Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 217 (2023) 570–583

Procedia Computer Science

www.elsevier.com/locate/procedia

4th International Conference on Industry 4.0 and Smart Manufacturing

A Review Study on ML-based Methods for Defect-Pattern Recognition in Wafer Maps

T. Theodosiou[a*], A. Rapti[a], K. Papageorgiou[a], T. Tziolas[a], E. Papageorgiou[a],

N. Dimitriou[b], G. Margetis[c], D. Tzovaras[b]

[a]*University of Thessalyirst affiliation, Dept. of Energy Systems, Gaiopolis, 41500 Larissa, Greece*
[b]*Centre for Research and Technology Hellas, Information Technologies Institute, 6th Km Charilaou-Thermi Road, 57001 Thermi-Thessaloniki, Greece*
[c] *Foundation for Research and Technology Hellas, Institute of Computer Science, N. Plastira 100, Vassilika Vouton, GR-700 13 Heraklion, Greece*

Abstract

The identification of defects plays a key role in the semiconductor industry as it can reduce production risks, minimize the effects of unexpected downtimes and optimize the production process. A literature review protocol is implemented and latest advances are reported in defect detection considering wafer maps towards quality control. In particular, the most recent works are outlined to demonstrate the use of AI-technologies in wafer maps defect detection. The popularity of these technologies is then presented in the form of visualizing graphs. This enables the identification of the most popular and most prominent ML-methods that can be exploited for the purposes of Industry 4.0.

Keywords: Artificial Intelligence; Machine Learning; Deep Learning; Defect Identification, Smart Manufacturing; Industry 4.0

1. Introduction

Common machine learning (ML) methods are reviewed in [1] for intelligent manufacturing, which triggers an extensive discussion on their strengths and weaknesses in a wide range of manufacturing applications. In a recent

* Corresponding author. Tel.: +30-2410-684-262 ; fax: +30-2410-684-262 .
E-mail address: dozius@uth.gr

1877-0509
10.1016/j.procs.2022.12.253

comparative review study on ML algorithms for smart manufacturing (SM), various well-known ML techniques, including Artificial Neural Network (ANN), Support Vector Machine (SVM) and Random Forest (RF), were implemented regarding industrial tool wear prediction [2]. Also, ML techniques including neural networks (NNs), fuzzy logic, genetic algorithms, and hybrid systems were reviewed for the decision making and monitoring of several industrial operations [3]. Concerning their structure, ML models are usually designed with shallow structures; such traditional models are ANN, SVM, logistic regression (LR), etc. In the case of ML models with limited handcrafted features, a decent performance is achieved in a variety of applications. However, the massive data in SM imposes a variety of challenges [4], such as the proliferation of multimodal data, high dimensionality of feature space, and multicollinearity among data measurements. These challenges seriously affect ML algorithms, greatly impeding their performance.

Deep Learning (DL) is an extension of ML and describes the ability of smart systems to imitate human brain functionality in tasks such as decision making and data processing. DL techniques enable people to (1) automatically learn from data, (2) detect underlying patterns, and eventually (3) make efficient decisions. Encapsulating automatic feature learning and high-volume modelling capabilities, DL provides an advanced analytics tool for SM in the big data era. It uses a cascade of layers of nonlinear processing to learn the representations of data corresponding to different levels of abstraction. The hidden patterns underneath are then identified and predicted through an end-to-end optimization. Thus, DL offers great potential to enhance data-driven manufacturing applications [5]. There are several review papers extracted from the related literature, which show the actual implementations of ML and DL methods in factory operations within the SM domain.

The authors in [6] performed a review study focused on the applications and challenges of ML techniques in SM. This study provides an overview regarding several ML algorithms (e.g. SVM, k-nearest neighbor, NN etc.) which bring notable improvements inside different manufacturing areas, such as optimization, quality control, prediction of failure, cost reduction and transparency. Future trends concerning the applications of ML in SM are also discussed.

Additionally, a systematic review on the application of ML for manufacturing processes was presented in [7]. This study focused on the efficient application of various DL models including Convolutional Neural Networks (CNNs) and other Deep NN architectures, in certain smart industry processes, such as image recognition and object detection, thereby enhancing industrial solutions.

In another study [8], a systematic literature review was conducted concerning the application of DL techniques to wafer maps defect detection. Typical DL models such as Recurrent NNs, Generative Adversarial Networks (GANs) and Deep CNNs play a key role in the automatic learning from data, thus, producing different levels of pattern recognition, mainly used for fault assessment and defect detection in semiconductor industry.

This review is conducted within the framework of the "OPTIMAI" project, oriented in defect detection in semiconductor industry considering wafer maps towards quality control. It is part of an extended literature survey on Artificial Intelligence (AI) and DL methods for defect pattern recognition, applied in certain OPTIMAI use cases in semiconductor wafers. The review was performed to investigate all the implemented methodologies so far in the domain of anomaly detection and fault recognition so as to help with the identification of root causes in the manufacturing process and the overall support of production optimization. The performed survey is referred to years from 2017 up to date.

2. Introduction

The present short review is based on the preferred reporting items for systematic reviews and meta-analyses (PRISMA) principles [9], as this process is suitable for well-structured articles research. According to this method, the survey was carried out using a predefined question which leads to the identification of the studies included in the survey. These studies are collected, analyzed and critically assessed.

2.1. Literature Search

In this review study, Scopus was the extracted database for the conducted literature search. Scopus is the only database exploited for the purposes of the current research for certain reasons. First, the presented research work constitutes a short review study which should comply with paper size limitations. This prevents investigating

extensively and in-depth the domain of defect-pattern recognition in wafer maps exploiting ML-based methods. An extended version is being prepared as an extension of this work. Secondly, Scopus is a well-known database, accessible by most scientists, which provides freely interdisciplinary scientific data and literature across all research areas. Its board members include an international group of scientists, researchers and librarians of all major scientific disciplines who can ensure the high quality and authenticity of the published articles. Thus, present work can be accredited with meeting all the scientific requirements and complying with ethical principles such as the accuracy of scientific knowledge.

