Deep Learning-Based Surface Defect Detection in Steel Products Using Convolutional Neural Networks

Irfan Ullah Khan[1], Nida Aslam[1*], Menna Aboulnour[1], Asma Bashamakh[1], Fatima Alghool[2], Noorah Alsuwayan[2], Rawaa Alturaif[3], Hina Gull[4], Sardar Zafar Iqbal[4], Tariq Hussain[5]

[1] Department of Computer Science, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia
[2] Solutions Delivery Department, King Fahd University of Petroleum and Minerals, Dhahran 34463, Saudi Arabia
[3] D&IT Strategy and Investment Department, Saudi Aramco, Dhahran 34481, Saudi Arabia
[4] Department of Computer Information Systems, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia
[5] School of Computer Science and Technology, Zhejiang Gongshang University, Hangzhou 310018, China

Corresponding Author Email: naslam@iau.edu.sa

ABSTRACT

In mechanical engineering, monitoring steel surface defects is crucial for ensuring the quality of industrial products, as these defects account for over 90% of flaws in steel items. Traditional manual inspection methods are time-consuming and may overlook some defects. To address these challenges, this study introduces an automated deep learning (DL) model for continuous monitoring of steel surface defects using real-world images from the Industrial Machine Tool Component Surface Defect (IMTCSD) dataset, which includes 1,104 three-channel images, 394 of which are categorized as exhibiting "pitting" damage. This study evaluated several Convolutional Neural Network (CNN) classifiers: EfficientNetB3, ResNet-50, and MobileNetV2, to determine the most effective model for defect detection. EfficientNetB3 is distinguished by its scalable architecture that adapts efficiently across various image dimensions, making it ideal for high-accuracy applications on limited computational resources. ResNet-50 uses residual connections to maintain performance in deeper networks by facilitating smooth gradient flow, yet it requires more computational power. MobileNetV2, designed for real-time applications on devices with limited resources, uses lightweight depthwise separable convolutions. The performance of these models was assessed using accuracy, recall, precision, specificity, F1-score, and AUC metrics. EfficientNetB3 emerged as the best performing model, achieving an accuracy of 0.981, specificity of 0.975, recall of 0.987, precision of 0.975, and an F1-score of 0.982. This model proved effective in detecting defects even on dirty surfaces, demonstrating its potential to significantly enhance quality control in industrial settings.

1. INTRODUCTION

Steel manufacturing plays a vital role in the quality of industrial production. However, steel defects could go unnoticed and could cause problems in the quality of the end product. The surface defect issue is responsible for over 90% of the defects found in steel products [1]. Recently, steel surface defect detection is increasingly attracting interest and has been proven to ensure products meet the industry's quality standards [2]. A defect can be described as the blemish deficiency of an area compared to a regular sample. The process of surface defect detection can be explained as finding color impurity, scratches, apertures, or damage spots on a test sample [3, 4]. In addition, several environmental factors can hinder the surface defect detection process, such as the reflection of lights and illumination, leading to increased difficulty in the detection process [5]. Typically, defects in

steel are identified manually. However, this approach is cumbersome and may not identify all defects. Moreover, manual process suffers from numerous disadvantages including low accuracy, efficiency, and sampling rate, as well as high labor intensity [6]. Also worth mentioning, some defects may go unnoticed, leading to companies losing their customers' trust [2]. In contrast, the use of an automated image-based model for surface defect detection can alleviate the drawbacks of manual examination such as poor resource utilization, low accuracy, and its time-consuming nature [6, 7]. Furthermore, it provides a faster and more detailed inspection than current approaches. Automated surface defect detection models typically have two functionalities: defects segmentation and defects processing which involves features extraction and defect classification [8].

Multiple image processing approaches for surface defect detection in images have been proposed, namely segmentation

and thresholding-based techniques utilizing some operators like the Sobel filter and the canny edge detector. Nevertheless, these are considered to be pure image processing techniques that can handle only a limited number of simple instances. In recent years, feature-based techniques in addition to machine learning (ML) specifically CNN models have gained increased attention in the field of surface defect detection, through the use of image processing methods for feature extraction. Initially some studies only focused on ML algorithms in this field, such as Support Vector Machine (SVM) and Logistic Regression (LR) [6]. However, most research has shifted to CNN which is now widely implemented in surface defect detection.

Several review-based studies analyzed current research directions and highlighted the challenges faced when dealing with steel images. Luo et al. [9] reviewed the existing surface defect detection technologies and categorized them with respect to the image features and nature of algorithms into four groups which are statistical, spectral, model-based, and ML. Moreover, Neogi et al. [10] aimed to cover the different issues regarding automatic steel surface defect detection and classification systems. Their findings suggest that steel surface images are difficult to handle due to many factors including vibration, highly variant illumination, large amounts of noise due to surface scale, etc. Finally, Qi et al. [11] discussed the current applications of CNNs in a range of industrial settings for surface defect detection tasks. Their findings include the expected future challenges that will be faced which include the need for autonomous system that that can detect small defects, and future trends such as the use of transfer learning, lightweight networks, and generalized defect detection models.

The motivation and contribution of our study can be summarized as:

-The increased attention to automating manual procedures in industrial production such as steel surface defect detection.

-Developing a model that detects defects even in polluted surfaces in order to speed up defect detection.

-The development of a model that can detect and classify defects with relatively high accuracy for better product quality and less cost.

-Finally, exploring a dataset that possesses great potential and developing the model with enhanced results compared with the baseline models. As per the authors knowledge the dataset used in the current hasn't been used before for the classification task.

