IFERP connecting engineers... developing research

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)

Vol 10, Issue 8, August 2023

SM-CapsNet: A SMOTE Based Capsule Network Approach for Semiconductor Wafer Defect Detection

[1] Anmol Bhasin, [2] Ramachandran Shekar, [3] Dr. Sanjeev Sofat

[1] Punjab Engineering College, Chandigarh, India
[2] Intel Technologies India Pvt. Ltd., India
[3] Professor, Department of Computer Science and Engineering, Punjab Engineering College, Chandigarh, India
Corresponding Author Email: [1] anmolbhasin14@gmail.com, [2] shekar.ramachandran@intel.com, [3] sanjeevsofat@pec.edu.in

Abstract— ***The detection of defects in semiconductor wafers is of paramount importance in guaranteeing the quality and dependability of silicon wafers. However, AI based detection of defects in wafers is challenging due to imbalanced datasets, where the number of defective samples is considerably lower than non-defective wafer bin maps. The study introduces SM-CapsNet, an approach that tackles the issue of class imbalance by utilizing SMOTE (Synthetic Minority Over-sampling Technique) alongside a modified capsule network. This combination enhances the classification model's ability to localize defects effectively. The methodology is evaluated on the widely used WM-811K dataset, which contains bin-map images of semiconductor wafers captured under various conditions. The experiments show that the proposed method outperformed standard neural networks based on accuracy and F1-score metrics. It is also noted that using SMOTE for data generation requires less time and resources compared to traditional data augmentation techniques like GANs and CAEs. Thus, SM-CapsNet depicts potential for accurate semiconductor wafer defect detection to improve silicon wafer production yields in the in the semiconductor industry.***

Index Terms— ***Semiconductor Wafer Defect Detection, SMOTE, Class Imbalance, Capsule Networks, WM-811K Semiconductor Wafers, AI Applications in Industry, Artificial Intelligence, Deep Learning.***

I. INTRODUCTION

Semiconductor wafers are foundational components in the production of integrated circuits (ICs) and microelectronic devices. They serve as the substrate on which intricate circuitry and electronic components are built. The wafer fabrication process is complex, involving multiple stages which are performed in controlled environments of semiconductor labs [1]. However, despite stringent quality control measures, defects can occur due to a variety of factors throughout the fabrication process including Inadequate cleaning and polishing procedures, improper deposition, misalignment during photoresist exposure and others. Defects can also be introduced through equipment malfunctions, contamination, process variations, or human errors [2]. As such, accurate detection of these defects in the wafer maps at the time of testing becomes of a vital importance. Wafer defects exhibit distinct patterns, which experts analyze through wafer bin maps for detection and categorization. However, this manual inspection process is time-consuming and lacks precision, impacting efficiency. As the demand for semiconductors continues to surge, the importance of automatic wafer defect detection grows. While various machine learning techniques have been proposed to accurately detect defects, their practical implementation faces limitations due to the scarcity of labeled and clean data. Consequently, research in semiconductor wafer defect detection aims to address the gaps in existing methodologies and develop production-ready solutions for efficient wafer testing, ensuring improved detection efficiency while maintaining quality standards.

II. LITERATURE REVIEW

The detection and classification of defects in semiconductor wafer maps (WMs) is a critical task in the semiconductor wafer fabrication and manufacturing industry. Manual classification of defects is challenging and time-consuming, making the automation of this process desirable. In recent years, deep learning models have shown great potential for automating defect classification in WMs. Several research papers propose innovative methods and techniques to address this problem. Here is a concise literature review summarizing the key findings and approaches discussed in these papers:

Tello [3] proposed a dual-stage transfer learning mechanism for defect classification in WMs. Their approach utilized a pretrained model for feature extraction, followed by training smaller models on these features. The study demonstrated the effectiveness of transfer learning in reducing overfitting and producing a generalized model.

