IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 25, 2021, accepted August 4, 2021, date of publication August 18, 2021, date of current version August 27, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3106171

A Systematic Review of Deep Learning for Silicon Wafer Defect Recognition

UZMA BATOOL[1,2], MOHD IBRAHIM SHAPIAI[1], (Member, IEEE), MUHAMMAD TAHIR[3], ZOOL HILMI ISMAIL[1], (Senior Member, IEEE), NOOR JANNAH ZAKARIA[1], AND AHMED ELFAKHARANY[1]

[1]Centre for Artificial Intelligence and Robotics iKohza, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur 54100, Malaysia
[2]Department of Computer Science, University of Wah, Wah 47040, Pakistan
[3]Department of Chemical Engineering, School of Chemical and Energy Engineering, Faculty of Engineering, Universiti Teknologi Malaysia (UTM), Skudai, Johor 81310, Malaysia

Corresponding author: Mohd Ibrahim Shapiai (md_ibrahim83@utm.my)

This work was supported in part by the Ministry of Higher Education of Malaysia through Malaysia Laboratories for Academia–Business Collaboration (MyLAB) under Grant JPT.S(BKPI)2000/016/018/07 Jld.26 (8), and in part by the Universiti Teknologi Malaysia under Grant Q.K130000.2543.21H17 for the project ''Imbalanced strategy for wafer defect classification using fully convolutional neural network.

ABSTRACT Advancements in technology have made deep learning a hot research area, and we see its applications in various fields. Its widespread use in silicon wafer defect recognition is replacing traditional machine learning and image processing methods of defect monitoring. This article presents a review of the deep learning methods employed for wafer map defect recognition. A systematic literature review (SLR) has been conducted to determine how the semiconductor industry is leveraged by deep learning research advancements for wafer defects recognition and analysis. Forty-four articles from well-known databases have been selected for this review. The articles' detailed study identified the prominent deep learning algorithms and network architectures for wafer map defect classification, clustering, feature extraction, and data synthesis. The identified learning algorithms are grouped as supervised learning, unsupervised learning, and hybrid learning. The network architectures include different forms of Convolutional Neural Network (CNN), Generative Adversarial Network (GAN), and Auto-encoder (AE). Various issues of multi-class and multi-label defects have been addressed, solving data unavailability, class imbalance, instance labeling, and unknown defects. For future directions, it is recommended to invest more efforts in the accuracy of the data generation procedures and the defect pattern recognition frameworks for defect monitoring in real industrial environments.

INDEX TERMS Wafer map defects, wafer bin map, defect recognition, deep learning, systematic literature review.

I. INTRODUCTION

Silicon chips are the backbone of the current digital era. The advancements in the emerging technologies of Internet of Things (IoT), Fifth Generation (5G) telecommunication networks, Artificial Intelligence (AI), and the automotive industry have propelled their consumption [1], [2]. Keeping up with the growing demand for semiconductor devices by embracing the efficient, most suitable manufacturing automation practices is more critical in present times. Like other manufacturing industries, semiconductor fabrication companies aim for maximum productivity by taking measures against yield limiting factors. Among several influencing parameters, wafer fabrication defects are significant [3]. Controlling the ratio of defective Integrated Circuits (ICs) determines an IC foundry or wafer fabrication facility (fab)'s productivity and indicates its control on the manufacturing processes [4]. The defects caused by the frontend operations in circuit fabrication reflect the manufacturing equipment and processes' flaws. They are characterized as random and systematic defects based on their originating factors [5]. Most wafers carry a mix of both types. Fig. 1 shows examples

The associate editor coordinating the review of this manuscript and approving it for publication was Zahid Akhtar.

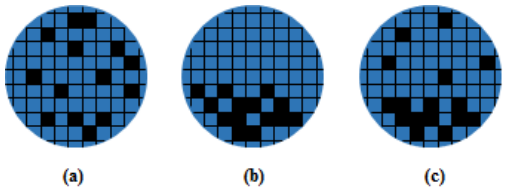

FIGURE 1. Defect types: (a) Random, (b) Systematic, and (c) Mix of both.

of wafer images having random and systematic defects and a combination of both. Random defects are attributed to the manufacturing environment and equipment faults. They disperse randomly on the wafer map, not appearing in a specific spatial pattern. The elimination of random defects requires equipment replacement and intense maintenance of the manufacturing environment, which is very expensive. Moreover, even in a near-sterile, perfect setting, the absence of dust particles cannot be guaranteed, and so the presence of random defects. Although they occur in acceptably small-scale, controlling them improves the fabrication yield [6]. Systematic defects are caused by the manufacturing processes and can be prevented by controlling the causing process. They appear in specific patterns forming a spatial cluster correlated to their causing factors. Thus, the identification and analysis of systematic defects facilitate process engineering and improve product quality and die yield by minimizing the rate of defective dies [7]. Defect recognition only targets systematic defects. Random defects are considered as noise since their presence adds to the difficulties of the systematic defect pattern analysis.

IC manufacturing process is divided into four basic steps: fabrication, probing, assembly, and final test [8]. As shown in Fig. 2, wafer defects are specified in the wafer probe or die soring phase, performed immediately after the fabrication. The functionality of integrated circuits is tested based on their electrical parameters, and a wafer map is generated, marking the exact locations of the failed circuits. Wafer maps are converted into wafer bin maps (WBMs), indicating the die having a circuit failure as a non-functional or defective die. Each die gets a bin color based on its category. The wafer bin map is, thus, an important artifact in the die manufacturing process, portraying all the normal and faulty dies fabricated on the wafer. It makes the basis for wafer defect analysis, pattern monitoring, root cause assessment, and process tracking.

Appropriate defect monitoring practices significantly affect the defect analysis and the overall yield at a fab. The conventional ways of defect monitoring rely on domain experts for manual inspections [9]. However, these traditional approaches are inadequate in their accuracy and efficiency. The cost of labor is also of concern. Another critical factor is the rapidly shrinking size of chips, which makes human intervention and manual defect monitoring out of the question. The semiconductor industry has constantly leveraged the latest available technological tools to reduce production costs and improve accuracy and efficiency. Hence, various computer-aided methods and automated visual inspections (AVI) have been widely adopted for defect monitoring, process improvement, product quality, and yield enhancement [10]. Image processing and machine learning methods are prominent in defect patterns analysis [11]–[13]. These methods provide cost-effective defect identification, improving accuracy and speed. However, the standard image processing and machine learning methods have their limitations. They are incapable of dealing with large-scale, low-quality and noisy data. Their reliance on feature engineering is the biggest restrain in their implementation, requiring domain expertise and experience in feature selection. Hence, data preprocessing under these methods extensively includes noise filtering, feature extraction and selection strategies. These intermediate steps are computational intensive. They cause distortion or loss of information, and thus reduce the accuracy of pattern recognition. The limitations of conventional methods have paved the way for deep learning in AVI. The breakthroughs of deep learning methods in defect monitoring is replacing the manual and traditional machine learning, making it a promising research area.

With the rapid development of technology in recent years, deep learning as a computer-aided algorithm has been actively realized in many fields [14]. Although it has been serving the semiconductor industry in many areas [15]–[17], deep learning's contribution in detecting and analyzing fabrication flaws is tremendous. Fabrication process defects appear in the form of specific patterns on the silicon substrate. These patterns are captured and reflected as wafer maps, Scanning Electron Microscope (SEM), and Energy Dispersive X-Ray Analyzer (EDX) images. As deep learning has proved to be a perfect tool for image processing and pattern recognition, it is best suited to identify defect patterns in such images. In addition, the automatic feature learning abilities of deep methods make them superior. Contrary to traditional machine learning's manual feature engineering, deep networks automatically learn meaningful features from raw data. Many nonlinear data processing units in the densely connected hidden layers of deep architectures enable feature extraction from raw input. Deep learning's performance for big datasets, noisy and incomplete data is surprisingly good [18], [19]. For these reasons, it has been successfully adopted for wafer map defect detection, identification, segmentation, and classification.

Although, many recent studies have demonstrated the effectiveness of deep learning for wafer defect pattern recognition, there is a lack of a comprehensive review of the field. The current study fills this gap by systematically reviewing the use of deep learning in silicon wafer defect recognition. This systematic literature review (SLR) investigated the deep learning algorithms and architectures employed to solve the specific issues of wafer defect pattern recognition. It covers the recent four years of literature published on wafer map defects and deep learning, providing an overview of what has been done in the field. The focus of the study has been on answering the specific questions about the learning

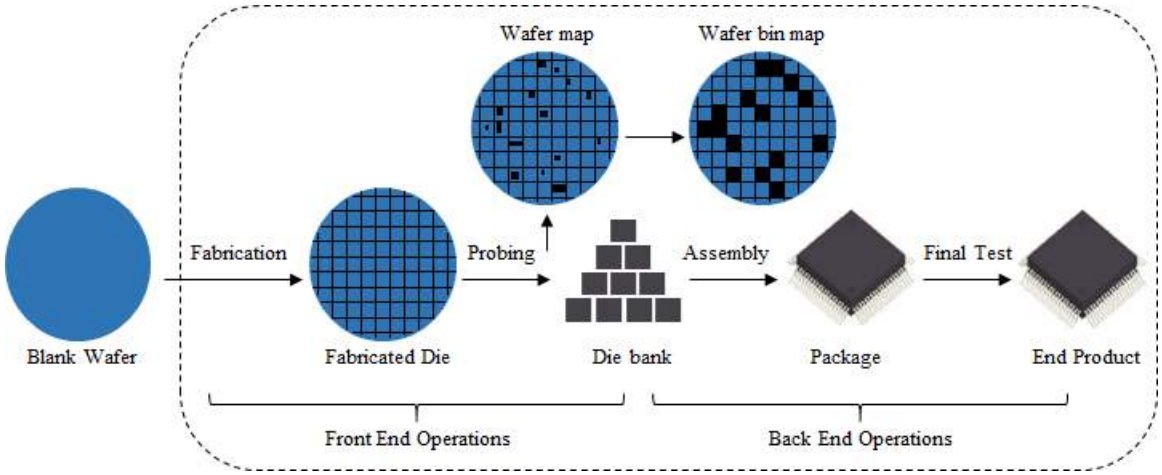

FIGURE 2. Four main stages of IC manufacturing; fabrication, probing, assembly, and final test.

algorithms, network architectures, and the datasets in the wafer map defect recognizing systems. The results drawn in this study highlight the challenges in implementing deep learning systems and devise the future research directions. It provides a reference for researchers and professionals working with wafer map defects, guiding them in improving the current systems and designing new defect-recognition frameworks.

The rest of the paper has been organized as follows: Section II describes the SLR method; the research questions, search terms, data sources, selection, and collection procedures. Section III represents the selected literature and its statistical analysis. Section IV holds the discussion, categorizes the literature to answer each question, critically analyzes the findings, and highlights specific issues. Finally, section V concludes the paper, giving some future work recommendations.

II. SYSTEMATIC LITERATURE REVIEW

The three-step process of plan, conduct, and report has been observed in conducting this SLR [20], [21]. In the planning phase, research questions were defined, and the review protocol was established, specifying the publication sources, search terms, and selection criteria. In the second step, the literature was collected following the review protocol. The selected literature was analyzed, extracting and synthesizing the required data to answer the questions. Finally, the review results were documented, addressing the research questions and the objectives of the SLR.

A. RESEARCH QUESTIONS

The main objective of this review was to determine how deep learning has been applied for wafer map defects. Furthermore, to look into the applications and how the defect recognition frameworks have been implemented using the deep networks. Thus, providing the knowledge of the current practices to building upon that for further improvement in the area. Therefore, the following three research questions (RQs) have been framed:

1) What learning algorithms have been applied?
2) What kind of architecture of the deep network has been employed?
3) What was the nature of the data used for the network training and testing?

While scanning the literature, a focused approach has been followed. Each article has been reviewed to answer the above questions. The gathered data has been reported in a comprehensive way to have a complete picture.

B. REVIEW PROTOCOL

Literature search sources, search terms, selection and rejection procedures adopted for this SLR are specified as followed:

1) SEARCH SOURCES

Four popular scientific databases; Scopus, IEEE Xplore, Web of Science, and Springer Link, were selected to extract the data.