The remainder of this paper is structured as follows: Section 2 discusses the literature review of related work. Section 3 covers the materials and methods including dataset description, preprocessing techniques, and classifiers applied. Section 4 contains the experimental setup, optimization, and results. Section 5 presents further results and discussion while Section 6 contains the conclusion and recommendation.

2. LITERATURE REVIEW

Numerous studies have been performed to propose different surface defect detection techniques to find the best method. Initially, few studies used Machine Learning (ML) algorithms for surface defects. However, due to the significance of CNN models in automated analysis of the images, recent studies have shifted towards CNN models.

One of the studies that utilized ML algorithms and achieved a relatively high performance was conducted by Gong et al. [12]. The dataset that was used contained 4,120 images which included six types of defects. They proposed a method for multi-class steel defect detection and classification based on Multiple Hyperspheres Support Vector Machine with additional information (MHSVM+). The MHSVM+ model utilizes additional information to identify hidden information in the defect dataset. The classifier achieved an overall accuracy of 0.973.

Recent studies have utilized DL models and the most commonly used dataset is the NEU dataset which contains 1,800 images of six different types of defects with 300 images for each type. He et al. [13] used this dataset to build an end-to-end Defect Detection Network (EDDN). The main features of the model include the integration of a ResNet that is robust in defect classification, a Region Proposal Network (RPN) for specific defect classification and localization, and a Multilevel-feature Fusion Network (MFN) for fusing features to support providing defect localization details. The experiments conducted using the EDDN with a mean Average Precision (mAP) of 0.823 and an accuracy of 0.996 for defect classification.

Furthermore, Lee et al. [1] used the same dataset and proposed a method for steel defect detection using deep structured learning with class activation maps. In their work, they built a Convolutional Neural Network (CNN) model to localize and analyze defect areas in images and compared its performance against the performance of SVM and Logistic Regression (LR) models. They concluded that the CNN model outperformed the two other ML algorithms and obtained an accuracy and F1-score of 0.994 and 0.99 respectively.

Likewise, Huang et al. [14] developed multiple models including one that is tested and trained using the NEU dataset. They proposed a CNN-based model that can inspect very small defects and does not require high-frequency CPUs to run. They used lightweight bottleneck a pyramid of lightweight kernels to generate powerful features without being computationally exhaustive. Moreover, the decoder was designed using a similar lightweight approach. Furthermore, the model includes Atrous Spatial Pyramid Pooling (ASPP) and depth-wise separable convolution layers. Using these lightweight designs greatly reduced redundant calculations and weights. The proposed model achieved an accuracy of 1.

In a similar study by Hao et al. [15], an advanced object detection approach was used to develop a steel surface Defect Inspection Network (DIN). A Deformable Convolution Network (DCN) was designed to enhance feature extraction quality and adapt to different defect shapes. In addition, a feature fusion network using balanced feature pyramids was used to create feature maps capable of inspecting multiple-size defects. Their proposed model achieved a mAP of 0.805.

Moreover Boudiaf et al. [16] developed a model for defect recognition based on the AlexNet CNN and SVM using NUE dataset. By using the FC7 layer of the AlexNet, they reduced the number of attributes to only 7% of the extracted features from each image. The model reached an accuracy of 0.997.

Conversely, some studies used NEU dataset along with other datasets to build their models. Fu et al. [17] used the NEU dataset for training. However, for testing, they used the NEU dataset along with a diversity-enhanced dataset they presented. They developed a CNN model to classify steel surface defects with a backbone architecture of SqueezeNet. Their aim was to reach a high performance using only a few training samples. Two techniques were introduced to enhance

the proposed model's accuracy in defect recognition. First, the fine-tuning of the pre-trained model was applied to accurately characterize defects. Second, a Multi-Receptive Field (MRF) module was incorporated to yield scale-dependent high-level features for more precise classification. The proposed method reached an accuracy of 1.

In a study by Lv et al. [5], they used two datasets, i.e., NEU dataset and their GC10-DET dataset. The GC10-DET dataset contains 3,570 images covering ten types of defects. They developed a model using an EDDN that is based on the Single Shot MultiBox Detector. In this approach, for every location found on the feature map, the model divides the defect bounding boxes into a group of default boxes. These default boxes have a range of scales and aspect ratios. The model reached a mAP of 0.724 for the NEU dataset and 0.651 for the GC10-DET dataset.

Similarly, Arikan et al. [18] used the NEU dataset for testing. However, they used their own dataset consisting of 22,000 images to build their models. They developed three CNN models specifically to handle real-time processing speed and capacity in surface defect detection systems. The best achieved accuracy was 0.98 using their dataset and 0.995 using the NEU dataset.

Recently Ibrahim and Tapamo [19] have proposed the deep learning model for the surface defect detection using NEU dataset. They have integrated transfer learning model and the CNN model. For the feature extraction VGG16 was used while for the classification CNN model was used. The study has achieved very significant results with an overall accuracy of 0.994 and F1 score of 0.994.

Additionally, another study was performed using NEU dataset [20]. The study has used several deep parallel attention CNN model for the surface defect detection. They have achieved results similar to the previous study [19] with 0.995

accuracy, 0.996 precision, and 0.998 accuracy. The study has also included the thermal maps to visualize the defects detect by the proposed model.

Recently, Li et al. [21] have used NUE-DET dataset to find the defects using two modules. First the DL model has been used for feature extraction and then the second module was used to optimize the extracted feature fusion. The model achieved an accuracy of 0.731 and mAP@0.5.

Another commonly used dataset is the Severstal dataset which contains 12,568 images of steel sheets. Wang et al. [22] used the Severstal dataset to develop a method that combines an object detection and a classification model. The improved Region-CNN model was used to detect a range of defects by including Spatial Pyramid Pooling (SPP) and Feature Pyramid Networks (FPN). In addition, the improved ResNet50-vd model was used to detect multiple shapes of defects using a DCN and enhanced cutouts. Finally, the best accuracy achieved by the proposed method was 0.982.