In [4], a novel scheme employing a Convolutional Autoencoder (CAE) with skip connections was introduced for WM classification. The method utilized the CAE to generate a concise representation of the input WMs, which was then used for defect categorization using a classifier. The inclusion of skip connections helped preserve minute details during the encoding and decoding processes.

[5] presented a novel approach using dual-channel CNNs for WM defect classification. The dual-channel CNN architecture consisted of two CNNs operating on separate channels, one containing original wafer map data and the other incorporating the concept of a difference map that pinpoints the location of defects. The two CNNs were trained in competition, and the final classification result was a combination of their outputs.

A lightweight model for structural wafer map defect detection was introduced in [6]. The proposed model utilized Conv-2D and FC layers to prioritize computational efficiency with fewer learning parameters. The study demonstrated the efficacy of the lightweight model in production environments for semiconductor defect detection.

[7] addressed the classification and detection of mixed defect patterns in a single wafer map using data augmentation techniques and mask Recurrent Convolution Network-based instance segmentation. The approach employed classical augmentation techniques and masking for accurate defect identification and classification. The study showcased the superiority of this method over traditional classification methods in terms of accuracy on large wafer map datasets.

A novel method using Single Shot Detector (SSD) was proposed in [8] for detecting mixed wafer defect patterns in a single wafer map. The method involved preprocessing the wafer maps to improve the detection of defect patterns in a single shot. The experiments demonstrated that the proposed method outperformed existing methods in terms of detection accuracy and computational efficiency.

Semi-supervised VAE and Ladder network approaches were compared in [9] for categorizing defects in wafer bin maps. Active learning and pseudo labelling techniques were employed to expedite the learning process and improve classification performance.

[10] presented a methodology involving training a prediction model using labeled samples and eliminating uncertain samples during the process. Unlabeled data was then classified using the prediction model, and an unsupervised learning system was used to extract additional fault patterns. The study successfully recognized 14 defect types and defined five new defect pattern types.

In [11], a hybrid approach combining deep learning and machine learning techniques was proposed for detecting both known and unknown defects in wafer maps. The approach utilized a CNN for feature extraction and a support vector machine (SVM) for defect identification and classification, achieving higher accuracy in defect detection.

A semi-supervised method incorporating label smoothing and CNN ensemble was suggested in [12]. The approach aimed to address CNN models' overfitting and promote more balanced predictions for training data.

[13] explored state-of-the-art transfer learning methods for the classification of wafer map defects. The authors conducted a comparative analysis of pretrained light deep learning models and investigated different fine-tuning strategies. The study demonstrated high accuracy in defect classification.

However, it is evident from the existing research that class imbalance of wafer defects occurs predominantly in public semiconductor wafer defect datasets along with several classification issues pertaining to neural network classification models [14].

III. PROPOSED METHODOLOGY

The proposed methodology SM-CapsNet is a combination SMOTE for handling class imbalance and a modified Capsule Network architecture to address translational invariance. The study aims to tackle the major research gaps from the existing literature survey given in section II of the paper to provide a comprehensive solution for the accurate classification of defects in semiconductor wafer maps.

A. Dataset

The research utilizes the widely recognized Wafer Map WM-811K dataset [15] as a benchmark for semiconductor wafer defect detection. The dataset includes 811 unique die wafer maps, representing wafers with grids of chips. The size of the maps varies, ranging from several hundred to several thousand chips. The dataset, previously analyzed in [16] and [17], encompasses nine categories, including eight defect types and one non-defect category. The defect categories consist of Center, Donut, Edge-Loc, Edge-Ring, Loc, Random, Scratch, and Near Full, each representing specific defect patterns on the wafer map. The ninth category, None, represents wafer maps without any defects. Fig. 1 depicts all nine wafer map defects denoted in WM-811K dataset.

The WM-811K dataset includes labels indicating the defect categories present on each chip or die in the wafer map. The dataset includes wafer binary map samples from fabrications, with 3.1% of the wafers (25,519 wafers) exhibiting real defects, 18.2% being wafers with no defects (147,431 wafers), and the remaining 78.7% (638,570 wafers) being unlabeled. The labels provide information about the type and location of defects. Among the labeled 21.3% data, a total of eight defects and one non-defect type wafer exists. Table I represents the percentage of samples present in each class in WM-811K dataset.

