Available online at www.sciencedirect.com

ScienceDirect

ICT Express 7 (2021) 535–539

www.elsevier.com/locate/icte

Evaluation of the machine learning classifier in wafer defects classification

Jessnor Arif Mat Jizat[a,*], Anwar P.P. Abdul Majeed[a,b], Ahmad Fakhri Ab. Nasir[a,b], Zahari Taha[c], Edmund Yuen[d]

[a] Faculty of Manufacturing & Mechatronic Engineering Technology (FTKPM), Universiti Malaysia Pahang, 26600 Pekan Pahang, Malaysia
[b] Centre for Software Development & Integrated Computing, Universiti Malaysia Pahang, 26600, Malaysia
[c] Fakultas Teknologi Industri, Universitas Islam Indonesia, Jl. Kaliurang KM. 14,5 Sleman Yogyakarta 55584, Indonesia
[d] Ideal Vision Integration Sdn Bhd, 02-25, Level 2, Setia Spice Canopy, Jln Tun Dr Awang, 11900 Bayan Lepas, Penang, Malaysia

Received 20 September 2020; received in revised form 28 December 2020; accepted 26 April 2021
Available online 3 May 2021

Abstract

In this paper, an evaluation of machine learning classifiers to be applied in wafer defect detection is described. The objective is to establish the best machine learning classifier for Wafer Defect Detection application. *k*-Nearest Neighbours (k-NN), Logistic Regression, Stochastic Gradient Descent, and Support Vector Machine were evaluated with 3 defects categories and one non-defect category. The key metrics for the evaluation are classification accuracy, classification precision and classification recall. 855 images were used to train, test and validate the classifier. Each image went through the embedding process by InceptionV3 algorithms before the evaluated classifier classifies the images.

Keywords: Logistic Regression; Stochastic Gradient Descend; Wafer defect detection

1. Introduction

In the semiconductor industry, wafer defect is a big issue where it can affect yield, company reputation as well as manufacturing processes. Commonly companies employed workers to do defect detection. However, training people to manual inspection on wafer defect proves to be time-consuming for the industry as it can take about 6–9 months of human workers training to achieve 90% accuracy in wafer defect detection. Nevertheless, within 15 months after the training had been completed, the manual inspection can drop to between 70%–85% accuracy due to several factors such as increased difficulties due to product evolution, demotivation due to mental fatigue, or process advancement [1]. Thus many industry players are adopting an automated defect detection system using machine vision with machine learning capabilities.

1.1. Related work

Automated optical inspection machine had evolved from labour-intensive manual inspection towards the fully automated machine with minimum human intervention. Image processing algorithm is the key driver [2–4] for the automated machine that is capable of detecting anomaly on the wafer under quality control. Any defects that appeared on the wafer can trigger a sensor to reject that particular wafer.

However, as the electronics industry evolves into nanoscale production, the requirement for wafer quality also increases. More complex defects detection, as well as faster decision machine, becomes the most sought after technology within the industry. With the advancement of computing technologies, Convolution Neural Network (CNN) offers improved performance in terms of faster image processing time and features extraction from the images [5]. CNN employs a technique to extract specifics features from images and clustered the images according to their extracted features. Thus, myriad of CNN based algorithm had been employed in various applications such as sports [6], healthcare [7], image processing [8] as well as in the semiconductor industry [9].

In order to reliably make a decision in a shorter time, CNN requires hundreds of images to be analysed before it can predict the classification of the image especially a new image that it never came across before. Thus, image databases on the internet offer a repository of images in various classifications open for any algorithms to do classification [10]. From here,

* Corresponding author.
E-mail address: jessnor@ump.edu.my (J.A. Mat Jizat).
Peer review under responsibility of The Korean Institute of Communications and Information Sciences (KICS).

https://doi.org/10.1016/j.icte.2021.04.007

transfer learning becomes a part of a machine learning algorithm where the algorithms were pre-trained with the images in a database and using the knowledge gained to predict the image classification.

Saqlain et al. [11] presented a work by extracting features certain features such as density-, geometry-, and radon-based from the raw image of wafer and had run four classifier algorithms. They combine the accuracy result by introducing a soft voting ensemble technique to increase accuracy. Furthermore, in a more recent paper, Saqlain et al. [12] introduced a new technique at increased accuracy and it is easy to deploy in a manufacturing plant.

