Scientific African 23 (2024) e02066

Contents lists available at ScienceDirect

Scientific African

journal homepage: www.elsevier.com/locate/sciaf

Transfer learning-based approach using new convolutional neural network classifier for steel surface defects classification

Alaa Aldein M.S. Ibrahim *, Jules R. Tapamo

Discipline of Electrical, Electronic and Computer Engineering, University of KwaZulu-Natal, Durban, KwaZulu-Natal 4041, South Africa

A R T I C L E I N F O

Editor: DR B Gyampoh

Keywords:
Convolutional neural network
Deep learning
Feature extraction
Machine learning
Steel defects
Transfer learning
VGG16

A B S T R A C T

Automatic surface defect detection of industrial products using visual inspection has progressively replaced manual defect detection of steel strips and become a necessary part of industrial product surface defect detection of steel strips. Various steel products exhibit a wide range of surface defects. Moreover, these defects show significant diversity and similarities, posing challenges in their classification. As a result, the models currently used for identifying these defects suffer from the challenge of low accuracy, which leaves ample opportunities for further enhancement. This paper aims to improve defect detection and classification accuracy using a new approach that combines part of a pre-trained VGG16 model as a feature extractor and a new convolutional neural network (CNN) as a classifier for classifying six types of defects appearing on steel surfaces. The experimental results have shown that our proposed method can effectively classify a variety of steel surface defects. A comparison with state-of-the-art methods shows the superiority of the proposed method.

Introduction

The detection and classification of surface defects are essential for ensuring the quality and reliability of industrial products. Defects on surfaces during manufacturing processes can result in significant financial losses, com- promised safety, and decreased product lifespan. Consequently, it has become crucial to develop and implement effective and accurate systems for detecting and classifying surface defects in a variety of products, such as steel [1,2], fabric [3], rail [4], and other surface items [5,6]. With the development of industrial automation and intelligent control in steel production, a defect detection system for steel surface has been widely used in the steel industry, which in turn led to improvement in the productivity of the company's steel strip. Presently, automatic inspection systems for steel surfaces are produced by many companies on the surface, such as Parsytec, Siemens-VAI, EES, Ma- tra, Sipar, etc, [7]. During steel production, defects such as Scratches, roll marks, inclusion, edge crack, crazing, and scales can be generated mainly due to environmental states. According to statistical data, more than 90% of defects occur on the surface of the steel due to a number of causes

Abbreviations: VGG, very deep convolutional networks designed by visual geometry group; NEU, northeastern university; SVM, support vector machine; NNC, nearest neighbor clustering; KNN, K-nearest neighbors algorithm; MLR, multiple linear regression; GLCM, gray level co-occurrence matrix; HOG, histogram of oriented gradients; AELTP, adaptive extended local ternary pattern; AECLBP, adjacent evaluation completed local binary patterns; CNN, convolutional neural network; AVDDCS, automatic visible defect detection and classification system; YOLOv4, you look only once network version 4; FPN, feature pyramid network; R-CNN, region convolutional neural network; DCNN, deep convolutional neural network WRN, wide residual networks; ReLU, rectified linear; Unit RF, random forest.

* Corresponding author.
E-mail address: alaaaldein@fashir.edu.sd (A.A.M.S. Ibrahim).

https://doi.org/10.1016/j.sciaf.2024.e02066
Received 16 September 2023; Received in revised form 25 November 2023; Accepted 9 January 2024
Available online 17 January 2024

like stresses, temperature, wearing and solidification, and corrosion manufacturing defects [8]. These defects may lead to product rejection by the customer, which causes a lot of financial damage to the production plant. In the past, surface defects inspection has been done by human inspectors by monitoring the differences in the surface appearance of the product with the naked eye. Automatic steel surface inspection systems have become mainstream in steel surface inspection, and it is an attractive alternative to human visual inspection. The main purpose is to prevent products containing defects from reaching customers. Recently, several attempts have been made to automatically detect and classify surface defects of steel strips [9–13]. They provided some guidelines for future studies and recommended suitable research directions. This paper aims to address the issue of low accuracy in automatically detecting and classifying surface defects on steel strips by employing deep learning techniques. The main objective is to enhance the accuracy of identifying and categorizing defects, specifically when confronted with some diversity or similarities in the defect types on steel surfaces. In order to achieve this, the study suggests a novel methodology that integrates a segment of the pre-trained VGG16 model as a feature extractor with a newly developed convolutional neural network (CNN) as a classifier. The main contributions of this paper are as follows:

- We enhanced defect detection and classification accuracy by introducing a novel deep-learning approach for automatic surface defect detection in steel strips. This method combines a part of the pre-trained VGG16 model as a feature extractor with a new convolutional neural network (CNN) for classification.
- Our study utilizes transfer learning to extract significant features from images of steel surfaces. We optimize the model's performance in steel surface defect classification by fine-tuning it using knowledge acquired from big datasets like ImageNet.
- The study addresses the issue of low accuracy in identifying steel surface defects, which often arise from their diverse and similar characteristics.
- We conducted a comparative analysis between the proposed method and existing studies based on machine learning and deep learning to demonstrate our proposed method's superior effectiveness compared to currently available methods.

The rest of the paper is organized as follows: Section 2 presents related works, including surface defects inspection on steel strips, deep learning, and transfer learning. Section 3 outlines details on the method proposed. The dataset used and experimental results achieved and their comparison with existing methods are discussed in Section 4. Finally, we draw a conclusion in Section 5.

Background and related works

Surface defects inspection on steel strips

It has always been a vision of steel manufacturing production to make a machine that can "see" and describe what it saw. The aim of defects classification for the steel surface is to identify the category of a defect on the steel surface to ensure the quality of steel surface products [14], so defect classification methods must have high efficiency and accuracy. Therefore, it is a challenge for scholars to improve efficiency and accuracy. Over the past few decades, numerous scholars have studied the detection and classification of defects on steel surfaces. The predominant approaches used in the detection and classification of steel surface defects involve traditional machine learning methods techniques such as the K-Nearest Neighbors (KNN) classification method [15], the Naive Bayesian method [16], and the support vector machine method [17]. However, in the case of potential diversity and similarity in some datasets, classifying accurately will be difficult using traditional classification methods. It may struggle with generalization when dealing with diverse or similar classes of defects. Gong et al. [18] proposed a multi-class classification model named MHSVM for steel surface defects, which adopts hyper-sphere as the decision boundary. This method is derived from THSVM and SVM. Two types of additional information have been used: local neighbor and local density to improve classification accuracy and reduce the adverse effect of noise. The classification accuracy achieved was 97.33%. However, solving quadratic programming of functions requires a lot of storage space. Wang et al. [19] proposed a method based on a guidance template to detect diverse types of defects strip steel surface defect detection. The average detection rate was achieved at 96.2% on a data set including 1500 test samples. Huijun et al. [20] proposed a support vector machine (SVM) classification model for identifying defects in steel strip surface images, using the geometric feature, grayscale feature and shape feature, extracted by combining the defect target image and its corresponding binary image. The classification accuracy achieved was 91.28%. Gong et al. [21] introduced a multi-class classifier called support vector hyper-spheres with insensitivity to noise (INSVHs) for classifying defects on steel plate surfaces. The accuracy achieved was 97.26%. Traditional machine learning algorithms mainly de- pend on manually designing features, where the features are designed and selected by human intervention. Determining suitable features can provide a formidable challenge when dealing with complex visual tasks such as surface defect identification. However, the limitations of traditional machine learning approaches in addressing the difficulties presented by diverse, similar, and complex classes in steel strip surface images have led to the use of deep learning and transfer learning techniques.