Predetermined eligibility criteria were applied on the search for articles and were based on the articles' title, abstract and keywords. Further investigation of the full texts was accomplished to verify that the articles initially located are in line with the inclusion criteria. The literature search strategy was conducted by utilizing the keywords including "wafer maps", "pattern recognition", "pattern classification", "deep learning" and "Machine learning", using the query string in Fig. 1; this query string was oriented in locating adequate numbers of relevant studies for inclusion to conduct a well-built research work with credible results.

```
( "wafer maps" AND ( "defect identification" OR
"pattern recognition" OR "pattern classification ")
AND ( "deep learning" OR "machine learning" ) )
```

Fig. 1. Query string employed in literature search.

2.2. Eligibility Criteria

The studies that are part of the short review meet the eligibility criteria incorporated, which include the reference of the author, the year of publication, the type of the article, the methodology applied that refers to AI, DL or ML, as well the scope and the industrial process in the manufacturing domain. The inclusion criteria include: (i) papers published between the 1st of January 2017 and the 15th of May 2022 (date of literature search)); (ii) the process of defect detection in wafer maps; (iii) AI-based algorithms, including both traditional ML and DL techniques incorporated for defect pattern recognition; (iv) ML-based methods in wafer maps defect detection.

The exclusion criteria were as follows: (i) articles published before 2017; (ii) articles not related to AI/ML; (iii) Websites and online material, student theses, book chapters, editorials, commentaries and non-original research articles, such as protocols, meta-analysis, etc.; (vi) journal or conference reviews; (v) articles not written in English.

2.3. AI Methods

In SM, AI has found significant applicability for processing and analysing big manufacturing data. The most popular DL methods are the DNNs, CNNs, Res-Nets, GANs, auto-encoders, mainly used for defect pattern recognition in semiconductor industry as regards wafer maps.

3. Results

After a thorough review of all mined articles, only those applying AI, ML, or DL methodologies for defect pattern recognition on wafer maps were retained. After the adoption of the PRISMA method and only those articles that were explicit to the subject of this short review were retrieved. The overall search process is graphically illustrated in Fig. 2.

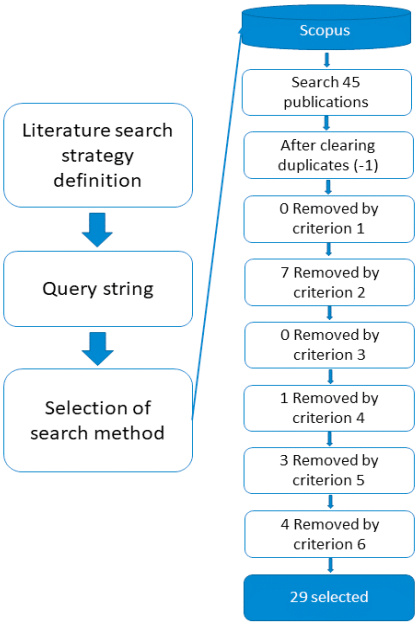

Fig. 2. Research process screening.

Finally, 29 research papers were selected for further analysis. The selected articles are listed in Table I accompanied by the year of publication, the method deployed, the scope of the study and the wafer map dataset in which the methodology was applied.

In this study, we focused on three main AI methods that were involved for wafer map defect pattern recognition, namely ML, DL and hybrid methods of these two. To this end, we categorized the literature results based on the aforementioned types of methods. The results can be further categorized by the datasets exploited for model's evaluation. Specifically, experimental evaluation was performed either on public datasets or on real wafer maps that haven't been publicly distributed.

Concerning public datasets, WM-811K and MixedWM38 were found in literature. WM-811K [10] was collected from 46,293 lots and consists of 811,457 real industrial wafer maps; approximately 20% of the wafer maps were annotated about their defect types by experts in the field where they defined 9 defect patterns. To alleviate single defect patterns, MixedWM38 [11] was introduced. This dataset consists of nine patterns of WM-811K and 29 combinations of them. To handle data imbalance, the producers of the dataset created synthetic images for class with minority samples. Thus, 38,000 samples of this dataset are based on both real and synthetic wafer maps. On the other hand, several works run experiments on different datasets than the previous mentioned, to examine customized solutions based on real and private datasets from semiconductor industries.

3.1. ML Methods

Starting with the ML approaches in the field of wafer map pattern recognition, a voting ensemble classifier with multi-type features to identify wafer map defect patterns is proposed in [12]. Four state-of-the-art ML base classifiers such as LR, RF, gradient boosting machine (GBM), and ANN were trained with the same set of features such as density and geometry, extracted from raw wafer images of WM-811K. An ensemble approach was proposed, collecting the best results of all classifiers and aggregating them to get the final classification result for all defect classes.

Using a clustering approach and specifically the density-based spatial clustering of applications with noise (DBSCAN) [13], the authors of [14] detected and classified the defected patterns of WM-811K. In their methodology, Cartesian and polar coordinates of defective and edge dies were extracted and used for DBSCAN clustering. Subsequently, to detect patterns, outliers were treated differently for each pattern, depending on whether they should

be removed or not. For classification, each wafer map was iteratively checked for each defect class. The classification results for only one type of pattern provided 100% recall and 85% precision.

