have proposed a robust and generalized wafer detection algorithm that surpasses existing algorithms and yields low computational costs to be deployed on edge devices.

URM STATEMENT

The authors acknowledge that the key author of this work meets the URM criteria of the ICLR 2024 Tiny Papers Track.

REFERENCES

Ramy Baly and Hazem Hajj. Wafer classification using support vector machines. *IEEE Transactions on Semiconductor Manufacturing*, 25(3):373–383, 2012. doi: 10.1109/TSM.2012.2196058.

Sejune Cheon, Hankang Lee, Chang Ouk Kim, and Seok Hyung Lee. Convolutional neural network for wafer surface defect classification and the detection of unknown defect class. *IEEE Transactions on Semiconductor Manufacturing*, 32(2):163–170, may 2019. doi: 10.1109/tsm.2019.2902657.

Jong-Chih Chien, Ming-Tao Wu, and Jiann-Der Lee. Inspection and classification of semiconductor wafer surface defects using cnn deep learning networks. *Applied Sciences*, 10(15):5340, 2020. doi: https://doi.org/10.3390/app10155340.

Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, 2017. doi: 10.1109/CVPR.2017.243.

Tsutomu Ishida, Izumi Nitta, Daisuke Fukuda, and Yuzi Kanazawa. Deep learning-based wafer-map failure pattern recognition framework. In *20th International Symposium on Quality Electronic Design (ISQED)*, pp. 291–297, 2019. doi: 10.1109/ISQED.2019.8697407.

YongSung Ji and Jee-Hyong Lee. Using gan to improve cnn performance of wafer map defect type classification : Yield enhancement. In *2020 31st Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, pp. 1–6, 2020. doi: 10.1109/ASMC49169.2020.9185193.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

Minghao Piao, Cheng Hao Jin, Jong Yun Lee, and Jeong-Yong Byun. Decision tree ensemble-based wafer map failure pattern recognition based on radon transform-based features. *IEEE Transactions on Semiconductor Manufacturing*, 31(2):250–257, 2018. doi: 10.1109/TSM.2018.2806931.

Muhammad Saqlain, Bilguun Jargalsaikhan, and Jong Yun Lee. A voting ensemble classifier for wafer map defect patterns identification in semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, 32(2):171–182, 2019. doi: 10.1109/TSM.2019.2904306.

Prashant P. Shinde, Priyadarshini P. Pai, and Shashishekar P. Adiga. Wafer defect localization and classification using deep learning techniques. *IEEE Access*, 10:39969–39974, 2022. doi: 10.1109/access.2022.3166512.

Ming-Ju Wu, Jyh-Shing R Jang, and Jui-Long Chen. Wafer map failure pattern recognition and similarity ranking for large-scale data sets. *IEEE Transactions on Semiconductor Manufacturing*, 28(1):1–12, 2014.

Jianbo Yu and Xiaolei Lu. Wafer map defect detection and recognition using joint local and nonlocal linear discriminant analysis. *IEEE Transactions on Semiconductor Manufacturing*, 29(1):33–43, 2016. doi: 10.1109/TSM.2015.2497264.

A IMPLEMENTATION DETAILS

We train the autoencoder using the Adam optimizer with minimization of the MSE loss. Further, the initial learning rate of the proposed CNN is set to 0.01 with an adaptive decay rate at the plateau of the categorical cross-entropy loss. The dataset is split into a training set comprising 70% and a

testing set comprising the remaining 30% data points. A 3-fold cross-validation approach is implemented, incorporating random data shuffling to mitigate any potential biases introduced during the training process. The batch size, optimizer, and epochs used for training are 1024, Adam, and 45, respectively. The configuration of both architectures used in the development of silicon wafer defect detection is shown in Table 2. AlexNet and DenseNet are trained from scratch using the same configuration used in the case of the proposed CNN model using 100 and 50 as the number of epochs, respectively.

