Journal of Intelligent Manufacturing (2025) 36:2163–2175
https://doi.org/10.1007/s10845-024-02351-0

CowSSL: contrastive open-world semi-supervised learning for wafer bin map

Insung Baek[1] · **Sung Jin Hwang**[2] · **Seoung Bum Kim**[1]

Received: 21 June 2023 / Accepted: 19 February 2024 / Published online: 29 March 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
In the semiconductor industry, wafer bin maps (WBMs) refer to image data that reveal the defect of each chip positioned on that wafer. The WBMs provide crucial information that can facilitate the identification of underlying causes for any defects present on a wafer. With the advent of artificial intelligence (AI), a significant amount of research has been conducted leveraging machine learning and deep learning techniques to automatically classify wafer bin map defects. Although there have been various attempts to enhance performance by using both unlabeled and labeled data, current research is constrained by its narrow focus on improving the detection of known defect patterns. However, in the real world, multiple novel patterns frequently arise that have not been previously encountered. Hence, AI models must exhibit the capacity to detect not only existing known defect patterns but also newly emerging defect patterns, while ensuring effective classification of these new patterns among themselves. In this study, we propose the contrastive open-world semi-supervised learning that can classify multiple novel patterns in WBMs simultaneously. We introduce a contrastive loss function to address the challenges associated with the existence of significantly fewer new defect patterns than existing patterns in the WBM problem. We confirm that the proposed methodology effectively detects and classifies diverse new patterns separately in real-world open data, WM-811 K. Moreover, we demonstrate that the proposed method outperforms other existing open-world semi-supervised learning in WBM classification.

Keywords Semiconductor manufacturing · Defect patterns classification · Wafer bin map · Open-world recognition · Semi-supervised learning · Contrastive learning

Introduction

In the semiconductor manufacturing industry, a wafer is a thin disk made of silicon and includes hundreds of semiconductor chips. In the wafer fabrication process, attaining optimal yield and quality requires minimizing the occurrence of defects, rapidly identifying the causative factors, and implementing corrective measures when defects arise (Choudhary et al., 2009; Jang et al., 2020; Jang & Lee, 2023; Kang, 2020). Thus, engineers conduct an array of inspections to assess the condition of the wafers and determine whether any defects are present. For example, electrical die sorting (EDS) serves as a method of verifying the proper functioning of each chip on a wafer, yielding wafer bin map (WBM) data. A WBM constitutes a format of image data that illustrates the position of each chip and designates its defective or non-defective status, as shown in Fig. 1. WBM patterns exhibit distinct variations depending on the underlying causes of the failure, serving as crucial indicators for estimating the origin of the malfunction (Jin et al., 2020; Kong & Ni, 2021). Hence, the classification of failure patterns in WBMs represents an essential process that serves as the foundation of yield quality management. Indeed, the manual classification of each

Insung Baek and Sung Jin Hwang have contributed equally to this work.

✉ Seoung Bum Kim
sbkim1@korea.ac.kr

Insung Baek
insung_baek01@korea.ac.kr

Sung Jin Hwang
sj90.hwang@samsung.com

[1] School of Industrial and Management Engineering, Korea University, Seoul 02841, Republic of Korea

[2] Memory Defect Science & Engineering Group, Samsung Electronics, Hwaseong 18448, Republic of Korea

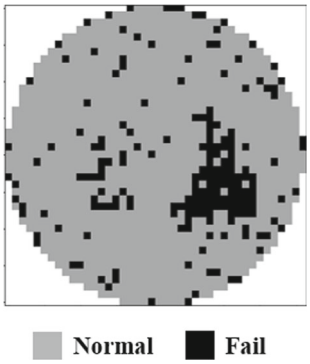

Fig. 1 Example of a wafer bin map

WBM, generated at high volumes, is a labor-intensive and costly endeavor for engineers, necessitating the development of a model capable of performing automated classification.

However, building an automated model to classify wafer defects poses a significant challenge because of the following characteristics of WBM. First, the scarcity of labeled data for model training poses a significant challenge (Kahng & Kim, 2021). Obtaining labels is costly and time-consuming because WBM generated in real industries necessitates the involvement of numerous specialized engineers to perform manual inspection and assign accurate labels. Moreover, maintaining a continuous labeling process for the vast amount of newly generated WBMs is nearly impossible because of accumulating tens of thousands of WBMs. Therefore, leveraging not only a limited pool of labeled data but also a considerable amount of unlabeled data becomes necessary. Second, actual WBMs have the potential to produce novel failure patterns that have not been previously identified (Jang et al., 2020; Jang & Lee, 2023; Kong & Ni, 2021). Engineers can promptly take appropriate measures when a previously known defect pattern emerges because of their accumulated experience in identifying the root cause of the defect and taking the appropriate actions to address it. However, when a novel defect pattern emerges, its root cause may not be immediately apparent. In such cases, tracing the cause of the defect is essential to mitigate the risk of potential quality incidents. Hence, developing a model capable of not only classifying known patterns but also detecting diverse emerging new defect patterns is necessary. Finally, although a new defect pattern may be identified, various patterns may exist in this new defect pattern. In the semiconductor industry, identifying and categorizing new defect patterns is crucial because engineers group wafers with the same new pattern to find process paths that cause defects. By tracing the root cause of the defect and implementing appropriate corrective measures when a defect pattern is clearly identified, it is possible to minimize the incidence of defects and prevent potential accidents from occurring. Consequently, an automated model for wafer defect classification must be capable of detecting diverse new patterns and classifying them into the appropriate categories. As shown in Fig. 2, developing a predictive model that can classify novel patterns using a significant amount of unlabeled data is essential to achieving this goal.

However, existing research on classifying WBM defect patterns has been limited to simply classifying defect patterns or detecting new ones (Hsu & Chien, 2022; Jang et al., 2020; Jang & Lee, 2023; Kahng & Kim, 2021; Kong & Ni, 2021). Therefore, in this study, we use open-world semi-supervised learning that can detect and classify new patterns by utilizing a substantial quantity of unlabeled data (Cao et al., 2021; Rizve et al., 2022a, 2022b). Open-world semi-supervised learning represents an approach that amalgamates open-world recognition and semi-supervised learning. Open-world recognition allows models to simultaneously detect and classify multiple unknown classes (Bendale & Boult, 2015; Parmar et al., 2023). This means that when a new defect pattern occurs in the WBM, it can be considered as multiple unknown classes. Semi-supervised learning uses both labeled and unlabeled data when training a model to improve prediction performance (Chapelle et al., 2009; Van Engelen & Hoos, 2020). This methodology can better reflect the characteristics of WBM data, in which a small number of labeled data and a large number of unlabeled data are collected.

In this research, we proposed an approach called contrastive open-world semi-supervised learning (CowSSL) for defect classification of WBMs. We built upon the work of Cao et al. (2021), which used three different loss functions to train a model on an image dataset containing labeled and unlabeled data classified as new classes. Nonetheless, the previous study exhibits limitations in demonstrating its performance for general image data where the number of data per class maintains balance. Our approach aims to address the challenges of extreme class imbalance in the semiconductor field, whereby the number of known pattern classes is overwhelmingly large compared to the number of new patterns that have not been identified. This often leads to overfitting to the overwhelmingly known classes, significantly reducing the performance of new pattern detection and classification accuracy. To overcome this challenge, we introduce a contrastive loss function to enhance the ability of the model to separate new patterns from existing patterns and distinguish among different new patterns (Chen et al., 2020a, 2020b; He et al., 2020). As a result, we propose a methodology that is capable of detecting and classifying new patterns, even when the number of new patterns is comparably small. The main contributions of this study can be summarized as follows:

- The proposed CowSSL method is designed based on the momentum contrastive learning (MoCo) algorithm (Chen et al., 2020a, 2020b; He et al., 2020), which uses negative