For the same dataset and task as the previous approach, an enhanced stacked denoising autoencoder (ESDAE) with manifold regularization was proposed in [15]. Wafer maps were denoised with median-filtering and four types of features (geometrical, gray, texture and projections) were extracted from each wafer map to be used as input in ESDAE. ESDAE consists of two autoencoders both having a single hidden layer. To handle imbalance, a cost sensitive learning method was developed to assign different costs to misclassifications of minority classes. The proposed system can recognize defect patterns with a 91.9% rate.

For a simulated dataset on analogue wafer test data and the task of defect pattern recognition, a comparison of three different ensemble approaches was examined in [16]. Specifically, the authors compared RF with ensemble RF-based approaches, namely: bagging, voting, and adaboost. All models achieved almost similar classification performance, with a slight increase in accuracy for adaboost and bagging.

Ensemble models were also used in the work of [17], which examined the recognition of the grid type defect pattern on a dataset of real wafer maps. In the proposed five-phase methodology, the first four steps were focused on enhancing the wafer maps to reveal grid patterns, while the last phase aimed to recognize the defect patterns with ML techniques. The ensemble models that were examined are adaboost with decision trees, RF, gradient boost, XGboost, extremely randomized trees, and bagging with decision tree. The features used in this method were density, radon, geometry and line. The highest classification accuracy (96.45%) was produced by the extremely randomized trees.

Another ML method was proposed in [18] which is based on semi-supervised learning, in which labeled data prone to misclassifications were excluded from the training. The first two phases of the proposed framework are dedicated to preprocessing (data cleansing, DBSCAN denoising) and feature extraction (geometric, statistics, radon). Phase three is about constructing the classifier, phases four and five are about clustering similar unlabeled data, while phase five involves enhanced labeling with experts. The bagging algorithm with 1000 decision trees was used as a classifier, while self-organized map was used for unsupervised learning in phase five. Experimental results in a modified WM-811K dataset with 5 new defect types produced an overall accuracy of 94.37%.

Subsequently, a few-shot learning method was also examined in the work of [19], based on a hybrid self-attention and prototype network. Prototypical networks differ from traditional DL methods as they adopt the meta-learning training paradigm to handle data scarcity learning. The proposed few-shot network for classifying defect patterns in both WM-811K and MixedWM38 datasets, involves three steps: feature embedding, feature representation and distance measurement. This method managed to produce high classification performance for most of the patterns even with a small amount of data.

To recognize defect patterns in WM-811K, a self-organizing incremental neural network (SOINN) was proposed in [20]. Radon features from wafer maps were used within the SOINN and were based on the input and the abilities of SOINN. The model determines if additional neurons are required and can identify new patterns that were not seen in the training phase. The proposed approach is useful for applications with dynamic changes.

Using a custom dataset of real wafer maps, the authors of [21] proposed a defect pattern recognition method with adaboost for scratch patterns. Out of the five-step methodology, the first four steps were dedicated in enhancing the patterns in wafer maps, by examining clustering and preprocessing techniques. In the final step, classification was performed with adaboost based on 500 decision trees and 0.7 learning rate. Feature extraction was performed for density, radon, geometry and line features. The results achieved over 89% recognition rate in scratch pattern and over 94% in common defect patterns.

3.2. DL Methods

The application of ML-based algorithms has the downside of the manual feature extraction process. Thus, DL methods emerged, that were focused on unsupervised feature extraction with CNNs. The authors in [22] proposed a CNN classifier for wafer map defect pattern recognition and image retrieval tasks. Their CNN classifier consists of three pairs of convolutional-pooling layers and two fully connected layers. The model was trained with simulated wafer maps and was tested in real wafer maps. It was proved that using only synthetic data for training the model can provide high classification accuracy with real wafer maps. Image retrieval was applied using the features from the first fully connected layer and the Hamming distance measure, giving 3.7% error rate.

In [23], the authors proposed a framework for wafer map pattern recognition based on CNN, and particularly VGG [24]. To increase the classification performance, they developed a rule-based denoising method for wafer-maps using the Hough transform. The accuracy and the macro F1-score presented in the WM-811K, are 94.7% and 73.67%, respectively.

Using a different approach, the authors in [25] presented a method for detecting and segmenting wafer map defect patterns with CNN autoencoders. They examined CNN autoencoder architectures based on the SegNet [26], the U-Net [27] and a fully convolutional network. Synthetic data were used to train the autoencoders, while the real data were used to evaluate the detecting and segmenting performances. The intersection over union formula was used as a metric for object detection performance. They demonstrated that their approach of using only synthetic data during training, is effective for detection of unseen defect patterns from real wafer maps.

An approach with GAN was presented in [28] to improve the classification performance in WM-811K. The authors proposed an adaptive balancing GAN (AdaBalGAN) with imbalanced learning. The framework consists of two parts: a categorical GAN for data generation and an adaptive generative controller for evaluating classification performance with respect to imbalanced learning ability and the sample size. The method achieved 96% classification accuracy.

Another CNN classification approach for the same task was proposed in [29]. The authors targeted on a small amount of wafer maps. Moreover, they analyzed popular CNN architectures such as VGG [24], ResNet [30] and MobileNet [31] to evaluate their classification performances. They used a composite small training dataset of both synthetic and real data. Real wafer maps were derived from WM-811K. The classification accuracy in the remaining data of WM-811K for this approach was 87.8%.

In a different method, the authors in [32] focused on a methodology that decreases the labelling needs for CNNs with the use of active learning. The proposed framework consists of four main phases that were constantly repeated in order to increase the classification performance. These phases involve: uncertainty estimation of the first results on an initial labeled dataset, query wafer selection based on the previous estimation, query wafer labeling by experts, and model update. Using a LeNet-5 [33] CNN architecture, they carried out experimental comparisons between various uncertainty estimation methods on the WM-811K dataset, to evaluate the classification performance of the proposed system.