2) SEARCH TERMS

The investigated topic combines three main search terms: 'deep learning', 'wafer map', and 'defects'. Each of the terms can be searched by multiple alternative words. The most relevant and commonly used applicable terms were selected and combined by the 'OR' operator. For example, to represent 'deep learning', three search terms were identified as 'deep learning', 'deep network', and 'deep architecture'. The other term for 'wafer map' was 'wafer bin map' and 'defect' was represented by its only main term. Individual search strings were concatenated by the 'AND' operator to form a search query. The wild card '$*$' has been added to include all verb forms of the key terms. Full text search has been employed to capture the maximum relevant literature. Complete search queries for each of the databases are shown in Fig. 3.

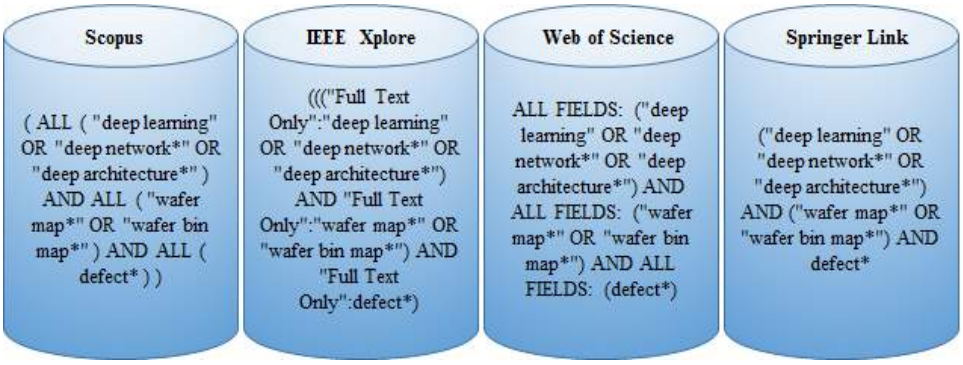

FIGURE 3. Search queries for each of the databases.

3) INCLUSION

This study is limited to the applications of deep learning for wafer bin map defects. All primary studies published in the English language, employing a deep learning algorithm for wafer map defect classification, identification, detection, segmentation, clustering, or any other task related to defect recognition, were included. For a broad search spectrum, no limits on the subject areas and time frame were imposed. However, since deep learning is an emerging field, the literature returned in response to the search queries, spanned over recent years; starting from 2017 onwards. The time period of the selected articles extends over four years 2017-2020. The chosen literature included journal articles, conference proceedings, and book sections on the explored topic.

4) EXCLUSION

This SLR includes studies on defects represented by wafer bin maps. The publications on other forms of images like SEM (scanning electron microscope) or EDX (energy dispersive X-ray) images, or any other format were not included. In addition, articles in languages other than English were excluded.

C. LITERATURE COLLECTION

The literature search was performed by supplying the search strings for each database, as shown in Fig. 3. Total 232 publications were returned as a response to these search queries. The search results from each database were assessed according to the predefined inclusion/exclusion criteria. In the initial screening, the review articles and non-English publications were excluded. Each article was evaluated based on its title, abstract and a quick review of text to decide its selection or rejection. This filtration reduced the number of articles to 104. After removing the duplicate articles, 57 publications were included in the full text assessment, and finally, 44 studies were selected to be the part of this SLR. The process of data selection has been shown by Preferred Reporting Items for Systematic Review and Meta-analysis (PRISMA) framework [22] in Fig. 4.

III. RESULTS

The selected publications are listed in Table 1 with the article title, publication year, source title, and the number of citations for each publication. Fig. 5 shows the distribution of the publications for 2017-2020. In the yearly distribution shown in Fig. 5(a), a sharp increase in the literature is noticeable, with only two publications in 2017 to 25 articles in 2020. Furthermore, as shown in Fig. 5(b), out of 44 articles, 26 were published in journals and 18 in conference proceedings. Fig. 5(c) shows the number of journal and conference publications each year. The distribution of articles in journals can be seen in Fig. 6. IEEE Transactions on Semiconductor Manufacturing is on the top of the list with seventeen publications. Journal of Intelligent Manufacturing has three articles, Applied Sciences contained two, and the other four journals have one publication each. Table 2 shows the names of the 18 conferences for conference articles.

IV. DISCUSSION

To answer the RQs, a detailed study of each publication has been conducted, extracting the required data. Each article was analyzed for the problem being solved, the main method, learning algorithm, network structure, data, and how data were prepared for the network training and testing. The findings for each RQ are explained in their respective sections as follows:

A. WHAT LEARNING ALGORITHMS HAVE BEEN APPLIED?

The selection of learning algorithm for a deep network depends on the nature of the problem and the data. Two main algorithms; supervised learning for labeled data and unsupervised learning for non-labeled data are the standard practices, respectively. Taking advantage of labeled and non-labeled data, a hybrid approach, combining supervised and unsupervised methods has also been a choice for more accurate analysis. This SLR has found supervised learning to be used maximum for wafer defect recognition. The obvious reason for this was having the labeled defect maps and aiming to categorize them into known classes. Unsupervised and hybrid learning have also been employed but comparatively

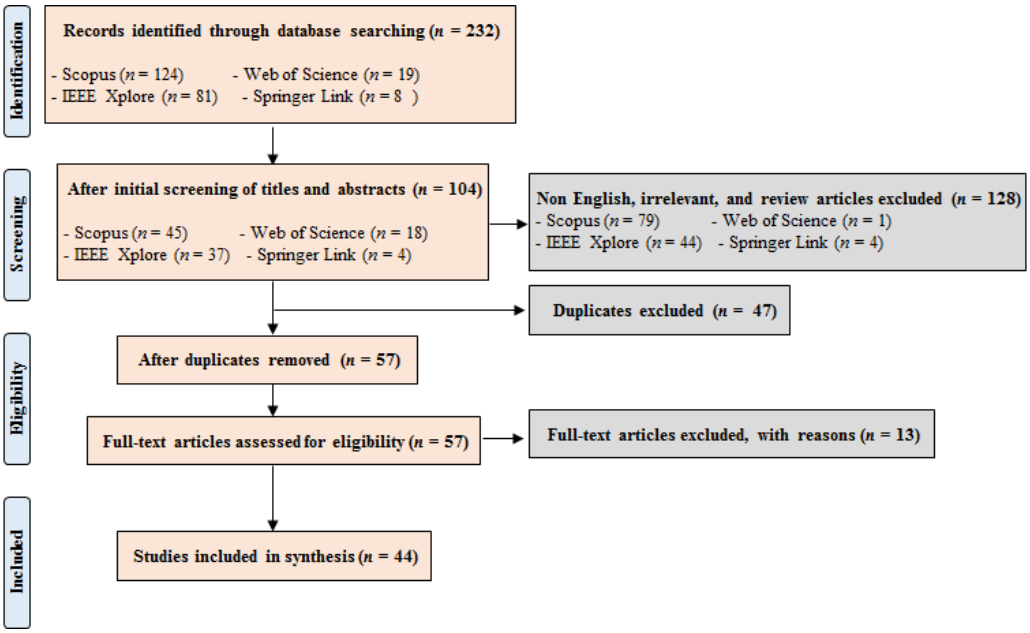

FIGURE 4. Preferred reporting items for systematic review and meta-analysis (PRISMA) diagram.

less frequently. Table 3 lists prominent references for each algorithm.

B. WHAT KIND OF ARCHITECTURE OF THE DEEP NETWORK HAS BEEN EMPLOYED?

There are three main deep networks employed for wafer map defect-recognition; Convolutional Neural Network (CNN), Generative Adversarial Network (GAN), and Auto-encoder (AE). We have categorized the literature based on them. The referring publications have been mentioned in the respective section of each. The publications which present a framework based on multiple architectures appear under all of them for complete characterization. The taxonomy of the literature followed in this SLR has been shown in Fig. 7. Following is the review of deep networks:

1) CONVOLUTIONAL NEURAL NETWORK

Since CNN is the most cited architecture, we have further sub-grouped the literature on CNN as custom-made CNN for single-label defect classification, CNN for multi-label defect classification, and pre-defined CNN and transfer learning. A detailed literature review of each topic is given below and a summary of publications in each category is provided in Tables 4, 5 and 6.

a: CUSTOM-MADE CNN FOR SINGLE-LABEL DEFECT CLASSIFICATION

Single label classification refers to the standard classification problem where a wafer map belongs to only one class. This section reviews custom build CNN for single defect classification, summarizing the important parameters in Table 4.

Based on the selected literature, the first mention of a CNN for wafer defect classification has been found as part of an auto disposition system. Lin *et al.* [23] proposed a system to reduce engineers' intervention in the trouble-shooting process. However, they provided only a brief description of their system, not giving the architecture design and details of the method. The second work in 2017 aimed at semiconductor yield enhancement by production inspection and monitoring. Nakata *et al.* [24] combined data mining and machine learning techniques in their framework for day-to-day tracking of manufacturing data. A CNN was used for long-term failure recurrence monitoring of defect patterns. It was a retrainable, one-class classifier with five layers whose parameters were determined empirically. They used to train it on the frequently occurring pattern, employing unlabeled images to monitor if that defect pattern emerges again. This limitation of the model for only one defect class was the shortcoming of the system. Also, retraining causes loss of resources, especially in the case of alternatively reoccurring defects.

From 2018 onwards, a continuous surge in literature making use of CNN has been observed. Most researchers presented standalone models, independently performing the whole task of classification [25]. We also see CNN combined with other machine learning algorithms in some architectures to accomplish the job [43].

Nakazawa and Kulkarni [25] proposed a standalone CNN model for wafer map defect classification and retrieval of similar maps. It aimed to perform well on the synthetic and real data and faster retrieval of images from the big library of wafer maps. They demonstrated the use of synthetic data in case of imbalanced or limited real data. Synthetic data were generated with more variations and different from the real

TABLE 1. Title, publication year, source, and the number of citations of the selected publications.

No.	Ref.	Publication Title	Year	Source Title	Cited By
1	[23]	Wafer pattern classification and auto disposition by machine learning	2017	2017 Joint International Symposium on e-Manufacturing and Design Collaboration (eMDC) & Semiconductor Manufacturing (ISSM)	-
2	[24]	A Comprehensive Big-Data-Based Monitoring System for Yield Enhancement in Semiconductor Manufacturing	2017	IEEE Transactions on Semiconductor Manufacturing	29
3	[25]	Wafer map defect pattern classification and image retrieval using convolutional neural network	2018	IEEE Transactions on Semiconductor Manufacturing	104
4	[26]	Deep-structured machine learning model for the recognition of mixed-defect patterns in semiconductor fabrication Processes	2018	IEEE Transactions on Semiconductor Manufacturing	40
5	[27]	Classification of Mixed-Type Defect Patterns in Wafer Bin Maps Using Convolutional Neural Networks	2018	IEEE Transactions on Semiconductor Manufacturing	51
6	[28]	Unsupervised Wafermap Patterns Clustering via Variational Autoencoders	2018	Proceedings of the International Joint Conference on Neural Networks	6
7	[29]	Anomaly detection and segmentation for wafer defect patterns using deep Convolutional Encoder-Decoder Neural Network Architectures in Semiconductor Manufacturing	2019	IEEE Transactions on Semiconductor Manufacturing	19
8	[30]	Bin2Vec: A better wafer bin map coloring scheme for comprehensible visualization and effective bad wafer classification	2019	Applied Sciences (Switzerland)	3
9	[31]	Deep Learning-Based Wafer-Map Failure Pattern Recognition Framework	2019	Proceedings - International Symposium on Quality Electronic Design, ISQED	7
10	[32]	Recognition and Location of Mixed-type Patterns in Wafer Bin Maps	2019	2019 IEEE International Conference on Smart Manufacturing, Industrial & Logistics Engineering (SMILE)	2
11	[33]	Stacked convolutional sparse denoising auto-encoder for identification of defect patterns in semiconductor wafer map	2019	Computers in Industry	15
12	[34]	Convolutional Neural Network for Semiconductor Wafer Defect Detection	2019	2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019	2
13	[35]	AdaBalGAN: An Improved Generative Adversarial Network with Imbalanced Learning for Wafer Defective Pattern Recognition	2019	IEEE Transactions on Semiconductor Manufacturing	20
14	[36]	Classification of wafer maps defect based on deep learning methods with small amount of data	2019	2019 International Conference on Engineering and Telecommunication, EnT 2019	2
15	[37]	Wafer defect pattern recognition and analysis based on convolutional neural network	2019	IEEE Transactions on Semiconductor Manufacturing	4
16	[38]	Wafer defect map classification using sparse convolutional networks	2019	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	2
17	[39]	Enhanced stacked denoising autoencoder-based feature learning for recognition of wafer map defects	2019	IEEE Transactions on Semiconductor Manufacturing	5
18	[40]	Silicon Wafer Map Defect Classification Using Deep Convolutional Neural Network with Data Augmentation	2019	2019 IEEE 5th International Conference on Computer and Communications, ICCC 2019	-
19	[41]	Wafer Map Defect Recognition Based on Deep Transfer Learning	2019	IEEE International Conference on Industrial Engineering and Engineering Management	-
20	[42]	A Semi-Supervised and Incremental Modeling Framework for Wafer Map Classification	2020	IEEE Transactions on Semiconductor Manufacturing	4
21	[43]	Wafer map defect pattern classification based on convolutional neural network features and error-correcting output codes	2020	Journal of Intelligent Manufacturing	3
22	[44]	Active Learning of Convolutional Neural Network for Cost-Effective Wafer Map Pattern Classification	2020	IEEE Transactions on Semiconductor Manufacturing	3
23	[45]	Mixed pattern recognition methodology on wafer maps with pre-trained convolutional neural networks	2020	ICAART 2020 - Proceedings of the 12th International Conference on Agents and Artificial Intelligence	-
24	[46]	Convolutional Neural Network for Imbalanced Data Classification of Silicon Wafer Defects	2020	Proceedings - 2020 16th IEEE International Colloquium on Signal Processing and its Applications, CSPA 2020	1
25	[47]	Discriminative feature learning and cluster-based defect label reconstruction for reducing uncertainty in wafer bin map labels	2020	Journal of Intelligent Manufacturing	3
26	[48]	Automatic reclaimed wafer classification using deep learning neural networks	2020	Symmetry	1