Other researchers have exploited the use of Dynamic Time Warping with different classifiers, i.e., kNN [13] and SVM [14] in order to improve the classification. Ruifang et al. [15] reported as high as 11 defect markers. They employed ZF-Net to extract the feature from dark field illuminated images. They also demonstrated a powerful GPU for an increased computing performance.

Logistic regression, support vector machine are among the most frequently used machine learning classifiers especially in wafer defect detection. As there are quite a number of machine learning classifiers, a reliable method is needed to evaluate the classifier performance.

Thus the objective of this paper is to establish the best machine learning classifier with known baseline parameter that works best with features extracted via InceptionV3 pre-trained CNN for wafer defect detection. This is important so that the researcher can focus on optimizing the specific classifier by tuning its hyperparameter. It is worth noting that the proposed pipeline i.e., the extraction of features via a pre-trained CNN model viz. Inception V3 along with the different classifiers evaluated has not yet been reported in the literature. This algorithm then will be deployed into a machine vision system that can reliably detect wafer defect.

2. Methodology

2.1. Machine learning classifier

Four Machine Learning Classifiers were chosen to be evaluated. The description of the classifiers is as follows:

- Logistic Regression — the logistic regression classification algorithm with ridge regularization.
- K-Nearest Neighbour — using five of the nearest neighbours with Euclidean distance (straight line distance between two points) and uniform weights (all points in each neighbours weighted equally).
- Stochastic Gradient Descend — using Ridge regularization with 0.0001 strength, constant learning rate with 0.01 initial rate and limited to 1000 iterations with tolerance 0.001.
- Support Vector Machine — using 1.00 cost (C), 0.10 epsilon distance, sigmoid kernel and limited to 100 iterations.

The classification accuracy, precision and recall of their classification result were used as evaluation criteria. Classification accuracy indicates the capability of the classifier to accurately predict the images according to its class. Meanwhile, classification precision deals with the proportion of the correctly predicted images and the predicted images in its class. On the other hand, classification recall deals with the proportion of the correctly predicted images and the actual images in its class. All these metrics will give a better understanding of the classifier predictive capability.

2.2. Hardware and software

The training and testing were conducted using a desktop with Intel(R) Core(TM) i7-6700U CPU 3.40 GHz processor with 16GB DDR3 RAM and NVIDIA GeForce GTX950 graphic card. It was run using Spyder Anaconda, a python programming software using Scikit-learn, Keras and Tensorflow libraries.

2.3. Dataset

The dataset is acquired from Idealvision Sdn Bhd, a machine vision company, using their own industrial machine vision platform Jaeger. The dataset is divided into three categories namely training, testing, and validation.

For training and testing, the images were separated according to their defect features, namely Bump, Burnt Mark and Foreign Object. A category of non-defect was also added for the training and testing. A total of 707 images were available for training and testing which consist of 557 for training and 150 for testing. This represented 80:20 training and testing proportion.

Another set of images was used for validation which consists of 168 images. These images in the validation category, contrary to training and testing image set, were not separated according to the defect features. Furthermore, validation dataset had significantly more “good” images in anticipation of more good images to be produced in actual automated optical inspection machine whereby in this case the good images class is about double from other classes.

In terms of image preprocessing, in order to gain maximum features extraction from each image, all 3 channel RGB were used. The images were neither converted to binary images nor to a greyscale value. The images dataset in each category was neither repeated, nor augmented, and was different from the other categories. All images were rescaled to [299,299,3] from [4096, 3072, 3] to cater to the computing performance of the desktop used as well as the required dimension for InceptionV3 image input. In Fig. 1, are shown samples of images used in the paper.

2.4. Training, testing and validation process

After the training and testing dataset was loaded to the software, each image went through an embedding process using InceptionV3 transfer learning algorithm. This embedding

