

PAPER • OPEN ACCESS

Surface Defect Detection: A feature-based transfer learning approach

To cite this article: Anwar P.P. Abdul Majeed *et al* 2024 *J. Phys.: Conf. Ser.* **2762** 012088

View the [article online](#) for updates and enhancements.

You may also like

- [Diagnosis of rectal cancer based on the Xception-MS network](#)
Sanli Yi, Yanrong Wei, Xiaomao Luo *et al.*
- [Interpretable surrogate models to approximate the predictions of convolutional neural networks in glaucoma diagnosis](#)
Jose Sigut, Francisco Fumero, Rafael Aray *et al.*
- [CAM-QUS guided self-tuning modular CNNs with multi-loss functions for fully automated breast lesion classification in ultrasound images](#)
Jarin Tasnim and Md Kamrul Hasan



The Electrochemical Society

Advancing solid state & electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research



Surface Defect Detection: A feature-based transfer learning approach

Anwar P.P. Abdul Majeed^{1*}, Muhammad Amirul Abdullah², Ahmad Fakhri Ab. Nasir³, Mohd Azraai Mohd Razman⁴, Wei Chen⁵ and Eng Hwa Yap¹

¹School of Robotics, XJTLU Entrepreneur College, Xi'an Jiaotong – Liverpool University (XJTLU), China

² Innovative Manufacturing, Mechatronics and Sports Laboratory, Faculty of Manufacturing and Mechatronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Malaysia

³ Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah, Malaysia

⁴ Innovative Manufacturing, Mechatronics and Sports Laboratory, Faculty of Manufacturing and Mechatronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Malaysia

⁵ School of Intelligent Manufacturing Ecosystem, XJTLU Entrepreneur College, Xi'an Jiaotong – Liverpool University (XJTLU), China

Corresponding author: Anwar.Majeed@xjtlu.edu.cn

Abstract. Surface defect detection is critical for maintaining product quality in manufacturing. In this work, we apply a feature-based transfer learning approach for surface defect classification on the NEU surface defect database. The database contains defects across 6 categories captured under various conditions. We utilised two pretrained convolutional neural network (CNN) architectures - VGG16 and InceptionV3 - by removing the final classification layer and using the CNN as a fixed feature extractor. The output feature vectors were classified using a logistic regression (LR) model. The data was split into train, validation, and test sets with a 70:15:15 ratio. The VGG16-LR model achieved classification accuracy (CA) of 100%, 98%, and 99% for the train, validation, and test sets respectively. The InceptionV3-LR model attained CA of 100%, 91%, and 92% for train, validation, and test. The results demonstrate the effectiveness of transfer learning with CNN feature extraction for surface defect detection on challenging multi-category industrial datasets. Further work includes tuning hyperparameters and evaluating additional architectures.

Keywords: Surface Defect Detection, Transfer Learning, Feature Extraction, Deep Learning

1. Introduction

Surface defects in manufactured products can lead to functional or aesthetic issues, making detection critical for quality control. Manual inspection is time-consuming, subjective, and inefficient for modern high-throughput manufacturing [1]–[3]. Automated computer vision techniques present a compelling solution. Recent work has shown success applying deep convolutional neural networks



(CNNs) for defect classification. However, training CNNs requires large labelled datasets which are costly to obtain in practice.

Transfer learning provides a method to leverage pretrained CNNs on general image datasets and transfer the learned features to new tasks with limited data. As reported in the literature, the aforesaid technique has been successfully employed on different applications [4]–[9]. With regards to defect detection, Tabl et al. [10] used a fine-tuned ResNet-50 CNN model to classify manufacturing defects as either normal or defective. The accuracy reported ranged widely from 48 to 96%, depending on batch size and epochs, indicating inconsistent performance.

Mat Jizat et al. [5] compared an InceptionV3 CNN paired with logistic regression (LR), k-nearest neighbors (kNN), SVM, and stochastic gradient descent (SGD) for classifying wafer defects. The InceptionV3-LR pipeline performed the best on both training and test data compared to the other methods investigated. Another study addressed common steel surface defects like scratches, pitting, inclusions, and patches to separate defective from non-defective surfaces [11]. The authors used a CNN with Xception architecture to detect these defects. It achieved reasonably good accuracy between 85 to 94%. Guan et al. [12] compared the efficacy of a new VSD network against VGG19 and ResNet on the NEU dataset. It was shown that the VSD network could achieve an overall accuracy of 89.86% on the validation dataset.

In this work, we present a feature-based transfer learning approach for multi-category surface defect classification on the Northeastern University (NEU) database. The key insight of the present study is to evaluate the efficacy of the pre-trained CNN models, i.e., VGG16 and InceptionV3 ability in extracting meaningful features for the classification of the defects.

2. Methodology

In this study, we conduct an in-depth investigation and comparison of two state-of-the-art pre-trained convolutional neural network (CNN) architectures – VGG16 and Inception V3 - for feature extraction and defect classification. The VGG16 was developed by the Visual Geometry Group at Oxford University. It consists of 16 layers and utilizes consecutive convolutional and max pooling layers of increasing depth to extract hierarchical features, with fully connected layers at the end for classification. We utilize a version of VGG16 without the fully connected layers, replacing them instead with a Logistic Regression (LR) classifier. It is worth noting that the original VGG16 model was pre-trained on millions of ImageNet images for object recognition tasks.

Conversely, the Inception V3 was developed by Google for image classification and recognition applications. It introduced a novel architectural block called the inception module, which contains convolutions of different sizes in parallel to capture features at multiple scales. Multiple inception modules are stacked for deeper representations. Similar to the VGG16 model, the Inception V3 was also pre-trained on the extensive ImageNet dataset. The key difference between the two architectures is that Inception V3 relies on convolutions in parallel for multi-scale feature learning, while VGG16 uses consecutive convolutional layers in a hierarchical fashion.

In our study, we leverage the transfer learning capacity of these two powerful pre-trained models by extracting features from the convolutional layers, which contain generic representations useful for many vision tasks. The extracted feature representations are then fed into a LR classifier for defect classification. Our experiments allow an in-depth analysis and comparison of the transferability of features from consecutive versus parallel convolutional architectures.

We evaluate the models on a surface defect classification dataset, particularly the NEU database (<https://www.kaggle.com/datasets/kaustubhdikshit/neu-surface-defect-database>). It consists of six different types of surface defects, i.e., crazing, inclusion, patches, pitted surface, rolled in scale and scratches, respectively [12]. Figure 1 illustrate samples of the different defects investigated. The dataset consists of 300 samples for each type of defects. It is worth noting that the 70:15:15 stratified split ratio was carried out for training, validation and testing dataset. A Python IDE, viz. Spyder (running on Python 3.7) was used to carry out the analysis. In the study, the default LR from the sklearn library hyperparameters were used whilst the Keras library was evoked for the VGG16 and

InceptionV3 architectures, respectively. The present study provides meaningful insights into the representational differences between these two CNN architectures and their ability to generalize for defect detection across different datasets. The classification accuracy (CA) as well as the confusion matrix (CM) are selected as the performance indicators to evaluate the performance of the formulated pipelines.

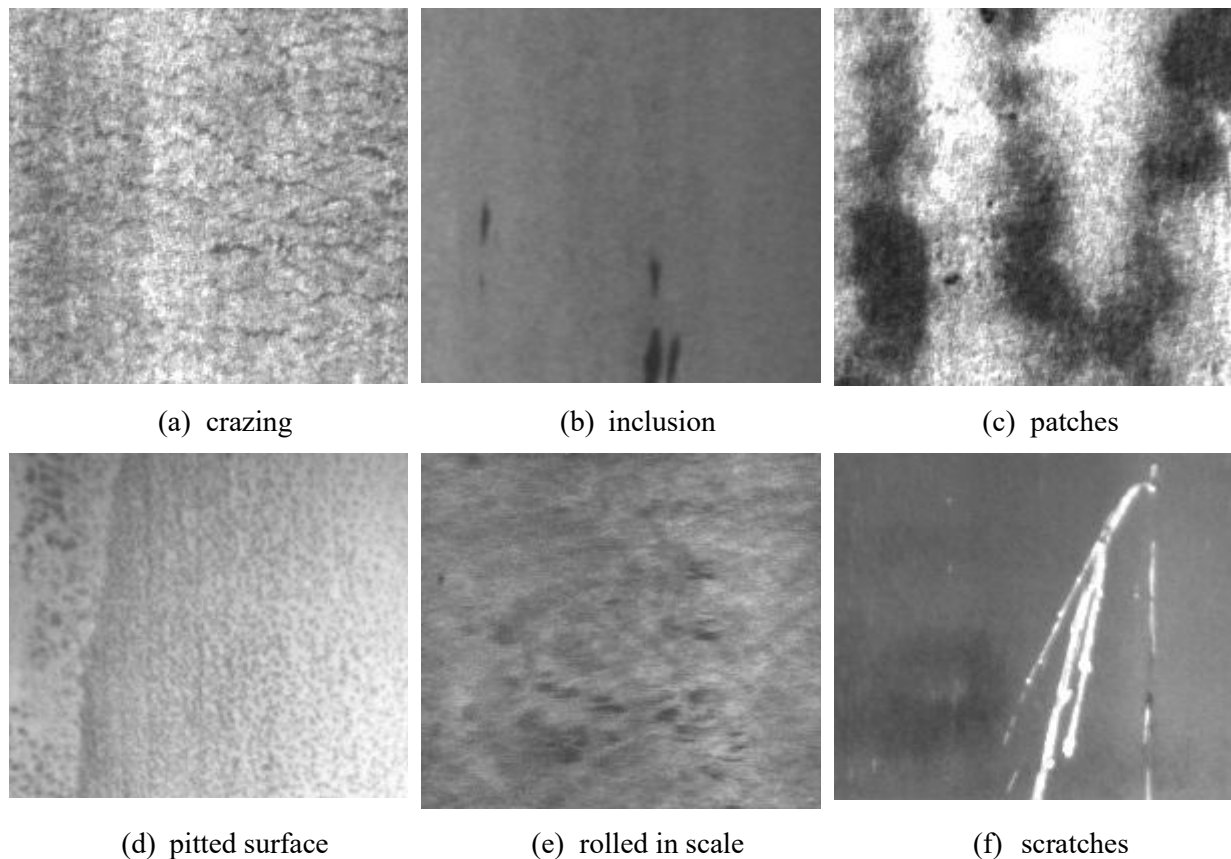


Figure 1. Sample defect images from the NEU database.

3. Results and Discussion

The results shown in Figure 2 demonstrate that using a pre-trained VGG16 model coupled with a logistic regression (LR) classifier was highly effective for classifying defects in this image dataset. This VGG16-LR pipeline achieved classification accuracies of 100%, 98% and 99% on the training, test and validation sets respectively, with no errors. This suggests the VGG16 model was able to extract meaningful features from the images that enabled accurate defect classification. However, combining the Inception V3 model with an LR classifier led to much poorer performance, with classification accuracies of only 91% and 92% on the test and validation sets, despite 100% training accuracy.

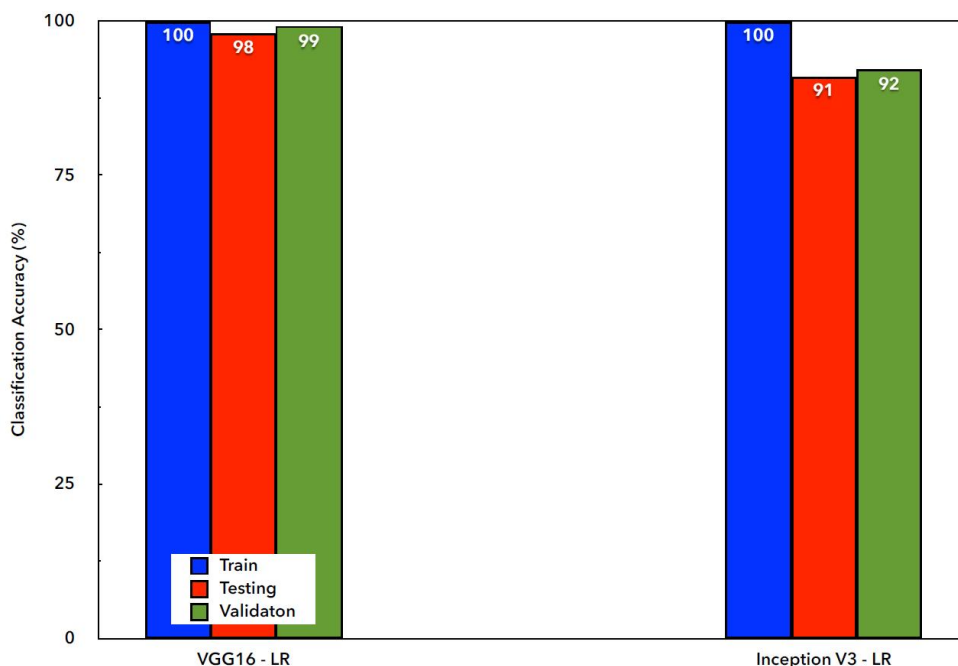


Figure 2. Pipeline performance.

Figure 3 (Testing CM) and Figure 4 (Validation CM) provides further evidence that the InceptionV3 model struggled to learn an effective feature representation compared to VGG16, leading to multiple misclassifications. The different defect classes, i.e., crazing, inclusion, patches, pitted surface, rolled in scale and scratches are denoted as 0, 1, 2, 3, 4 and 5, respectively. In summary, the VGG16 model demonstrated superior ability over InceptionV3 to extract discriminative features from this defect image dataset when used with an LR classifier. The VGG16-LR pipeline provided excellent defect classification performance, while the InceptionV3-LR pipeline had difficulty learning effective image features, resulting in poorer classification accuracy.

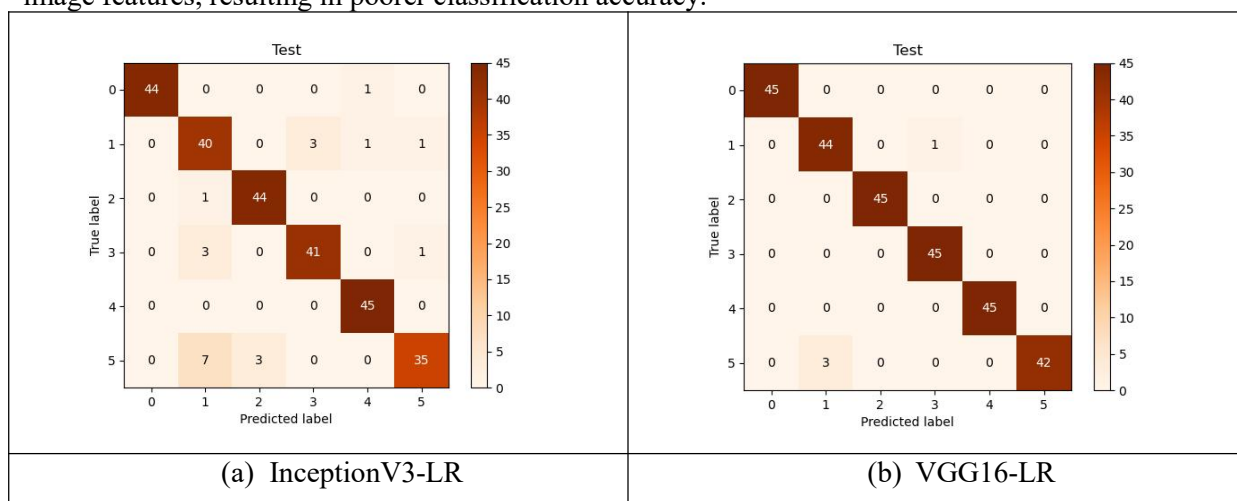


Figure 3. Confusion Matrix on the testing dataset

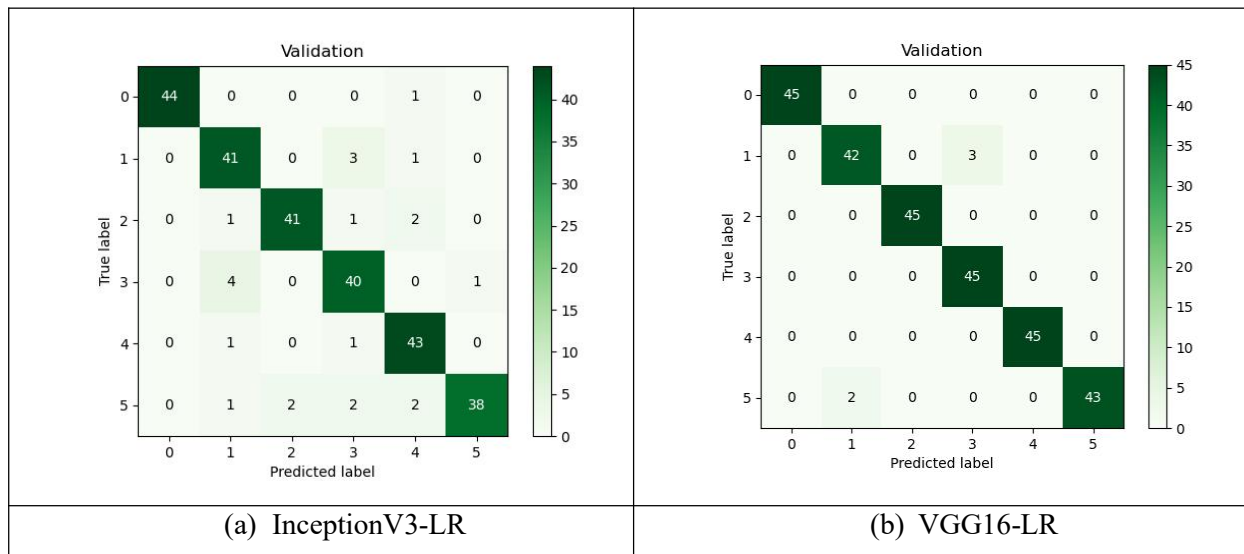


Figure 4. Confusion Matrix on the validation dataset.

4. Conclusion

In conclusion, the pre-trained VGG16 model coupled with a simple logistic regression classifier provides an excellent pipeline for classifying defects in this image dataset, significantly outperforming the InceptionV3-LR approach. The VGG16 model is able to extract highly discriminative features from the images, enabling near perfect classification accuracy. Moving forward, the VGG16-LR pipeline could be deployed for real-time defect detection in manufacturing quality control. Additionally, further hyperparameter tuning and model optimization could potentially improve performance even further. In addition, the effect of different pre-trained transfer learning models as well as the combination of different classifiers shall be explored. Beyond this specific application, the results highlight the importance of selecting an appropriate pre-trained model for a given computer vision task and dataset.

References

- [1] G. Fu *et al.*, “A deep-learning-based approach for fast and robust steel surface defects classification,” *Opt. Lasers Eng.*, vol. 121, 2019, doi: 10.1016/j.optlaseng.2019.05.005.
- [2] L. S. Xuen *et al.*, “The Classification of Wafer Defects: A Support Vector Machine with Different DenseNet Transfer Learning Models Evaluation,” in *Robot Intelligence Technology and Applications 7: Results from the 10th International Conference on Robot Intelligence Technology and Applications*, 2023, pp. 304–309.
- [3] Y. Hao, M. Xiang, and Z. Zhu, “A defect detection method for plastic gears based on deep learning and machine vision,” in *Fourteenth International Conference on Digital Image Processing (ICDIP 2022)*, 2022, vol. 12342, pp. 113–118.
- [4] J. A. Mat Jizat, A. P. P. Abdul Majeed, A. F. Ab. Nasir, Z. Taha, E. Yuen, and S. X. Lim, “Evaluation of the Transfer Learning Models in Wafer Defects Classification,” in *Recent Trends in Mechatronics Towards Industry 4.0: Selected Articles from iM3F 2020, Malaysia*, 2022, pp. 873–881.
- [5] J. A. Mat Jizat, A. P.P. Abdul Majeed, A. F. Ab. Nasir, Z. Taha, and E. Yuen, “Evaluation of the machine learning classifier in wafer defects classification,” *ICT Express*, vol. 7, no. 4, pp. 535–539, 2021, doi: 10.1016/j.ict.2021.04.007.
- [6] J. L. M. Kumar *et al.*, “The classification of EEG-based wink signals: A CWT-Transfer Learning pipeline,” *ICT Express*, 2021, doi: <https://doi.org/10.1016/j.ict.2021.01.004>.
- [7] J. L. Mahendra Kumar *et al.*, “The classification of EEG-based winking signals: a transfer

- learning and random forest pipeline,” *PeerJ*, vol. 9, p. e11182, Mar. 2021, doi: 10.7717/peerj.11182.
- [8] M. N. A. Shapiee, M. A. R. Ibrahim, M. A. Mohd Razman, M. A. Abdullah, R. M. Musa, and A. P. P. Abdul Majeed, “The Classification of Skateboarding Tricks by Means of the integration of Transfer Learning and Machine Learning Models,” in *Embracing Industry 4.0. Lecture Notes in Electrical Engineering, vol 678. Springer, Singapore*, vol. 678, A. K. M. Mohd Razman M., Mat Jizat J., Mat Yahya N., Myung H., Zainal Abidin A., Ed. 2020, pp. 219–226.
- [9] M. A. Abdullah *et al.*, “The classification of skateboarding tricks via transfer learning pipelines,” *PeerJ Comput. Sci.*, vol. 7, p. e680, 2021.
- [10] A. Abou Tabl, A. Alkhateeb, and W. ElMaraghy, “Deep Learning Method based on Big Data for Defects Detection in Manufacturing Systems Industry 4.0,” *Int. J. Ind. Sustain. Dev.*, vol. 2, no. 1, 2021, doi: 10.21608/ijisd.2021.145552.
- [11] V. F. Fadli and I. O. Herlistiono, “Steel surface defect detection using deep learning,” *Int. J. Innov. Sci. Res. Technol.*, vol. 5, pp. 244–250, 2020.
- [12] S. Guan, M. Lei, and H. Lu, “A steel surface defect recognition algorithm based on improved deep learning network model using feature visualization and quality evaluation,” *IEEE Access*, vol. 8, pp. 49885–49895, 2020.