TABLE 1. *(Continued.)* **Title, publication year, source, and the number of citations of the selected publications.**

27	[49]	A Deep Convolutional Neural Network for Wafer Defect Identification on an Imbalanced Dataset in Semiconductor Manufacturing Processes	2020	IEEE Transactions on Semiconductor Manufacturing	2
28	[50]	Variational Deep Clustering of Wafer Map Patterns	2020	IEEE Transactions on Semiconductor Manufacturing	2
29	[51]	A Neural-Network Approach to Better Diagnosis of Defect Pattern in Wafer Bin Map	2020	2020 China Semiconductor Technology International Conference (CSTIC)	-
30	[52]	Wafer Map Classifier using Deep Learning for Detecting Out-of-Distribution Failure Patterns	2020	Proceedings of the International Symposium on the Physical and Failure Analysis of Integrated Circuits, IPFA	-
31	[53]	Wafer map defect patterns classification using deep selective learning	2020	Proceedings - Design Automation Conference	2
32	[54]	Memory-Augmented Convolutional Neural Networks With Triplet Loss for Imbalanced Wafer Defect Pattern Classification	2020	IEEE Transactions on Semiconductor Manufacturing	1
33	[55]	A Light-Weight Neural Network for Wafer Map Classification Based on Data Augmentation	2020	IEEE Transactions on Semiconductor Manufacturing	-
34	[56]	Inspection and classification of semiconductor wafer surface defects using CNN deep learning networks	2020	Applied Sciences (Switzerland)	-
35	[57]	Two-Dimensional Principal Component Analysis-Based Convolutional Autoencoder for Wafer Map Defect Detection	2020	IEEE Transactions on Industrial Electronics	-
36	[58]	Using GAN to Improve CNN Performance of Wafer Map Defect Type Classification : Yield Enhancement	2020	2020 31st Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)	-
37	[59]	Deformable Convolutional Networks for Efficient Mixed-Type Wafer Defect Pattern Recognition	2020	IEEE Transactions on Semiconductor Manufacturing	1
38	[60]	Qualitative and Quantitative Analysis of Multi-Pattern Wafer Bin Maps	2020	IEEE Transactions on Semiconductor Manufacturing	-
39	[61]	Rotation-Invariant Wafer Map Pattern Classification With Convolutional Neural Networks	2020	IEEE Access	-
40	[62]	Oversampling based on data augmentation in convolutional neural network for silicon wafer defect classification	2020	Frontiers in Artificial Intelligence and Applications	-
41	[63]	Semi-Supervised Multi-Label Learning for Classification of Wafer Bin Maps With Mixed-Type Defect Patterns	2020	IEEE Transactions on Semiconductor Manufacturing	-
42	[64]	Ensemble convolutional neural networks with weighted majority for wafer bin map pattern classification	2020	Journal of Intelligent Manufacturing	-
43	[65]	A Defect Detection Model for Imbalanced Wafer Image Data Using CAE and Xception	2020	2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)	-
44	[66]	A Wafer Map Defect Pattern Classification Model Based on Deep Convolutional Neural Network	2020	2020 IEEE 15th International Conference on Solid-State & Integrated Circuit Technology (ICSICT)	-

types to cover the identification of the rare events. Although variations are good for a generalized model, synthetic images should be closer to the real images if data synthesis aims at addressing the imbalance. If similar colors were used to represent identical bin codes or defects, it will ease the process of pattern identification [30]. To prove their hypothesis, Kim *et al.* [30] presented a neural network-based bin coloring method and built a four-layered CNN to distinguish good and bad wafers. However, further classification of bad wafers into respective defect types was not done, which is required for complete classification and the defect root cause analysis. In both of these works, the data imbalance issue was solved by easy data manipulation. Instead of following conventional data sampling methods, Nakazawa and Kulkarni [25] produced an entirely new balanced dataset for CNN training. Kim *et al.* [30] followed a batch training for the original imbalanced data, sampling an equal number of good and bad wafers in each batch.

Considering defective and regular wafers as two major classes of patterned and non-patterned images, some studies employed binary classifiers to sort them out. Then, the patterned class was further classified into the respective types. Yu *et al.* [37] employed two CNNs in their framework; one for detection and the other for classification of defect patterns. The eight-layered detection model (binary classifier) sorted the defect images into patterned and non-patterned classes. The patterned type was further classified into one of the defect classes by the thirteen-layered classifier. Furthermore, they extracted the feature set from the classifier's fully connected layer and used it for the defect root cause analysis after reducing its dimensionality by PCA. They under-sampled the major class (non-pattern) to control imbalance and expanded the two min classes (Near-full and Donut) by adding rotated, scaled, and noisy images. Training images were randomly cropped and rotated. They employed L2 regularization to reduce overfitting due to non-uniform class distribution. However, for some of the classes, model performance was not so good.

Kong and Ni [42] adopted a small LeNet like CNN as binary classifier for patterned and non-patterned wafers. Defect patterns on wafers were referred to as gross failing areas (GFA). Their actual work lies under semi-supervised

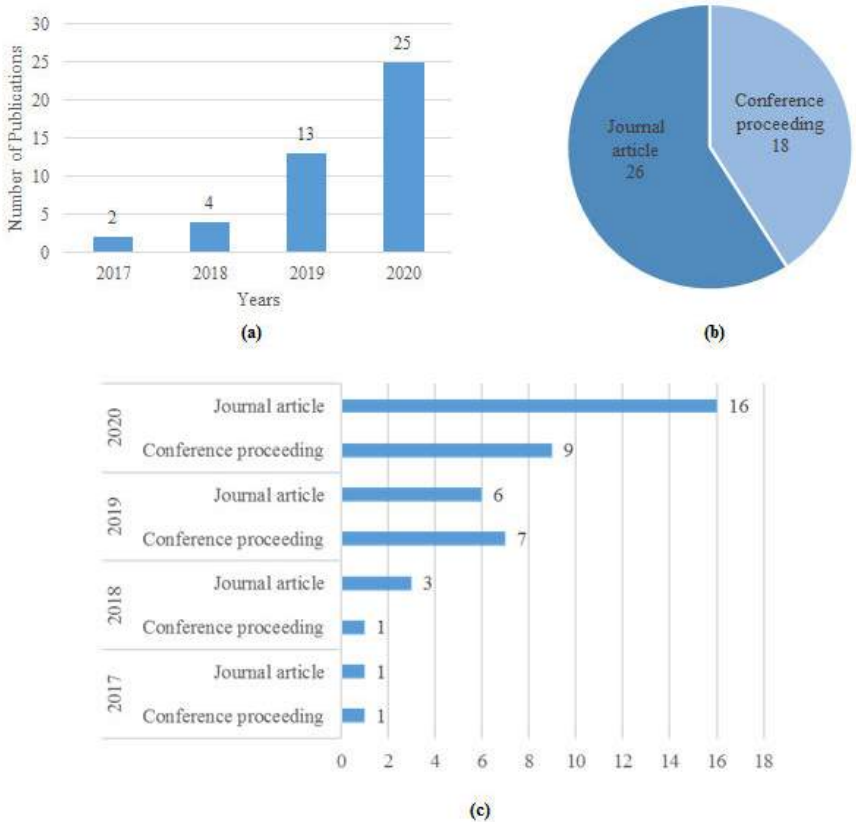

FIGURE 5. Publication distribution: (a) Publications in each year (b) Journal and conference publications (c) Journal and conference publications in each year.

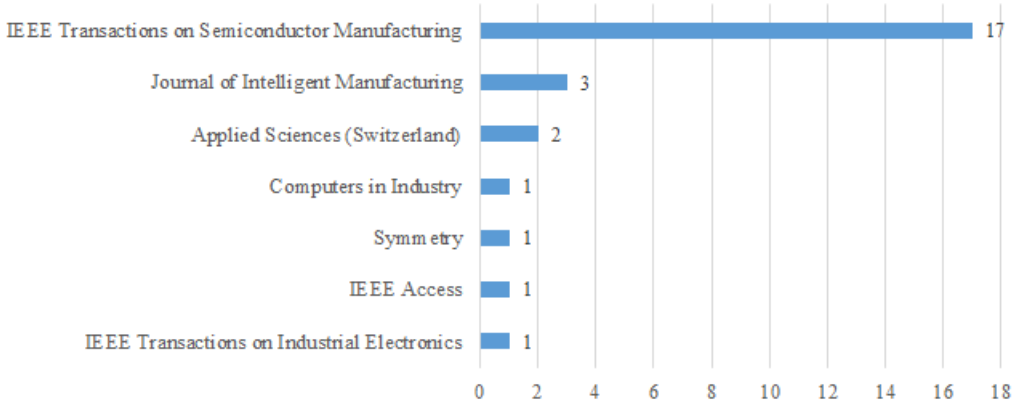

FIGURE 6. Number of articles in each journal, IEEE transactions on semiconductor manufacturing has highest number of publications.

and active learning, demonstrating the use of auto-encoders and ladder network. Jin *et al.* [43] used CNN as a features extractor and a combination of error-correcting output codes and support vector machine (SVM) as a classifier.

The complexity and nature of data are the crucial factors in determining the network architecture. di Bella *et al.* [38] adopted submanifold sparse convolutional network (SSCN) as a binary classifier for sparse data of larger images. Dataset was acquired from STMicroelectronics in Agrate Brianza, Italy, and contained 29,746 wafer maps in 12 classes. The images were extremely big, with a resolution of 20, 000 $\times$ 20.000 pixels. In another experiment with WM-811k, when the wafer map sizes were small, they adopted a standard CNN. In addition, oversampling was adopted to overcome the class imbalance. All classes were expanded by adding image transformations; rotations, horizontal flips, translation, noise injection and, random mixing of cropped images from less peculiar classes. Tsai and Lee [55] also presented separate

TABLE 2. Conference names for 18 conference articles.

No.	Conference Title
1	2017 Joint International Symposium on e-Manufacturing and Design Collaboration (eMDC) & Semiconductor Manufacturing (ISSM)
2	2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019
3	2019 IEEE 5th International Conference on Computer and Communications, ICCC 2019
4	2019 IEEE International Conference on Smart Manufacturing, Industrial & Logistics Engineering (SMILE)
5	2019 International Conference on Engineering and Telecommunication, EnT 2019
6	2020 31st Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)
7	2020 China Semiconductor Technology International Conference (CSTIC)
8	2020 IEEE 15th International Conference on Solid-State & Integrated Circuit Technology (ICSICT)
9	2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)
10	Frontiers in Artificial Intelligence and Applications
11	ICAART 2020 - Proceedings of the 12th International Conference on Agents and Artificial Intelligence
12	IEEE International Conference on Industrial Engineering and Engineering Management
13	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)
14	Proceedings - 2020 16th IEEE International Colloquium on Signal Processing and its Applications, CSPA 2020
15	Proceedings - Design Automation Conference
16	Proceedings - International Symposium on Quality Electronic Design, ISQED
17	Proceedings of the International Joint Conference on Neural Networks
18	Proceedings of the International Symposium on the Physical and Failure Analysis of Integrated Circuits, IPFA

TABLE 3. Learning algorithms cited in literature.