Similarly, Konovalenko et al. [23] also used the Severstal dataset combined with more images to obtain a total of 87,704 images. They developed a method for classifying three types of rolled metal surface defects using residual CNN based on the ResNet50 NN and Stochastic Gradient Descent (SGD) for optimization and a binary loss function. This model reached a general accuracy of 0.969.

Another study by Abu et al. [2], used both the Severstal and NEU datasets. They used four types of transfer learning models, namely VGG16, MobileNet, DenseNet121, and ResNet101, to develop a model for steel surface detection. These models were tested for binary classification, using the Severstal dataset, and for multiclass classification, using the NEU dataset. Their results demonstrated that the MobileNet approach achieved the highest results of 0.804 for the binary classification and 0.969 for the multiclass classification.

Table 1. Summary of the literature review

Ref.	Technique	Dataset	Result
[12]	MHSVM+	4,120 images	Accuracy= 0.973
[13]	EDDN	NEU dataset	mAP= 0.823 for defect detection Accuracy= 0.996
[1]	CNN	NEU dataset	Accuracy= 0.994 F1-score= 0.99
[14]	CNN	NEU dataset	Accuracy = 1
[15]	DIN	NEU dataset	mAP = 0.805
[16]	AlexNet CNN and SVM	NEU dataset	Accuracy = 0.997
[17]	CNN	NEU dataset, 1,800 images diversity-enhanced dataset, 5,400 images	Accuracy = 1
[5]	EDDN	NEU dataset, 1,800 images GC10-DET dataset, 3,570 images	mAP = 0.724 for the NEU dataset mAP = 0.651 for the GC10-DET
[18]	CNN	NEU dataset, 1,800 images Private dataset, 22,000 images	Accuracy = 0.98 using their dataset Accuracy = 0.995 using NEU dataset
[19]	CNN + VGG16	NEU dataset	Accuracy=0.994 F1 score=0.994
[20]	CNN	NEU dataset	Accuracy=0.995 Precision=0.996 AUC=0.998
[21]	DL	NEU dataset	accuracy =0.731 and mAP@0.5
[22]	Region-CNN2, ResNet50-vd	Severstal dataset, 12,568 images	Accuracy = 0.982
[23]	Residual CNN	Severstal dataset combined with more images 87,704 images	Accuracy = 0.969
[2]	MobileNet	Severstal and NEU dataset	Accuracy = 0.804 for binary classification Accuracy = 0.969 for multiclass classification
[24]	CNN	22,408 rail track images	Accuracy= 0.92
[25]	CASAE, compact CNN	50 images	Accuracy= 0.868
[26]	CAE-SGAN	21,000 images	Accuracy = 0.986
[6]	Regression-based model	AigleRN dataset, 38 images DAGM2007 dataset, 11,500 images Capacitor dataset, 3,839 images	F1-score= 0.938 for the AigleRN dataset F1-score= 0.915 for the DAGM2007 dataset Accuracy= 0.92 for the capacitor dataset

[27]	CNN	RSDDs,NRSD-MN NEU-DET,BSData	mAP = 0.985 for the private dataset mAP = 0.867 for the RSDDs dataset mAP = 0.82 for the BSData dataset mAP = 0.810 for the NRSD-MN dataset mAP = 0.746 for the NEU-DET

Comparatively, some studies used other datasets to develop their models. Faghih-Roohi et al. [24] proposed a deep CNN to automate the detection of defects in rail surfaces. The dataset used to develop the model contains 22,408 rail track images that were manually labeled as normal or categorized into 1 of 5 types of defects. Furthermore, a mini-batch gradient descent method was used to optimize the entire network. They performed a comparative analysis of three deep CNN architectures and concluded that the largest architectures outperformed the two smaller ones by achieving an accuracy of 0.92.

Another approach was introduced by Tao et al. [25], which is a two-step approach to automatically detect metallic defects by segmenting and classifying them. The dataset that was used in this study consisted of 50 metallic defect images. For segmenting defects, the Cascaded Autoencoder (CASAE) architecture was developed. Classification of the defected areas of the segmented images into their distinct classes was performed using compact CNN. The proposed method achieved an accuracy of 0.868.

Similarly, semi-supervised learning method was proposed by Di et al. [26] to classify the defects. The proposed method is CAE-SGAN which is based on Convolutional Autoencoder (CAE) and semi-supervised Generative Adversarial Networks (SGAN). A passthrough layer was used to assist the CAE in extracting features. The dataset used in this study which contains around 21,000 images part of which were randomly generated using the SGAN. The highest classification rate reached was 0.986.

In another approach, He and Liu [6] aimed to detect and classify industrial defects using a four-stage framework. The proposed method is a regression-based model which uses CNN to predict the severity of the defects. Furthermore, the model classifies the defects into their respective types and reduces false positive rate. They tested their model on three datasets, AigleRN dataset containing 38 images, DAGM2007 dataset containing 11,500 images, and an in-house capacitor dataset containing about 3,839 images. The model achieved an F1-score of 0.938 using the AigleRN dataset, and 0.915 using

the DAGM2007 dataset. As for the capacitor dataset, the model achieved an overall accuracy of 0.92.

Recently, Li et al. [27] have used path aggregation network (PANet) to identify the defects in the steel surface. The study has utilized augmentation techniques for model generalization. They models were trained and tested using five datasets such as RSDDs, NRSD-MN, NEU-DET, BSData and one private dataset. The aim of the study was detection, and the highest mAP@.5 of 0.985 was achieved using their own dataset, 0.82 mAP was BSData.