Deep learning

Recently, deep learning has attracted increasing attention. The consequence has been that many researchers have replaced traditional machine learning methods with deep learning approaches for surface defect detection and classification. Zhao et al. [22] proposed an improved YOLOv4 architecture using a feature pyramid network (FPN) module. The average detection accuracy showed 92.5%, where the model only tested on three defect types in the NEU dataset (patches, scratches, crazing). However, the NEU dataset includes six types of defects; therefore, the authors expected more improvements if several defects could be involved. He et al. [23]

Table 1
The body architecture of the original VGG16 network.

Layer (type)	Output shape	Parameters
block1-conv1 (Conv2D)	(None,224,22, 64)	1792
block1-conv2 (Conv2D)	(None,224,22, 64)	36928
block1-pool (MaxPooling2D)	(None,112,112, 64)	0
block2-conv1 (Conv2D)	(None,112,112, 128)	73856
block2-conv2 (Conv2D)	(None,112,112, 128)	147584
block2-pool (MaxPooling2D)	(None,56,56, 128)	0
block3-conv1 (Conv2D)	(None,56,56, 256)	295168
block3-conv2 (Conv2D)	(None,56,56, 256)	590080
block3-conv3 (Conv2D)	(None,56,56, 256)	590080
block3-pool (MaxPooling2D)	(None,28,28,256)	0
block4-conv1 (Conv2D)	(None,28,28, 512)	1180160
block4-conv2 (Conv2D)	(None,28,28, 512)	2359808
block4-conv3 (Conv2D)	(None,28,28, 512)	2359808
block4-pool (MaxPooling2D)	(None,14,14, 512)	0
block5-conv1 (Conv2D)	(None,14,14, 512)	2359808
block5-conv2 (Conv2D)	(None,14,14, 512)	2359808
block5-conv3 (Conv2D)	(None,14,14, 512)	2359808
block5-pool (MaxPooling2D)	(None,7,7, 512)	0
flatten (Flatten)	(None, 25088)	0
fc1 (Dense)	(None, 4096)	102764544
fc2 (Dense)	(None, 4096)	16781312
predictions(Dense)	(None, 1000)	4097000

proposed a deep learning network system by fusing the multilevel features to obtain the specific category and detailed location of defects in the steel plates. Singla et al. [24] implemented a deep-learning framework for steel classification. In order to extract features from grayscale images to classify the inputs, a convolutional neural network was used for six types of steel surface defects, and the accuracy achieved was 95.6%. He et al. [25] proposed a semi-supervised learning approach using multi-training of two networks: residual network and categorized generative adversarial network to classify steel surface defects. The generative adversarial network has been used to generate many unlabeled samples, and then two classifiers based on different learning strategies have been used. To detect periodic defects for roll marks on hot-rolled steel plates. Liu et al. [26] proposed a periodical defect detection method using a convolutional neural network for feature extraction and long short-term memory for defect recognition. The detection success rate of this method has achieved 81.9%. In order to use the previous information for more accuracy, the method is enhanced with the attention mechanism, and the detection success rate of the improved method was 86.2%. However, the attention mechanism increases the complexity of the algorithm. Jiangyun et al. [27] improved the You Only Look Once (YOLO) network to detect six types of cold-rolled defects in steel strips. Chen et al. [8] proposed a new ensemble approach based on deep CNN. The authors trained three different DCNNs algorithms and used an average to combine their output of them to increase the recognition rate for steel surface defects. As a consequence, the computation was too large. For example, the model size of the WRN-28-10 network was more than 1.2 GB. It takes a long time to train.