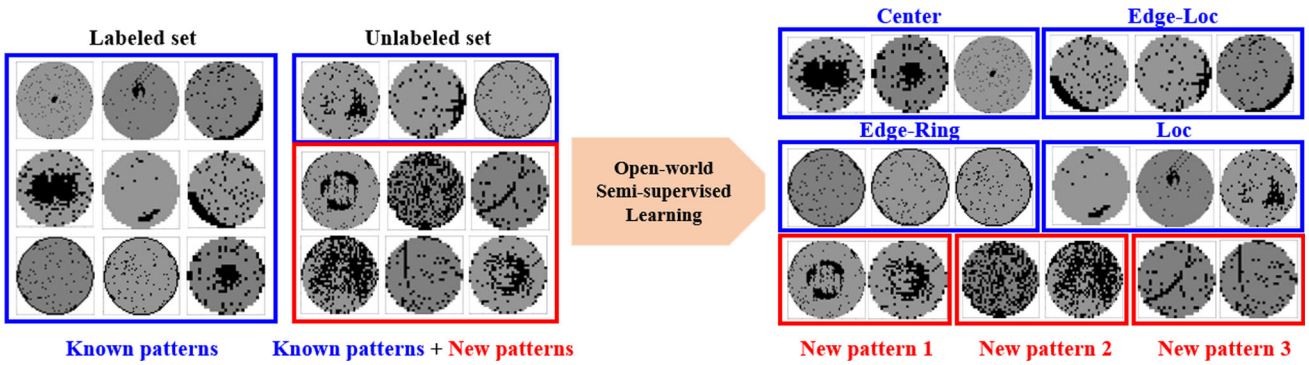

Fig. 2 Schematization of open-world semi-supervised learning

samples to enhance the detection of very few novel patterns in the WBMs. It has been demonstrated that known and new patterns can be efficiently segregated because the probability of selecting a small number of novel patterns as negative samples is minimal.

- The effectiveness of the proposed CowSSL loss function that integrates the contrastive loss function into the open-world semi-supervised loss function is validated in the classification of novel patterns corresponding to few data. The proposed CowSSL loss function enhances the classification performance of novel patterns in WBMs by clustering new patterns with similar characteristics in the embedding space, while simultaneously separating other new patterns exhibiting dissimilar characteristics.

The reminder of the paper is organized as follows. "Related works" Section consists of a review of the studies on WBM classification and open-world semi-supervised learning. "Methodology" Section illustrates the details of the proposed CowSSL method. "Experiments and results" Section presents the qualitative and quantitative experimental results. Finally, "Results" Section contains our concluding remarks and directions for future research.

Related works

Wafer bin map classification

The WBM classification problem holds significant importance in the semiconductor industry and has been widely studied using a variety of algorithms. Recently, the evolution of machine learning has led to the exploration of this problem using several models with remarkable classification performance. Wu et al. (2015) released the wafer bin map data "WM-811K," publicly available. The authors employed a support vector machine algorithm to analyze this dataset. Piao et al. (2018) extracted wafer features using a radon transform and used a multiple decision tree ensemble method for classification. Nonetheless, these techniques suffer from the limitation of reduced classification accuracy with increasing intricacy of the wafer pattern shape. Recently, there has been significant progress in the utilization of convolutional neural networks (CNN) to analyze WBM image data. Lee et al. (2017) introduced an FDC-CNN model aimed at classifying one normal class and five fault classes using wafer data collected from an on-site chemical vapor deposition (CVD) tool. The model achieved an impressive accuracy of 97.9%. Nakazawa and Kulkarni generated a set of 22 defect classes using synthetic wafers and subsequently trained a CNN model using 15,400 samples as the training data. The model achieved a high classification accuracy of 98.2%. Kyeong and Kim (2018) proposed an approach for defect classification that considers multiple patterns on a wafer, as opposed to traditional methods that focus on a single pattern. However, the aforementioned studies are based on supervised learning, which requires a considerable amount of labeled data for training. However, obtaining adequate labeled data in the semiconductor industry requires significant resources. Therefore, there is a pressing need for a methodology that can effectively utilize the abundant amounts of unlabeled data present for the classification of WBM failure patterns.

Shim et al. (2020) demonstrated that high classification performance can be achieved with a limited amount of labeled data by utilizing uncertainty-based sample selection techniques that prioritize samples having the greatest impact on performance. Kahng and Kim (2021) proposed a self-supervised learning model involving CNNs and pretext-invariant representation learning to effectively classify WBM defect patterns. Although these studies have the advantage of utilizing unlabeled data, they do not consider the possibility of new patterns that may emerge in practice. Jang et al. (2020) introduced the support weighted ensemble model for detecting a new pattern. Jang and Lee (2023) proposed an open-set recognition model that employs probability scores derived from the calculated reconstruction errors and random network errors. Although these studies have achieved success

in detecting new patterns in WBM, they fall short in classifying multiple new patterns simultaneously. Jin et al. (2019) introduced a pattern clustering method for WBMs by using DBSCAN (density based spatial clustering of applications with noise). Park et al. (2021) used a classification technique that involves label reassignment based on a Siamese network approach. Although pattern clustering and label reassignment techniques have demonstrated the potential for identifying new patterns, there is a dearth of open-world learning studies that cater to the specific characteristics of WBM, such as similarity of features across classes and smaller amounts of new pattern data. In this study, we proposed the CowSSL method, which effectively addresses the challenges posed by similarity among WBM classes and a limited number of new pattern data, enabling the accurate classification of multiple new patterns.

Open-world semi-supervised learning

The present study addresses a multifaceted problem, aiming to develop a model that not only accurately classifies WBM data but also leverages unlabeled data, effectively detects novel patterns, and distinguishes between these patterns across multiple classes. Achieving these objectives requires the integration of various research areas in machine learning. Machine learning methodologies that leverage both labeled and unlabeled data include semi-supervised learning, open-set recognition for detecting new patterns, and open-world recognition for classifying each class within new patterns. Traditional semi-supervised learning studies have conducted under the assumption that labeled and unlabeled data share an equal number of classes (Chapelle et al., 2009). However, this assumption is frequently violated in practice, as unlabeled data may include new classes heretofore not considered. Scheirer et al. (2012) proposed an open-set recognition approach that assigns samples with distinct features from the recognized classes to novel classes. Although such an approach is capable of detecting new patterns, it is not equipped to perform classification between these new classes. Bendale et al. (2015) introduced the notion of open-world recognition that addresses these limitations by enabling new classes to be identified through human intervention, empowering the model to continually learn new classes. However, these methods ultimately necessitate human intervention.

Cao et al. (2021) presented the ORCA (open-world with uncertainty-based adaptive margin) algorithm that integrates three distinct loss functions to effectively discriminate between multiple novel patterns in unlabeled data. Following pre-training with SimCLR (simple framework for contrastive learning of visual representations), the model was trained using a loss function that combined three distinct loss functions, exhibiting excellent performance across a diverse range of datasets, including MNIST, CIFAR10, CIFAR100, and ImageNet. The ORCA algorithm is based on an open-world semi-supervised learning approach and is noteworthy for its capability to identify and differentiate new classes exclusively through model learning, without the need for human intervention. Rizve et al. (2022a, 2022b) noted that the OpenLDN (learning to discover novel classes for open-world semi-supervised learning) method, which involves pre-training with ORCA, is computationally demanding. They proposed an enhancement that incorporates pairwise similarity loss and cross-entropy loss with pseudo-labeling, resulting in improved performance and reduced computational requirements. In the same year, Rizve et al. (2022a, 2022b) proposed TRSSL (toward realistic semi-supervised learning), which employed pseudo-labeling for training purposes and optimized the Sinkhorn–Knopp algorithm to enhance the probability distribution between classes. They also presented a method for scaling the uncertainty of the results as the model underwent repeated training to achieve superior performance. These studies all exhibited strong performance on conventional image datasets, such as MNIST, CIFAR10, CIFAR100, and ImageNet. However, unlike conventional image data, WBMs exhibit a considerably lower proportion of data points as unknown classes relative to known classes, in addition to the relatively comparable patterns across classes. Thus, in this study, we develop an approach that merges open-world semi-supervised learning with contrastive learning to effectively discriminate a limited number of novel patterns at once, while simultaneously accounting for the unique attributes of WBMs.