Contrary to the previous approach, CNN with selective learning was presented in [34] as well as data augmentation with autoencoders. The CNN classifier consists of three pairs of convolutional and pooling layers and one fully connected layer. The output of the model has two heads: one for a prediction and one for a selection function. Selection function is a binary reject option that indicates the risk of misclassification. With this option, the model is allowed to abstain from prediction when there is an unpredictable sample. The accuracies achieved with and without the reject option in WM-811K were 99% and 94% respectively.

In the next work [35], a totally different dataset for defect pattern recognition was assessed. Specifically, real wafer images were used as wafer maps in order to recognize 11 common defect patterns. The images were transformed with Radon transform while both transformations and original data were combined in a CNN classifier to improve recognition accuracy. The proposed model (RadonNet) is a combination of VGG16 [24], Inception [36] and ResNet [30]. Using data augmentation techniques based on basic image manipulations, the authors achieved 98.5% accuracy.

Subsequently, the authors of [11] introduced a dataset for mixed type defect patterns naming it as MixedWM38. There are 38 defect patterns in this dataset, based on both real and simulated wafer maps. Moreover, they proposed a deformable CNN of five blocks to classify the mixed defect patterns. The first block uses convolutional and batch normalization layers. In the next three blocks, deformable convolutions were added, while the last block consisted of two fully connected layers. The model recognized the mixed-type defect patterns with 93.2% average accuracy.

The authors in [37] proposed a DL-based CNN for automatic wafer defect identification (CNN-WDI). A 2D CNN model was constructed for WDI with one input layer, eight convolutional layers (CL) comprising batch normalization (BN) [38], padding and Rectified Linear Unit (ReLU) activation [39], five pool layers (PL) [33] (four stacking pairs of Conv-Pool-Conv), one dropout layer, two fully connect (FC) layers, and one output layer. Experimental results showed that the CNN-WDI model outperforms previously proposed models such as WMFPR, DTE-WMFPR, and WMDPI, using the same dataset, in terms of classification accuracy (96.2%).

The abilities of transfer learning for wafer map defect recognition were examined in [40]. The authors proposed a method, where the 29-layer DCNN model was trained with the MNIST [33] dataset and fine-tuned with data from

WM-811K. Transfer learning aims to minimize the needs for large datasets. Thus, the authors performed fine-tuning on their model with small training datasets derived from the semiconductor benchmark dataset. The provided accuracy in a small and balanced dataset of seven defect patterns was above 94.9%.

In the next work [41] defect detection for mixed-type defect patterns with an enhanced Mask R-CNN [42] was performed. The authors used the ResNet architecture as the backbone feature extraction network and further improved the Mask R-CNN with the use of the feature pyramid network and the soft-NMS algorithm. The experimental results in the MixedWM38 showed that their improved model had higher detection precision of wafer maps compared to the Mask R-CNN. In addition, soft-NMS provided higher detection precision in testing than the non-maximum suppression algorithm.

Another CNN classifier for WM-811K was proposed in [43]. This approach differs from other CNN implementations as it uses few-shot learning to alleviate imbalance learning by sampling evenly the different classes in a training batch. Furthermore, self-supervised learning technique was examined to exploit the unlabeled data that exist in the dataset. Thus, the proposed classification framework incorporates few-shot loss and self-supervised loss for labeled and unlabeled data, respectively. Inception blocks [36] were also adopted for the CNN architecture. The experimental results were further compared to ML approaches demonstrating higher classification performance.

In [44], popular CNN classifiers and out-of-distribution learning were examined for wafer map pattern identification, in a custom dataset of real wafer maps. Four popular CNN architectures namely VGG16, ResNet, MobilenetV2 [45] and shufflenet [46] were explored. For out-of-distribution detection, a threshold in the softmax output was utilized to determine that these are undefined data. All models showed high detection accuracy for out-of-distribution data with ResNet achieving the highest classification performance.

A CNN and a variational autoencoder for solving the data imbalance of WM-811K with data augmentation, were proposed in [47]. In this approach, the variational autoencoder generates similar wafer maps, while a deep CNN with convolutional, pooling, and batch normalization layers and leaky ReLU activation function, performs classification. The proposed method showed very high accuracy (99.19%).

3.3. Hybrid methods

CNNs “black box” work principle of unsupervised feature extraction has the risk that the most representative features may not be extracted during training. Thus, several works examined both DL and ML methods and combinations of them to increase the recognition accuracy. In [48] they examined classifiers based on a CNN, a SVM, an adaptive boosting (Adaboost) and an extreme gradient boosting (XGboost). For the non-CNN models, feature extraction with singular value decomposition was performed. In addition, GridSearchCV approach was used to find the optimization parameters for these ML algorithms. To tackle overfitting, data augmentation techniques with geometric transformations were used. CNN achieved the highest accuracy of 99.2% in a real wafer map dataset of TSMC 300mm fab.

The authors in [49] proposed a hybrid method with stacked convolutional sparse denoising autoencoder (SCSDAE), for the defect pattern identification task in WM-811K. Stacked sparse autoencoders were used for feature learning on a stochastically corrupted input. The SCSDAE architecture consists of two CSDAEs with convolution and pooling layers and one final classification layer with an SVM classifier. Dropout layer with 0.9 probability was used for regularization. The recognition accuracy this method achieved was 94.75%.

A combination of CNN, error correcting output codes (ECOC) and SVM for wafer map defect pattern recognition was presented in [50]. The extracted features of the CNN were used to train the ECOC model and consist of SVM binary classifiers. Experiments were conducted in WM-811K, and their approach showed accuracy of 98.43%

A hybrid method that combines principal component analysis and convolutional autoencoder (PCACAE) for the semiconductor benchmark dataset was proposed in [51]. The PCACAE was trained layer-wise and has two convolutional modules with convolutional pooling and batch normalization layers, and an output module with two fully connected layers. Feature extraction was further performed with a conditional two-dimensional principal component analysis (C2DPCA) and then features were cascaded with the autoencoder. This method achieved 97.27% accuracy and outperformed popular CNN classifiers.