Learning Algorithm	References
Supervised learning	[23], [24], [25], [26], [27], [30], [31], [32], [34], [36], [37], [38], [40], [41], [43], [44], [45], [46], [48], [49], [52], [53], [54], [55], [56], [58], [59], [60], [61], [62], [64], [65], [66]
Unsupervised learning	[28], [29], [39], [47], [50]
Hybrid learning	[33], [35], [42], [51], [57], [63]

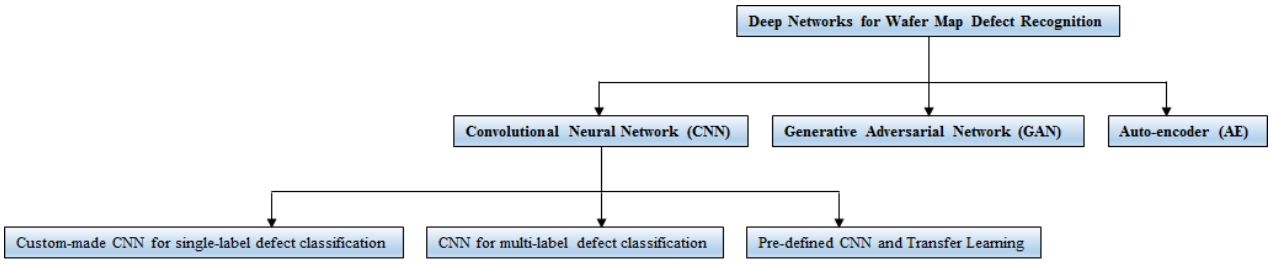

FIGURE 7. Taxonomy of literature made in this study.

models for two different real datasets because of the nature of the data. A lightweight CNN with fewer parameters for WM-811K was based on depthwise separable convolutions, a concept borrowed from MobileNet V1 and V2. The network for the 21-defect dataset was based on the Residual network. A convolutional auto-encoder (CAE) was used to oversample images in each class. The synthesized images further went through rotations to develop a more robust training set.

Many researchers employed data undersampling for handling imbalance. Shawon *et al.* [40] used a 3-layered CNN for the WM-811k benchmark and resolved the imbalance by data augmentation, adding 2000 images in each type. A CAE was used for data synthesis. However, they dropped the max class (non-pattern), including only eight patterned types in classifier training. Batool *et al.* [46] introduced an eight-layered CNN for wafer defect classification. They employed under-sampling by taking an equal number of images from each class and dropping the min class (Near-Full). Also, image transformations were added to bring variations in the training data. Kim *et al.* [52] performed wafer map classification by applying an out-of-distribution or out-of-domain image detection technique. The CNN architecture

was based on VGG16 for the dataset of real binary wafer maps of 13 classes. Class sizes were downsampled to the min class to have balanced data. Du and Shi [66] proposed a CNN with global average pooling, replacing flattened and fully connected layers to reduce the parameter set for faster training. They excluded the major class (non-pattern) in model training to reduce its impact on the data distribution. Although data under-sampling may ease the implementation or, in some cases, be needed for model comparison, excluding some of the classes from the available dataset means providing an incomplete solution. Major and minor classes are particularly significant in quantifying how much biased a model is. So, their inclusion in training is critical in building a fair model.

To construct a well-performing model and reduce the labeling cost of wafers, Shim *et al.* [44] introduced active learning in CNN training. The proposed defect pattern classification system provided a means for cost effective intelligent labeling and imbalance management. They employed a small LeNet-5 like structure as a classifier and trained it incrementally. Each model building iteration followed four steps: uncertainty estimation, wafer query selection, labeling, and model updation. Bayesian Active Learning by Disagreement (BALD) and Mean Standard Deviation (Mean-STD) were selected empirically for uncertainty estimation. The diversified top-K selection was employed as query wafer selection for imbalance alleviation.

Shih *et al.* [48] worked on defect classification of reclaimed wafers [67] to reduce manufacturing costs by recovering reclaimed wafers. They employed and compared three networks, multilayer perceptron (MLP), CNN, and Residual Network (ResNet), experimenting with different structures of the networks to find suitable architecture. The defect patterns were analyzed to determine whether the wafers could be re-polished or not, also specifying the exact defect type. The employed models were compared based on the error loss, accuracy, training time, and the number of parameters. CNN with properly designed kernels and structures ranked higher in this comparison because of its lower error rate, higher accuracy, fewer parameters, and lesser training and validation time. ResNet stood in the middle and MLP at the bottom.

Data oversampling is done mainly by data augmentation techniques. Also, CAE and GAN have been employed for data synthesis. Saqlain *et al.* [49] presented an eight-layered CNN, addressing class imbalance by oversampling through data augmentation. Alawieh *et al.* [53] adopted deep selective learning to construct a CNN with prediction and selection output. The CNN model was equipped with a reject option, abstaining predictions for samples with high misclassification risk. The model was built on WM-811K, synthesizing data by a CAE in underrepresented classes. The training set was further augmented by adding transformations of rotated images to solve the imbalance. Ji and Lee [58] demonstrated the performance of CNN on multiple datasets generated by GAN, augmentation of image transformations, and the combination of both. To increase the quantity and diversity of the scarce training data, Kang [61] demonstrated the effectiveness of rotation-based data augmentation in wafer map pattern classification, adopting a LeNet5 like CNN. Batool *et al.* [62] performed augmentation by rotating and flipping images for imbalance management and robust training.

Chien *et al.* [56] demonstrated and compared two ways of building a CNN for wafer defect classification: a custom-made CNN, carefully designed model for the specific dataset, and a pre-trained CNN fine-tuned by transfer learning. Both methods were equally good and performed better than the machine learning algorithms; SVM, logistic regression, random forest, and soft voting ensemble. A predefined model, faster-R-CNN pre-trained on COCO and KITTI datasets was used in the study.

This section has reviewed the literature on single defect wafer classification by custom-built CNN. We have seen that different configurations of the CNN have been employed; as a part of the more extensive framework or a single complete classifier. The model design was based on the nature of the data. CNN training was done on real and synthetic images. Various methods for data synthesis have been employed, including GAN and CAE. In addition, multiple ways of imbalanced data handling have been adopted. A summary of the literature included in this section has been presented in Table 4. A wafer may carry more than one defect, and a rigorous analysis requires all of them to be identified accurately. In the next section, a review of the multi defect wafers has been presented.

b: CNN FOR MULTI-LABEL DEFECT CLASSIFICATION

Identification of multiple defects on a wafer is critical for accurate defect classification and root cause analysis. Leaving some unidentified defects may lead to incomplete or inaccurate root cause analysis, affecting the fabrication process improvement and IC yield. This section reviews the literature on multi-defect wafer classification.

A multi-defect recognition system based on a randomized general regression network (RGRN) and CNN was presented by Tello *et al.* [26]. First, random defects were removed by a spatial filter in the data preprocessing step, and data were divided into single and multi-defect subgroups by a splitter (Information Gain theory). Then RGRN was used for single and CNN for mixed defects identification. Real and synthetic data with seven defect types (three basic and four mixed types) were employed in the experiments. In their configuration, splitter is the leading player with maximum accuracy of 95%. However, RGRN and CNN performance is relative as they depend on the splitter for their data input. Following the CNN-only approach, Kyeong and Kim [27] proposed a mixed defect classification system consisting of four CNN models for four basic defect types of the dataset. They demonstrated that a separate model for each type performs better than a single model for all types. The proposed model was able to identify 16 defect types as combinations of

TABLE 4. Literature on custom-made CNN for single-label defect classification.

No.	Ref.	Authors	Year	Learning Method	Model	Problem Solved	Data		No. of Classes	Augmentation			Imbalance Management		
							Synthetic	Real		Robust Training	Data Generation	Under-Sampling	Over-sampling	Weighted Cost	Others
1	[23]	Kung *et al.*	2017	Supervised	CNN	Auto disposition	-	-	-	-	-	-	-	-	-
2	[24]	Nakata *et al.*	2017	Supervised	CNN	Yield enhancement monitoring	-	✓	-	-	-	-	-	-	Single class classifier
3	[25]	Nakazawa and Kulkarni	2018	Supervised	CNN	Classification and Retrieval, rare event detection	✓	✓	22	x	x	x	x	x	Balanced training set
4	[30]	Kim *et al.*	2019	Supervised	NN, CNN	Bin Coloring, Classification	x	✓	-	-	-	x	x	x	Balanced batch
5	[37]	Yu, Xu, and Wang	2019	Supervised	CNN and PCA	Classification and root cause analysis	x	✓ WM-811k	9	✓	x	✓	✓	x	-
6	[42]	Kong and Ni	2020	Hybrid (Semi-supervised)	CNN Ladder network, autoencoder	Classification, Labeling	x	✓ (2 real with 22 classes)	22	-	-	-	-	-	✓
7	[43]	Jin *et al.*	2020	Supervised	CNN-ECOC-SVM	Classification	x	✓ WM-811k (Near-full excluded)	8	x	x	✓	✓	x	x
8	[38]	di Bella *et al.*	2019	Supervised	Submanifold Sparse Convolutional Network (SSCN), CNN	Classification	x	✓ 2 real, WM-811k	12, 9	x	✓	x	✓	x	-
9	[55]	Tsai and Lee	2020	Supervised	CNN, CAE	Classification	x	✓ WM-811k 21-defect dataset	9, 21	✓	✓	x	✓	x	-
10	[40]	Shawon *et al.*	2019	Supervised	CNN, CAE	Detection and Classification	x	✓ WM-811k	8 (Non excluded)	x	✓	✓	✓	x	-
11	[46]	Batool *et al.*	2020	Supervised	CNN	Classification	x	✓ WM-811k	9	✓	✓	✓	x	x	x
12	[52]	Kim, Cho and Lee	2020	Supervised	CNN	Classification	x	✓	13	x	x	✓	x	x	-
13	[66]	Du and Shi	2020	Supervised	CNN	Classification	x	✓ WM-811K	8 (Non excluded)	x	x	x	x	x	-
14	[44]	Shim *et al.*	2020	Supervised	CNN	Classification	x	✓ WM-811k	9	-	-	-	-	-	✓
15	[48]	Shih, Hsu, and Tien	2020	Supervised	CNN, Resnet 34, MLP	Classification	x	✓	10	x	✓	x	✓	x	-
16	[49]	Saqlain, Abbas, and Lee	2020	Supervised	CNN	Classification	x	✓ WM-811k	9	✓	✓	x	✓	x	-
17	[53]	Alawieh, Boning and Pan	2020	Supervised	CNN, CAE (data generation)	Classification	✓	WM-811k	9	x	✓	x	✓	x	-
18	[58]	Ji and Lee	2020	Supervised	CNN, GAN	Classification	✓	✓ WM-811k	10	x	✓	x	✓	x	-
19	[61]	Kang	2020	Supervised	CNN	Classification	x	✓ WM-811k	9	-	✓	-	-	-	-
20	[62]	Batool *et al.*	2020	Supervised	CNN	Classification	x	✓ WM-811k	9	✓	✓	x	✓	x	-
21	[56]	Chien, Wu, and Lee	2020	Supervised	CNN And Faster RCNN	Classification	x	✓	4	x	x	x	x	x	-

four basic types. Synthetic data were used for model training and testing. However, only six real wafers were employed for testing, a significantly small set for a thorough model evaluation. The system was not equipped with a feature to return the count of defects if the same defect appeared multiple times on a wafer. The ensemble of models is easier to adopt for new merging defects and addresses class imbalance, but it means putting more than the required resources for

the task. Taking this into account, Devika and George [34] optimally trained a single CNN on four basic types of single defect patterns that could detect combinations of the basic types. For each defect type of the real dataset, more images were generated by drawing patterns in paint that is a very naïve approach for such a sensitive task. Also, they did not provide a comparison of their model against any benchmark.

To classify multi-defect wafers with overlapping and non-overlapping patterns, Kong and Ni [32] presented a multi-step system. First, they employed a binary CNN to classify wafers with overlapping and non-overlapping patterns. The wafer maps with a single pattern or non-overlapping mixed-type pattern were segmented into single pattern maps through a seed filling algorithm and then classified by a CNN. Then, overlapped patterns were classified by a primitive template matching technique. Their subsequent work [60] investigated more potential ways by employing UNet and CNN in the classification system to locate and segment the patterned groups. Overlapping patterns were unwrapped and segmented into single patterns and then classified. A real dataset with seven classes of single patterns was used for model evaluation. The superimposition of single patterns generated new multi-pattern types. The combination of UNet (for defect boundary segmentation) with CNN (for classification) improved the overall system performance.