Table I .WM-811K: Labeled Data Distribution

Label	Original Data Count	Data %
None	**147,431**	**85.2%**
Edge-Ring	9680	**5.6%**
Edge-Loc	5189	**3.0%**
Center	4294	**2.5%**
Loc	3593	**2.1%**
Scratch	1193	**0.7%**

Label	Original Data Count	Data %
Random	866	$\mathbf{0.5\%}$
Donut	555	$\mathbf{0.3\%}$
Near Full	149	$\mathbf{0.1\%}$

To tackle the imbalance in minority and majority classes, this study proposes the use of Synthetic Minority Over-sampling Technique (SMOTE) for data augmentation.

B. Synthetic Minority Over-sampling Technique (SMOTE)

SMOTE (Synthetic Minority Over-sampling Technique) is a widely used data augmentation method designed to address class imbalance in datasets. Its purpose is to create synthetic samples for the minority class, thereby improving the balance between different classes [18]. The SMOTE algorithm operates by interpolating new instances between existing samples of the minority class. To generate synthetic samples, SMOTE selects a minority class instance and identifies its k nearest neighbors using a distance metric. The algorithm is defined as follows-

1. Select a minority class instance from the dataset in question. Let's denote the selected minority class instance as x_i, where i represents the index of the instance in the dataset.
2. Determine the k nearest neighbors of the selected instance. Using a distance metric, such as Euclidean distance, the distances between x_i and all other instances in the dataset are calculated. The k nearest neighbors are the instances with the smallest distances to x_i. Denote the set of k nearest neighbors as $NN(x_i)$.
3. Randomly choose one of the nearest neighbors. From the set of k nearest neighbors, we randomly select one instance. Let's denote the randomly selected neighbor as x_n, where n represents the index of the neighbor instance.
4. Generate a synthetic sample by interpolating between the selected instance and the chosen neighbor. To create a synthetic sample, the difference vector (d) between x_i and x_n, denoted as

$$d = x_n - x_i$$

Then, a random number between 0 and 1, denoted as λ is chosen. The synthetic sample, denoted as x_{new}, is generated as follows:

$$x_{new} = x_i + \lambda \times d$$

Here, λ controls the position of the synthetic sample between x_i and x_n. By varying λ from 0 to 1, we obtain different synthetic samples along the line segment connecting the original instance and the neighbor instance [19]

5. Repeat steps 1-4 until the required balance among classes is achieved. Steps 1-5 are repeated until minority class is sufficiently represented in the dataset, achieving the necessary class balance.

Parameter k, representing the number of nearest neighbors, is typically chosen based on the characteristics of the dataset and with respect to large modulations, a larger value of k provides generalized samples. Different choices of k impact the density of synthetic samples and the degree of oversampling [19].

To ensure that the generated samples are comparable enough to be used in the augmented dataset, similarity between the generated samples and the synthetic samples is calculated using PSNR ratio and cosine similarity. By applying SMOTE, the researchers intend to provide a more balanced representation of the different defect classes, thereby improving the performance of the classification model.

C. Proposed Capsule Network

The introduction of Convolutional Neural Networks (CNNs) revolutionized computer vision by enabling the extraction of important visual features and accurate image classification. However, CNNs have limitations, including the loss of positional information and the misclassification of samples [20]. To address these challenges, this research proposes the use of Capsule Networks, a novel deep learning architecture inspired by the human brain's hierarchical connections. Capsule Networks mimic how the brain learns and interprets visual information by capturing orientation, size, and presence through capsules instead of neurons [21]. This approach preserves hierarchical relationships between objects and parts, leading to improved image classification and segmentation performance. By leveraging the benefits of Capsule Networks, this research aims to overcome the limitations of CNNs in spatially invariant image classification, specifically in wafer-map defect classification. Table II describes the proposed architecture of Capsule Network that consists of initial Convolution 2D layers to narrow down the feature space of wafer maps, followed by primary capsule layers and digit capsule layers to obtain routing by agreement in the consecutive feature layers.