Lastly, a hybrid method is proposed in [52] for the automation of wafer map pattern classification in semiconductor manufacturing. This method describes a stacking ensemble that combines base classifiers and a meta-classifier. Base

classifiers are built based on the manual feature extraction (MFE), which transforms a wafer map into a fixed-length vector of handcrafted features and then builds an off-the-shelf classifier, and CNN approaches. The meta-classifier is trained to combine the different strengths of the heterogeneous base classifiers [53]. The proposed method yielded an improved classification performance compared to the base classifiers and other baseline methods, outperforming the best model between the MFE+FNN and the CNN for most defect classes, with respect to F1 scores. The model can recognize the wafer map failure types while it automatically extracts its features.

Table 1. Results of Literature Search sored by Year.

No	Author	Year	AI/DL Method	Scope	Dataset
1	(Nakazawa and Kulkarni 2018) [22]	2018	CNN	WM defect pattern recognition and image retrieval	Simulated, real WM
2	(Saqlain, Jargalsaikhan & Lee., 2019) [12]	2019	Voting ensemble classifier consisting of LR, RF, GBM and ANN	WM defect pattern recognition	WM-811K
3	(Ishida et al. 2019) [23]	2019	CNN based on VGGnet	WM defect pattern recognition	WM-811K
4	(Nakazawa and Kulkarni 2019) [25]	2019	CNN autoencoders	WM defect detection and segmentation	Synthetic, real WM
5	(Yuan-Fu 2019) [48]	2019	CNN, SVM, adaboost and XGboost	WM defect pattern recognition	TSMC 300mm fab
6	(Yu, Zheng, and Liu 2019) [49]	2019	Stacked convolutional sparse denoising autoencoder with SVM	WM defect pattern recognition	Simulated, WM-811K
7	(Jin et al. 2019) [14]	2019	DBSCAN	WM defect pattern recognition and detection	WM-811K
8	(Wang et al. 2019) [28]	2019	GAN	WM defect pattern recognition	WM-811K
9	(Yu 2019) [15]	2019	Enhanced stacked denoising autoencoder with manifold regularization	WM defect pattern recognition	WM-811K
10	(Maksim et al. 2019) [29]	2019	CNN	WM defect pattern recognition	Synthetic, WM-811K
11	(Shim, Kang, and Cho 2020) [32]	2020	CNN with active learning	Cost-effective WM defect pattern recognition	WM-811K
12	(Alawieh, Boning, and Pan 2020) [34]	2020	CNN with selective learning	WM defect pattern recognition	WM-811K
13	(Yuan-Fu and Min 2020) [35]	2020	CNN	WM defect pattern recognition	Images from real wafers
14	(Wang et al. 2020) [11]	2020	Deformable CNN	WM defect pattern recognition	MixedWM38
15	(Jin et al. 2020) [50]	2020	CNN with error-correcting output codes and SVM	WM defect pattern recognition	WM-811K
16	(Saqlain, Abbas & Lee, 2020) [37]	2020	CNN	WM defect pattern recognition	WM-811K
17	(Abdullah, Rahman, and Akhter 2021) [16]	2021	Ensemble approach with bagging, boosting, voting, adaboost	WM defect pattern recognition	Simulated

No	Author	Year	AI/DL Method	Scope	Dataset
18	(Liao et al. 2021) [17]	2021	Adaboost with decision tree, random forest, gradient boost, XGboost, extremely randomized trees, and bagging with decision tree	WM grid defect pattern recognition	Real wafer maps
19	(Chen et al. 2021) [18]	2021	Bootstrap aggregating and Self-Organizing Maps	WM defect pattern recognition and enhanced labeling	WM-811K
20	(Chen et al. 2021) [40]	2021	CNN based on transfer learning	WM defect pattern recognition	WM-811K
21	(Li and Wang 2021) [41]	2021	Improved Mask R-CNN	WM defect pattern detection	MixedWM38
22	(Geng et al. 2021) [43]	2021	CNN with few-shot and self-supervised learning	WM defect pattern recognition	WM-811K
23	(Yu et al. 2021) [19]	2021	Prototypical network	WM defect pattern recognition	WM-811K, MixedWM38
24	(Yu et al. 2021) [20]	2021	Self-organizing incremental neural network	WM defect pattern recognition	WM-811K
25	(Li et al. 2021) [21]	2021	Adaboost	WM defect pattern recognition	Real wafer maps
26	(Kim, Cho, and Lee 2021) [44]	2021	CNNs	WM defect pattern recognition	Real wafer maps
27	(Yu and Liu 2021) [51]	2021	PCACAE	WM defect pattern recognition	WM-811K
28	(Wang et al. 2021) [47]	2021	CNN	WM defect pattern recognition	WM-811K
29	(Kang & Kang, 2021) [52]	2021	Stacking ensemble, based on MFE and CNN.	WM defect pattern recognition	WM-811K

4. Discussion

After a thorough literature analysis regarding the utilization of various AI-technologies reported in the beforementioned reviewed state of the art, the distribution of the three groups of employed AI-technologies for defect pattern recognition in wafer maps is depicted in Fig. 3.

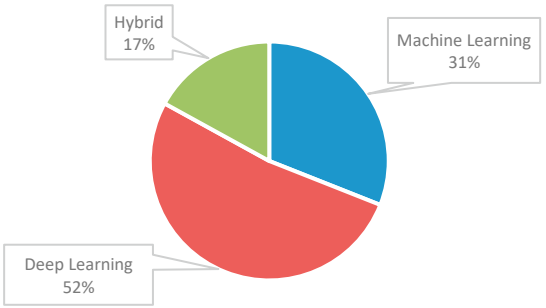

Fig. 3. Preference of the three AI-based methodologies.