Some studies introduced variations in the standard CNN architecture. Byun and Baek [45] used a CAE to initialize CNN weights. The model was trained on single-type defect map data and tested on the combination of single and mixed type patterns. In testing, defect categories were distinguished based on their probability and a threshold value. Eight classes from the WM-811k were used as the primary single defect types, excluding the non-pattern class. Five new classes of mixed defects were produced by mixing center, scratch, edge-loc, and edge-ring patterns. The model performed well for single defects but not for the mixed types. Hyun and Kim [54] proposed a memory-augmented CNN with triplet loss for highly imbalanced WBM data comprising mixed-type defect patterns. Rather than a classifier, CNN trained by triple loss acted as an embedding function to map high dimensional WBMs to low dimensional. For class imbalance management, a key-value memory module fixed the same amount of memory for each class. The model was evaluated on the synthetic dataset of 16 classes; one non-pattern, four single and eleven mixed types with different levels of imbalance, and train/test configuration. Wang *et al.* [59] proposed a deformable convolutional network for mixed-type defect patterns by selectively sampling and extracting high-quality features from mixed wafer defects. They also introduced a public domain dataset ''MixedWM38'', having 38 types of wafer maps. Zhuang *et al.* [51] employed a deep belief network (DBN), investigating hybrid learning for the task. They trained an ensemble of six DBNs for six classes of defects in the real wafer map data and tested it on the wafers having single and mixed type defects. Lee and Kim [63] presented a semi-supervised deep convolutional generative model using labeled and non-labeled data.

This section provided a review of the literature on multi-defect wafer map classification. Several articles have addressed the multi-defects problem, employing CNN in some way. Standalone CNN, an ensemble of CNN, CNN with variations and combined with other algorithms. Real and synthetic, both sorts of data were employed for model training and validation. Since multi-defects are less frequent than single defects in the real data, multi-defect wafer maps were generated for system evaluation in almost every work. Many researchers generated balanced training sets, and less attention has been paid to address the imbalanced nature of the data. Furthermore, the scope of the studies was limited to find a maximum of two defects on a wafer, which is not a realistic approach. A model should be more generalized in identifying all defect patterns on a wafer and should report if multiple defects of the same type are present on a wafer map. Table 5 shows a summary of the findings in this section.

c: PRE-DEFINED CNN AND TRANSFER LEARNING

Designing a well-suited CNN and training it for a specific task requires a certain level of expertise. Network training is resource-intensive; demanding time, computational resources, and a large training dataset. Pre-defined standard networks are a solution to the problem. They offer ease of use and a confidence level in the network architecture. Furthermore, adopting pre-trained networks through transfer learning is a less resource-intensive way of having a trained model than training a new model from scratch. In this section, we review the literature on pre-defined and pre-trained networks for wafer map defect recognition.

Ishida *et al.* [31] compared VGG, AlexNet, and GoogLeNet, suggesting data augmentation with noise reduction to achieve higher recognition rates. Among the three models, VGG performed best for the underlying data. They applied Hough transformation for denoising random defect patterns on greyscale wafer images during the data preparation phase. Then, the clearer images went through rotation and flipping, producing variations of the quality images and thus augmenting the data with better object instances. The training benchmark dataset WM-811K contained nine defect classes, but the deep network was trained as a binary classifier for a single target pattern only. All other images were considered belonging to the non-target class. A cost function was also employed for imbalance addressing, assigning a weighted cost to the target class samples in the backpropagation phase of the learning algorithm. The weighted cost was proportional to the class imbalance ratio, i.e., the ratio of the target and non-target class sizes. In all the cited literature, this is the only work addressing imbalance by a cost-sensitive network; an idea that is worth investigating more. Hsu and Chien [64] presented an end-to-end ensemble model of LeNet, AlexNet, and GoogLeNet for WM-811k. A weighted majority function was employed for prediction output, giving more weightage to high performing base model.

TABLE 5. Literature on multi-label defect classification.

No	Ref.	Authors	Year	Learning Method	Model/Method	Problem	Real (Classes)	Synthetic (Classes)	Imbalance Control
1	[26]	Tello *et al.*	2018	Supervised	CNN, RGRN, IG	Classification	✓ (7)	✓	x
2	[27]	Kyeong and Kim	2018	Supervised	CNN ensemble	Classification	✓ (4)	✓ (16)	✓
3	[34]	Devika and George	2019	Supervised	CNN	Classification	✓ (4)	✓	-
4	[32]	Kong and Ni	2019	Supervised	CNN, seed filling, pattern matching	Classification	✓ 2 datasets 31x31(12), 38x38 (9)	✓ 31x31(5), 38x38 (5)	-
5	[60]	Kong and Ni	2020	Supervised	CNN, Unet (Boundary detection)	Classification	✓ (7)	✓ (8)	-
6	[45]	Byun and Baek	2020	Supervised	CNN, CAE (weight initialization)	Classification	✓ (8) WM-811k	✓ (5)	x
7	[54]	Hyun and Kim	2020	Supervised	CNN (mapping function)	Classification	✓ (4)	✓ (16)	yes
8	[59]	Wang *et al.*	2020	Supervised	Deformable CNN	Classification	✓ (38)	✓ (38)	yes
9	[51]	Zhuang *et al.*	2020	Hybrid	DBN	Classification	✓ (6)	x	-
10	[63]	Lee and Kim	2020	Hybrid	CNN	Classification	✓ (4)	✓ (16)	x

Maksim *et al.* [36] selected VGG-19, ResNet-34, ResNet-50, and MobileNetV2 for comparison and demonstrated their capabilities for wafer defect classification. To overcome the scarcity of real data, they used synthetic data along with the original. The generated images were based on six defect classes of WM-811k. Models were first pre-trained and validated on the synthetic images and then trained again on composite data of real and synthetic images, but only real images were used for testing. Trained in this manner, RestNet-50 stood out to be the best among all the comparing networks. Shih *et al.* [48] employed ResNet-34 to experiment with reclaimed wafers defect classification. Cha *et al.* [65] used Xception model for wafer defect classification and a CAE to generate more images, solving the imbalanced issue of WM-811k.

There are only a few studies on transfer learning. Shen and Yu [41] employed DenseNet169 with and without transfer learning and found transfer learning a better approach than the only pre-defined model. The DenseNet169 was originally trained on the Imagenet dataset. Their work demonstrated that transfer learning is a faster and more accurate way of wafer map defect classification. Chien *et al.* [56] compared faster-R-CNN pre-trained on COCO and KITTI datasets and found the KITTI pre-trained model better than the other. A balanced data from the four classes of WM-811K was selected for model retraining.

Park *et al.* [47] presented an unsupervised clustering method for wafer defects, employing a Siamese network for feature learning. The main objective of their work was to reduce the uncertainty in the manual labeling of defect classes. Their class label reconstruction method was based on discriminative feature learning of the Siamese network, repeated cross-learning of the class label reconstruction, and Gaussian means (G-means) clustering. The method was verified on WM-811k and discovered new defect types in the data which were not known previously when only the manual labeling by engineers was in effect.

This section has reviewed the work employing pre-trained CNN and transfer learning for wafer map defects. The prominent networks cited in the literature are ResNet, AlexNet, GoogLeNet, and LeNet. It has been noticed that transfer learning from standard datasets in other domains to wafer defects data provides better results. Table 6 shows a summary of the findings in this section.

2) GENERATIVE ADVERSARIAL NETWORK (GAN)

GAN is a hybrid deep architecture, consisting on 'generator' and 'discriminator' components. It is well-known for data generation features, but it has not been used much for silicon wafer defects. Only two studies investigated GAN for wafer defect classification and data generation. Wang *et al.* [35] presented an adaptive balancing generative network (AdaBal-GAN) solving class imbalance, simultaneously performing data generation and classification. Ji and Lee [58] demonstrated the performance of CNN on multiple datasets generated by GAN, classical augmentation, and their combination. The model performed better on the data produced by GAN compared to the classical augmentation. Table 7. lists the publications on GAN.

3) AUTO_ENCODER (AE)

Several AE types have been employed for various tasks in wafer defect recognition; feature learning, defect segmentation, clustering, and classification. Tulala *et al.* [28] used variational autoencoders for feature extraction from the wafer map patterns. On the resultant feature set, k-means clustering was applied to divide wafer maps into disjoint clusters. Yu *et al.* [33] used stacked convolutional sparse denoising auto-encoders (SCSDAE) for effective feature learning from simulated and real data, employing sampling methods for

TABLE 6. Literature on pre-defined and pre-trained CNN.

No.	Ref.	Authors	Year	Learning Method	Model	Problem Solved	Data		No. of Classes	Augmentation		Imbalance Management			Others
							Synthetic	Real		Robust Training	Data Generation	Under-Sampling	Over-sampling	Weighted Cost	
1	[31]	Ishida *et al.*	2019	Supervised	VGG, AlexNet, and GoogLeNet	Classification	x	✓	9	x	✓	x	✓	✓	x
2	[64]	Hsu and Chien	2020	Supervised	Ensemble of LeNet, AlexNet, and GoogLeNet	Classification	x	✓ WM-811k	9	x	x	x	x	x	✓
3	[36]	Maksim *et al.*	2019	Supervised	ResNet-50	Classification	✓	✓	6	x	x	x	✓	x	x
4	[48]	Shih, Hsu and Tien	2020	Supervised	CNN, Resnet 34, MLP	Classification	x	✓	10	x	✓	x	✓	x	x
5	[65]	Cha *et al.*	2020	Supervised	Xception, CAE (data generation)	Classification	x	✓ WM-811k	9	x	✓	x	✓	x	x
6	[41]	Shen and Yu	2020	Supervised and Transfer	DenseNet169	Classification	x	✓ WM-811k	9	x	x	x	x	x	x
7	[56]	Chien, Wu, and Lee	2020	Supervised and Transfer	CNN, and Faster RCNN	Classification	x	✓	4	-	-	-	-	-	-
8	[47]	Park, Jang and Kim	2020	Unsupervised	Siamese network (discriminative features), G-means clustering	Uncertain class label reconstruction		✓ WM-811k	9	-	✓	x	✓	x	x

TABLE 7. Literature on generative adversarial network (GAN).

No.	Reference	Authors	Year	Learning Method	Model	Problem Solved	Data		No. of Classes	Augmentation		Imbalance Management			
							Synthetic	Real		Robust Training	Data Generation	Under-Sampling	Over-sampling	Weighted Cost	
1	[35]	Wang *et al.*	2019	Hybrid	GAN	Classification	x	✓ WM-811k	9	x	x	x	✓	x	
2	[58]	Ji and Lee	2020	Supervised	CNN, GAN	Classification	x	✓ WM-811k	10	x	✓	x	✓	x	

imbalance mitigation. Extensive experimentations demonstrated the model's effectiveness on various data compositions. Yu [39] combined stacked denoising auto-encoders and manifold regularization for robust feature learning capabilities. Hwang and Kim [50] implemented a variational autoencoder with Gaussian mixture distribution to extract more suitable features for clustering. A Dirichlet process was further applied for automated one-step clustering. Addressing class imbalance and classification, Yu and Liu [57] proposed a two-dimensional principal component analysis-based convolutional auto-encoder (PCACAE) for wafer map defect recognition. Unsupervised AE was combined with conditional two-dimensional principal component analysis (C2DPCA) for feature extraction. The framework was compared with other deep neural networks in terms of accuracy and per iteration training time. The performance was better, especially for the small classes.

Nakazawa and Kulkarni [29] presented SegNet, U-Net, and FCN based autoencoders for detection and segmentation of abnormal wafer map defect patterns. SegNet and U-net based networks showed similar training accuracy, while FCN based architecture performed lower than them.

AE has also been used in wafer defect labeling. Kong and Ni [42] introduced semisupervised and active learning in wafer defect classification to facilitate wafer map labeling. They studied two semisupervised models; the Ladder network and the semisupervised variational autoencoder (SVAE). Active learning and pseudo labeling were adopted for incremental modeling. Comparison with a standard CNN on two real datasets showed the superiority of SVAE over the ladder network and the standard CNN. However, there is a need to experiment with other semisupervised models and different CNN architectures in this area. A summary of the AE publications has been provided in Table 8.