Transfer learning

Transfer learning is a machine learning technique where a model trained on one task is re-purposed on a new related task, which is presently very common in deep learning because it can train models based on deep neural networks with relatively small datasets. Using transfer learning can effectively reduce the difficulty of training, particularly when using a small training set [28]. Several works have been conducted for steel surface defect detection and classification using transfer learning. Wang et al. [29] combined an enhanced, faster region convolutional neural network (faster R-CNN) and improved ResNet50 to reduce the average running time and improve the detection accuracy of the steel surface dataset. The accuracy of this method reached 98.2%. However, the dataset used needed improvement before being passed to the model because some samples with defects were put in the samples without defects, and some samples had missing markers. Fu et al. [30] proposed an automatic visual recognition of steel surface defects. The pre-trained SqueezeNet has been adapted as the backbone architecture on a diverse dataset of steel surface defects with severe camera noise, motion blur, and nonuniform illumination. The accuracy was 97.5%. Fadli et al. [31] used two deep learning models, VGG16 and VGG19, for image recognition of six different types of surface defects (Inclusion, Crazing, Scratches, Rolled, Pitted, and Patches). The performance of the two models was compared, and they achieved accuracies of 93.30% and 97.20% for VGG16 and VGG19, respectively. Nagy et al. [32] proposed a new architecture by combining EfficientNet deep neural networks and randomized classifiers to solve the classification of steel surface defects of the two datasets of steel surface defects. Abu et al. [33] developed deep learning models to detect and evaluate steel surface defects using transfer learning techniques, in which four types of transfer learning models: MobileNet, ResNet, DenseNet, and VGG, have been studied. MobileNet achieved the highest detection rate with 80.41% when using the SEVERSTAL dataset and 96.94% using NEU dataset.

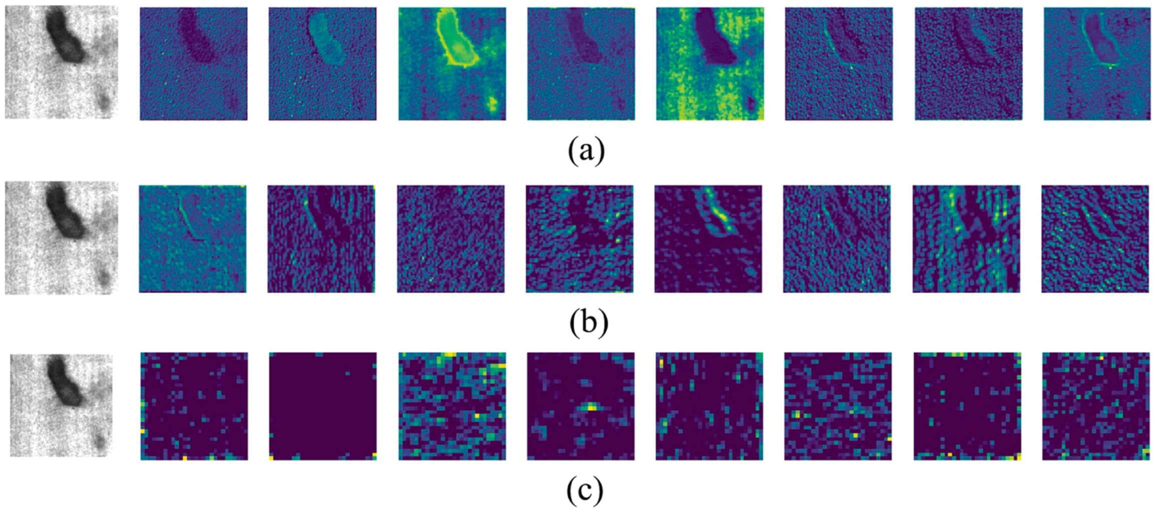

Fig. 1. Part of the feature maps visualization of the pre-trained vgg16 model. (a) When the image of the steel surface passes through the first pool layer. (b) When the image of the steel surface passes through the second pool layer. (c) When the image of the steel surface passes through the third pool layer.

Materials and methods

This section outlines the methodology employed to tackle the task of detecting and classifying defects on steel surfaces. The process comprises two key components: feature extraction and classification.

Features extraction

VGG16 is a Convolutional Neural Network (CNN) model designed for image classification tasks, proposed by Zisserman and Simonyan [34]. It comprises of 16 convolutional and fully-connected layers. The architecture is characterized by a consistent structure, with 3 × 3 kernels and a stride of 1 for convolutional layers and 2 × 2 kernels with a stride of 2 for pooling layers. This model employs convolutional and max-pooling layers for feature extraction from input images, followed by fully-connected layers for prediction. Table 1 shows the architecture of the original VGG16 network.