More specifically, Fig. 3 initially provides a general visual distribution of the AI methods that researchers focused on, for the task of identifying defect patterns in wafer maps. DL approaches are the most popular choices (52%) as they succeeded in increasing both the preprocessing efficiency and the classification accuracy. ML methods are also

widely used with 31% presence in the results. The complexity of combining different models, features and hyperparameters held back most of the researchers and thus only 17% of the works focused on hybrid models [54,55].

Most of the methods examined herein, performed on the WM-811K (55%) public dataset as shown in Fig. 4. The WM-811K dataset is the world's largest known dataset of wafer maps, available to the public, including 811,457 wafer maps collected from 46,393 lots in real-world fabrication. Thus, it has been extensively used by researchers for pattern recognition tasks, and more specifically to train and test their models. Therefore, the numerical predominance of the WM-811K dataset against all the rest datasets utilized for the purposes of this work is fully accepted and justified. The other public MixedWM38 dataset hasn't received much attention in literature, as it provides 7% presence as standalone and 3% along with the WM-811K. Moreover, a considerable amount of works (35%) focused on customized datasets of wafer maps that kept private from the research community.

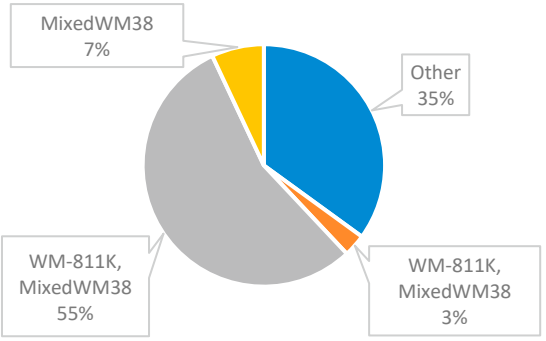

Fig. 4. Usage of datasets in literature.

The distribution of publications according to the year that have been published is shown in Fig. 5. It is observed that there was a downtrend in the research for this topic in the year 2020. This led to research gaps, that scientists focused on in the next year. No publications for the current year (2022) that match the eligibility criteria have been found in Scopus.

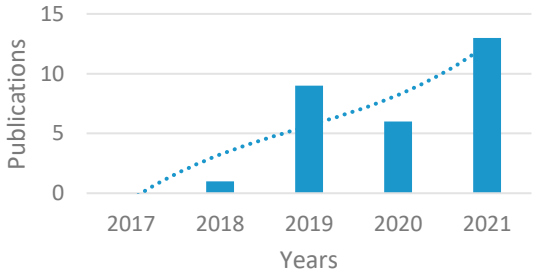

Fig. 5. Publications per year.

The findings of this research study showed that CNNs are the most popular architecture, followed by ANNs and other ensemble methods utilizing Adaboost, all having a significant contribution in Industry 4.0. The exploitation of these AI-based technologies towards defect-pattern recognition in the wafer maps domain spark certain challenges such as explainability, quality of training, interoperability, privacy and security. These challenges hold a significant part in the exploitation of these models and can be considered future trends in this realm, which need to be further and exclusively investigated and dealt with in future extensions of this work.

5. Conclusion

This paper reviews the current state of the art and the existing results reported in research articles regarding AI-based defect recognition methods in wafer maps of the semiconductor manufacturing domain. It is obvious that many

AI technologies have already been successfully incorporated within the framework for defect pattern recognition in wafer maps. ML and especially DL have taken tremendous steps towards performance while they enhance the state of the art in this quality inspection topic for smart manufacturing.

Acknowledgement

This work has been supported by the European Commission through project OPTIMAI funded by the European Union (H2020-NMBP-TR-IND-2020-singlestage, Topic: DT-FOF-11-2020, GA 958264). The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the European Commission.

References

[1] Lade P, Ghosh R, Srinivasan S. Manufacturing Analytics and Industrial Internet of Things. IEEE Intell Syst 2017;32:74–9. https://doi.org/10.1109/MIS.2017.49.
[2] Wu D, Jennings C, Terpenny J, Gao RX, Kumara S. A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests. Journal of Manufacturing Science and Engineering 2017;139:071018. https://doi.org/10.1115/1.4036350.
[3] Helu M, Libes D, Lubell J, Lyons K, Morris KC. Enabling Smart Manufacturing Technologies for Decision-Making Support. Volume 1B: 36th Computers and Information in Engineering Conference, Charlotte, North Carolina, USA: American Society of Mechanical Engineers; 2016, p. V01BT02A035. https://doi.org/10.1115/DETC2016-59721.
[4] Wuest T, Weimer D, Irgens C, Thoben K-D. Machine learning in manufacturing: advantages, challenges, and applications. Production & Manufacturing Research 2016;4:23–45. https://doi.org/10.1080/21693277.2016.1192517.
[5] Wang J, Ma Y, Zhang L, Gao RX, Wu D. Deep learning for smart manufacturing: Methods and applications. Journal of Manufacturing Systems 2018;48:144–56. https://doi.org/10.1016/j.jmsy.2018.01.003.
[6] Bajic B, Cosic I, Lazarevic M, Sremcev N, Rikalovic A. Machine Learning Techniques for Smart Manufacturing: Applications and Challenges in Industry 4.0, 2018.
[7] Fahle S, Prinz C, Kuhlenkötter B. Systematic review on machine learning (ML) methods for manufacturing processes – Identifying artificial intelligence (AI) methods for field application. Procedia CIRP 2020;93:413–8. https://doi.org/10.1016/j.procir.2020.04.109.
[8] Batool U, Shapiai MI, Tahir M, Ismail ZH, Zakaria NJ, Elfakharany A. A Systematic Review of Deep Learning for Silicon Wafer Defect Recognition. IEEE Access 2021;9:116572–93. https://doi.org/10.1109/ACCESS.2021.3106171.
[9] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. International Journal of Surgery 2010;8:336–41. https://doi.org/10.1016/j.ijsu.2010.02.007.
[10] M. Wu, J. R. Jang, J. Chen. Wafer Map Failure Pattern Recognition and Similarity Ranking for Large-Scale Data Sets. IEEE Transactions on Semiconductor Manufacturing 2015;28:1–12. https://doi.org/10.1109/TSM.2014.2364237.
[11] Wang J, Xu C, Yang Z, Zhang J, Li X. Deformable Convolutional Networks for Efficient Mixed-Type Wafer Defect Pattern Recognition. IEEE Transactions on Semiconductor Manufacturing 2020;33:587–96. https://doi.org/10.1109/TSM.2020.3020985.
[12] Saqlain M, Jargalsaikhan B, Lee JY. A Voting Ensemble Classifier for Wafer Map Defect Patterns Identification in Semiconductor Manufacturing. IEEE Trans Semicond Manufact 2019;32:171–82. https://doi.org/10.1109/TSM.2019.2904306.
[13] Ester M, Kriegel H-P, Sander J, Xu X, others. A density-based algorithm for discovering clusters in large spatial databases with noise. kdd, vol. 96, 1996, p. 226–31.
[14] Jin CH, Na HJ, Piao M, Pok G, Ryu KH. A Novel DBSCAN-Based Defect Pattern Detection and Classification Framework for Wafer Bin Map. IEEE Transactions on Semiconductor Manufacturing 2019;32:286–92. https://doi.org/10.1109/TSM.2019.2916835.