4) CRITICAL SUMMARY OF THE RQ2 FINDINGS

A recap of RQ2 shows CNN, GAN, and AE are the principal components of any deep framework for wafer map defect recognition. CNN has been the most dominantly used network for classification. Network architectures were designed to address the needs of the specific datasets of wafer defect maps. Datasets vary in their types of defects, the number of defect classes, and class sizes. Also, no standard convention for defect names has been followed in the industry. Accessibility of data is also an issue. So, every study tried to come up with a solution for the available data. No general-purpose architecture is possible in this scenario. Usually, small CNNs with three to eight layers have been employed with standard convolutional and pooling layers. Some of the studies experimented with network regularization methods, batch normalization, and dropout. Pre-defined networks and

TABLE 8. Literature on auto-encoder (AE).

No.	Ref.	Authors	Year	Learning Method	Model	Problem Solved	Data		No. of Classes	Augmentation		Imbalance Management		
							Synthetic	Real		Robust Training	Data Generation	Under-Sampling	Over-sampling	Weighted Cost
1	[28]	Tulala *et al.*	2018	Unsupervised	Variational Autoencoders (VAE), K-means	Clustering, Feature extraction	x	✓	8	-	-	-	-	-
2	[33]	Yu, Zheng and Liu	2019	Hybrid	Stacked Convolutional Sparse Denoising Auto-encoders	Classification	✓	✓ WM-811k	9	X	X	✓	✓	X
3	[39]	Yu	2019	Unsupervised	Stacked denoising Auto-encoders and Manifold regularization	Classification	x	✓ WM-811k	8	x	x	x	x	✓
4	[50]	Hwang and Kim	2020	Unsupervised	Gaussian model, variational autoencoder, Dirichlet	Clustering	x	✓	-	-	-	-	-	-
5	[57]	Yu and Liu	2020	Hybrid (Semi-supervised)	Two dimensional Principal Component Analysis Convolutional Auto-Encoder (PCACAE)	Classification	x	✓ WM-811k	9	-	-	-	-	-
6	[29]	Nakazawa and Kulkarni	2019	Unsupervised	CAE	Detection and segmentation	✓	✓ (Only for testing)	-	-	-	-	-	-
7	[42]	Kong and Ni	2020	Hybrid (Semi-supervised)	SVAE, Ladder network	Classification, Labeling	x	✓ (2 real)	22	-	-	-	-	-

TABLE 9. Network architectures.

Model	Type	Task	References
CNN	Custom-made	Single defect classification	[23], [24], [25], [30], [37], [38], [40], [42], [44], [46], [48], [49], [52], [53], [55], [56], [58], [61], [62], [66]
		Multi-label defect classification	[26], [27], [32], [34], [45], [51], [54], [59], [60], [63]
		Feature Extraction	[43]
	Pre-defined	Single defect classification	[31], [36], [41], [47], [48], [56], [58], [64] [59], [60], [61], [62], [64], [65], [66]
GAN	Advanced Balancing	Single defect classification	[35]
	Standard GAN	Single defect classification	[58]
AE	Stacked Convolutional Sparse Denoising	Single defect classification	[33]
	Variational	Feature extraction	[28], [50]
	Semi-supervised Variational	Single defect classification	[42]
	Stacked denoising	Single defect classification	[39]
	Convolutional	Single defect classification	[57]

transfer learning have also been practiced. GAN has been used for data synthesis and classification. AE has been used for CNN weights initialization, features extraction, classification, clustering, and data synthesis. The combination of these architectures has also been investigated. However, there is still a need to experiment with network configurations, investigating the effects of the cost function, learning rate, and other network parameters. Also, there is a lack of defect monitoring systems tested in real environments and verified by domain experts. A summary of the architectures and references cited in answering RQ2 is presented in Table 9.

C. WHAT WAS THE NATURE OF THE DATA USED FOR THE NETWORK TRAINING AND TESTING?

A review of the sources and types of the wafer map defect data for the deep network training and evaluation has been given below:

1) REAL AND SYNTHETIC DATA

Two types of datasets have been employed for the deep model evaluation; real and synthetic. Since real defect data from fab is not accessible, most researchers build their models on the public domain data or artificially synthesized dataset. Public domain or open-source datasets act as a benchmark in model comparison. Synthetic data are frequent, especially for multi-defect applications where the real dataset lacks wafers with multiple defects.

2) DATA AUGMENTATION

Various ways of data synthesis have been followed for generating wafer maps. In some cases, defect maps were produced from scratch by statistical methods, following the probability distribution algorithms. In others, newly generated images were based on the actual wafer map images, created by following the data augmentation techniques. The primary purpose of data augmentation has been twofold; balancing class distribution in the real data and generating a more comprehensive training set, reducing overfitting, and producing a more generalized model.

3) IMBALANCE ADDRESSING STRATEGY

Wafer map data is imbalanced by nature; non-pattern class carries maximum instances, and defect classes have variable frequencies. Therefore, consistent use of data-level imbalance addressing methods has been observed in the preprocessing phase. The reason may be their ease of

TABLE 10. Summary of learning outcomes in RQs 1, 2, and 3.

	No.	Ref.	Learning Method	Model	Problem Solved	Data	No. of Classes	Augmentation	Imbalance Management
						Real / Synthetic		Robust Training / Data generation	Under-sampling / Oversampling / Weighted cost / Others
Custom-made CNN for Single-label defect classification	1	[23]	Supervised	CNN	Auto disposition	-	-	-	-
	2	[24]	Supervised	CNN	Yield enhancement monitoring	Real	-	-	✓
	3	[25]	Supervised	CNN	Classification and Retrieval, rare event detection	Real (Private) Synthetic (Private)	22	x	✓
	4	[30]	Supervised	NN, CNN	Bin Coloring, Classification	Real (private)	-	-	✓
	5	[37]	Supervised	CNN and PCA	Classification and root cause analysis	Real (WM-811k)	9	Robust Training	✓
	6	[42]	Hybrid (Semi-supervised)	CNN Ladder network, autoencoder	Classification, Labeling	2 Real (Private)	22	-	✓
	7	[43]	Supervised	CNN-ECOC-SVM	Classification	Real (WM-811k)	8 (Near-full excluded)	x	Under-sampling + Oversampling
	8	[38]	Supervised	Submanifold Sparse Convolutional Network (SSCN), CNN	Classification	(12, 9)	Data generation	Under-sampling + Oversampling	(12, 9)
	9	[55]	Supervised	CNN, CAE	Classification	Real (WM-811k) Real (Private-21-defect dataset)	9, 21	Robust training and Data generation	x
	10	[40]	Supervised	CNN, CAE	Detection and Classification	Real (WM-811k)	8 (Non-pattern excluded)	Data generation	Under-sampling + Oversampling
	11	[46]	Supervised	CNN	Classification	Real (WM-811k)	9	Robust training and Data generation	Under-sampling
	12	[52]	Supervised	CNN	Classification	Real (Private)	13	x	Under-sampling
	13	[66]	Supervised	CNN	Classification	Real (WM-811k)	8 (Non excluded)	x	x
	14	[44]	Supervised	CNN	Classification	Real (WM-811k)	9	-	✓
	15	[48]	Supervised	CNN, Resnet 34, MLP	Classification	Real (Private)	10	Data generation	Oversampling
	16	[49]	Supervised	CNN	Classification	Real (WM-811k)	9	Robust training and Data generation	Oversampling
	17	[53]	Supervised	CNN, CAE (data generation)	Classification	Real (WM-811k)	9	Data generation	Oversampling
	18	[58]	Supervised	CNN, GAN	Classification	Real (WM-811k) Synthetic (Private)	10	Data generation	Oversampling

TABLE 10. *(Continued.)* **Summary of learning outcomes in RQs 1, 2, and 3.**

	No.	Ref.	Learning Method	Model	Problem Solved	Data	No. of Classes	Augmentation	Imbalance Management
						Real / Synthetic		Robust Training / Data generation	Under-sampling / Oversampling / Weighted cost / Others
CNN for multi-label defect classification	19	[61]	Supervised	CNN	Classification	Real (WM-811k)	9	Data generation	-
	20	[62]	Supervised	CNN	Classification	Real (WM-811k)	9	Robust training and Data generation	Oversampling
	21	[56]	Supervised	CNN And Faster RCNN	Classification	Real (Private)	4	x	x
	1	[26]	Supervised	CNN, RGRN, IG	Classification	Real and Synthetic	7	-	-
	2	[27]	Supervised	CNN ensemble	Classification	Real and Synthetic	(4, 16)	-	✓
	3	[34]	Supervised	CNN	Classification	Real and Synthetic	4	-	-
	4	[32]	Supervised	CNN, seed filling, pattern matching	Classification	Real and Synthetic (2 datasets)	(12, 5), (9, 5)	-	-
	5	[60]	Supervised	CNN Unet (Boundary detection)	Classification	Real and synthetic	(7, 8)	-	-
	6	[45]	Supervised	CNN, CAE for weight initialization	Classification	Real and Synthetic (WM-811k)	(8, 5)	-	-
	7	[54]	Supervised	CNN (mapping function)	Classification	Real and Synthetic	(4, 16)	-	✓
	8	[59]	Supervised	Deformable CNN	Classification	Real and Synthetic	38	-	✓
	9	[51]	Hybrid	DBN	Classification	Real	6	-	-
	10	[63]	Hybrid (Semi-supervised)	CNN	Classification	Real and Synthetic	(4, 16)	-	-
Pre-defined CNN and transfer learning	1	[31]	Supervised	VGG, AlexNet, and GoogLeNet	Classification	Real (Private)	9	Data generation	Oversampling and weighted cost
	2	[64]	Supervised	Ensemble of LeNet, AlexNet, and GoogLeNet	Classification	Real (WM-811k)	9	x	x
	3	[36]	Supervised	ResNet-50	Classification	Real (Private) Synthetic (Private)	6	x	Oversampling
	4	[48]	Supervised	CNN, Resnet 34, MLP	Classification	Real (Private)	10	Data generation	Oversampling
	5	[65]	Supervised	Xception, CAE (data generation)	Classification	Real (WM-811k)	9	Data generation	Oversampling
	6	[41]	Supervised and Transfer	DenseNet169	Classification	Real (WM-811k)	9	x	x

TABLE 10. *(Continued.)* **Summary of learning outcomes in RQs 1, 2, and 3.**

	No.	Ref.	Learning Method	Model	Problem Solved	Data	No. of Classes	Augmentation	Imbalance Management
						Real / Synthetic		Robust Training / Data generation	Under-sampling / Oversampling / Weighted cost / Others
	7	[56]	Supervised and Transfer	CNN, and Faster RCNN	Classification	Real (Private)	4	-	-
	8	[47]	Unsupervised	Siamese network (discriminative features), G-means clustering	Uncertain class label reconstruction	Real (WM-811k)	9	Data generation	Oversampling
GAN	1	[35]	Hybrid	GAN	Classification	Real (WM-811k)	9	x	Oversampling
GAN	2	[58]	Supervised	CNN, GAN	Classification	Real (WM-811k)	10	Data generation	Oversampling
AE	1	[28]	Unsupervised	Variational Autoencoders (VAE), K-means	Clustering, Feature extraction	Real (Private)	8	-	-
AE	2	[33]	Hybrid	Stacked Convolutional Sparse Denoising Auto-encoders	Classification	Real (WM-811k) Synthetic (Private)	9	x	Undersampling and Oversampling
AE	3	[39]	Unsupervised	Stacked denoising Auto-encoders and Manifold regularization	Classification	Real (WM-811k)	8	x	Weighted cost
AE	4	[50]	Unsupervised	Gaussian model, variational autoencoder, Dirichlet	Clustering	Real (Private)	-	-	-
AE	5	[57]	Hybrid (Semi-supervised)	Two dimensional Principal Component Analysis Convolutional Auto-Encoder (PCACAE)	Classification	Real (WM-811k)	9	-	-
AE	6	[29]	Unsupervised	CAE	Detection and segmentation	Synthetic (Private) Real (Private-testing only)	-	-	-
AE	7	[42]	Hybrid (Semi-supervised)	SVAE, Ladder network	Classification, Labeling	2 Real (Private)	22	-	-

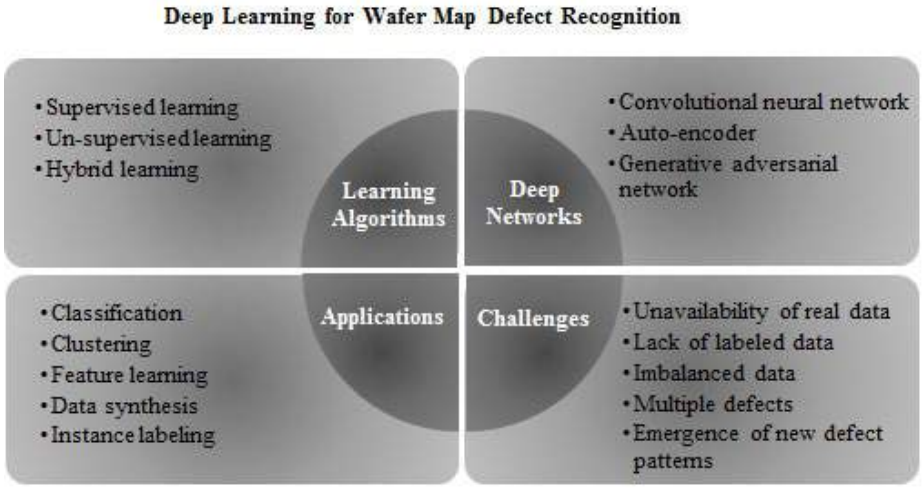

FIGURE 8. A summary of the algorithms, networks, applications and challenges of deep learning for wafer map defect recognition.