The classification process begins with an input image, denoted as o, progressing through successive layers represented by $\alpha^{[i]}$. The mathematical representation of the VGG16 model is given by:

$$z^{[i]} = h\left(\alpha^{[i]}\right),\ \alpha^{[i]} = \left(W_c^{[i]}\alpha^{[i-1]} + \beta^{[i]}\right),\quad \alpha^{[0]} = o \tag{1}$$

Here, $W_c^{[i]}$ denotes the weights of the convolutional layer, and $\beta^{[i]}$ represents components of the bias vector. Eq. (1) encompasses convolution, bias addition, and ReLU activation, transitioning from the input image o to $\alpha^{[1]}$. Each filter in $W_c^{[1]}$ is denoted as $W_{c(j)}^{[1]}$, where j is the filter number.

The ReLU activation function $h(x)$ is defined by:

$$h(x) = \begin{cases} x \ if \ x > 0 \\ 0 \ otherwise \end{cases} \tag{2}$$

The transition from the input o to the first convolution layer, $\alpha^{[o]}$, involves applying the convolution function with padding. Padding is crucial to maintaining spatial dimensions and preventing loss of information.

In addition to activation functions, pooling layers in VGG16 down-sample input tensors reduces spatial dimensions. The max-pooling function, defined in Eq. (3), computes down-sampled features by taking the maximum value within a window:

$$a_j \ = \ \max(u(n,\ n)) \tag{3}$$

Here, n represents the pooling size, and $u(n, n)$ is a window function. The VGG16 model employs a flat layer to transform the output of convolutional layers into a 1D vector. Finally, the architecture concludes with a Softmax function in the last layer, converting real values into a probability distribution over predefined classes.

A specific part of the pre-trained VGG16 model has been chosen for this study. This strategy enables us to use the knowledge acquired by the pre- trained model on a large dataset while also allowing us to customize the model for our specific purpose by changing the architecture and modifying the weights. To choose the appropriate part of VGG16 to use as a feature extractor, we input an image of a steel surface into a pre-trained VGG16 model and observe the feature maps generated as the image progresses through the first, second, and third pool layers. Fig. 1(c). shows the result. Analysis of these feature maps revealed that the layers before the

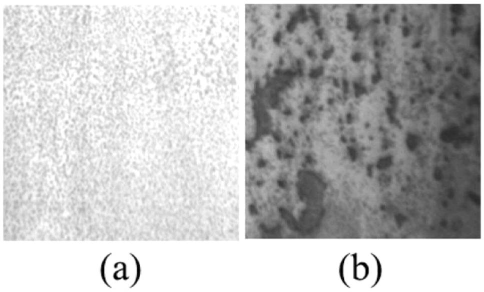

Fig. 2. The diversity sample in one class of pitted surface in the NEU dataset.

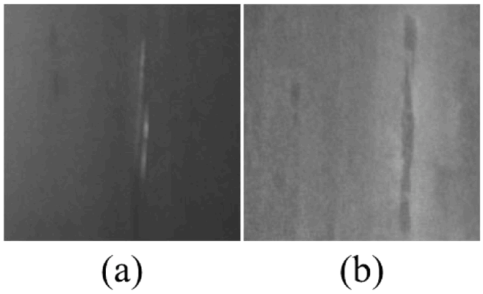

Fig. 3. The similarity sample between the two classes of the NEU dataset: (a) scratches, (b) inclusion.