Capsule networks calculate the classification loss in a different way then conventional CNNs. Margin loss is employed for accurate image classification by comparing the predicted probabilities of correct and incorrect classes. It is calculated by summing the squared differences between the predicted probability and a margin hyperparameter for both correct and incorrect classes. The margin loss equation is as follows:

$$L_{margin} = \sum [t_i \times \max(0, m^+ - a_j)^2 + \lambda(1 - t_{i)} \times \max(0, a_j - m^-)^2]$$

where t_i is the target label activation and a_j is the j^{th} capsule activation for i^{th} class, m^+ is the margin for positive

ISSN (Online) 2394-6849
International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 10, Issue 8, August 2023

class, m^- is the margin for negative class and λ is the regularization parameter [30]. Reconstruction loss is utilized as a training mechanism to ensure effective encoding of input image features. It measures the squared differences between the input image and the image generated by the capsule network. It is basically calculated as the L2 norm. The overall loss, denoted as L, is a combination of both the margin loss and the reconstruction loss, weighted by a factor α:

$$L = L_{margin} + \alpha(L_{reconstruction})$$

The weighting factor α balances the contribution of each loss term, allowing for effective optimization of the capsule network during training.

Thus, using a combination of the above losses capsule network overcomes the loss of positional information in CNNs, preserves hierarchical relationships, and provides improved handling of spatial invariance.

Table II . Proposed Capsule Network Architecture

Input	Operator	c	k	s	n
$256^2 \times 3$	Conv 2D	32	7	5	
$50^2 \times 32$	Conv 2D	64	3	1	-
$24^2 \times 64$	Primary	8	5	2	-
$10^2 \times 8 \times 8$	Conv Capsules	5	-	-	8
$800 \times 8D\ Vectors$	Digit Caps	-	-	-	16
$1^2 \times 16 \times 5$	Linear	-	-	-	-
$1^2 \times 512$	Linear	-	-	-	-
$1^2 \times 1024$	Linear	-	-	-	-
$1^2 \times 196608$	Activation (Sigmoid)	-	-		-

IV. EXPERIMENTAL SETUP

In this section, the experiments conducted to evaluate the effectiveness of proposed model against the state-of-the-art classification models using two subsets of WM-811K dataset- one with SMOTE generated data and the other baseline data with classical augmentation techniques are described.

Multiclass SMOTE (Synthetic Minority Over-Sampling Technique) is used to produce random samples of the minority value classes based on selection k-nearest numbers as described in section III-(B). Table III depicts the parameters used to generate SMOTE samples for minority classes.

Table III. Configuration for SMOTE Sample Generation

Parameters	Value
k (k-nearest neighbor)	5
Random Sampling State	45
m (m-neighbor)	3
No of parallel jobs	5

A sampling strategy of ratio of samples required to be generated and the actual samples in majority class is used here to generate 5000 samples for each of the minority class.

Experiments are conducted to perform a comparative analysis of semiconductor wafer defect dataset using three different classification networks namely- MobileNetV2, ImageNet, and proposed Capsule Network (SM-CapsNet). The dataset consists of 5000 colored wafer bin maps from each of the 8 defect classes and one non-defect class with an image size of 256 x 256 pixels. The training parameters described in Table IV are kept constant for the experiments while the network model for evaluation changes in each set. The training is conducted on a dual Nvidia A100 GPU.

Table IV. Training Configuration for Classification Networks

Parameters	Value
Epochs	30
Learning Rate	0.01
Batch Size	32
Optimizer	SGD
Input Image Size	3 x 256 x 256

V. RESULTS AND ANALYSIS

Synthetic Minority Over-sampling Technique (SMOTE) is used to generate samples for the minority classes in the dataset including Center, Donut, Loc, Random, Scratch, Near full. Similarity of the generated samples is calculated using PSNR ratio and cosine similarity as described in Section III B of the paper. PSNR ratio greater 30dB and a cosine similarity value closer to 1 depict higher similarity among the samples and original dataset. Table 3 represents the average PSNR ratio and cosine similarity for a random of 100 generated samples with actual samples of each minority class.