TABLE 11. Challenges and solutions.

Challenge	Solution	Reference
Unavailability of real data	Synthetically generated data	[25], [26], [27], [33], [34], [36], [45], [59]
Lack of Labeled data / limitations of manual labeling	Clustering, Semi supervised learning	[28], [42], [44], [54], [57]
Imbalance class distribution	imbalance addressing mechanism	[31], [33], [35], [36], [39], [43], [44], [46], [49], [55], [58],[62]
Multiple defect patterns on a single wafer	Multi-label classification	[26], [27], [34], [45], [59]
The emergence of new defect patterns	Anomaly detection and selective learning	[29], [53]

implementation as compared to the algorithmic-level imbalance handling strategies. Also, it is customary to generate more data by augmentation as deep networks like abundance of training data. So, it is a natural tendency to opt for oversampling for class equalization and robust training.

D. LEARNING OUTCOMES FROM THE RQs

A review of the RQs shows that deep learning has made a strong presence in silicon wafer defect recognition in a short span of only four years. A massive volume of research covering all aspects of defect pattern recognition with deep learning has been contributed. We see simple feature extractors, small binary class detectors and complex classifiers, a combination of various architectures, and learning algorithms. Table 10. summarizes the learning outcomes from the RQs 1, 2 and 3.

E. GENERAL DISCUSSION ON ADDRESSING THE SPECIFIC ISSUES

Table 11. lists the challenges tried to be solved by the studies presented in this review. On the top of the list is the unavailability of real data for research and development. Lack of labeled data because of the limitations of manual data labeling is also a hurdle. Imbalanced class distribution of real data is a major issue that needs to be solved while working with defect data. Identification of multiple defect patterns on a single wafer and the emergence of new defect patterns that were unknown at the model training time are also big challenges for the researchers.

V. CONCLUSION AND FUTURE RECOMMENDATIONS

We conclude this SLR by summarizing the findings and giving some future directions. A complete picture of what has been learnt so far is shown in Fig. 8. It shows the learning algorithms, network types, applications of deep learning for wafer map defects, and the major challenges in this field.

A. CONCLUSION

The publications included in this SLR have employed various deep learning algorithms and architectures, depending on the scope of the research and the availability of data. Every paper investigated some aspects of the wafer defect recognition with deep learning. Alongside the general issues of classification and clustering, more specific issues of class imbalance, instance labeling, data synthesis, and feature extraction have been addressed. We see a range of learning algorithms, including supervised learning, unsupervised learning, and hybrid learning. The types of deep networks and architectures comprise CNN, GAN, and various kinds of AEs. The use of pre-trained networks through transfer learning has also been explored widely. Some of the studies investigated the combination of multiple deep algorithms and other machine learning techniques.

The choice of a particular method has been subjected to the availability of the data and the research objectives. Some of the models have been used more than others. We observed that CNN is the most preferred deep network. However, the frequent use does not mean that it is the best choice for the task, as in some cases, other architectures like AE and GAN applied for the problem performed better than CNN.

The fusion of deep algorithms with other machine learning and image processing techniques was also promising.

We believe this SLR will pave the way for further research on developing more effective solutions for the wafer map defect-recognition problem. There is a need for more accurate frameworks able to be applied in actual industrial setups.

In our future work, we aim to build on the outcomes of this study, focusing on designing deep classifiers for wafer map defects taking into account the class imbalance of the data.

B. FUTURE DIRECTIONS AND RECOMMENDATIONS

Based on the findings of this SLR, following directions should be focused for further contribution to the field:

1) Providing real data for the research purpose. Accurately labeled defect data for deep network training is a must for exact feature learning. The industry should take a step forward, engaging more collaborations from academia.
2) Producing synthetic data similar to the real defect patterns. The artificial wafers should reflect more realistic trends of class sizes and defects morphology to develop more effective models to handle the real data.
3) More imbalance management methods should be examined, for example, algorithmic level methods, cost-sensitive networks, and new cost functions.
4) More attention should be given to evaluate network parameters; number of layers, activation and loss functions, kernel and stride size, and regularization methods.
5) Fusion of networks and transfer learning have shown better performance and need to be investigated more.
6) More accurate defect identification systems. The accuracy of the defect learning networks should be improved in working conditions to allow real-time identification.
7) Reduction in the computational burden. Efficient learning algorithms should be developed to reduce training time, memory, and processing resources.

REFERENCES

[1] T. Zanni, L. Clark, C. Gentle, S. Lohokare, and S. Jones, "Semiconductors: As the backbone of the connected world, the industry' s future is bright," in *Proc. 14th Annu. Global Semiconductor Outlook Rep. (KMPG)*, 2019, pp. 1–24.

[2] *Semiconductors—The Next Wave, Opportunities and Winning Strategies for Semiconductor Companies*, Deloitte, New York, NY, USA, 2019.

[3] L. Milor, "A survey of yield modeling and yield enhancement methods," *IEEE Trans. Semicond. Manuf.*, vol. 26, no. 2, pp. 196–213, May 2013.

[4] J.-S. Kim, S.-J. Jang, T.-W. Kim, H.-J. Lee, and J.-B. Lee, "A productivity-oriented wafer map optimization using yield model based on machine learning," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 1, pp. 39–47, Feb. 2019.

[5] T. Yuan, W. Kuo, and S. J. Bae, "Detection of spatial defect patterns generated in semiconductor fabrication processes," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 3, pp. 392–403, Aug. 2011.

[6] C.-W. Liu and C.-F. Chien, "An intelligent system for wafer bin map defect diagnosis: An empirical study for semiconductor manufacturing," *Eng. Appl. Artif. Intell.*, vol. 26, nos. 5–6, pp. 1479–1486, May 2013.

[7] Y. S. Jeong, S. J. Kim, and M. K. Jeong, "Automatic identification of defect patterns in semiconductor wafer maps using spatial correlogram and dynamic time warping," *IEEE Trans. Semicond. Manuf.*, vol. 21, no. 4, pp. 625–637, Nov. 2008.

[8] L. Monch, J. W. Fowler, and S. Mason, *Production Planning and Control for Semiconductor Wafer Fabrication Facilities: Modeling, Analysis, and Systems*, vol. 52. New York, NY, USA: Springer, 2012. [Online]. Available: https://www.springer.com/gp/book/9781461444718

[9] C.-Y. Hsu, W.-J. Chen, and J.-C. Chien, "Similarity matching of wafer bin maps for manufacturing intelligence to empower industry 3.5 for semiconductor manufacturing," *Comput. Ind. Eng.*, vol. 142, Apr. 2020, Art. no. 106358.

[10] S.-H. Huang and Y.-C. Pan, "Automated visual inspection in the semiconductor industry: A survey," *Comput. Ind.*, vol. 66, pp. 1–10, Jan. 2015.

[11] A. R. Mickelson, *Defect Recognition and Image Processing in Semiconductors 1995*. USA, 1996. [Online]. Available: https://www.osti.gov/biblio/405505-defect-recognition-image-processing-semiconductors

[12] J. Doneker and I. Rechenberg. (1998). *Defect Recognition and Image Processing in Semiconductors 1997: Proceedings of the Seventh International Conference on Defect Recognition and Image Processing in Semiconductors (DRIP VII)*. Accessed: May 9, 2021. [Online]. Available: https://books.google.com.pk/books?hl=en&lr=&id=ymeS1HIt1YYC&oi=fnd&pg=PR3&dq=Defect+recognition+and+image+processing+in+semiconductors+1995&ots=UoRqWbHZbh&sig=IqNm8SoA52-x1RSYm_jOh88z82Q#v=onepage&q=Defectrecognition and image processing in semiconductor

[13] F. Adly, P. D. Yoo, S. Muhaidat, and Y. Al-Hammadi, "Machine-learning-based identification of defect patterns in semiconductor wafer maps: An overview and proposal," in *Proc. IEEE Int. Parallel Distrib. Process. Symp. Workshops (IPDPS)*, May 2014, pp. 420–429.

[14] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, "A survey of deep learning and its applications: A new paradigm to machine learning," *Arch. Comput. Methods Eng.*, vol. 27, no. 4, pp. 1071–1092, Sep. 2020.

[15] P. Chung and S. Y. Sohn, "Early detection of valuable patents using a deep learning model: Case of semiconductor industry," *Technol. Forecasting Social Change*, vol. 158, Sep. 2020, Art. no. 120146.

[16] H. Kim, D.-E. Lim, and S. Lee, "Deep learning-based dynamic scheduling for semiconductor manufacturing with high uncertainty of automated material handling system capability," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 1, pp. 13–22, Feb. 2020.

[17] F. Beuth, T. Schlosser, M. Friedrich, and D. Kowerko, "Improving automated visual fault detection by combining a biologically plausible model of visual attention with deep learning," in *Proc. 46th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2020, pp. 5323–5330.

[18] S. Gupta and A. Gupta, "Dealing with noise problem in machine learning data-sets: A systematic review," *Procedia Comput. Sci.*, vol. 161, pp. 466–474, Jan. 2019.

[19] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *J. Big Data*, vol. 2, no. 1, p. 1, Dec. 2015.

[20] S. Keele, "Guidelines for performing systematic literature reviews in software engineering," Keele Univ., Durham Univ. Joint Rep., U.K., Tech. Rep. EBSE-2007-01, 2007. [Online]. Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.117.471&rep=rep1&type=pdf and https://www.bibsonomy.org/bibtex/aed0229656ada843d3e3f24e5e5c9eb9

[21] B. Kitchenham, R. Pretorius, D. Budgen, O. P. Brereton, M. Turner, M. Niazi, and S. Linkman, "Systematic literature reviews in software engineering—A tertiary study," *Inf. Softw. Technol.*, vol. 52, no. 8, pp. 792–805, 2010, doi: 10.1016/j.infsof.2010.03.006.

[22] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *PLoS Med.*, vol. 6, no. 7, Jul. 2009, Art. no. e1000097.

[23] J. Lin, J. F. Kung, P. Cheng, A. Hwu, C. T. Wang, and Y. B. Hsu, "Wafer pattern classification and auto disposition by machine learning," in *Proc. Joint Int. Symp. e-Manuf. Design Collaboration (eMDC) Semiconductor Manuf. (ISSM)*, 2017, pp. 3–5.

[24] K. Nakata, R. Orihara, Y. Mizuoka, and K. Takagi, "A comprehensive big-data-based monitoring system for yield enhancement in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 4, pp. 339–344, Nov. 2017.

[25] T. Nakazawa and D. V. Kulkarni, "Wafer map defect pattern classification and image retrieval using convolutional neural network," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 2, pp. 309–314, May 2018.

[26] 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 Trans. Semicond. Manuf.*, vol. 31, no. 2, pp. 315–322, May 2018.