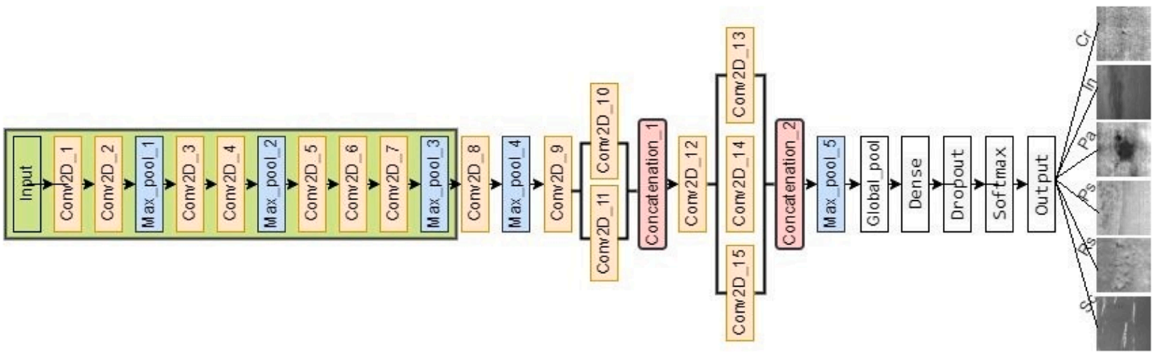

Fig. 4. The architecture of the proposed method for steel surface defect classification.

second pooling layer extracted simple information from the image. It was noted that the layers before the third pooling layer extracted more sophisticated image features, making them the best option for extracting features. Based on these observations, the decision was made to use the layers preceding the third pooling layer of VGG16 as the feature extractor, which is filled with green in Fig. 4.

Classifier

The classification component of our methodology involves a newly designed CNN classifier. This classifier is designed to improve the accuracy of defect classification, particularly in the presence of diversity and similarity within some defect classes. Fig. 2. shows the diversity sample in one class of steel surface defect, and the similarity between two classes of steel surface defect is shown in Fig. 3. The CNN classifier is comprised of 8 convolutional layers, 8 Rectified Linear Unit (ReLU) activation layers, two max-pooling layers, two depth concatenation layers, one global pooling layer, one dense layer, one dropout layer, and one softmax layer. The complete structure, which includes the selected part of pre-trained VGG16 and our designed CNN classifier, is illustrated in Fig 4. For accurate defect classification, a specific approach is taken to limit the number of convolutional and pooling layers in the first layers of the classifier's architecture. Subsequently, the number of convolutional and pooling layers is gradually increased. The architecture of the proposed classifier consists of two combined convolutional layers. The initial combined layer comprises two convolutional layers and two Rectified Linear Unit (ReLU) layers, whereas the subsequent composite layer consists of three convolutional layers and three ReLU levels. After each combined layer, a concatenation layer is employed to merge the features obtained from the previous layers. The