Methodology

Data preprocessing

The dataset we used for our experimentation, WM-811K, is publicly available and encompasses 811,457 WBMs from the real-world. The WM-811K dataset, introduced by Wu et al. (2015) is renowned for its reliability because it originates from real-world fabrication processes and has been meticulously labeled by domain experts. This inspection and labeling process is shown in Fig. 3. The WM-811K dataset has been widely used in many studies (Hsu & Chien, 2022; Kahng & Kim, 2021; Piao et al., 2018; Wu et al., 2015). Among these, 172,950 wafers were annotated according to defect pattern and were leveraged for experimental validation. Furthermore, we omitted the none class, which does not have any distinct pattern, from our study. The dataset comprised eight types of patterns including center, edge-loc, edge-ring, loc, scratch, random, donut, and near-full. WBM data is classified as either ‘normal’ or ‘fail based on the inspection of multiple chips on a single wafer. Each chip

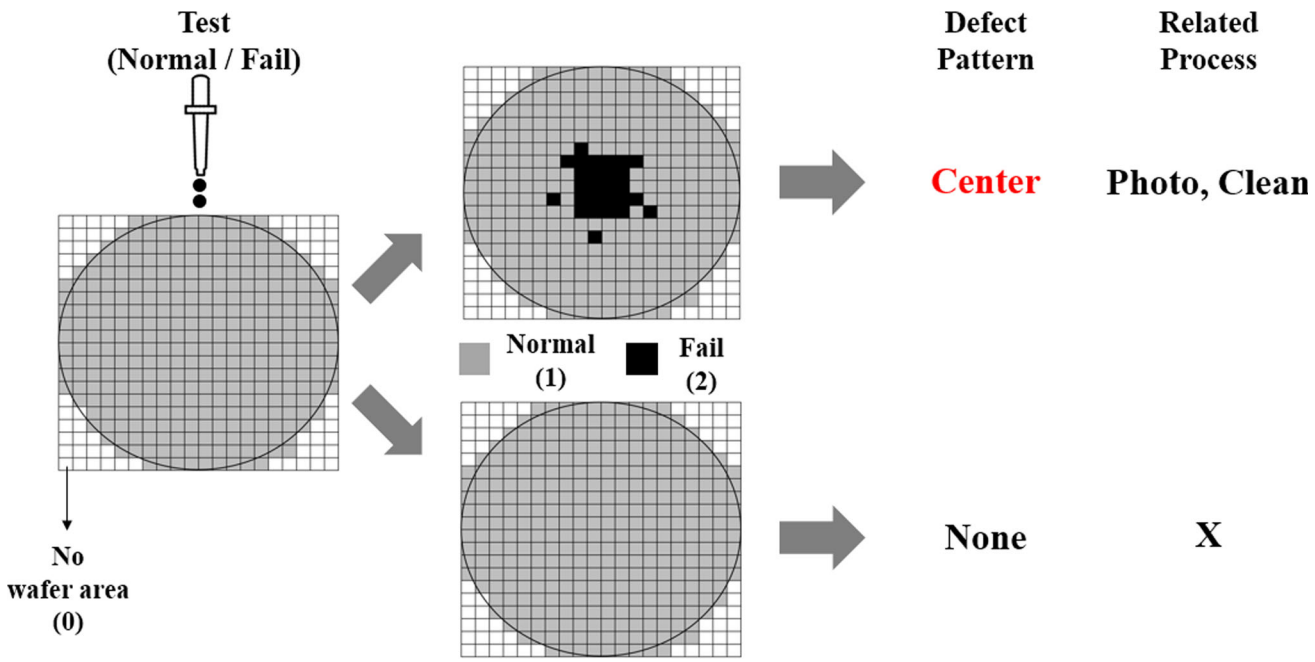

Fig. 3 Test process to determine WBM defect pattern

is assigned a value of 1 for 'normal' or 2 for 'fail' results. Throughout this process, the defect patterns are categorized based on how the 'fail' chips cluster together. These defective patterns are classified by domain experts according to their shape. For example, a 'center' pattern indicates that defective chips are concentrated in the middle of the WBM, forming a central shape. Conversely, an 'edge-ring' pattern describes a form where bad chips are arranged in the shape of a ring around the outer edge of the WBM. Given that the near-full pattern is nearly identical to the random pattern, it is deemed unnecessary and subsequently removed in this study. Ultimately, as illustrated in Fig. 4, known patterns encompassed four classes, and new patterns composed three classes, namely, scratch, random, and donut. Among various combinations of selecting known and unknown patters, our choice is based on industry practices in the semiconductor field. Given the ample availability of known pattern data, we selected a sizable sample of edge-ring, edge-loc, center, and loc patterns. In contrast, because of the small amount of data for the new patterns, we selected smaller sample sizes for the scratch, random, and donut classes. To standardize the data, we initially resized each wafer to a consistent size of 32×32. Because the pixel value represents categorical data that is categorized as 0 (no wafer area), 1 (good chip), and 2 (bad chip), we converted it into three channels through one-hot encoding. Finally, we formed a $3 \times 32 \times 32$ dimensional image serving as input data.

We partitioned the total WM-811 K dataset into training and test sets, with 90% of the data used for training and the remaining 10% for testing. We further divided the training data into labeled and unlabeled sets, with only 20% of the data labeled and the remaining 80% characterized as unlabeled data by masking the original labeling data. Table 1 shows the number of labeled, unlabeled, and test data records per class. As shown in Table 1, the number of samples in the three new classes, scratch, random, and donut, is smaller than that of the known classes. Finally, we experimentally verified that the proposed methodology can accurately detect and classify new patterns in a situation in which the amount of unlabeled data is relatively larger than the amount of labeled data and the number of new patterns is smaller than the number of known patterns.

CowSSL

Before the training process, we conducted learning with a pre-trained model using the SimCLR algorithm to learn the underlying features of the data (Chen et al., 2020a, 2020b). For this purpose, we utilized the entire training dataset as input data for the pre-trained model, without the presence of any labels. The SimCLR model was designed to learn from the similarity score of input pairs of samples in which high similarity between the two samples indicates that the presence of similar features. When training with a pre-trained model, generating improved representations becomes possible. This not only improves the models in classifying known WBM patterns but also helps them perform better in identifying new WBM patterns.

The designed model in our study learns by utilizing both labeled and unlabeled data, building on a pre-trained