Table V. Average PSNR and Average Cosine Similarity Comparison of Generated and Original Samples

Minority Label	PSNR Ratio (dB)	Cosine Similarity
Center	32.12	0.894
Donut	31.45	0.940
Random	32.04	0.950
Loc	30.85	0.929
Scratch	31.78	0.937
NearFull	31.21	0.887

Further, a comparison is drawn between the conventional image generation techniques – GANs and CAEs, and SMOTE. Resource utilization and training duration for GANs [22] and CAEs [4] is compared with SMOTE. It is noted that since SMOTE does not involve learning via training, it results in optimization of resources utilized. Table II gives the gist of the above optimization flow.

IFERP connecting engineers... developing research

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)

Vol 10, Issue 8, August 2023

Upon the successful generation of SMOTE samples, we use the augmented balanced subset of WM-811K dataset for classification of defects in semiconductor wafers. This study aims at providing a comparative analysis of state-of-the-art classification methods including MobileNetV2 [23], ImageNet [24] and the proposed Capsule Network for classification on balanced subset of WM-811K dataset and an unbalanced subset of the dataset where the minority classes are not augmented.

Table VI. Resource Comparison between SMOTE, and GANs and CAEs

	NVIDIA A100 GPUs x2		Intel Core i9 Processor CPU	
Method	**Train Time (mins)**	**Utilization**	**Train Time (mins)**	**Utilization**
Style GANs	135	80%	786	89%
CAEs	124	70%	600	95%
SMOTE	4	11%	10	60%

Figure 2-(a) and 2-(b) depict the variation of accuracy in percentage when MobileNetV2 is used for classification on Wafer-Map data without SMOTE generated samples and with SMOTE augmentation respectively.

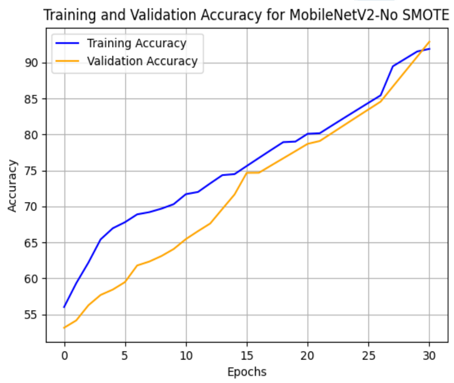

Fig. 2 Variation of Accuracy for MobileNetV2
(a) On Baseline WM-811K Dataset without SMOTE

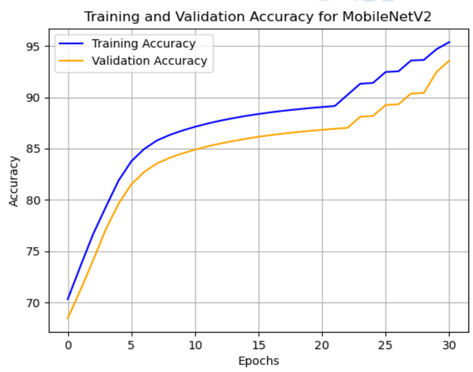

Fig. 2 Variation of Accuracy for MobileNetV2
(b) On WM-811K Dataset with SMOTE

Figure 3-(a) and 3-(b) depict the variation of accuracy in percentage when ImageNet is used for classification on Wafer-Map data without SMOTE generated samples and with SMOTE augmentation respectively.