[15] Yu J. Enhanced Stacked Denoising Autoencoder-Based Feature Learning for Recognition of Wafer Map Defects. IEEE Transactions on Semiconductor Manufacturing 2019;32:613–24. https://doi.org/10.1109/TSM.2019.2940334.
[16] Abdullah Md, Rahman MdH, Akhter S. Pattern Recognition in Analog Wafermaps with Multiple Ensemble Approaches. 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), 2021, p. 587–91. https://doi.org/10.1109/ICREST51555.2021.9331084.
[17] Liao PY-Y, Shu-Min Li K, Chen LL-Y, Wang S-J, Huang AY-A, Chau-Cheung Cheng K, et al. WGrid: Wafermap Grid Pattern Recognition with Machine Learning Techniques. 2021 IEEE International Test Conference (ITC), 2021, p. 309–13. https://doi.org/10.1109/ITC50571.2021.00043.
[18] Chen LL-Y, Shu-Min Li K, Jiang X-H, Wang S-J, Huang AY-A, Chen JE, et al. Semi-Supervised Framework for Wafer Defect Pattern Recognition with Enhanced Labeling. 2021 IEEE International Test Conference (ITC), 2021, p. 208–12. https://doi.org/10.1109/ITC50571.2021.00029.
[19] Yu N, Chen H, Xu Q, Hasan MM. Wafer map defect recognition with few shot learning based on hybrid self-attention mechanism and prototype network. 2021 China Automation Congress (CAC), 2021, p. 4128–34. https://doi.org/10.1109/CAC53003.2021.9727346.
[20] Yu N, Xu Q, Hasan MM, Jiang K. Wafer Map Defect Pattern Recognition Based on Self-Organizing Incremental Neural Network. 2021 China Automation Congress (CAC), 2021, p. 5617–22. https://doi.org/10.1109/CAC53003.2021.9728184.
[21] Li KS-M, Liao PY-Y, Cheng KC-C, Chen LL-Y, Wang S-J, Huang AY-A, et al. Hidden Wafer Scratch Defects Projection for Diagnosis and Quality Enhancement. IEEE Transactions on Semiconductor Manufacturing 2021;34:9–16. https://doi.org/10.1109/TSM.2020.3040998.
[22] Nakazawa T, Kulkarni DV. Wafer Map Defect Pattern Classification and Image Retrieval Using Convolutional Neural Network. IEEE Transactions on Semiconductor Manufacturing 2018;31:309–14. https://doi.org/10.1109/TSM.2018.2795466.
[23] Ishida T, Nitta I, Fukuda D, Kanazawa Y. Deep Learning-Based Wafer-Map Failure Pattern Recognition Framework. 20th International Symposium on Quality Electronic Design (ISQED), 2019, p. 291–7. https://doi.org/10.1109/ISQED.2019.8697407.
[24] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, arXiv; 2014, p. 1409–556.
[25] Nakazawa T, Kulkarni DV. Anomaly Detection and Segmentation for Wafer Defect Patterns Using Deep Convolutional Encoder–Decoder Neural Network Architectures in Semiconductor Manufacturing. IEEE Transactions on Semiconductor Manufacturing 2019;32:250–6. https://doi.org/10.1109/TSM.2019.2897690.
[26] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 2017;39:2481–95.
[27] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Cham: Springer International Publishing; 2015, p. 234–41.
[28] Wang J, Yang Z, Zhang J, Zhang Q, Chien W-TK. AdaBalGAN: An Improved Generative Adversarial Network With Imbalanced Learning for Wafer Defective Pattern Recognition. IEEE Transactions on Semiconductor Manufacturing 2019;32:310–9. https://doi.org/10.1109/TSM.2019.2925361.
[29] Maksim K, Kirill B, Eduard Z, Nikita G, Aleksandr B, Arina L, et al. Classification of Wafer Maps Defect Based on Deep Learning Methods With Small Amount of Data. 2019 International Conference on Engineering and Telecommunication (EnT), 2019, p. 1–5. https://doi.org/10.1109/EnT47717.2019.9030550.
[30] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, p. 770–8. https://doi.org/10.1109/CVPR.2016.90.
[31] Xie L, Huang R, Cao Z. Detection and Classification of Defect Patterns in Optical Inspection Using Support Vector Machines. In: Huang D-S, Bevilacqua V, Figueroa JC, Premaratne P, editors. Intelligent Computing Theories, vol. 7995, Berlin, Heidelberg: Springer Berlin Heidelberg; 2013, p. 376–84. https://doi.org/10.1007/978-3-642-39479-9_45.
[32] Shim J, Kang S, Cho S. Active Learning of Convolutional Neural Network for Cost-Effective Wafer Map Pattern Classification. IEEE Transactions on Semiconductor Manufacturing 2020;33:258–66.