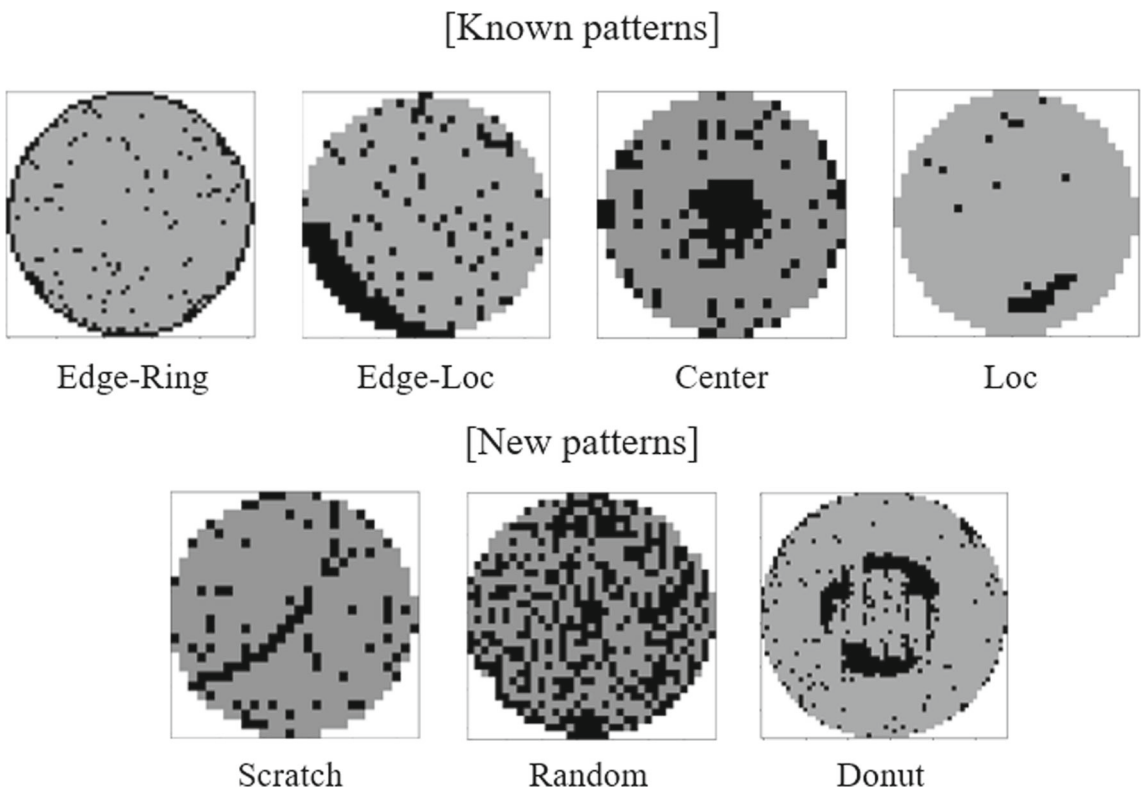

Fig. 4 Examples of known and new patterns of WBM data in our study

Table 1 Description of the dataset. New patterns are not included in the labeled data

Known/new	Patterns (classes)	Total	Labeled	Unlabeled	Test
Known patterns	Edge-Ring	9680	1746	6980	954
	Edge-Loc	5189	938	3740	511
	Center	4294	771	3100	423
	Loc	3593	649	2590	354
New patterns	Scratch	1193	0	1076	117
	Random	866		781	85
	Donut	555		500	55
Total		25,370	4104	18,767	2499

model using SimCLR. In addition, we designed a model that uses a ResNet-18 structure and employs a key encoder to apply a contrastive loss function based on the MoCo algorithm (Chen et al., 2020a, 2020b; He et al., 2020). Existing open-world semi-supervised learning algorithms, such as ORCA, OpenLDN, and TRSSL, used the ResNet-18 model, except for the ImageNet-100 dataset. To ensure an equitable comparison with these existing algorithms, we also used the ResNet-18 model. Figure 5 shows an overview of the proposed CowSSL algorithm which was learned via the weighted sum of four loss functions. As shown in Fig. 5, SimCLR pretrain was performed excluding labels from all data (labeled dataset and unlabeled dataset) to learn image features. Subsequently, we trained a query encoder based on the ResNet-18 model using the initial weight values from the pre-trained model. During this phase, we used the ORCA methodology, which can leverage both labeled and unlabeled datasets. Additionally, to improve the classification of new defective patterns, we used a key encoder that uses negative samples. During the learning process, the weight of the key encoder was updated at each iteration using the momentum update method, which combined the weight of the query encoder and the key encoder. The key encoder was exclusively used for extracting feature vectors from negative samples during model training and was not used during the actual testing stage. Below, we explained a more detailed explanation of the four losses used in CowSSL. Three of the proposed loss functions came from the ORCA

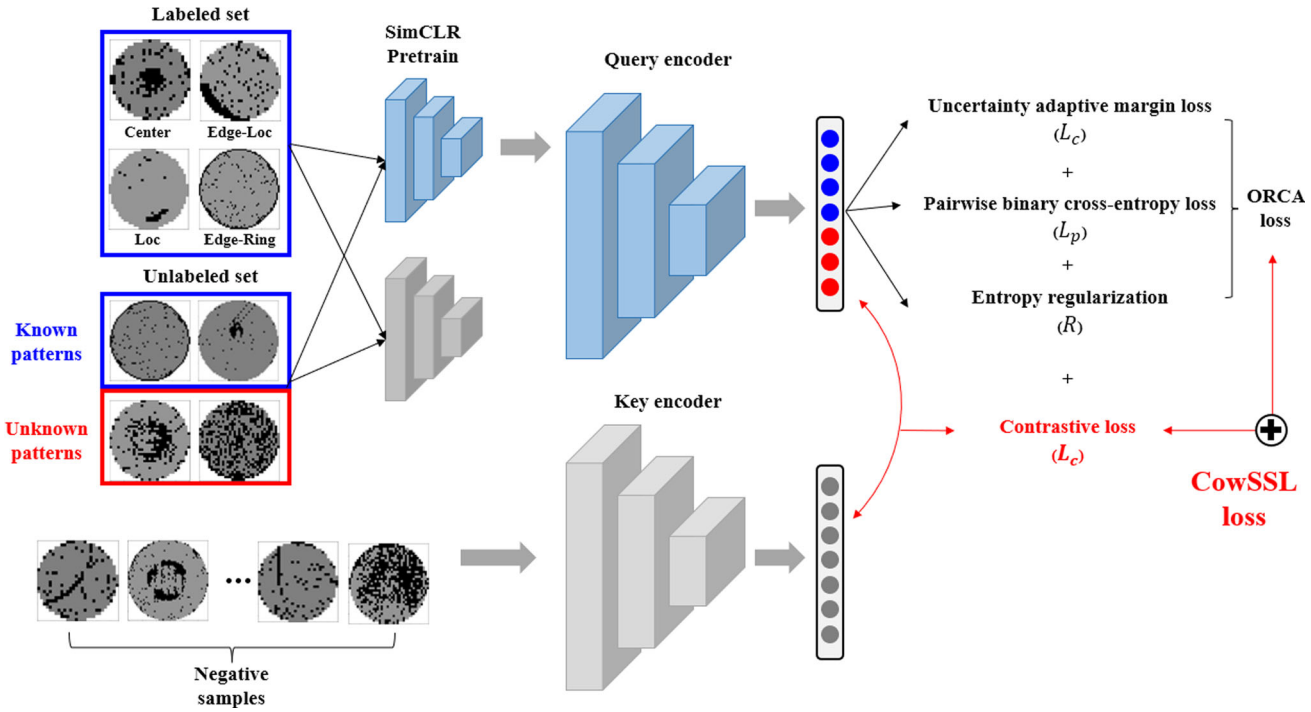

Fig. 5 Overview of the CowSSL algorithm. (SimCLR pretraining is conducted on all datasets without the use of labels. Subsequently, a query encoder based on the ORCA methodology is trained using the pre-trained weights as initial values. Furthermore, to enhance the performance of new pattern classification, a key encoder is used with negative samples. Consequently, CowSSL is trained using a total of four loss functions)

method and included the following (Cao et al., 2021): (1) The supervised objective with uncertainty adaptive margin loss function (L_s) used the uncertainty of the model probability values within the cross-entropy loss to distinguish new patterns from known patterns. (2) The pairwise binary cross-entropy (BCE) loss function (L_p) grouped similar patterns by facilitating the bringing together of the two pairs with the closest cosine similarity in the batch. (3) Entropy regularization (R) ensured that the model did not overfit to any specific class. In the ORCA research, just these three loss functions were satisfactory at classifying new patterns but only for regular image data in which the number of classes was the same for all classes and there were clear differences between the features of each class. However, for WBM semiconductor data, the performance in classifying new patterns was not satisfactory because of the large number of known patterns and relatively few new patterns. Therefore, we designed the CowSSL loss function, adding a contrastive loss function (α_c) to the ORCA loss function, which improves the ability to distinguish a relatively small number of new patterns from the existing known patterns. In this study, we used a contrastive loss function that enables contrastive learning by making a different negative sample from the input data (Chen et al., 2020a, 2020b; He et al., 2020). Finally, we proposed the CowSSL loss function as a mix of four loss functions: three ORCA loss functions and one contrastive loss function, as shown in Eq. (1). α_s, α_p, α_r and α_c are hyperparameters. By optimizing the model using this loss function, both known and new patterns in WBMs can be classified at once.

$$L = \alpha_s L_s + \alpha_p L_p + \alpha_r R + \alpha_c L_c. \tag{1}$$

Each component of the proposed loss function is described in detail below. First, the uncertainty adaptive margin loss function (L_s) considers the uncertainty adaptive margin calculated from the model's output probability value within the general cross-entropy loss, which is used to learn labeled data with known patterns. This loss function aids in distinguishing new patterns from known patterns by using an adaptive margin. In Eq. (3), $\bar{\mu}$ is calculated by taking the average of the largest probability values output by the model, subtracted by one. The larger the uncertainty value is, the more likely the model is to predict multiple patterns with equal probabilities, rather than a single pattern with high probability. Ultimately, when substituting Eq. (3) into Eq. (2), if the model predicts a new pattern that differs from the known pattern, the probability values become more evenly distributed, causing the uncertainty to increase accordingly and enabling the identification of the new pattern. In Eq. (2), y_i represents

the labeled data and W denotes optimized weights. Furthermore, the feature representation vector is denoted by z_i, λ is a hyperparameter, and s means a scaling parameter.

$$L_s = \frac{1}{n} \sum_{z_i \in Z_l} -log \frac{e^{s(W_{y_i}^T \cdot z_i + \lambda \bar{\mu})}}{e^{s(W_{y_i}^T \cdot z_i + \lambda \bar{\mu})} + \sum_{j \neq i} e^{s W_{y_i}^T \cdot z_i}}, \tag{2}$$

$$\bar{\mu} = \frac{1}{|D_u|} \sum_{x_i \in D_u} 1 - \max_k (Pr(Y = k | X = x_i)). \tag{3}$$

The second pairwise BCE loss function uses both labeled and unlabeled data, selecting the pair of samples in the batch with the highest cosine similarity in their representations. The goal is to minimize the BCE loss function for these samples, encouraging the model to classify them as the same pattern. This implies that the representation encourages samples with the same pattern to be clustered together through the clustering of close neighbors. For the unlabeled data, the loss function is calculated based on the nearest sample. However, for labeled data, because there is a ground truth, the loss function is computed by comparing it to the ground truth, rather than comparing it to a separate nearby sample. In conclusion, the pairwise loss function, as defined in Eq. (4), facilitates the clustering of different patterns together. In Eq. (5), Z_l indicates the feature representations in the labeled dataset, Z_u denotes the feature representations in the unlabeled dataset, m denotes the number of labeled data, and n indicates the number of unlabeled data.

$$L_p = \frac{1}{m+n} \sum_{\substack{z_i,\, z_i' \in \\ (Z_l \cup Z_u,\, Z_l' \cup Z_u')}} -log < \sigma\left(W^T \cdot z_i\right), \sigma\left(W^T \cdot z_i'\right) > . \tag{4}$$

The third regularization term, as given in Eq. (5), serves as a regulatory measure to prevent the model from overfitting and classifying everything into a single class during training. This calculation employs the Kullback–Leibler (KL) divergence to ensure that the model's predicted probability distribution remains similar to the prior probability distribution.

$$R = KL\left(\frac{1}{m+n} \sum_{z_i \in (Z_l \cup Z_u)} \sigma\left(W^T \cdot z_i\right) \| P(\bar{y})\right). \tag{5}$$

Finally, we applied a contrast loss function based on the MoCo algorithm to improve the model's ability to discriminate a small number of new patterns from many known patterns and to classify new patterns (Chen et al., 2020a, 2020b; He et al., 2020). The MoCo algorithm is one of the self-supervised learning methods that uses a contrastive loss with negative samples. This approach facilitates the learning of the features of input data without relying on labels. As the learning progresses, samples with similar features in the feature space tend to be together, while samples with dissimilar features tend to move farther apart. As shown in Fig. 6, the contrastive loss function performs contrastive learning by employing a query encoder and a key encoder. In this process, data that has undergone two augmentations on the same sample is placed into positive pairs. One of these pairs is input to the query encoder to create a query vector, and the other is fed into the key encoder to generate a positive key. In addition, some of the remaining samples are chosen as negative samples, and their representations are computed using the key encoder. These representations are then stored as a dictionary of negative keys. The query and positive keys are trained to become closer together in the feature space, whereas the query and negative keys are trained to move further apart. An important aspect to note here is that the negative sample selection comes from all other samples, excluding the sample itself, which has undergone different augmentations. Consequently, data from a new class containing a small number of samples are less likely to be selected as a negative sample, whereas data from a known class containing a large number of samples are more likely to be chosen as a negative sample. Therefore, the majority class (known class) is more likely to be selected as a negative sample and is further distanced from the minority class (new class), thereby increasing the probability of separation between the two classes. Additionally, classification performance among minority classes improves because a small number of new classes exhibit a higher likelihood of crowding out other new classes as negative samples. Consequently, the contrastive loss function proves highly effective in separating new classes from known classes in WBM data and classifying new patterns. In Eq. (6), q represents a query representation, k^+ denotes the representation of the positive key, and k^- refers to the representation of the negative keys. τ is a temperature hyperparameter.

$$L_c = -log \frac{\exp(q \cdot k^+/\tau)}{\exp(q \cdot k^+/\tau) + \sum_{k^-} \exp(q \cdot k^-/\tau)}. \tag{6}$$

Experiments and results

Experimental setting

To optimize the model, we used stochastic gradient descent with momentum as the algorithm, with the following hyperparameters: learning rate at 0.005, momentum at 0.9, and weight decay. The batch size for the input data was 512, and we used a key dictionary size of 2,048 for MoCo. We set α_s, α_p, and α_r are 1.0 like the ORCA paper. α_c set to