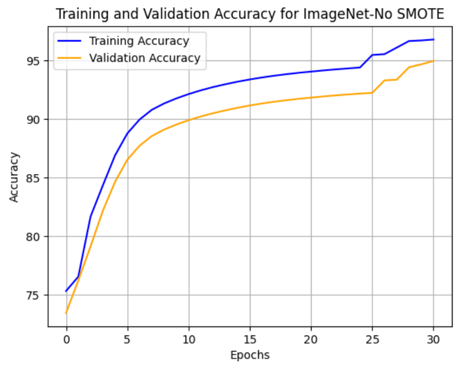

Fig. 3 Variation of Accuracy for ImageNet
(a) On Baseline WM-811K Dataset without SMOTE

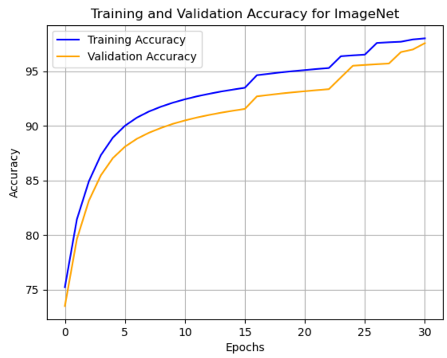

Fig. 3 Variation of Accuracy for ImageNet
(b) On WM-811K Dataset with SMOTE

Figure 4-(a) and 4-(b) depict the variation of accuracy in percentage when proposed Capsule Network is used for classification on Wafer-Map data without SMOTE generated samples and with SMOTE augmentation respectively. As is depicted by the results, SMOTE balanced data performs better classification of semiconductor defects when trained on each of the MobileNetV2, ImageNet and CapsNet than its counter data that is not balanced using synthetic minority oversampling.

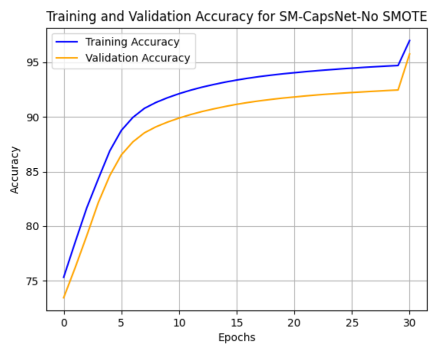

Fig. 4 Variation of Accuracy for Proposed Capsule Network
(a) On Baseline WM-811K Dataset without SMOTE

ISSN (Online) 2394-6849
International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 10, Issue 8, August 2023

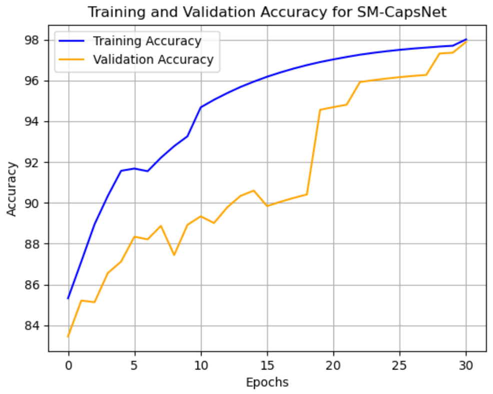

Fig. 4 Variation of Accuracy for Proposed Capsule Network (b) On WM-811K Dataset with SMOTE

Based on the experimental setup, Table IV showcases the maximum accuracies for baseline and improved methodology using SMOTE, the training and validation losses for each case and the F1 Score for minority classes. It is noted that proposed SM-CapsNet when clubbed with SMOTE beats the extensive state of the art ImageNet model for semiconductor wafer defect detection problem by a small margin of 0.25%. However, a jump of 0.009 is observed in terms of F1 scores of minority classes which along with resource optimization depicted in Table III make SM-CapsNet a suitable method for wafer defect detection in semiconductor manufacturing in production.

For the baseline dataset, classical augmentation techniques are used in order to generate a dataset with equal samples for all classes. However, due to the presence of limited samples for minority defects classical augmentation techniques perform substandard as compared to SMOTE where the minority class is scaled to the majority class using better sample generation.