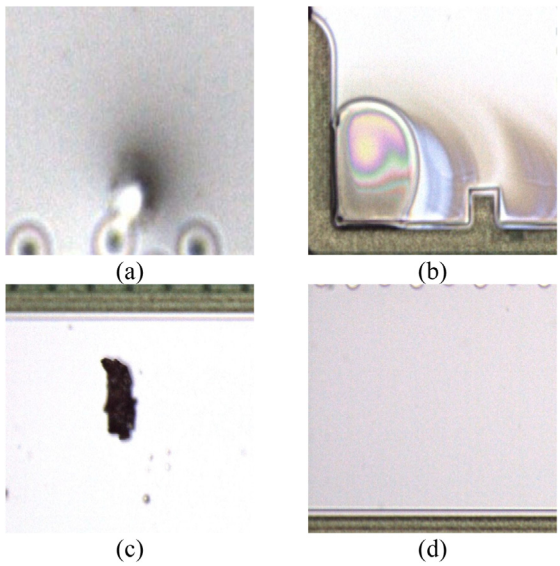

Fig. 1. Sample images from each category (a) Bump (b) Burnt Mark (c) Foreign Object (d) Good.

process calculated feature vector in each image and embedded more data numbers into the image based on the specific feature present in the image. From the enhanced data images, a machine learning classifier from Section 1 was selected to train and test the classification capability. In order to mitigate overfitting, the data went through a 10-fold cross-validation process. After the cross-validation process, the software generates a confusion matrix to indicate the performance of the training.

For the validation process, the best classifier from training and testing was chosen. The image also went through the embedding process using InceptionV3 transfer learning algorithm. The images were never introduced to the algorithm before. Using the trained classifier, machine learning performed a prediction for each image to classify the image accordingly. A table of confusion matrix was tabulated from the prediction. (*See Table* 2.)

The workflow diagram is shown in Fig. 2.

3. Result and discussion

Fig. 3 shows the comparison of the average accuracy performance of four different machine learning classifier models in terms of wafer defect classification. Out of the four machine learning classifiers evaluated, Logistic Regression classifier gives the best classification accuracy with 86.0% during training and 88.0% during testing while *k*-Nearest Neighbours had the worst classification accuracy performance with 72.2% in training and 74.0% in testing. The Stochastic Gradient Descent had a similar score to Logistic Regression with a classification accuracy of 85.8% during training and 87.3% during testing. These classifiers Logistic Regression and Stochastic Gradient Descent requires further optimization to increase their classification accuracy tailored for wafer defect detection.

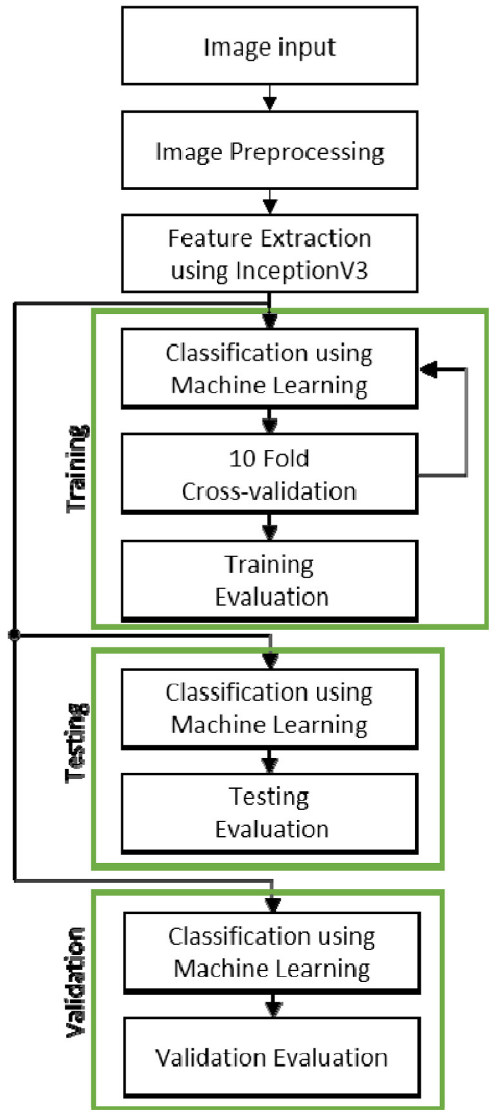

Fig. 2. Workflow diagram for training, testing and validation.

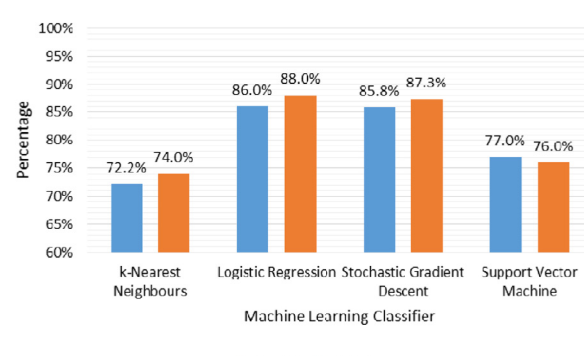

Fig. 3. Comparison of average classification accuracy performance.