In this study we aim to latest open-source dataset. Therefore, we have used IMTCSD dataset provided by Schlagenhauf and Landwehr [28]. This dataset provides images collected under real-world conditions. Furthermore, the dataset will enable us to perform classification without the need to collect different datasets or perform manual labeling on the images. To further enhance defect detection systems, we aim to enable detection even on polluted surfaces to allow early detection. Table 1 contains the summary of the previous related studies.

3. MATERIALS AND METHODS

CNN is the most widely used neural network model in the field of image classification, as with CNN. Presently, there are numerous types of CNN architectures such as ResNet, AlexNet, EfficientNet, DenseNet, MobileNet, U-Net, etc. In this paper, we aim to perform a comparative analysis among three CNN-based classifiers, namely EfficientNetB3, ResNet-50, and MobileNetV2, to find and construct an enhanced classifier for component surface defect detection. The CNN model will go through several steps, starting with filtering or convolution, nonlinearity using ReLU function, pooling to reduce the size of the activation cards, and finally classification [16]. To achieve this, we will use the Industrial Machine Tool Component Surface Defect (IMTCSD) dataset that provides an updated dataset of real-world defect images to enable building a reliable and robust model.
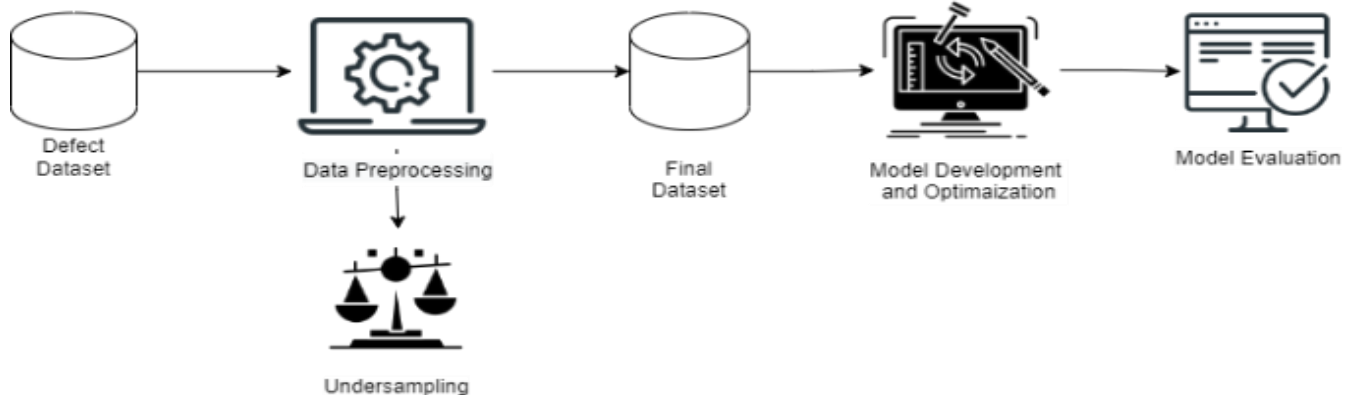

Figure 1. Proposed methodology pipeline

The proposed method to build a component surface defect detection model will be achieved in three steps: preprocessing the IMTCSD using the necessary procedures, developing, and

training the models, and model testing as demonstrated in Figure 1. The dataset will be divided into 80-20% holdout split for training and testing, respectively. Furthermore, the model's

performance will be evaluated in terms of accuracy, F1-score, recall, precision, specificity, and Area Under the Curve (AUC).

3.1 Description of the dataset

The IMTCSD dataset [28] was produced under a real-world setup consisting of real-world examples that are hard to classify. The dataset includes 1,104 channel-3 images with 394 images categorized as "pitting" surface damage type. Moreover, it contains labeled images, to prevent the need for substantial domain knowledge. Furthermore, the dataset reflects a small interclass variance as well as a large intraclass variance. One of the unique features of this dataset is that it includes data depicting various failure stages, making it possible to use it for wear prognosis and to detect defects at an early stage. Hence, the dataset could be employed as a benchmark dataset to develop defect classification and prognostic models for the industrial setting. Consequently, this is the first practical dataset that allows segmentation, classification, and defect prediction from a single source. One common problem in ML is data imbalance when one class has a larger number of observation than the other. To address this issue This issue can be solved using several techniques, the most common is data under-sampling. Undersampling was performed using State of the Art (SOTA) approach. In this approach clustering was used to find the highly similar images. Highly similar images were removed from the majority class [29]. Before applying under-sampling, the dataset consisted of 1,104 samples and after the under-sampling, the number of samples decreased to 788 samples with 394 samples in each class. Figure 2 demonstrates the number of samples per category before and after applying the data undersampling.
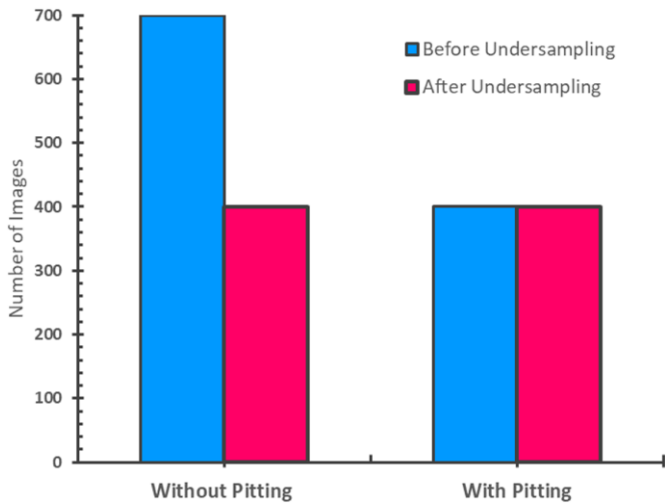

Figure 2. Dataset statistics before and after applying SOTA undersampling

3.2 Description of the classifiers

The section below presents the description of the proposed techniques.

3.2.1 EfficientNetB3

EfficientNet is a family of models based on the CNN architecture with the base model EfficientNet-B0 which was developed using a multi-objective neural architecture search that aims to increase accuracy and floating-point operations.

The full family scales from B1 to B7, we employed EfficientNetB3 for this study. The EfficientNetB3 model is a more powerful scaling method that is based on uniformly scaling the depth, width, and resolution dimensions in a standard way instead of the traditional arbitrary scaling. A set of fixed scaling coefficients, ø, is used to perform compound scaling of the three dimensions as shown in Eq. (1). This compound coefficient represents the number of more computational resources that can still be used for model scaling.

Depth: $d = \alpha^{ø}$, Width: $w = \beta^{ø}$, Resolution: $w = \beta^{ø}, r = \gamma^{ø}$.

$$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2, \alpha \geq 1, \beta \geq 1, \gamma \geq 1 \qquad (1)$$

Hence, if the aim is to use 2^N times high computational resources, the network depth, width, and resolution are increased. The scaling is rationalized based on the perception that the bigger the input image is the more layers needed in the network to boost the receptive field and the more channels needed to secure more clear patterns.

Moreover, EfficientNetB3 utilizes transfer learning to save time and resources. Thus, it often outperforms other Convolutional Neural Networks (ConvNets) such as AlexNet, ResNet, and DenseNet [30].

3.2.2 ResNet-50

The Residual Network (ResNet-50) is one of the most popular neural network architectures. Its main advantage is its ability to overcome the vanishing gradient problem that has led to the failure of many previous CNN architectures as they unnecessarily keep learning deeper and deeper while the gradients are close to zero meaning the network isn't learning anymore. The vanishing gradient problem is overcome by ResNet-50 by using the skip connection technique also known as identity mapping.

This technique basically adds the output from a preceding layer, *x*, with a desirable accuracy to the following layer and adds only one layer. However, sometimes the dimensions in one layer, *x*, do not match the dimensions of the following layer, *F(x)*. In this case, as in Eq. (2) the identity mapping is multiplied by a *W*, a linear projection, to expand the channels of shortcuts to match the residual. Hence, the network learns at a more rapid rate. In addition, this technique enables combining the input from *x* and the input from *F(x)* into a single input to the following layer and allows the network to explore larger feature spaces [31].

$$y = F(x, \{W_i\} + W_s x) \qquad (2)$$

3.2.3 MobileNetV2

Considering the complexity and the large number of parameters used in CNN models, it demands powerful hardware availability to run smoothly, which limits their application area. The MobileNetV2 is a lightweight deep neural network proposed by Tan et al. [32] from Google's research team with the aim to allow DL models to run on mobile and embedded devices. It was the first mobile-friendly computer vision family of models for TensorFlow, as it acknowledges the limitation in resources in these devices yet still aims to achieve high accuracy rates. Moreover, this model's main aim is to minimize resource usage as it requires less space, power, and time. Hence, it reduces latency and amplifies learning speed. In addition, the model uses depth-

wise separable convolutions which reduces the number of parameters needed, making the model a lighter deep neural network compared to other ConvNets with regular convolutions. Hence, it is a popular large-scale model designed to fit a vast range of applications, e.g., segmentation, detection, and classification [32].

## 4. EXPERIMENTAL SETUP AND RESULTS

The proposed models were built using Google Colab's GPU using Python 3.8.0. We applied holdout method for partitioning the dataset into 80% for training and 20% for testing. In addition to the classifiers, some of the additional layers were added to the classifiers are shown in Table 2. The final classification model architecture is shown in Figure 3, and the classifiers parameters are shown in Table 3.

Experiments were performed in the original dataset and the undersampled dataset. Table 4 represents the performance metrics of all classifiers before and after applying data undersampling. The accuracy values of EfficientNetB3 increased after undersampling the data. However, ResNet-50 and MobileNetV2 accuracy slightly decreased. As shown in the Table 4, the highest obtained accuracy is 0.991, which was achieved by the ResNet-50 and MobileNetV2 classifiers before applying undersampling. As for after applying undersampling, all three classifiers achieved the highest accuracy of 0.981.

Table 2. Layers parameters in the three proposed classifiers

| Layer	Number of Neurons	Activation Function
Dense	128	ReLu
Dense	64	ReLu
Dense	1	Sigmoid
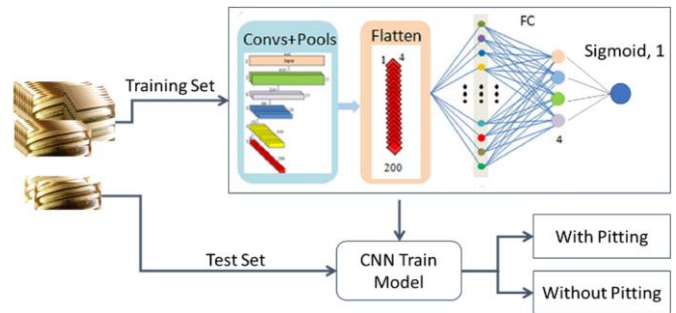
Figure 3. Deep learning model for defect detection

Table 3. Parameters for the three proposed classifiers

Model	Parameters	Value
	Batch Size	16
EfficientNetB3	Loss Function	Binary cross-entropy
	Optimizer	Adam
ResNet-50	Epochs	200
		monitor = 'val_accuracy'
		factor = 0.3
MobileNetV2	ReduceLROnPlateau	patience = 3
		min_lr = 0.000001

Table 4. Classifier performance comparison before and after applying data undersampling

Sampling	Classifier	Accuracy	Precision	Specificity	Recall	F1	AUC
Without Sampling	EfficientNetB3	0.968	0.986	0.993	0.923	0.956	0.989
	ResNet-50	0.991	0.986	0.992	0.986	0.989	0.992
	MobileNetV2	0.991	0.989	0.988	0.989	0.991	0.999
With Sampling	EfficientNetB3	0.981	0.975	0.975	0.987	0.982	0.992
	ResNet-50	0.981	0.985	0.990	0.985	0.975	0.998
	MobileNet V2	0.981	0.977	0.960	0.988	0.980	0.985
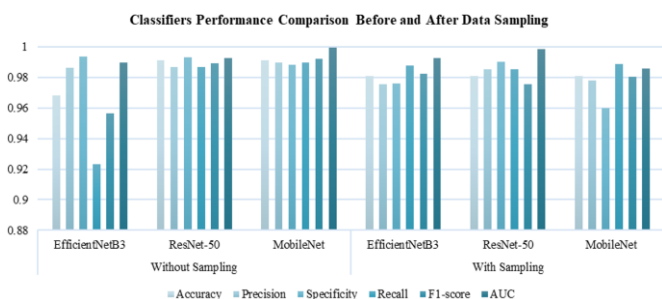
Figure 4. Classifiers' performance before and after applying data undersampling

Figure 4 shows the difference in performance before and after data sampling for all the three classifiers. Before applying data sampling, the EfficientNetB3 recall was notably lower. A low recall score implies that the false-negative rate is high, which means the model classified many instances belonging to the 'with pitting' class as 'without pitting'. In an industrial setting, this would cause an increase in the financial costs because many defective surfaces would go unnoticed and therefore would not solve our initial problem of identifying defective surfaces. However, the EfficientNetB3 performance significantly improved after data sampling. The accuracy and AUC slightly increased whereas the recall and F1-score showed a significant increase. On the other hand, the precision and specificity slightly decreased. Moreover, the ResNet-50 classifier maintained similarly excellent performance through the two scenarios, in which the AUC increased after applying data sampling, and reached a value of 0.9984. After applying data sampling, all classifiers obtained the same accuracy of 0.981 where EfficientNetB3 and ResNet-50 achieved competitive performance with regard to the rest of the evaluation metrics. However, the MobileNetV2 classifier's specificity decreased after sampling, which might cause unnecessary costs for manufacturers since the classifier will

predict some undamaged metals as damaged. In addition, the ResNet-50 performance slightly surpasses the EfficientNetB3 classifier regarding the precision, specificity, and AUC. However, EfficientNetB3 achieved a slightly higher recall of 0.987 and since a higher recall value will provide fewer missed defected metals, we can conclude that the EfficientNetB3 classifier can be employed in industrial settings and contribute significantly to identifying metallic surface defects.

5. DISCUSSION

Deep Learning (DL) has been widely used for the visual data [33, 34]. Nowadays, technology plays an essential role in determining product quality in industrial production. In steel manufacturing, defects can go unnoticed which can lead to a decline in product quality. A defect can be defined as an imperfection in a specific area when compared to an undamaged sample [3]. The typical practice is to identify these defects manually, which is a tiresome and error-prone task. However, by using intelligent autonomous methods we can avoid most of the manual inspection's disadvantages such as low accuracy, efficiency, and sampling rate [6].

Table 4 shows the EfficientNetB3 model achieved the best overall performance after applying undersampling and obtained an accuracy of 0.981.

The issue of undetected metal defects can lead to severe consequences, including substantial financial costs. Therefore, it is critical that classifiers minimize the false negative rate, ideally aiming for zero to ensure no defects go unnoticed. Figure 5 presents the confusion matrices for the three classifiers—EfficientNetB3, MobileNetV2, and ResNet-50—after applying undersampling techniques. Both MobileNetV2 and EfficientNetB3 recorded only one false negative and two false positives, whereas ResNet-50 registered two false negatives and one false positive. The former scenario is preferable, as missing fewer defects is crucial to avoid increasing manufacturing costs.
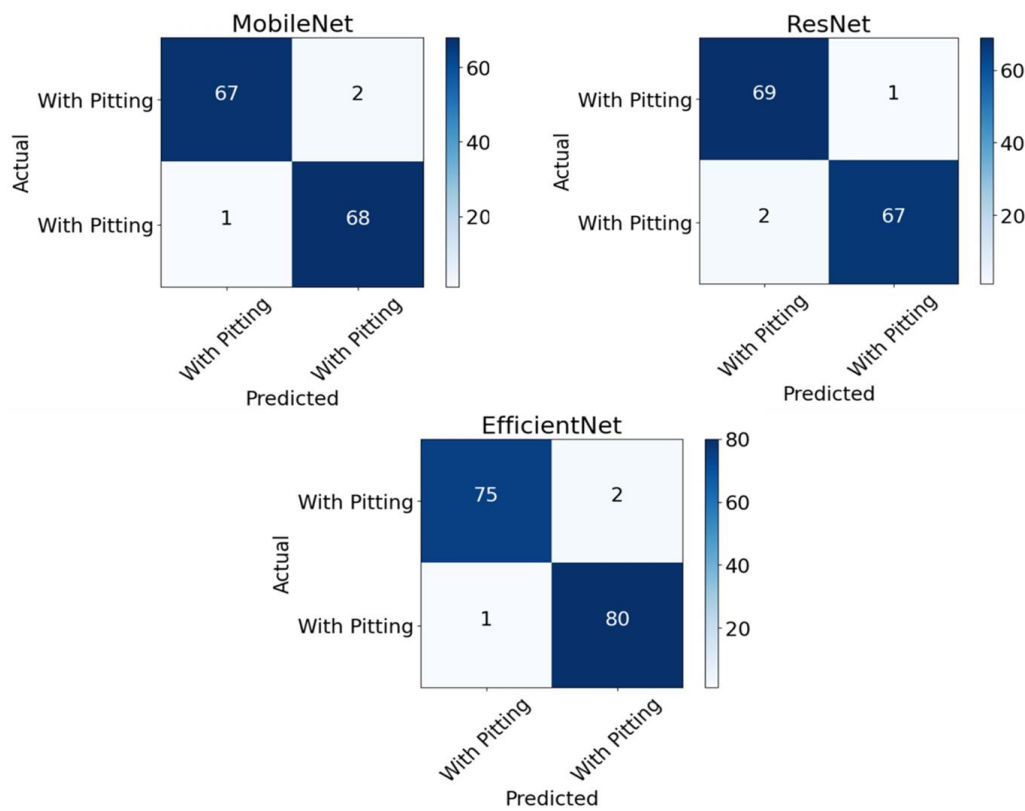