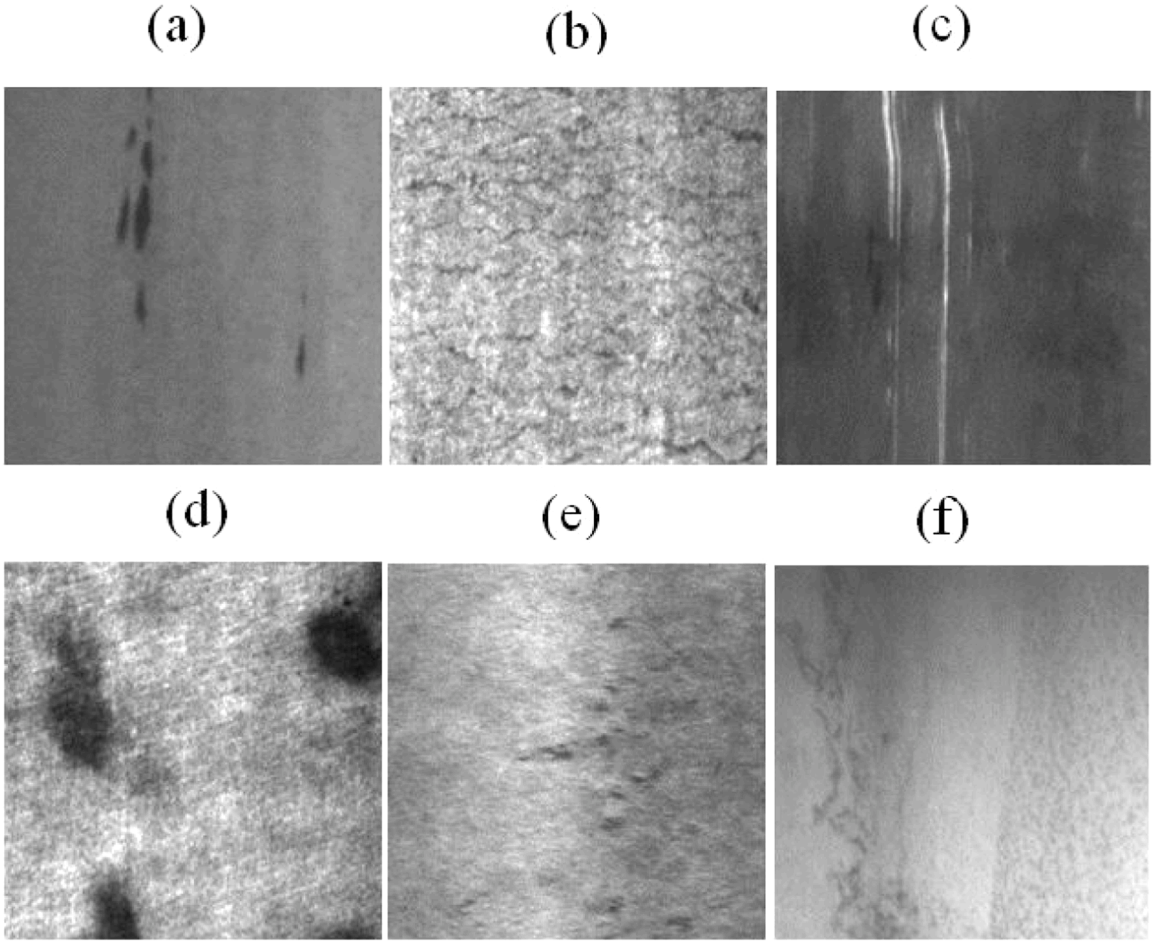

Fig. 5. Samples for the six types of defect classes of the NEU dataset: (a) inclusion (In), (b) crazing (Cr), (c) scratches (Sc), (d) patches (Pa), (e) rolled-in scale (Rs), (f) pitted surface (Ps).

novelty of our study resides in the architecture of our CNN classifier, specifically designed to handle the diversity and similarity within certain defect classes present in the dataset.

Experimental results and discussion

Dataset

We performed experiments using the Northeastern University dataset (NEU), a benchmark dataset containing surface imperfections found on hot- rolling steel strips [35]. The NEU dataset comprises six categories of surface defects found on hot-rolled steel strips. These categories include patches, crazing, inclusion, rolled-in scale, pitted surface, and scratches. The dataset consists of 1800 grayscale images that have been labeled, with each image having a resolution of 200 $\times$ 200 pixels. There are a total of 300 examples for each category of defect. Fig. 5 shows representative images from the NEU dataset.

Training

To conform to the input dimensions of the VGG16 neural network, we modified the image dataset by resizing it from 200 $\times$ 200 pixels to 224 $\times$ 224 pixels. The dataset was divided into training, validation, and testing. The training set comprised 864 images, while the test set had 360 images, and the remaining images were allocated to the validation set. The learning rate was set at 0.0001, and a batch size of 40 was used due to the relatively small dataset. We implemented a dropout rate of 0.5 on the fully connected layers and used the Adaptive Moment Estimation (Adam) optimizer. We did not employ any data augmentation strategy in our study to assess the effectiveness of our suggested method when applied to a limited training dataset. The model attained high classification accuracy without employing data augmentation. The experiments were conducted utilizing an Intel Core(TM) i5-7200 CPU with 16GB of RAM, an Nvidia GeForce 940MX graphics card, and a 2.60 GHz processor.