[27] K. Kyeong and H. Kim, "Classification of mixed-type defect patterns in wafer bin maps using convolutional neural networks," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 3, pp. 395–402, Aug. 2018.
[28] P. Tulala, H. Mahyar, E. Ghalebi, and R. Grosu, "Unsupervised wafermap patterns clustering via variational autoencoders," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.
[29] T. Nakazawa and D. V. Kulkarni, "Anomaly detection and segmentation for wafer defect patterns using deep convolutional encoder–decoder neural network architectures in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 2, pp. 250–256, May 2019.
[30] J. Kim, H. Kim, J. Park, K. Mo, and P. Kang, "Bin2Vec: A better wafer bin map coloring scheme for comprehensible visualization and effective bad wafer classification," *Appl. Sci.*, vol. 9, no. 3, p. 597, Feb. 2019.
[31] T. Ishida, I. Nitta, D. Fukuda, and Y. Kanazawa, "Deep learning-based wafer-map failure pattern recognition framework," in *Proc. 20th Int. Symp. Qual. Electron. Design (ISQED)*, Mar. 2019, pp. 291–297.
[32] Y. Kong and D. Ni, "Recognition and location of mixed-type patterns in wafer bin maps," in *Proc. IEEE Int. Conf. Smart Manuf., Ind. Logistics Eng. (SMILE)*, Apr. 2019, pp. 4–8.
[33] J. Yu, X. Zheng, and J. Liu, "Stacked convolutional sparse denoising auto-encoder for identification of defect patterns in semiconductor wafer map," *Comput. Ind.*, vol. 109, pp. 121–133, Aug. 2019.
[34] B. Devika and N. George, "Convolutional neural network for semiconductor wafer defect detection," in *Proc. 10th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2019, pp. 1–6.
[35] J. Wang, Z. Yang, J. Zhang, Q. Zhang, and W.-T.-K. Chien, "AdaBalGAN: An improved generative adversarial network with imbalanced learning for wafer defective pattern recognition," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 3, pp. 310–319, Aug. 2019.
[36] K. Maksim, B. Kirill, Z. Eduard, G. Nikita, B. Aleksandr, L. Arina, S. Vladislav, M. Daniil, and K. Nikolay, "Classification of wafer maps defect based on deep learning methods with small amount of data," in *Proc. Int. Conf. Eng. Telecommun. (EnT)*, Nov. 2019, pp. 1–5.
[37] N. Yu, Q. Xu, and H. Wang, "Wafer defect pattern recognition and analysis based on convolutional neural network," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 4, pp. 566–573, Nov. 2019.
[38] R. di Bella, D. Carrera, B. Rossi, P. Fragneto, and G. Boracchi, "Wafer defect map classification using sparse convolutional networks," in *Proc. Int. Conf. Image Anal. Process.*, 2019, pp. 125–136.
[39] J. Yu, "Enhanced stacked denoising autoencoder-based feature learning for recognition of wafer map defects," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 4, pp. 613–624, Nov. 2019.
[40] A. Shawon, M. O. Faruk, M. B. Habib, and A. M. Khan, "Silicon wafer map defect classification using deep convolutional neural network with data augmentation," in *Proc. IEEE 5th Int. Conf. Comput. Commun. (ICCC)*, Dec. 2019, pp. 1995–1999.
[41] Z. Shen and J. Yu, "Wafer map defect recognition based on deep transfer learning," in *Proc. IEEE Int. Conf. Ind. Eng. Eng. Manage. (IEEM)*, Dec. 2019, pp. 1568–1572.
[42] Y. Kong and D. Ni, "A semi-supervised and incremental modeling framework for wafer map classification," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 1, pp. 62–71, Feb. 2020.
[43] C. H. Jin, H.-J. Kim, Y. Piao, M. Li, and M. Piao, "Wafer map defect pattern classification based on convolutional neural network features and error-correcting output codes," *J. Intell. Manuf.*, vol. 31, no. 8, pp. 1861–1875, Dec. 2020.
[44] J. Shim, S. Kang, and S. Cho, "Active learning of convolutional neural network for cost-effective wafer map pattern classification," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 2, pp. 258–266, May 2020.
[45] Y. Byun and J.-G. Baek, "Mixed pattern recognition methodology on wafer maps with pre-trained convolutional neural networks," in *Proc. 12th Int. Conf. Agents Artif. Intell. (ICAART)*, 2020, pp. 974–979.
[46] U. Batool, M. I. Shapiai, H. Fauzi, and J. X. Fong, "Convolutional neural network for imbalanced data classification of silicon wafer defects," in *Proc. 16th IEEE Int. Colloq. Signal Process. Appl. (CSPA)*, Feb. 2020, pp. 230–235.
[47] S. Park, J. Jang, and C. O. Kim, "Discriminative feature learning and cluster-based defect label reconstruction for reducing uncertainty in wafer bin map labels," *J. Intell. Manuf.*, vol. 32, no. 1, pp. 251–263, Jan. 2021.
[48] P.-C. Shih, C.-C. Hsu, and F.-C. Tien, "Automatic reclaimed wafer classification using deep learning neural networks," *Symmetry*, vol. 12, no. 5, pp. 1–19, 2020.
[49] M. Saqlain, Q. Abbas, and J. Y. Lee, "A deep convolutional neural network for wafer defect identification on an imbalanced dataset in semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 3, pp. 436–444, Aug. 2020.
[50] J. Hwang and H. Kim, "Variational deep clustering of wafer map patterns," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 3, pp. 466–475, Aug. 2020.
[51] J. Zhuang, G. Mao, Y. Wang, X. Chen, and Z. Wei, "A neural-network approach to better diagnosis of defect pattern in wafer bin map," in *Proc. China Semiconductor Technol. Int. Conf. (CSTIC)*, Jun. 2020, pp. 1–3.
[52] Y. Kim, D. Cho, and J.-H. Lee, "Wafer map classifier using deep learning for detecting out-of-distribution failure patterns," in *Proc. IEEE Int. Symp. Phys. Failure Anal. Integr. Circuits (IPFA)*, Jul. 2020, pp. 1–5.
[53] M. B. Alawieh, D. Boning, and D. Z. Pan, "Wafer map defect patterns classification using deep selective learning," in *Proc. 57th ACM/IEEE Design Automat. Conf. (DAC)*, Jul. 2020, pp. 1–6.
[54] Y. Hyun and H. Kim, "Memory-augmented convolutional neural networks with triplet loss for imbalanced wafer defect pattern classification," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 4, pp. 622–634, Nov. 2020.
[55] T.-H. Tsai and Y.-C. Lee, "A light-weight neural network for wafer map classification based on data augmentation," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 4, pp. 663–672, Nov. 2020.
[56] J. C. Chien, M. T. Wu, and J. D. Lee, "Inspection and classification of semiconductor wafer surface defects using CNN deep learning networks," *Appl. Sci.*, vol. 10, no. 15, pp. 1–13, 2020.
[57] J. Yu and J. Liu, "Two-dimensional principal component analysis-based convolutional autoencoder for wafer map defect detection," *IEEE Trans. Ind. Electron.*, vol. 68, no. 9, pp. 8789–8797, Sep. 2021.
[58] Y. Ji and J.-H. Lee, "Using GAN to improve CNN performance of wafer map defect type classification: Yield enhancement," in *Proc. 31st Annu. SEMI Adv. Semiconductor Manuf. Conf. (ASMC)*, Aug. 2020, pp. 1–6.
[59] J. Wang, C. Xu, Z. Yang, J. Zhang, and X. Li, "Deformable convolutional networks for efficient mixed-type wafer defect pattern recognition," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 4, pp. 587–596, Nov. 2020.
[60] Y. Kong and D. Ni, "Qualitative and quantitative analysis of multi-pattern wafer bin maps," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 4, pp. 578–586, Nov. 2020.
[61] S. Kang, "Rotation-invariant wafer map pattern classification with convolutional neural networks," *IEEE Access*, vol. 8, pp. 170650–170658, 2020.
[62] U. Batool, M. I. Shapiai, N. Ismail, H. Fauzi, and S. Salleh, "Oversampling based on data augmentation in convolutional neural network for silicon wafer defect classification," in *Frontiers in Artificial Intelligence and Applications*, vol. 327. Amsterdam, The Netherlands: IOS Press, 2020, pp. 3–12. [Online]. Available: https://ebooks.iospress.nl/volumearticle/55467
[63] H. Lee and H. Kim, "Semi-supervised multi-label learning for classification of wafer bin maps with mixed-type defect patterns," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 4, pp. 653–662, Nov. 2020.
[64] C.-Y. Hsu and J.-C. Chien, "Ensemble convolutional neural networks with weighted majority for wafer bin map pattern classification," *J. Intell. Manuf.*, pp. 1–14, Oct. 2020, doi: 10.1007/s10845-020-01687-7.
[65] J. Cha, S. Oh, D. Kim, and J. Jeong, "A defect detection model for imbalanced wafer image data using CAE and xception," in *Proc. Int. Conf. Intell. Data Sci. Technol. Appl. (IDSTA)*, Oct. 2020, pp. 28–33.
[66] D. Du and Z. Shi, "A wafer map defect pattern classification model based on deep convolutional neural network," in *Proc. IEEE 15th Int. Conf. Solid-State Integr. Circuit Technol. (ICSICT)*, Nov. 2020, pp. 2–4.
[67] M. B. Korzenski and P. Jiang, "Wafer reclaim," in *Handbook for Cleaning for Semiconductor Manufacturing: Fundamentals and Applications*. Hoboken, NJ, USA: Wiley, 2011, pp. 473–500.

UZMA BATOOL received the B.S. degree in computer science from the University of the Punjab, Lahore, Pakistan, and the M.S. degree in computer science from the National University of Computer and Emerging Sciences, Islamabad, Pakistan. She is currently pursuing the Ph.D. degree with Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia. She has over a decade of experience in the software industry and academia. Her research interests include pattern recognition using deep learning algorithms, data mining, machine learning, and image processing.

MOHD IBRAHIM SHAPIAI (Member, IEEE) received the M.Eng. degree from the University of York, U.K., in 2007, and the Ph.D. degree in the area of machine learning from the Universiti Teknologi Malaysia, in 2013. From March 2010 to April 2010, he was a Visiting Researcher with the Graduate School of Information, Production and Systems, Waseda University, Japan, under the supervision of Dr. Junzo Watada, and the Faculty of Engineering, Leeds University, U.K., from June 2012 to July 2012, under the supervision of Dr. Vassili Toropov. He is currently a Senior Lecturer with the Universiti Teknologi Malaysia and a Researcher with the Center of Artificial Intelligence and Robotics (CAIRO). He has also been appointed as a Certified NVIDIA Deep Learning Instructor. His research interests include artificial intelligence, machine learning, brain–computer interface, and swarm intelligence.

MUHAMMAD TAHIR received the Ph.D. degree in chemical engineering from the University of Technology Malaysia (UTM), Malaysia. He is currently an Assistant Professor and the Head of the Photocatalysis Laboratory, School of Chemical and Energy Engineering, University of Technology Malaysia. He is an active researcher in the areas of modeling and simulation, heterogeneous photocatalysis, design of new functional materials, and reaction engineering for greenhouse gas conversion and hydrogen production application. His current research interest includes the simulation, DFT, and development of advanced structured materials for environmental and energy applications.

ZOOL HILMI ISMAIL (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees in mechatronics engineering from the Universiti Teknologi Malaysia, Skudai, Johor, Malaysia, in 2005 and 2007, respectively, and the Ph.D. degree in electrical engineering from Heriot-Watt University, Edinburgh, U.K., in 2011. He was appointed as a Senior Lecturer with the Universiti Teknologi Malaysia (UTM), Kuala Lumpur, in 2011. He was also appointed as a Visiting Researcher with Kyoto University and Jordan University of Science and Technology, in 2014 and 2016, respectively. He is currently a Research Member of Malaysia-Japan International Institute of Technology and the Center for Artificial Intelligence and Robotics, UTM. His current research interests include the area of edge computing, model-predictive control, path planning, and task allocation based on deep-reinforcement learning. He is a member and a Registered Professional Engineer of the Board of Engineers Malaysia, a member of the Society for Underwater Technology, The Institution of Engineering and Technology, the Institute of Electrical and Electronics Engineers—Oceanic Engineering Society, and the Asian Control Association, and a member and a Registered Chartered Marine Engineer of the Institute of Marine Engineering, Science and Technology. In 2017, he was appointed as one of the committee members of International RoboCup@Home Education. He has many awards from International RoboCup Competitions (Service Robot Category).

NOOR JANNAH ZAKARIA received the B.E. and M.Phil. degrees in electronic system engineering from Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia, where she is currently pursuing the Ph.D. degree. Her research interests include artificial intelligence, supervised learning, deep learning, reinforcement learning, and object detection in autonomous vehicle field of interest.

AHMED ELFAKHARANY received the B.Sc. degree in mechatronics from the Arab Academy for Science, Technology and Maritime Transport, Egypt, and the M.Phil. degree from Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia.