Steel Surface Defect Classification Via Deep Learning

Mustafa Mert Tunalı
Computer Engineering
Mef University
Istanbul, Türkiye
tunalim@mef.edu.tr

Ahmet Yıldız
Computer Engineering
Mef University
Istanbul, Türkiye
yildizah@mef.edu.tr

Tuna Çakar
Computer Engineering
Mef University Tam Finans Agm
Istanbul, Türkiye
cakart@mef.edu.tr

2022 7th International Conference on Computer Science and Engineering (UBMK) | 978-1-6654-7010-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/UBMK55850.2022.9919470

***Abstract*—Deep learning and image processing methods have taken place in many parts of our lives, as well as in the quality control stages of production lines. The aim of this study is to train and use a deep learning model to improve quality management using limited data and computing power. To achieve that, deep learning for quality control models were trained by classifying six different steel surface defect images in the NEU-DET dataset. Xception, ResNetV2 152, VGG19 and InceptionV3 architectures were used to train the model. High accuracy was obtained with both Xception and ResNetV2 152.**

***Keywords* —Quality Control, Steel Surface Defect, Image Classification, Deep Learning, NEU-DET, Xception, Inception V3, ResNet152 V2, VGG19**

I. Introduction

Artificial Intelligence technologies continue to develop day by day and take more place in our lives. Many companies are taking advantage of this change and developing artificial intelligence applications using the latest machine learning algorithms to market their products and improve customer relationships. On the other hand, the use of artificial intelligence applications has fallen behind when we look at the production stages of products that use the latest technology deep learning methods in the marketing phase [1].

The most appropriate part where artificial intelligence applications can be used in the production phase is the quality management and quality control departments, because the cost of eliminating defective products is quite high when it is done by human power [2]. If adequate quality control is not achieved, this has the potential to generate higher costs, given the high current shipping and storage costs. For this reason, quality control has begun to play an increasingly critical role, especially for companies operating in production.

Considering the studies in the academic literature, certain studies on computer vision and its direct use in industrial processes draw attention with their findings [3], [4]. Neural networks were used jointly in these studies and supervised learning approaches were determined and relatively high results were obtained on labeled data [5]–[7]. However, the literature on error detection developed with similar processes is quite limited, but it is expected that such studies will become widespread and increase their impact power in the near future [8]. The basic approach put forward is that the defects in processed materials can be resolved by computer vision methods, and there are certain computer vision-based modeling studies supporting this in the literature [9], [10].

Currently, there are studies in the field of quality control where object detection is performed by applying computer vision methods on some datasets (see e.g Srdoč et al. [11]). Moreover image processing methods are insufficient in different scenarios and different conditions in the industry. By using deep learning methods, algorithms that make accurate classifications can be created even in different scenarios and conditions. The aim of this article is to improve performance by improving quality management, to provide consumers with less faulty and problematic products by using deep learning methods with NEU-DET dataset.

II. Materials and Methods

A. Data Collection and Pre-Processing

The steel surface images used for this study were obtained from the NEU-DET dataset. This dataset includes 6 classes such as crazing, inclusion, patches, pitted surface, rolled-in scale, and scratches [11]. The total number of 1800 images is divided into 70% training, 15% validation and 15% testing. The total of the training data are 1260, validation data are 270 and the test data are 270. The collected images have a resolution of 200 x 200 pixels (px). However, the images were re-scaled to 299 x 299 px to stablish a base size. Since the number of data in the dataset is limited for this type of deep learning algorithm, data augmentation techniques were applied to avoid over-fitting and data size has been increased. The images in the dataset were rotated randomly on the horizontal and vertical axis, and rotated 20 degrees clockwise and counterclockwise. The pixel values of the images were scaled between 0 and 1 for an easier convergence of the model.

B. Model Architectures

Convolutional Neural Networks (CNNs) are widely used in image classification problems [12], [13]. CNN-based models learn useful features to reduce the error rate [14]. In this work, it uses state-of-the-art CNN architectures such as InceptionV3 [15], Xception [16], VGG [17] and ResNet [18] to predict steel surface defects. All models used have pre-trained ImageNet [19] weights. Since transfer learning is already efficient, the layer of models other than the Batch-Normalization layers

978-1-6654-7010-0/22/$31.00 ©2022 IEEE

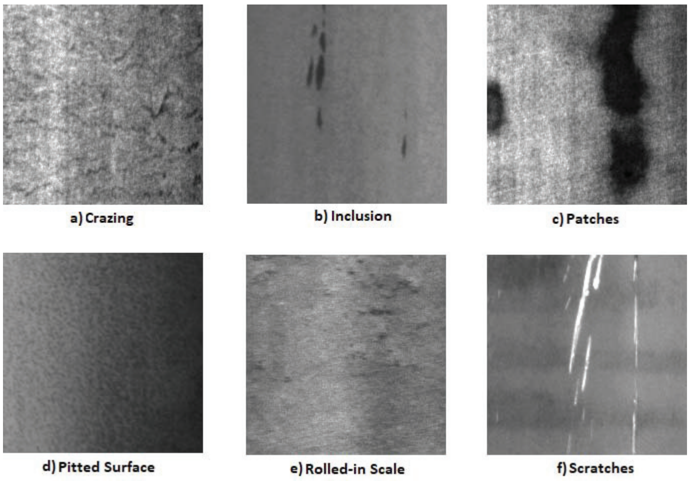

Fig. 1. Dataset examples

is frozen to avoid over-fitting [20]. To minimize the errors of deep learning models, Adam and SDG optimizers were used for each model in 4 different combinations with learning rates of 0.01 and 0.001. Considering all these parameters, experiments were done on 16 different combinations with four different state-of-the-art architectures, two different optimizers and two different learning rates, for 50 epochs.

C. Implementation

All implementations were done using Python language version 3.8 and TensorFlow 2.7, Matplotlib 3.5. Xception and InceptionV3 models were trained using NVIDIA Tesla K80 Graphic Processing Unit (GPU) on Ubuntu 18.04, and ResNet152 V2, VGG19 models were trained on Google Colab Pro environment using Tesla V100 GPU.

III. Metrics

The receiver operating characteristic and the area under the curve (Receiver Operating Characteristic - ROC and Area Under Curve - AUC) were used to evaluate the success of the model. The ROC curve is a plot of sensitivity to 1-specificity at various threshold settings for binary classifiers. The numerical value of the area under the AUC is defined and used by the model as a metric to measure successful decision making. However, this study is a multi-class problem, so the ROC curve is calculated as one versus stationary. This indicates that each class is evaluated against all other classes combined. To calculate the percentage of predicted values that matches with real values for one-hot encoded labels, categorical accuracy was used.

Categorical accuracy was calculated as:

$$\frac{1}{N}\sum_{i=1}^{N}\delta_i \tag{1}$$

where $\delta_i = \{1, 0\}$ if $\hat{y}[i] == y[i]$, N is the number of samples.

The AUC is calculated as follows:

$$AUC = \int TPR\,\mathrm{d}(\mathrm{FPR}) \tag{2}$$

The representations of equation 2 above are as follows:

- TPR: True Positive Ratio
- FPR: False Positive Ratio

A. Model Outputs and Evaluation

The pilot study was about classifying imperfections in the steel surface with limited data available using an artificial intelligence model. Among the 16 experiments performed to classify images, the results of the three most unsuccessful models are shown in Table I and the three successful ones are shown in Table II.

TABLE I
FAILED MODEL RESULTS

Model Name	Parameters and Results			
	Optimizer	Loss	Accuracy	AUC
VGG19	SGD @ 0.001	0.91	0.805	0.981
InceptionV3	Adam @ 0.01	0.61	0.74	0.725
InceptionV3	SGD @ 0.01	0.42	0.6870	0.88

TABLE II
SUCCESSFUL MODEL RESULTS

Model Name	Parameters and Results			
	Optimizer	Loss	Accuracy	AUC
Xception	**SGD @ 0.01**	**0.0201**	**0.9926**	**1.0**
ResNet152 V2	SGD @ 0.001	0.026	0.9926	1.0
VGG19	Adam @ 0.01	0.045	0.9816	0.99

As can be seen in table II, the Xception model, which is used with the SGD optimizer with a learning rate of 0.01 is the most successful model among 16 experiments. A total of 270 images was used as test data and the categorical accuracy was calculated as 99.26% and the AUC score as 100%. In addition, the ResNet152 V2 model, which has more parameters than Xception can be also accepted as the best model with a categorical accuracy of 99.26% and AUC score of 100%. As can be understood from table II, SGD optimizer found the best models. Moreover, VGG19 model is the third successful model with different optimizer Adam with a learning rate of 0.01. In addition, All these architectures are state-of-the-art models with high success rates.

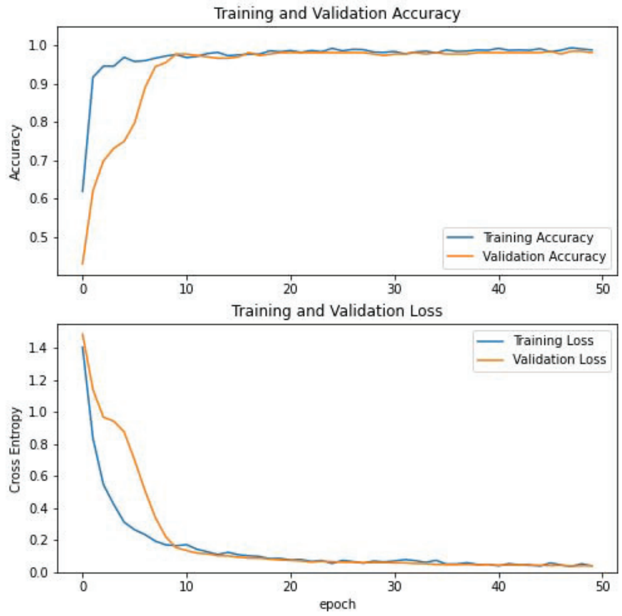

Fig. 2. Training and testing history for the most successful model Xception

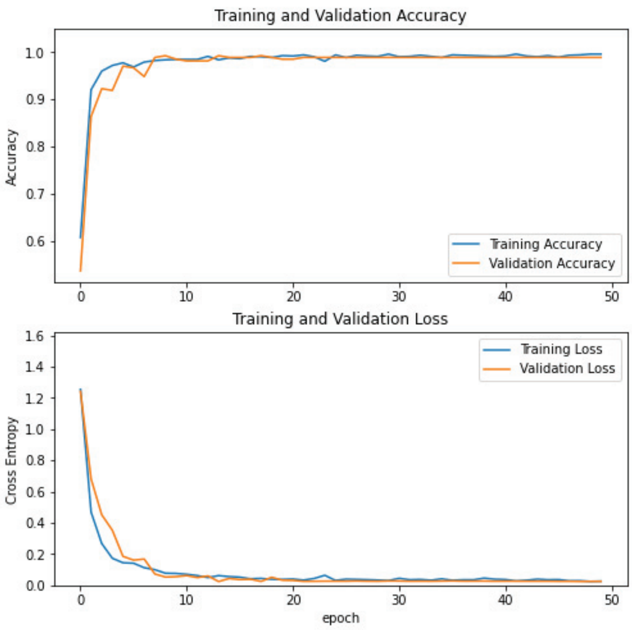

Fig. 3. Training and validation history for ResNet152V2 model

When the training and validation loss graph in Figure 2 is examined, it can be seen that training and validation categorical accuracy increases while training and validation loss decreases and approaches to zero. This graph proves that the Xception model converges successfully and there are no cases of over-fitting or under-fitting on the model architecture.

Furthermore, training history of the second successful model which is ResNet152 V2 can be examined in Figure 3. It can be seen that the model converges like the Xception model without under-fitting or over-fitting.

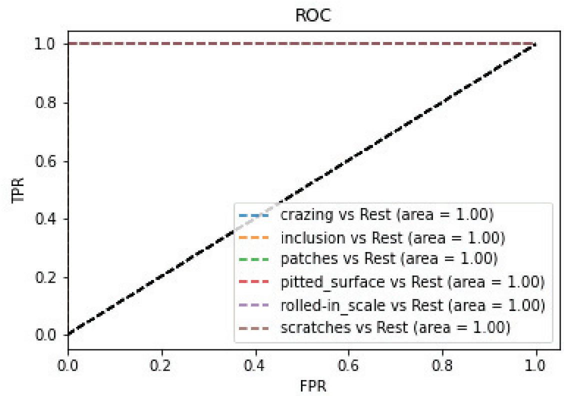

Fig. 4. ROC curve for both Xception and ResNet152 V2 models

The other metric ROC measures the performance of the model. From Figure 4, it can bee seen that the model has a strong attitude for each class and the precision value is close to 1.

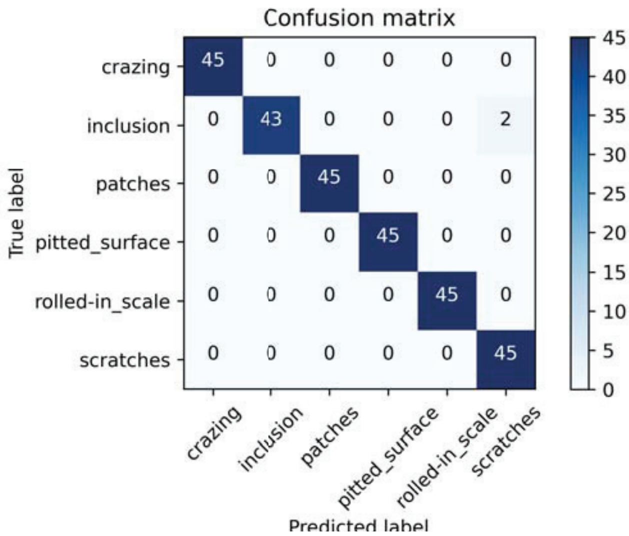

Fig. 5. Confusion Matrix of validation data for Xception model

Moreover, the Confusion Matrix metric visualizes and summarizes the performance of the model. As can be examined in Figure 5, there are 2 wrong predictions in the validation set where the categorical accuracy is 0.985. The model predicted the image as scratches while the correct label is inclusion.

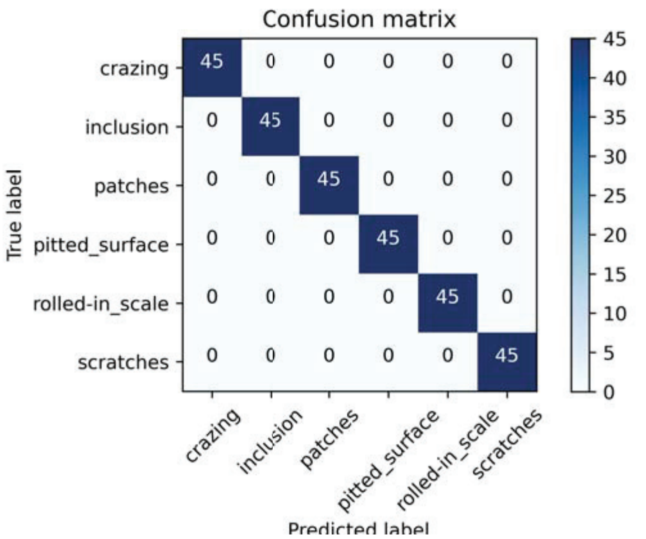

Fig. 6. Confusion Matrix of test data for both Xception and Resnet152 V2 models

According to Figure 6, it shows that the Xception model predicts a total of 270 images in 6 different classes accurately.

B. Discussion and Conclusion

As stated at the beginning of the study, especially in the next 20 years, the most appropriate part where artificial intelligence applications can be used in the production phase is the quality management and quality control departments, because of the operational cost of eliminating defective products is relatively high when it is done with human power [2]. If adequate quality control is not achieved, it has the potential to generate higher costs, given the excess of current transport and storage costs. Therefore, today, quality control has begun to evolve into a role that requires technological competence and sufficient equipment. It is possible that such technology-based systems will become an integral part of production processes in the next 30 years.

Currently, there are studies in the field of quality control where object detection is performed by applying computer vision methods on datasets (see eg Lv et al. [11]). Computer vision methods are insufficient in different scenarios and different conditions in the industry. By using deep learning methods, algorithms that make accurate detections can be created even in different scenarios and conditions. The purpose of this article is to improve quality management, to provide consumers with products that are less likely to be faulty using the NEU-DET dataset and deep learning methods. When the studies in the related literature are examined, the literature on error or malfunction detection is relatively limited, but it is expected that the studies in this field will gain momentum due to the high potential of industrial added value that successful applications will provide.

This study classifies the damages on steel surfaces using the NEU-DET dataset, pointing to the potential of quality control in this area using deep learning and artificial intelligence methods. When the research findings are evaluated, Xception, InceptionV3, VGG19 and ResNet152 V2 model architectures together with Adam and SGD optimizers, learning rates of 0.01 and 0.001 were used in order to find the optimal converged model. Thus, 16 different experiments were performed. Experiment outputs were compared according to AUC and accuracy values. Moreover, these test data were applied after the training was completed and the purpose here is to measure the performance of the model against data that it has never seen during the training. The SGD optimizer was the most successful model with the Xception model architecture with a learning rate of 0.01 with an categorical accuracy of 99.26% and an AUC score of 100%.

A study on another NEU-DET dataset [11] focused on object detection rather than image classification, and Faster-CNN [10], YOLO-V2 [23], YOLO-V3 [24] and VGG16 [25] large models was measured separately for each defect class and the overall predictive value of the best model was 72.4%. When this study and our methods and findings are compared, they have revealed structures that will consume a lot of energy using greater models, and their estimation rates are too low to be used in the industry. Moreover, in another study [26], it can be seen that the CNNs in the model they developed over the classification problem are more successful than the Support Vector Machine (SVM) [27]. The fact that the classification models developed within the scope of this study, work with less computing power and produce fast inference results that can be contributed to the production line in the industry.

In future research, the dataset can be expanded by performing real-time tests using the best-structured model structure, and also by adding a class of defect-free surface photos, this deep learning model can be used in a real production line. Thus, the quality control can be increased by using the machine power more effectively.

REFERENCES

[1] P. Santhanam, “Quality Management of Machine Learning Systems,” Jun. 2020. [Online]. Available: https://arxiv.org/pdf/2006.09529.

[2] A. GORCHET, “Deep Learning Has Reinvented Quality Control in Manufacturing—but It Hasn’t Gone Far Enough,” IEEE Spectrum, Nov. 20, 2020. https://spectrum.ieee.org/deep-learning-has-reinvented-quality-control-in-manufacturingbut-it-hasnt-gone-far-enough (accessed Nov. 07, 2021).

[3] Y.LeCun,L.Bottou,Y.Bengio,andP.Haffner,“Gradient-basedlearning applied to document recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, November 1998.

[4] S. Behnke, Hierarchical Neural Networks for Image Interpretation, ser. Lecture Notes in Computer Science. Springer, 2003, vol. 2766.

[5] P. Simard, D. Steinkraus, and J. Platt, “Best practices for convolutional neural networks applied to visual document analysis,” in Seventh International Conference on Document Analysis and Recognition, 2003, pp. 958–963.

[6] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, “Flexible, high performance convolutional neural networks for image classification,” in International Joint Conference on Artificial Intelligence, 2011, pp. 1237–1242.

[7] L. Martins, F. Pa anddua, and P. Almeida, “Automatic detection of surface defects on rolled steel using computer vision and artificial neural networks,” in IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society, 2010, pp. 1081 –1086.

[8] T. K. Leung and J. Malik, “Representing and recognizing the visual appearance of materials using three-dimensional textons,” International Journal of Computer Vision, vol. 43, no. 1, pp. 29–44, 2001.

[9] M. Varma and A. Zisserman, “Texture classification: Are filter banks necessary?” in IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2003, pp. 691–698.
[10] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” arXiv.org, 06-Jan-2016. [Online]. Available: https://arxiv.org/abs/1506.01497.
[11] X. Lv, F. Duan, J. Jiang, X. Fu, and L. Gan, “Deep Metallic Surface Defect Detection: The New Benchmark and Detection Network,” Sensors, vol. 20, no. 6, p. 1562, Mar. 2020, doi: 10.3390/s20061562.
[12] K. Pai and A. Giridharan, “Convolutional Neural Networks for classifying skin lesions,” IEEE Xplore, Oct. 01, 2019. https://ieeexplore.ieee.org/abstract/document/8929461 (accessed Nov. 07, 2021).
[13] W. Rawat and Z. Wang, “Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review,” Neural Computation, vol. 29, no. 9, pp. 2352–2449, Sep. 2017, doi: 10.1162/necoa00990.
[14] Trolice MP, Curchoe C, Quaas AM. Artificial intelligence—the future is now. J Assist Reprod Genet [Internet]. 2021 Jul 1 [cited 2021 Aug 30];38(7):1607–12. Available from: https://doi.org/10.1007/s10815-021-02272-4
[15] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. Las Vegas, NV, USA: IEEE; 2016 [cited 2021 Jul 26]. p. 2818–26. Available from: http://ieeexplore.ieee.org/document/7780677/
[16] Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. In: 2017 IEEE Conferenceon Computer Vision and Pattern Recognition (CVPR). 2017. p. 1800–7.
[17] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv14091556 Cs[Internet]. 2015 Apr 10 [cited 2021 Sep 4]; Available from: http://arxiv.org/abs/1409.1556
[18] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” arXiv:1512.03385 [cs], Dec. 2015, [Online]. Available: https://arxiv.org/abs/1512.03385?context=cs
[19] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” 2009 IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2009, doi: 10.1109/cvpr.2009.5206848.
[20] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” arXiv.org, 2015. https://arxiv.org/abs/1502.03167.
[21] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” arXiv.org, 2014. https://arxiv.org/abs/1412.6980.
[22] K. Fukushima, “Neocognitron: A self-organizing neural network for a mechanism of pattern recognition unaffected by shift in position,” Biological Cybernetics, vol. 36, no. 4, pp. 193–202, 1980.
[23] J. Redmon and A. Farhadi, “YOLO9000: Better, Faster, Stronger,” arXiv:1612.08242 [cs], Dec. 2016, [Online]. Available: https://arxiv.org/abs/1612.08242v1.
[24] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” arXiv.org, 2018. https://arxiv.org/abs/1804.02767.
[25] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv.org, 10-Apr-2015. [Online]. Available: https://arxiv.org/abs/1409.1556.
[26] J. Masci, U. Meier, D. Ciresan, J. Schmidhuber, and G. Fricout, “Steel defect classification with Max-Pooling Convolutional Neural Networks,” IEEE Xplore, Jun. 01, 2012. https://ieeexplore.ieee.org/abstract/document/6252468/ (accessed Feb. 08, 2022).
[27] Cristianini N., Ricci E. (2008) Support Vector Machines. In: Kao MY. (eds) Encyclopedia of Algorithms. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30162-4_415