Figure 5. (a) MobileNetV2 confusion matrix, (b) ResNet-50 confusion matrix, (c) EfficientNetB3 confusion matrix

The IMTCSD dataset, recently utilized for a detection task, achieved a mean Average Precision (mAP) of 0.82 [27]. Unlike previous studies that focused on detection, this study has performed classification and achieved significant results. A key advantage of the current dataset is its versatility, supporting not just classification but also detection and segmentation tasks, enhancing its utility in addressing industrial quality control challenges.

6. CONCLUSION

In this paper, a comparative analysis was conducted among three CNN-based classifiers, namely EfficientNetB3, ResNet-50, and MobileNetV2, to construct an enhanced classifier for defect detection on industrial machine tools. The IMTCSD dataset was used to develop the models to enable them to detect defects even on polluted surfaces. The proposed method is composed mainly of three basic steps: preprocessing the IMTCSD dataset using the necessary techniques, initiating the experiments, and training the models, and model testing. Data undersampling was applied to eliminate the data imbalance. The effectiveness of the proposed method is evaluated in terms of accuracy, F1-score, recall, precision, specificity, and AUC. The best performing model was EfficientNetB3 which achieved an accuracy of 0.981, recall of 0.987, precision of 0.975, and F1-score of 0.982. Nevertheless, the study has achieved significant results but there is still room for further improvement. In future, we hope to apply the vision transformers and also investigate the performance of the

proposed algorithms on other real dataset for predicting the remaining lifetime of the components based on the classification of its defect severity stages.

REFERENCES

[1] Lee, S.Y., Tama, B.A., Moon, S.J., Lee, S. (2019). Steel surface defect diagnostics using deep convolutional neural network and class activation map. Applied Sciences, 9(24): 5449. https://doi.org/10.3390/app9245449

[2] Abu, M., Amir, A., Lean, Y.H., Zahri, N.A.H., Azemi, S.A. (2021). The performance analysis of transfer learning for steel defect detection by using deep learning. Journal of Physics: Conference Series, 1755(1): 012041. https://doi.org/10.1088/1742-6596/1755/1/012041

[3] Chen, Y., Ding, Y., Zhao, F., Zhang, E., Wu, Z., Shao, L. (2021). Surface defect detection methods for industrial products: A review. Applied Sciences, 11(16): 7657. https://doi.org/10.3390/app11167657

[4] Saberironaghi, A., Ren, J., El-Gindy, M. (2023). Defect detection methods for industrial products using deep learning techniques: A review. Algorithms, 16(2): 95. https://doi.org/10.3390/a16020095.

[5] Lv, X., Duan, F., Jiang, J.J., Fu, X., Gan, L. (2020). Deep metallic surface defect detection: The new benchmark and detection network. Sensors, 20(6): 1562. https://doi.org/10.3390/s20061562.

[6] He, Z., Liu, Q. (2020). Deep regression neural network for industrial surface defect detection. IEEE Access, 8: 35583-35591. https://doi.org/10.1109/ACCESS.2020.2975030

[7] Ibrahim, A.A.M., Tapamo, J.R. (2024). A survey of vision-based methods for surface defects' detection and classification in steel products. Informatics, 11(2): 25. https://doi.org/10.3390/informatics11020025

[8] Xiao, M., Jiang, M., Li, G., Xie, L., Yi, L. (2017). An evolutionary classifier for steel surface defects with small sample set. EURASIP Journal on Image and Video Processing, 2017(1): 1-13. https://doi.org/10.1186/s13640-017-0197-y

[9] Luo, Q., Fang, X., Liu, L., Yang, C., Sun, Y. (2020). Automated visual defect detection for flat steel surface: A survey. IEEE Transactions on Instrumentation and Measurement, 69(3): 626-644. https://doi.org/10.1109/TIM.2019.2963555

[10] Neogi, N., Mohanta, D.K., Dutta, P.K. (2014). Review of vision-based steel surface inspection systems. EURASIP Journal on Image and Video Processing, 2014: 1-19. https://doi.org/10.1186/1687-5281-2014-50

[11] Qi, S., Yang, J., Zhong, Z. (2020). A review on industrial surface defect detection based on deep learning technology. In Proceedings of the 2020 3rd International Conference on Machine Learning and Machine Intelligence, Hangzhou, China, pp. 24-30. https://doi.org/10.1145/3426826.3426832

[12] Gong, R., Wu, C., Chu, M. (2018). Steel surface defect classification using multiple hyper-spheres support vector machine with additional information. Chemometrics and Intelligent Laboratory Systems, 172: 109-117. https://doi.org/10.1016/j.chemolab.2017.11.018.