https://doi.org/10.1109/TSM.2020.2974867.
[33] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. IEEE Proceedings, vol. 86, 1989.
[34] Alawieh MB, Boning D, Pan DZ. Wafer Map Defect Patterns Classification using Deep Selective Learning. 2020 57th ACM/IEEE Design Automation Conference (DAC), 2020, p. 1–6. https://doi.org/10.1109/DAC18072.2020.9218580.
[35] Yuan-Fu Y, Min S. Double Feature Extraction Method for Wafer Map Classification Based on Convolution Neural Network. 2020 31st Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC), 2020, p. 1–6. https://doi.org/10.1109/ASMC49169.2020.9185393.
[36] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. Thirty-first AAAI conference on artificial intelligence, 2017.
[37] Saqlain M, Abbas Q, Lee JY. A Deep Convolutional Neural Network for Wafer Defect Identification on an Imbalanced Dataset in Semiconductor Manufacturing Processes. IEEE Trans Semicond Manufact 2020;33:436–44. https://doi.org/10.1109/TSM.2020.2994357.
[38] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. 32nd International Conference on Machine Learning, ICML 2015, 2015.
[39] Glorot X, Bordes A, Bengio Y. Deep Sparse Rectifier Neural Networks. vol. 15. 2010.
[40] Chen S, Zhang Y, Yi M, Ma J, Hou X. Wafer maps defect recognition based on transfer learning of handwritten pre-training network. 2021 International Symposium on Computer Technology and Information Science (ISCTIS), 2021, p. 280–3. https://doi.org/10.1109/ISCTIS51085.2021.00064.
[41] Li Y, Wang J. A Defect Detection Method Based on Improved Mask R-CNN for Wafer Maps. 2021 International Conference on Computer Network, Electronic and Automation (ICCNEA), 2021, p. 133–7. https://doi.org/10.1109/ICCNEA53019.2021.00038.
[42] He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. Proceedings of the IEEE international conference on computer vision, 2017, p. 2961–9.
[43] Geng H, Yang F, Zeng X, Yu B. When Wafer Failure Pattern Classification Meets Few-shot Learning and Self-Supervised Learning. 2021 IEEE/ACM International Conference On Computer Aided Design (ICCAD), 2021, p. 1–8. https://doi.org/10.1109/ICCAD51958.2021.9643518.
[44] Kim Y, Cho D, Lee J-H. Wafer defect pattern classification with detecting out-of-distribution. Microelectronics Reliability 2021;122:114157. https://doi.org/10.1016/j.microrel.2021.114157.
[45] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. ArXiv:180104381 [Cs] 2019.
[46] Zhang X, Zhou X, Lin M, Sun J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, p. 6848–56. https://doi.org/10.1109/CVPR.2018.00716.
[47] Wang S, Zhong Z, Zhao Y, Zuo L. A Variational Autoencoder Enhanced Deep Learning Model for Wafer Defect Imbalanced Classification. IEEE Trans Compon, Packag Manufact Technol 2021;11:2055–60. https://doi.org/10.1109/TCPMT.2021.3126083.
[48] Yuan-Fu Y. A Deep Learning Model for Identification of Defect Patterns in Semiconductor Wafer Map. 2019 30th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC), 2019, p. 1–6. https://doi.org/10.1109/ASMC.2019.8791815.
[49] Yu J, Zheng X, Liu J. Stacked convolutional sparse denoising auto-encoder for identification of defect patterns in semiconductor wafer map. Computers in Industry 2019;109:121–33. https://doi.org/10.1016/j.compind.2019.04.015.
[50] Jin CH, Kim H-J, Piao Y, Li M, Piao M. Wafer map defect pattern classification based on convolutional neural network features and error-correcting output codes. Journal of Intelligent Manufacturing 2020;31:1861–75. https://doi.org/10.1007/s10845-020-01540-x.
[51] Yu J, Liu J. Two-Dimensional Principal Component Analysis-Based Convolutional Autoencoder for Wafer Map Defect Detection. IEEE Transactions on Industrial Electronics 2021;68:8789–97. https://doi.org/10.1109/TIE.2020.3013492.
[52] Kang H, Kang S. A stacking ensemble classifier with handcrafted and convolutional features for wafer map pattern classification. Computers in Industry 2021;129:103450. https://doi.org/10.1016/j.compind.2021.103450.

[53] Rokach L. Ensemble-based classifiers. Artif Intell Rev 2010;33:1–39. https://doi.org/10.1007/s10462-009-9124-7.
[54] Probst P, Boulesteix A-L, Bischl B. Tunability: Importance of Hyperparameters of Machine Learning Algorithms. J Mach Learn Res 2019;20:1934–65.
[55] Feurer M, Hutter F. Hyperparameter Optimization. In: Hutter F, Kotthoff L, Vanschoren J, editors. Automated Machine Learning: Methods, Systems, Challenges, Cham: Springer International Publishing; 2019, p. 3–33. https://doi.org/10.1007/978-3-030-05318-5_1.