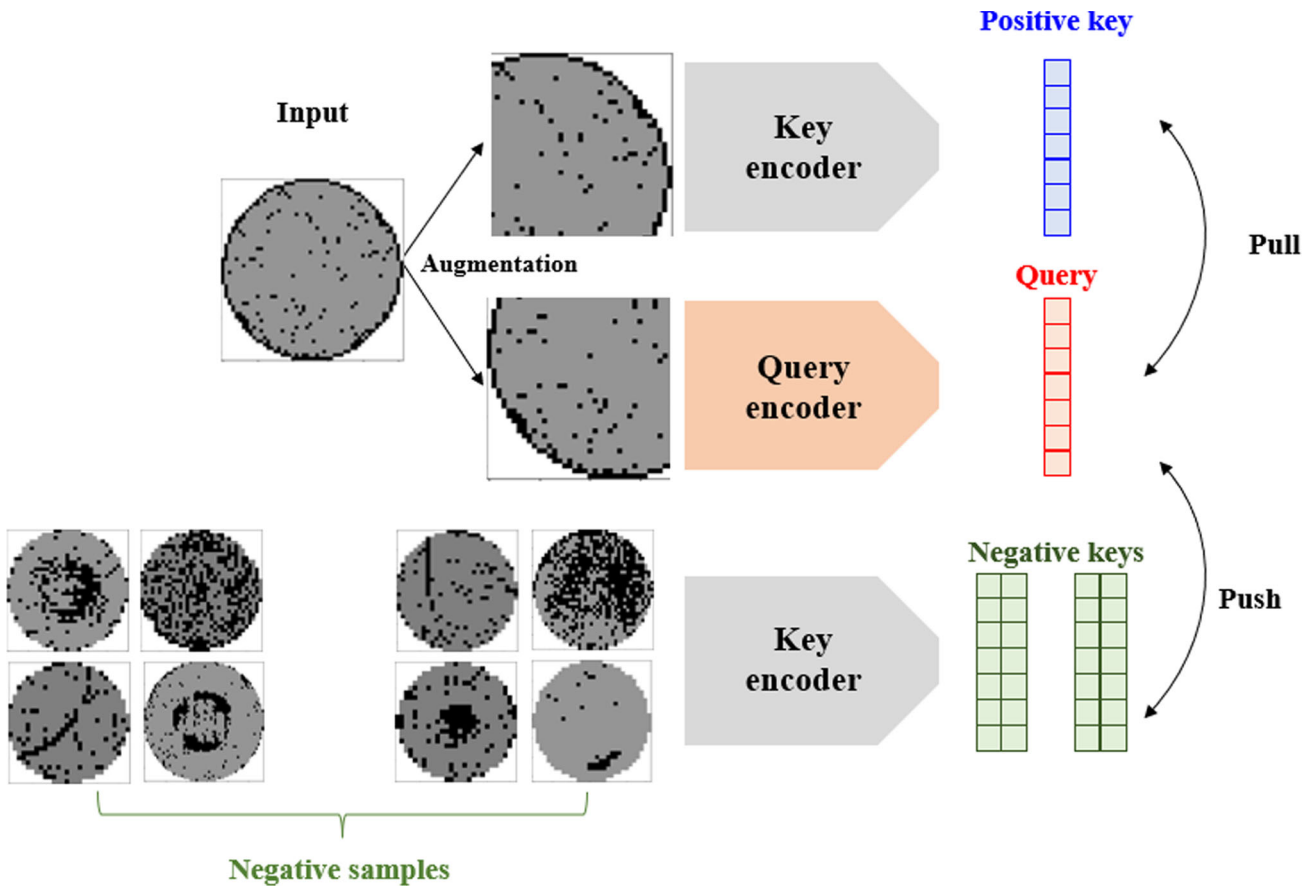

Fig. 6 Structure of MoCo using WBM data

Table 2 The confusion matrix for calculating evaluation indicators

		Ground truth	
		True	False
Predicted	Positive	True positives	False positives
	Negative	False negatives	True negatives

0.7 in first 100 epochs and 0.3 in last 50 epochs. We ran a total of 150 epochs, and all experiments were repeated ten times with different randomization settings for each sampling strategy, with the average value used. Our experiments were implemented with PyTorch 1.8.0 and conducted on a single NVIDIA GeForce RTX 3090 GPU and an i-9 10900KF CPU. We have mainly used the F1-score for comparison because the 'WM-811K' dataset used in our study exhibits significant data imbalance. The most prevalent pattern, 'edge-ring,' representing 38.1% of the entire dataset, while the least prevalent, 'donut,' constitutes only 2.2%. Therefore, in the context of this data imbalance, performance evaluation is conducted using the F1-score, as it takes into account both recall and precision. Table 2 shows a standard confusion matrix. The F1-score is calculated as the harmonic mean of precision and recall, as shown in Eq. (9), with a value between 0 and 1. The higher F1-score correspond to better classification performance of the model. We optimized the process of fitting the final clustered class of the model against the correct answer using the Hungarian algorithm.

$$\text{Precision} = \frac{True\ Positives}{True\ Positives + False\ Positives}, \tag{7}$$

$$\text{Recall} = \frac{True\ Positives}{True\ Positives + False\ Negatives}, \tag{8}$$

$$\text{F1 - score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \tag{9}$$

Results

In this section, we discuss the classification performance of both the open-world semi-supervised learning methodology and the proposed CowSSL methodology when applied to existing and new patterns. In conclusion, our proposed

Table 3 Comparison of F1-score between the open-world semi-supervised methodologies

Method	Known patterns (classes)				New patterns (classes)			Total
	Center	Edge-Loc	Edge-Ring	Loc	Scratch	Random	Donut	Average
OpenLDN Rizve et al. (2022a, 2022b)	0.9107 (0.01)	0.8276 (0.03)	**0.9663 (0.01)**	0.7125 (0.01)	0.0016 (0.00)	0.1014 (0.16)	0.0034 (0.01)	0.5034
TRSSL Rizve et al. (2022a, 2022b)	0.8394 (0.04)	0.7892 (0.03)	0.7422 (0.05)	0.6715 (0.02)	0.2397 (0.04)	0.2785 (0.09)	0.0119 (0.03)	0.5103
ORCA Cao et al. (2021)	0.9393 (0.01)	0.8457 (0.01)	0.9427 (0.01)	0.7355 (0.01)	0.2204 (0.25)	0.7934 (0.10)	0.4637 (0.38)	0.7058
CowSSL (proposed)	**0.9397 (0.01)**	**0.8508 (0.01)**	0.9576 (0.01)	**0.7387 (0.01)**	**0.4588 (0.02)**	**0.8057 (0.04)**	**0.6404 (0.12)**	**0.7702**

This scenario consists of four known patterns (center, edge-loc, edge-ring, loc) and three new patterns (scratch, random, donut)
Bold values indicate the highest F1 score among the methods

CowSSL outperformed other methodologies in terms of performance. Table 3 displays the F1-scores for each pattern across different methodologies. Upon examining the overall average values, the traditional ORCA achieved a score of 0.7058, whereas the performance significantly improved to 0.7702 with our proposed method, surpassing both TRSSL and OpenLDN. It is evident that known patterns exhibit a consistent level of performance across all models, whereas new patterns demonstrate substantial divergence. The OpenLDN model predicted the novel pattern as a solitary cluster, whereas TRSSL exhibited poor performance by detecting only two out of the three new patterns. While the existing ORCA seemed to detect three new patterns on average, we observed that it underperformed the proposed CowsSSL and exhibited a very large standard deviation. This implies that the probability of correctly detecting all of the new patterns is highly variable from experiment to experiment. However, the proposed CowSSL method demonstrated excellent performance in correctly classifying all the new patterns. Notably, the CowSSL excelled in accurately identifying the scratch pattern, which is highly similar to the loc pattern and can be challenging to differentiate. Additionally, it exhibited the highest accuracy in classifying the donut pattern, despite its relatively small sample size. Consequently, our findings demonstrate that the contrast loss function integrated into the proposed CowSSL successfully separates the minority patterns from the majority patterns, thereby facilitating a more precise classification of the novel patterns.

We conducted supplementary experiments to validate the suitability of our proposed methodology. For this purpose, we deliberately created a scenario where the known pattern dataset is large, while the new pattern dataset is small. We selected center, edge-loc, and edge-ring patterns because of their relatively abundant data as known patterns, while loc, scratch, random, and donut patterns were chosen as the new patterns. As illustrated in Table 4, the results demonstrated that the proposed CowSSL algorithm exhibits the highest performance. These results were very similar to those in Table 3. It was confirmed that the proposed CowSSL in preparation for OpenLDN, TRSSL, and ORCA classified each of the four new patterns well. Consequently, our findings confirmed that the CowSSL method effectively classifies the small number of new patterns in the WBM dataset. The proposed CowSSL methodology applies contrastive loss to ORCA, an open-world semi-supervised learning algorithm. To ascertain whether the application of contrastive loss improves the classification of new patterns with low quantities in other open-world semi-supervised learning methods, we conducted an experiment in which contrastive loss was applied to OpenLDN. Table 5 presents the performance results of OpenLDN both with and without contrastive loss. Incorporating contrastive loss into the OpenLDN method resulted in an average performance enhancement. Notably, a significant performance improvement was observed in the 'random' class among new patterns. We confirmed the effectiveness of contrastive loss in accurately separating a small number of novel patterns in the WBM. We verified the learning efficiency of the proposed CowSSL by assessing learning times. As presented in Table 6, it is evident that the introduction of contrastive loss to ORCA does not result in a significant change in learning time. Moreover, we observed that the time required for one epoch was shorter compared to OpenLDN and TRSSL.