However, classification accuracy for wafer defect will be biased towards non-defect classification since each wafer manufacturing process strives for zero defect. Thus, the non-defect wafer will always inundate the image samples as wafer defect

Table 1
Precision–recall performance for machine learning classifier.

		Precision	Recall
Training	*k*-Nearest Neighbours	74.3%	72.2%
	Logistic Regression	86.4%	86.0%
	Stochastic Gradient Descent	86.1%	85.8%
	Support Vector Machine	81.0%	77.0%
Testing	*k*-Nearest Neighbours	76.9%	74.0%
	Logistic Regression	88.5%	88.0%
	Stochastic Gradient Descent	88.7%	87.3%
	Support Vector Machine	82.8%	76.0%

Table 2
Confusion matrix for Logistic Regression Prediction using validation dataset.

		Predicted				
		Bump	BM	F. Obj	Good	$\sum$
Actual	Bump	29	0	0	5	34
	BM	0	31	0	3	34
	F. Obj.	0	0	31	3	34
	Good	3	5	3	55	66
	$\sum$	32	36	34	66	168

BM = Burnt Mark, F. Obj = Foreign Object.

is an occasional occurrence. Consequently, a classification accuracy prediction is not sufficient to evaluate the defect detection system. The system must also be evaluated in terms of Precision–Recall. Table 1 shows the training mean precision–recall performance and testing mean precision–recall performance across the categories of the machine learning classifiers.

From Table 1 we can learn that all classifier models give a higher percentage of precision compared to recall. A higher precision shows all the classifiers demonstrate a bias towards false negative. False-negative means some of the good wafers being wasted and this might affect the wafer yield. Small percentage false negative bias is always preferable compared to false-positive bias where defective wafer may be classified as good. The worst performance in this evaluation was Support Vector Machine scoring more than 4% difference between Precision and Recall both during training and during testing. Coupled with data from Fig. 2, where its accuracy during testing is lower than during training shows that Support Vector Machine may not be reliable to detection wafer defects. K-Nearest Neighbours performed slightly better than Support Vector Machine in terms of precision–recall performance where the difference between training and testing is about 2%. However, the classification accuracy for k-Nearest Neighbours is less than 75%. The next classifier, Stochastic Gradient Descent, had scores with the best performance in terms of precision–recall with a difference of 0.3% and 0.4% during training and testing respectively. However, in terms of classification accuracy SGD only improved by 1.5% between training and testing. Thus taking into account classification accuracy, classification precision and classification recall from training and testing, Logistic Regression performed the best overall where the difference between training and testing for accuracy is about 2%, and the difference between precision and recall for training and test is 0.4% and 0.5% respectively.

In order to understand the performance of Logistic Regression classifier better, validation dataset is used to predict the classification of the image. Table 2 shows the confusion matrix for the wafer defect detection using Logistic Regression classifier during the validation process.

From Table 2, we can calculate average classification accuracy, average classification precision and average classification recall. Average classification accuracy is defined as the total number of correct classifications over the total number of image samples. Average classification precision is defined as an average of correct prediction of each class over a total number of predictions of each class. While average classification recall is defined as an average of a correctly identified image of each class over the actual images in each class. For this validation process, classification accuracy is calculated to be 86.9% while average precision is 87.8% and average recall is 87.7%. This validation performance is indeed similar to the training and testing performance.

It is worth noting that precision–recall performance is almost equal. An equal precision–recall performance shows a balance misclassification. Evidently, by referring to Table 2, we can see that misclassification occurs mainly on "good" prediction as well as misclassification of actual "good" images. Investigation on these misclassification images shows that the defects were tiny compared to the whole images, positioned at the image edges, or occurred in a cluttered area.

4. Conclusion

From the experiment above, it can be established that Logistic Regression classifier is the best classifier to run a wafer defect detection at 86.9% accuracy. This classifier performed together with an untuned InceptionV3 Transfer Learning. Further optimization can be performed to increase the accuracy. However, it is best to remember that the more the image samples obtained for training and testing the more accurate the classifier can be. It is also worth noting that the specific defects are known beforehand and a new class of images may not be detected. Future works, shall incorporate other pre-trained CNN models for feature extraction as well as investigate the effect of hyperparameter optimization towards the classification accuracy for detecting the defects .