Table VII. Comparative Analysis of SMOTE and Baseline Methods

Model	SMOTE	Training Accuracy	Validation Accuracy	F1 Score	Training Loss	Validation Loss
MobileNetV2	No	93.57%	92.91%	0.876	0.904	0.756
	Yes	95.38%	93.59%	0.899	0.456	0.254
Image Net	No	96.78%	94.95%	0.9	0.956	0.662
	Yes	98.03%	97.59%	0.917	0.404	0.456
Proposed Capsule Network	No	97.01%	**95.76%**	**0.921**	0.678	**0.704**
	Yes	**98.99%**	**97.84%**	**0.926**	0.046	**0.045**

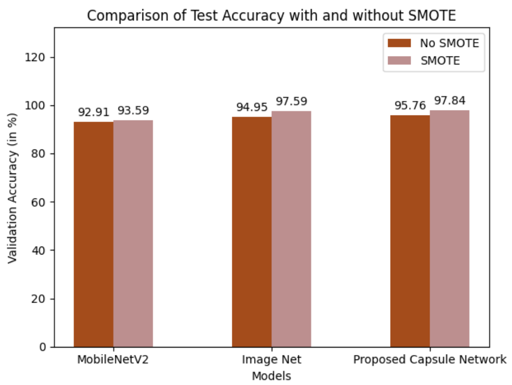

Fig 5. Comparative Analysis of Model Accuracies with SMOTE/CAT data

Fig. 5 represents a comparative analysis of accuracies on classical augmentation of data using rotation, flipping and other image transformations, and accuracies of various classification methods on data generated with the help of SMOTE.

VI. CONCLUSION

The proposed solution uses a combination of SMOTE for generation of balanced labeled dataset. Further, the proposed study uses capsule networks for dynamic routing and defect detection as the base classification network Further a comparative analysis with various state of the art deep neural networks is performed that reveals that our presented method, SM-CapsNet with SMOTE and Capsule Network as the classification model closely outperforms traditional transfer learning mechanisms on WM-811K balanced dataset. The analysis of resources and training time required on each resource for SMOTE generation shows that SMOTE can be a light-weight solution for data augmentation in comparison with GANs and other classical data augmentation techniques.

In future, a semi-supervised approach can be used to leverage the entire available wafer bin map samples rather than just the labeled subset of WM-811K data.

REFERENCES

[1] J.W.Fowler, L.Mönchand T. Ponsignon, "DISCRETE-EVENT SIMULATION FOR SEMICONDUCTOR WAFER FABRICATION FACILITIES: A TUTORIAL.," International Journal of Industrial Engineering, vol. 22, 2015.

[2] S. Maldonado, "The importance of new “sand-to-silicon” processes for the rapid future increase of photovoltaics," ACS Energy Letters, vol. 5, p. 3628–3632, 2020

[3] G. Tello, O. Y. Al-Jarrah, P. D. Yoo, Y. Al-Hammadi, S. Muhaidat and U. Lee, "Deep-structured machine learning model for the recognition of mixed-defect patterns in semiconductor fabrication processes," IEEE Transactions on Semiconductor Manufacturing, vol. 31, p. 315–322, 2018.

[4] J. Cha, J. Park and J. Jeong, "A Novel Defect Classification Scheme Based on Convolutional Autoencoder with Skip Connection in Semiconductor Manufacturing," in 2022 24th International Conference on Advanced Communication Technology (ICACT), 2022.

[5] T.-H. Tsai and Y.-C. Lee, "A light-weight neural network for wafer map classification based on data augmentation," IEEE Transactions on Semiconductor Manufacturing, vol. 33, p. 663–672, 2020.

[6] J. Yu and J. Liu, "Two-dimensional principal component analysis-based convolutional autoencoder for wafer map defect detection," IEEE Transactions on Industrial Electronics, vol. 68, p. 8789–8797, 2020.

[7] W. Shin, H. Kahng and S. B. Kim, "Mixup-based classification of mixed-type defect patterns in wafer bin maps," Computers & Industrial Engineering, vol. 167, p. 107996, 2022.