[13] He, Y., Song, K., Meng, Q., Yan, Y. (2019). An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. IEEE Transactions on Instrumentation and Measurement, 69(4): 1493-1504. https://doi.org/10.1109/TIM.2019.2915404

[14] Huang, Y., Qiu, C., Wang, X., Wang, S., Yuan, K. (2020). A compact convolutional neural network for surface defect inspection. Sensors, 20(7): 1974. https://doi.org/10.3390/s20071974.

[15] Hao, R., Lu, B., Cheng, Y., Li, X., Huang, B. (2021). A steel surface defect inspection approach towards smart industrial monitoring. Journal of Intelligent Manufacturing, 32: 1833-1843. https://doi.org/10.1007/s10845-020-01670-2

[16] Boudiaf, A., Benlahmidi, S., Harrar, K., Zaghdoudi, R. (2022). Classification of surface defects on steel strip images using convolution neural network and support vector machine. Journal of Failure Analysis and Prevention, 22(2): 531-541. https://doi.org/10.1007/s11668-022-01344-6

[17] Fu, G., Sun, P., Zhu, W., Yang, J., Cao, Y., Yang, M.Y., Cao, Y. (2019). A deep-learning-based approach for fast and robust steel surface defects classification. Optics and Lasers in Engineering, 121: 397-405. https://doi.org/10.1016/j.optlaseng.2019.05.005

[18] Arikan, S., Varanasi, K., Stricker, D. (2019). Surface defect classification in real-time using convolutional neural networks. arXiv preprint arXiv:1904.04671. https://doi.org/10.48550/arXiv.1904.04671

[19] Ibrahim, A.A.M., Tapamo, J.R. (2024). Transfer learning-based approach using new convolutional neural network classifier for steel surface defects classification. Scientific African, 23: e02066. https://doi.org/10.1016/j.sciaf.2024.e02066.

[20] Zhao, Y., Sun, X., Yang, J. (2023). Automatic recognition of surface defects of hot rolled strip steel based on deep parallel attention convolution neural network. Materials Letters, 353: 135313. https://doi.org/10.1016/j.matlet.2023.135313

[21] Li, Z., Wei, X., Hassaballah, M., Li, Y., Jiang, X. (2024). A deep learning model for steel surface defect detection. Complex & Intelligent Systems, 10(1): 885-897. https://doi.org/10.1007/s40747-023-01180-7

[22] Wang, S., Xia, X., Ye, L., Yang, B. (2021). Automatic detection and classification of steel surface defect using deep convolutional neural networks. Metals, 11(3): 388. https://doi.org/10.3390/met11030388

[23] Konovalenko, I., Maruschak, P., Brezinová, J., Viňáš, J., Brezina, J. (2020). Steel surface defect classification using deep residual neural network. Metals, 10(6): 846. https://doi.org/10.3390/met10060846

[24] Faghih-Roohi, S., Hajizadeh, S., Núñez, A., Babuska, R., De Schutter, B. (2016). Deep convolutional neural networks for detection of rail surface defects. In 2016 International Joint Conference on Neural Networks (IJCNN): Vancouver, BC, Canada, pp. 2584-2589. https://doi.org/10.1109/IJCNN.2016.7727522

[25] Tao, X., Zhang, D., Ma, W., Liu, X., Xu, D. (2018). Automatic metallic surface defect detection and recognition with convolutional neural networks. Applied Sciences, 8(9): 1575. https://doi.org/10.3390/app8091575.

[26] Di, H., Ke, X., Peng, Z., Dongdong, Z. (2019). Surface defect classification of steels with a new semi-supervised learning method. Optics and Lasers in Engineering, 117:

40-48. https://doi.org/10.1016/j.optlaseng.2019.01.011

[27] Li, G., Shao, R., Wan, H., Zhou, M., Li, M. (2022). A model for surface defect detection of industrial products based on attention augmentation. Computational Intelligence and Neuroscience, 2022(1): 9577096. https://doi.org/10.1155/2022/9577096

[28] Schlagenhauf, T., Landwehr, M. (2021). Industrial machine tool component surface defect dataset. Data in Brief, 39: 107643. https://doi.org/10.1016/j.dib.2021.107643

[29] Lehmann, D., Ebner, M. (2022). Subclass-based undersampling for class-imbalanced image classification. In Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2022), pp. 493-500. https://doi.org/10.5220/0010841100003124

[30] Tan, M., Le, Q.V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946. https://doi.org/10.48550/arXiv.1905.11946

[31] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778. https://doi.org/10.1109/CVPR.2016.90

[32] Tan, M.X., Le, Q.V. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. https://doi.org/10.48550/arXiv.1905.11946

[33] Mbilong, P.M., Aarab, Z., Belouadha, F.Z., Kabbaj, M.I. (2023). Enhancing fault detection in CNC machinery: A deep learning and genetic algorithm approach. Ingénierie des Systèmes d'Information, 28(5): 1361-1375. https://doi.org/10.18280/isi.280525

[34] Aslam, N., Khan, I.U., Albahussain, T.I., Almousa, N.F., Alolayan, M.O., Almousa, S.A., Alwhebi, M.E. (2022). MEDeep: A deep learning based model for memotion analysis. Mathematical Modelling of Engineering Problems, 9(2): 533-538. https://doi.org/10.18280/mmep.090232