Table 2: Configuration of the autoencoder and the proposed CNN

Model	**Configuration**
Autoencoder	Conv($3 \times 3 \times 64$), ReLU, Conv($3 \times 3 \times 128$), ReLU, Conv($3 \times 3 \times 256$), ReLU, MaxPool(3×3), ConvTranspose($3 \times 3 \times 128$), ReLU, ConvTranspose($3 \times 3 \times 64$), ReLU, UpSampling(3×3), ConvTranspose($3 \times 3 \times 3$), sigmoid
Proposed CNN	Conv($3 \times 3 \times 16$), ReLU, Conv($3 \times 3 \times 64$), ReLU, Conv($3 \times 3 \times 128$), ReLU, Conv($3 \times 3 \times 256$), ReLU, Flatten, FullyConnected(512), FullyConnected(128), FullyConnected(9), SoftMax

B FIGURES

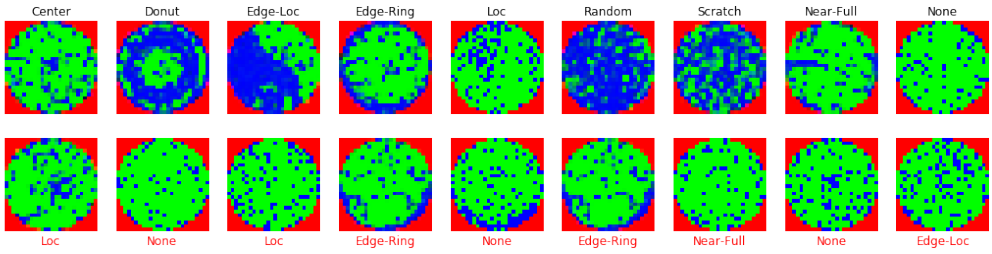

Figure 3: Correctly and incorrectly classified samples for some classes.

In the above figure, we show the correctly classified samples in the first row with their respective classes and the incorrectly classified samples in the second row along with the class in which they are misclassified. For classes with no misclassified points, the samples from other classes are included below, depending on their frequency of misclassification. It is evident from the figure mentioned above that misclassified samples exhibit properties characteristic of both the actual class and the misclassified class. Consequently, instances may be misclassified due to the overlapping characteristics shared between the actual and misclassified classes. In addition, there may be instances where the wafer map corresponds to a class that has not been explicitly labeled yet or does not strictly classify to one of the existing classes. As we have confined the training process to the provided labels, misclassifications can occur when encountering unclassified classes.

Table 3: Comparison of class-wise F1-scores of the proposed algorithm with SOTAs.

Classes	AlexNet	Saqlain et al. (2019)	DenseNet	Shinde et al. (2022)	**Proposed**
Center	98.9	89.8	94.3	98.0	99.3
Donut	100.0	86.0	99.4	96.0	100.0
Edge-Ring	97.2	96.3	99.2	99.0	98.6
Near-Full	100.0	96.5	100.0	100.0	99.3
Random	98.0	92.4	98.9	94.0	97.3
Edge-Local	92.7	79.9	93.0	95.0	99.9
Scratch	96.1	53.3	96.1	93.0	99.4
Local	94.7	67.0	93.6	93.0	98.3
None	94.9	98.8	96.2	97.0	99.6
Overall	**97.0**	**96.8**	**96.5**	**95.7**	**98.8**

C Ablation Study: Classwise Performance Comparison

To quantify our observations, we have compared the F1 scores of the proposed algorithm of each class with the SOTAs. From the results reported in Table 3, it is observed that while the existing works are found highly effective in detecting a few classes, their performances are not consistent across the classes. For example, the algorithm by Shinde et al. (2022) has achieved 100% performance on Near-full classes but shows significantly lower accuracies on multiple classes such as scratch and local. The proposed not only outperforms the existing work in terms of overall performance but yields consistent results across the classes reflecting lower signs of biasness towards any particular class.

D Abbreviations

IC	Integrated Circuits
SEM	Scanning Electron Microscope
CNN	Convolutional Neural Network
DCNN	Deep Convolutional Neural Network
ADC	Automatic Defect Classification
ANN	Artificial Neural Network
VGG	Visual Geometry Group Net
YOLO	You only look once
ReLU	Rectified Linear Unit
CBAM	Convolutional Block Attention Module
JLNDA-FD	Fisher-discriminant-based Joint Local and Nonlocal Linear Discriminant Analysis