To better visualize this, we examined the representation of each model. Figure 7 displays a representation of the OpenLDN, TRSSL, ORCA, and CowSSL models through t-SNE (t-distributed stochastic neighbor embedding) analysis. t-SNE is a methodology used to reduce the dimensionality of high-dimensional data to facilitate the visualization. Our objective of using t-SNE visualization was aimed at evaluating the degree to which defect patterns with distinct features could be effectively separated in the feature space. When the representation predicted by the model is drawn for the test

Table 4 Comparison of F1-score between the open-world semi-supervised methodologies

Method	Known patterns (classes)			New patterns (classes)				Total
	Center	Edge-Loc	Edge-Ring	Loc	Scratch	Random	Donut	Average
OpenLDN Rizve et al. (2022a, 2022b)	0.8080 (0.03)	0.6827 (0.01)	**0.9681 (0.01)**	0.0398 (0.03)	0.0098 (0.01)	0.0000 (0.00)	0.0676 (0.12)	0.3680
TRSSL Rizve et al. (2022a, 2022b)	0.7925 (0.01)	0.6877 (0.06)	0.7760 (0.03)	0.3381 (0.03)	0.0549 (0.09)	0.2841 (0.09)	0.0921 (0.11)	0.4322
ORCA Cao et al. (2021)	0.9322 (0.01)	**0.8468 (0.01)**	0.9697 (0.01)	0.4147 (0.03)	0.3303 (0.03)	0.6760 (0.03)	0.1043 (0.11)	0.6106
CowSSL (proposed)	**0.9380 (0.01)**	0.8459 (0.01)	0.9679 (0.00)	**0.4336 (0.01)**	**0.3752 (0.04)**	**0.7541 (0.05)**	**0.4111 (0.05)**	**0.6751**

This scenario consists of three known patterns (center, edge-loc, edge-ring) and four new patterns (loc, scratch, random, donut)
Bold values indicate the highest F1 score among the methods

Table 5 Comparison of F1-score between the OpenLDN and OpenLDN(with contrastive loss)

Method	Known patterns (classes)				New patterns (classes)			Total
	Center	Edge-Loc	Edge-Ring	Loc	Scratch	Random	Donut	Average
OpenLDN	0.9107 (0.01)	0.8276 (0.03)	**0.9663 (0.01)**	**0.7125 (0.01)**	0.002 (0.00)	0.1014 (0.16)	**0.0034 (0.01)**	0.5034
TRSSL (with contrastive loss)	**0.9289 (0.01)**	**0.8285 (0.02)**	0.9231 (0.05)	0.6883 (0.03)	**0.0696 (0.07)**	**0.4840 (0.17)**	0.0000 (0.00)	**0.5604**

Bold values indicate the highest F1 score among the methods

Table 6 Comparison of time required for one epoch training by methodologies

	OpenLDN	TRSSL	ORCA	CowSSL (proposed)
Time	47 s	2100 s	37 s	38 s

sample for each class, it can be seen that they are distinguished from each other. In particular, when examining the representation visualization results for all models, we confirmed that a relatively large number of known patterns were well distinguished. However, as shown in Table 3, OpenLDN and TRSSL were unable to distinguish new patterns, which are small in number. In the case of ORCA, clusters of new patterns are segregated compared to known patterns, but it is evident that new patterns are clustered together, as illustrated by the blue circle. This suggests that in situations with severe class imbalance like WBM, distinguishing small patterns using conventional loss functions is challenging. On the other hand, the proposed CowSSL was more effective at separating random and donut patterns corresponding to the new, small-number patterns, as shown by the red circles. Consequently, our findings demonstrate that the contrastive loss function proposed in this study effectively distances new patterns from the majority of known patterns and separates between new patterns, resulting in enhanced discrimination of the minority new patterns.

Conclusions

With tens of thousands of WBM data pieces produced each day, it is crucial to quickly detect and identify new defects. In the industrial field, spotting new patterns during the fabrication process is important because they can indicate a new issue that was not previously known. Neglecting these issues can lead to major quality incidents. Therefore, obtaining new pattern samples that have the same morphology is vital to determining the common history of problematic wafers. Our study shows that it is possible to detect and classify new patterns within large amounts of unlabeled data using the contrastive loss function, even if the number of new patterns is tiny. This is the first study to apply open-world semi-supervised learning to WBM data, demonstrating that it is possible to learn a robust model despite the challenges of class imbalance and unlabeled data. In particular, our proposed CowSSL algorithms demonstrate the

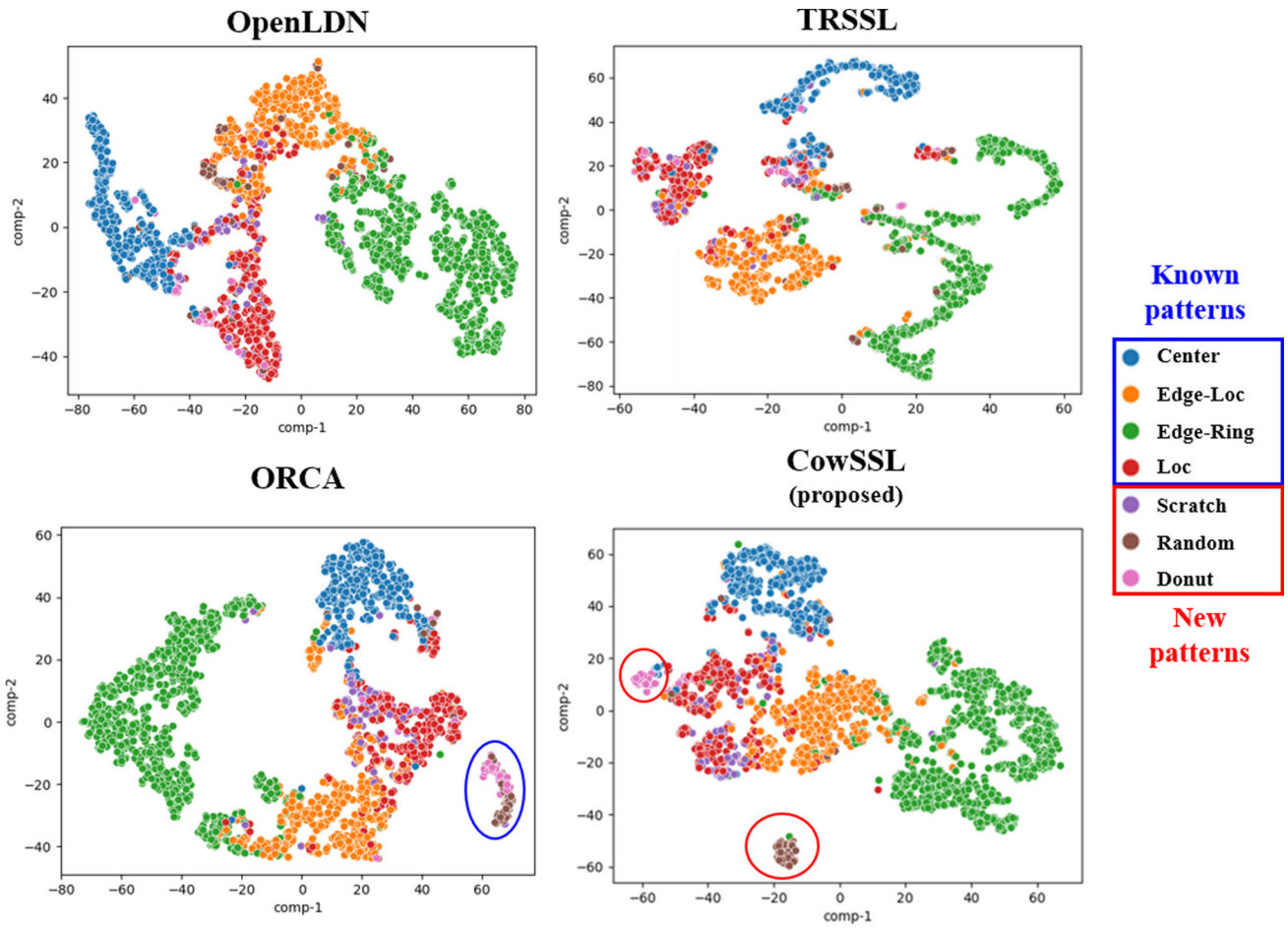

Fig. 7 Comparing representations of OpenLDN, TRSSL, ORCA, and CowSSL using t-SNE

improved classification of a small number of new patterns using the MoCo-based contrastive loss. This improvement is because of the lower likelihood of selecting a small number of new patterns data as the negative sample. Consequently, in comparison to existing open-world semi-supervised learning methods, we substantiate both quantitatively and qualitatively that our approach effectively separates and classifies each minority new pattern.