CRediT authorship contribution statement

Jessnor Arif Mat Jizat: Writing - original draft, Investigation. **Anwar P.P. Abdul Majeed:** Methodology, Formal analysis. **Ahmad Fakhri Ab. Nasir:** Writing - review & editing. **Zahari Taha:** Conceptualization, Project administration. **Edmund Yuen:** Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to thank IdealVision Sdn Bhd for providing the image dataset to make this evaluation possible as well as Universiti Malaysia Pahang for funding the study via UIC200815 and RDU202404.

References

[1] E. Tuv, M. Guven, P. Ennis, D.H.L. Lee, IT@Intel White Paper: Faster, more Accurate Defect Classification using Machine Learning, 2018, https://www.intel.com/content/dam/www/public/us/en/documents/best-practices/faster-more-accurate-defect-classification-using-machine-vision-paper.pdf.

[2] S.H. Huang, Y.C. Pan, Automated visual inspection in the semiconductor industry: A survey, Comput. Ind. (2015) http://dx.doi.org/10.1016/j.compind.2014.10.006.

[3] K.P. White, B. Kundu, C.M. Mastrangelo, Classification of defect clusters on semiconductor wafers via the hough transformation, IEEE Trans. Semicond. Manuf. 21 (2008) 272–277, http://dx.doi.org/10.1109/TSM.2008.2000269.

[4] H.D. Lin, S.W. Chiu, Flaw detection of domed surfaces in LED packages by machine vision system, Expert Syst. Appl. (2011) http://dx.doi.org/10.1016/j.eswa.2011.05.080.

[5] Y. LeCun, et al., Generalization and network design strategies, in: Connect Perspect, 1989.

[6] K. Rangasamy, M.A. As'ari, N.A. Rahmad, N.F. Ghazali, Hockey activity recognition using pre-trained deep learning model, ICT Express (2020) http://dx.doi.org/10.1016/j.icte.2020.04.013.

[7] G. Swapna, R. Vinayakumar, K.P. Soman, Diabetes detection using deep learning algorithms, ICT Express (2018) http://dx.doi.org/10.1016/j.icte.2018.10.005.

[8] K. Jhang, H. Kang, H. Kwon, CNN training for age group prediction in an illumination condition, ICT Express (2020) http://dx.doi.org/10.1016/j.icte.2020.05.001.

[9] Y. Yuan-Fu, A deep learning model for identification of defect patterns in semiconductor wafer map, in: ASMC (Advanced Semicond. Manuf. Conf. Proc. doi:https://doi.org/10.1109/ASMC.2019.8791815.

[10] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge, Int. J. Comput. Vis. (2015) http://dx.doi.org/10.1007/s11263-015-0816-y.

[11] M. Saqlain, B. Jargalsaikhan, J.Y. Lee, A voting ensemble classifier for wafer map defect patterns identification in semiconductor manufacturing, IEEE Trans. Semicond. Manuf. 32 (2019) 171–182, http://dx.doi.org/10.1109/TSM.2019.2904306.

[12] M. Saqlain, Q. Abbas, J.Y. Lee, A deep convolutional neural network for wafer defect identification on an imbalanced dataset in semiconductor manufacturing processes, IEEE Trans. Semicond. Manuf. 33 (2020) 436–444, http://dx.doi.org/10.1109/TSM.2020.2994357.

[13] Y.S. Jeong, S.J. Kim, M.K. Jeong, Automatic identification of defect patterns in semiconductor wafer maps using spatial correlogram and dynamic time warping, IEEE Trans. Semicond. Manuf. (2008) http://dx.doi.org/10.1109/TSM.2008.2005375.

[14] Y.S. Jeong, Semiconductor wafer defect classification using support vector machine with weighted dynamic time warping kernel function, ind, Eng. Manag. Syst. (2017) http://dx.doi.org/10.7232/iems.2017.16.3.420.

[15] R. Ye, C.S. Pan, M. Chang, Q. Yu, Intelligent defect classification system based on deep learning, Adv. Mech. Eng. (2018) http://dx.doi.org/10.1177/1687814018766682.