[8] S. Manivannan, "An ensemble-based deep semi-supervised learning for the classification of Wafer Bin Maps defect patterns," Computers & Industrial Engineering, vol. 172, p. 108614, 2022.

[9] K. N. T. I. T. H. K. a. S. Y. Imoto, "A CNN-based transfer learning method for defect classification in semiconductor manufacturing," in IEEE Transactions on Semiconductor Manufacturing, 32(4), pp.455-459, 2019.

[10] K. S.-M. Li, X.-H. Jiang, L. L.-Y. Chen, S.-J. Wang, A. Y.-A. Huang, J. E. Chen, H.-C. Liang and C.-L. Hsu, "Wafer Defect Pattern Labeling and Recognition Using Semi-Supervised Learning," IEEE Transactions on Semiconductor Manufacturing, vol. 35, p. 291–299, 2022.

[11] T. a. B. K. Kim, "Advances in machine learning and deep learning applications towards wafer map defect recognition and classification: a review.," Journal of Intelligent Manufacturing, pp. 1-33, 2022.

[12] F. López de la Rosa, J. L. Gómez-Sirvent, C. Kofler, R. Morales and A. Fernández-Caballero, "Detection of Unknown Defects in Semiconductor Materials from a Hybrid Deep and Machine Learning Approach," in International Work-Conference on the Interplay Between Natural and Artificial Computation, 2022.

[13] R. Doss, J. Ramakrishnan, S. Kavitha, S. Ramkumar, G. Charlyn Pushpa Latha and K. Ramaswamy, "Classification of Silicon (Si) Wafer Material Defects in Semiconductor Choosers using a Deep Learning ShuffleNet-v2-CNN Model.," Advances in Materials Science & Engineering, vol. 2022, 2022.

[14] P. Bhatnagar, T. Arora and R. Chaujar, "Semiconductor Wafer Map Defect Classification Using Transfer Learning," in 2022 IEEE Delhi Section Conference (DELCON), 2022.

[15] MIRLab, MIR-WM811K: Dataset for wafer map failure pattern recognition, 2015.

[16] J. A. Mat Jizat, A. P. P. Abdul Majeed, A. F. Ab. Nasir, Z. Taha, E. Yuen and S. X. Lim, "Evaluation of the Transfer Learning Models in Wafer Defects Classification," in Recent Trends in Mechatronics Towards Industry 4.0: Selected Articles from iM3F 2020, Malaysia, 2022

[17] M.-J. Wu, J.-S. R. Jang and J.-L. Chen, "Wafer map failure pattern recognition and similarity ranking for large-scale data sets," IEEE Transactions on Semiconductor Manufacturing, vol. 28, p. 1–12, 2014.

[18] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal of artificial intelligence research, vol. 16, p. 321–357, 2002.

[19] J. Chen, H. Huang, A. G. Cohn, D. Zhang and M. Zhou, "Machine learning-based classification of rock discontinuity trace: SMOTE oversampling integrated with GBT ensemble learning," International Journal of Mining Science and Technology, vol. 32, p. 309–322, 2022.

[20] J. D. W. S. R. L. L. L. K. a. F.-F. L. Deng, "Imagenet: A large-scale hierarchical image database," in IEEE Conference on Computer Vision and Pattern Recognition, 2009, June.

[21] S. Sabour, N. Frosst and G. E. Hinton, "Dynamic routing between capsules Advances in Neural Information Processing Systems 30," in Proceedings of the Annual Conference on Neural Information Processing Systems

[22] S. K. C. a. P. D. Chen, "Detecting and measuring defects in wafer die using gan and yolov3," in Applied Sciences, 2020.

[23] J. D. W. S. R. L. L. L. K. a. F.-F. L. Deng, "Imagenet: A large-scale hierarchical image database," in IEEE Conference on Computer Vision and Pattern Recognition, 2009, June.

[24] P. P. Shinde, P. P. Pai and S. P. Adiga, "Wafer Defect Localization and Classification Using Deep Learning Techniques," IEEE Access, vol. 10, p. 39969–39974, 2022.