In future research, we aim to investigate methods for selecting negative samples with distinctly different characteristics from the input data when employing a contrastive loss function to enhance the classification of new patterns. The currently proposed contrastive loss function merely considers all samples, excluding the input data, as candidates for negative samples during selection. However, this approach has a limitation in that negative samples may encompass data from the same class as the input data. Therefore, by developing a method to select negative samples exhibiting significantly distinct characteristics from the input data, we anticipate that the contrastive loss function will work more effectively in detecting and classifying new patterns. Furthermore, the proposed methodology holds promise for the classification of mixed patterns in the WBM dataset. These mixed patterns can be intentionally designed to appear as new patterns. However, our current methodology was developed with a focus on single patterns where the distinction between existing defect patterns and new defect patterns was relatively clear. Therefore, as part of our future research, we intend to develop a model capable of effectively classifying mixed patterns by incorporating known single patterns and utilizing them as training data.

Author contributions Conceptualization: IB, SH, SBK. Formal analysis: IB, SH. Methodology: IB, SH, SBK. Supervision: SBK. Validation: IB, SH. Writing—original draft: IB, SH. Writing—review & editing: SBKim.

Funding This work was supported by Samsung Advanced Institute of Technology, Brain Korea 21 FOUR, the Ministry of Science and ICT (MSIT) in Korea under the ITRC support program supervised by the Institute for Information Communication Technology Planning and Evaluation (IITP-2020-0-01749), and the National Research Foundation of Korea grant funded by the Korea government (RS-2022-00144190).

Data availability All data are fully available without restriction. The dataset used in this study is available from the following website: https://www.kaggle.com/datasets/qingyi/wm811k-wafer-map.

Declarations

Competing interest The authors have declared that no competing interests exist.

References

Bendale, A., & Boult, T. (2015). Towards open world recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1893–1902).

Cao, K., Brbic, M., & Leskovec, J. (2021). Open-world semi-supervised learning. arXiv preprint arXiv:2102.03526. https://doi.org/10.48550/arXiv.2102.03526.

Chapelle, O., Scholkopf, B., & Zien, A. (2009). Semi-supervised learning In: chapelle, O., Scholkopf, B., & Zien, A. (Eds), (2006) [book reviews]. *IEEE Transactions on Neural Networks* 20 (3):542–542. https://doi.org/10.1109/TNN.2009.2015974.

Chen, X., Fan, H., Girshick, R., & He, K. (2020a). Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297. https://doi.org/10.48550/arXiv.2003.04297.

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020b). A simple framework for contrastive learning of visual representations. *International conference on machine learning* (pp. 1597–1607). PMLR.

Choudhary, A. K., Harding, J. A., & Tiwari, M. K. (2009). Data mining in manufacturing: A review based on the kind of knowledge. *Journal of Intelligent Manufacturing*, *20*, 501–521. https://doi.org/10.1007/s10845-008-0145-x.

He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9729–9738). https://doi.org/10.48550/arXiv.1911.05722.

Hsu, C. Y., & Chien, J. C. (2022). Ensemble convolutional neural networks with weighted majority for wafer bin map pattern classification. *Journal of Intelligent Manufacturing*, *33*(3), 831–844. https://doi.org/10.1007/s10845-020-01687-7.

Jang, J., & Lee, G. T. (2023). Decision fusion approach for detecting unknown wafer bin map patterns based on a deep multitask learning model. *Expert Systems with Applications*, *215*, 119363. https://doi.org/10.1016/j.eswa.2022.119363.

Jang, J., Seo, M., & Kim, C. O. (2020). Support weighted ensemble model for open set recognition of wafer map defects. *IEEE Transactions on Semiconductor Manufacturing*, *33*(4), 635–643. https://doi.org/10.1109/TSM.2020.3012183.

Jin, C. H., Na, H. J., Piao, M., Pok, G., & Ryu, K. H. (2019). A novel DBSCAN-based defect pattern detection and classification framework for wafer bin map. *IEEE Transactions on Semiconductor Manufacturing*, *32*(3), 286–292. https://doi.org/10.1109/TSM.2019.2916835.

Jin, C. H., Kim, H. J., Piao, Y., Li, M., & Piao, M. (2020). Wafer map defect pattern classification based on convolutional neural network features and error-correcting output codes. *Journal of Intelligent Manufacturing*, *31*(8), 1861–1875. https://doi.org/10.1007/s10845-020-01540-x.

Kahng, H., & Kim, S. B. (2021). Self-supervised representation learning for wafer bin map defect pattern classification. *IEEE Transactions on Semiconductor Manufacturing*, *34*(1), 74–86. https://doi.org/10.1109/TSM.2020.3038165.

Kang, S. (2020). Joint modeling of classification and regression for improving faulty wafer detection in semiconductor manufacturing. *Journal of Intelligent Manufacturing*, *31*(2), 319–326. https://doi.org/10.1007/s10845-018-1447-2.

Kong, Y., & Ni, D. (2021). A one-shot learning approach for similarity retrieval of wafer bin maps with unknown failure pattern. *IEEE Transactions on Semiconductor Manufacturing*, *35*(1), 40–49. https://doi.org/10.1109/TSM.2021.3123290.

Kyeong, K., & Kim, H. (2018). Classification of mixed-type defect patterns in wafer bin maps using convolutional neural networks. *IEEE Transactions on Semiconductor Manufacturing*, *31*(3), 395–402. https://doi.org/10.1109/TSM.2018.2841416.

Lee, K. B., Cheon, S., & Kim, C. O. (2017). A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes. *IEEE Transactions on Semiconductor Manufacturing, 30*(2), 135–142. https://doi.org/10.1109/TSM.2017.2676245

Park, S., Jang, J., & Kim, C. O. (2021). Discriminative feature learning and cluster-based defect label reconstruction for reducing uncertainty in wafer bin map labels. *Journal of Intelligent Manufacturing*, *32*, 251–263. https://doi.org/10.1007/s10845-020-01571-4.

Parmar, J., Chouhan, S., Raychoudhury, V., & Rathore, S. (2023). Open-world machine learning: Applications, challenges, and opportunities. *ACM Computing Surveys*, *55*(10), 1–37. https://doi.org/10.1145/3561381.

Piao, M., Jin, C. H., Lee, J. Y., & Byun, J. Y. (2018). Decision tree ensemble-based wafer map failure pattern recognition based on radon transform-based features. *IEEE Transactions on Semiconductor Manufacturing*, *31*(2), 250–257. https://doi.org/10.1109/TSM.2018.2806931.

Rizve, M. N., Kardan, N., & Shah, M. (2022a). Towards realistic semi-supervised learning. *European Conference on Computer Vision* (pp. 437–455). Cham: Springer https://doi.org/10.1007/978-3-031-19821-2_25.

Rizve, M. N., Kardan, N., Khan, S., Shahbaz Khan, F., & Shah, M. (2022b). Openldn: Learning to discover novel classes for open-world semi-supervised learning. *European Conference on Computer Vision* (pp. 382–401). Cham: Springer https://doi.org/10.1007/978-3-031-19821-2_22.

Scheirer, W. J., de Rezende Rocha, A., Sapkota, A., & Boult, T. E. (2012). Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(7), 1757–1772. https://doi.org/10.1109/TPAMI.2012.256.

Shim, J., Kang, S., & Cho, S. (2020). Active learning of convolutional neural network for cost-effective wafer map pattern classification. *IEEE Transactions on Semiconductor Manufacturing*, *33*(2), 258–266. https://doi.org/10.1109/TSM.2020.2974867.

Van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine Learning*, *109*(2), 373–440. https://doi.org/10.1007/s10994-019-05855-6.

Wu, M. J., Jang, J. S. R., & Chen, J. L. (2015). Wafer map failure pattern recognition and similarity ranking for large-scale data sets. *IEEE Transactions on Semiconductor Manufacturing*, *28*(1), 1–12. https://doi.org/10.1109/TSM.2014.2364237.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.