Table 2
The experimental results.

Type of defect	Precision	Recall	F1 score	Samples
Crazing	100%	100%	100%	60
Inclusion	0.98%	0.98%	0.98%	60
Patches	100%	100%	100%	60
Pitted Surface	0.98%	100%	0.99%	60
Rolled-in-Scales	100%	100%	100%	60
Scratches	100%	0.98%	0.99%	60

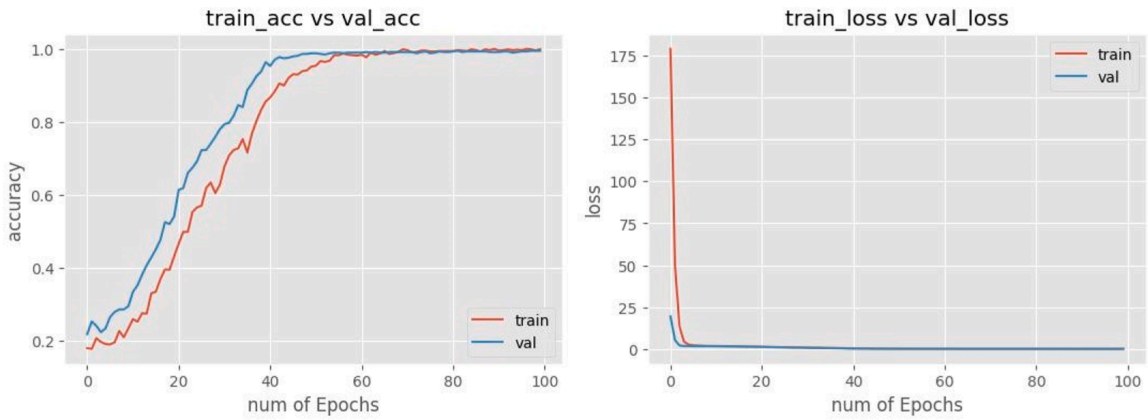

Fig. 6. Training and validation graphs of the proposed method for steel surface defects classification.

Results

We assessed the efficiency of our model on the test set by measuring accuracy and categorical cross-entropy loss as critical measures. We used formulas 4 and 5 to capture and portray the model's performance accurately [36]. The output of the softmax function corresponds to the probability assigned to each class, where the class with the highest probability is considered the predicted class.

$$Accuracy = \frac{Number\ of\ Correctly\ Classified\ Images}{Total\ Number\ of\ Input\ Images} \times\ 100\% \tag{4}$$

$$Loss = \ -\sum_{k=i}^{n} \widehat{t}_i\ \log f(z_i), \tag{5}$$

When n is equal to 6, $\hat{ti}$ is set to 1 if the label of the input image matches i; otherwise, $\hat{ti}$ is set to 0. The confidence score $f(z_i)$ is derived through the softmax function, which is defined as:

$$f(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}, \tag{6}$$

Where z_i is the number i output value. Additionally, this research article presents experimental metrics for classifying six categories of defects on steel surfaces. The metrics that have been introduced include Precision, Recall, and F1 score. The metrics used evaluate the proposed method's efficacy in accurately classifying various defects.

Precision is a metric that measures the degree of accuracy in positive predictions [37]. The calculation involves determining the ratio of true positive instances to the overall number of instances that were predicted as positive. It can be defined as:

$$\text{Precision} = \ \frac{TP}{TP + FP} \tag{7}$$

The measure of recall evaluates the model's ability to identify all positive instances correctly [37]. The calculation determines the ratio of true positive cases to the total number of positive instances. The Recall is defined as:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

The F1 score can be defined as the harmonic mean of precision and recall [37]. The metric provides a comprehensive evaluation of a model's efficacy by considering both the occurrence of false positives and false negatives. It is calculated using the formula:

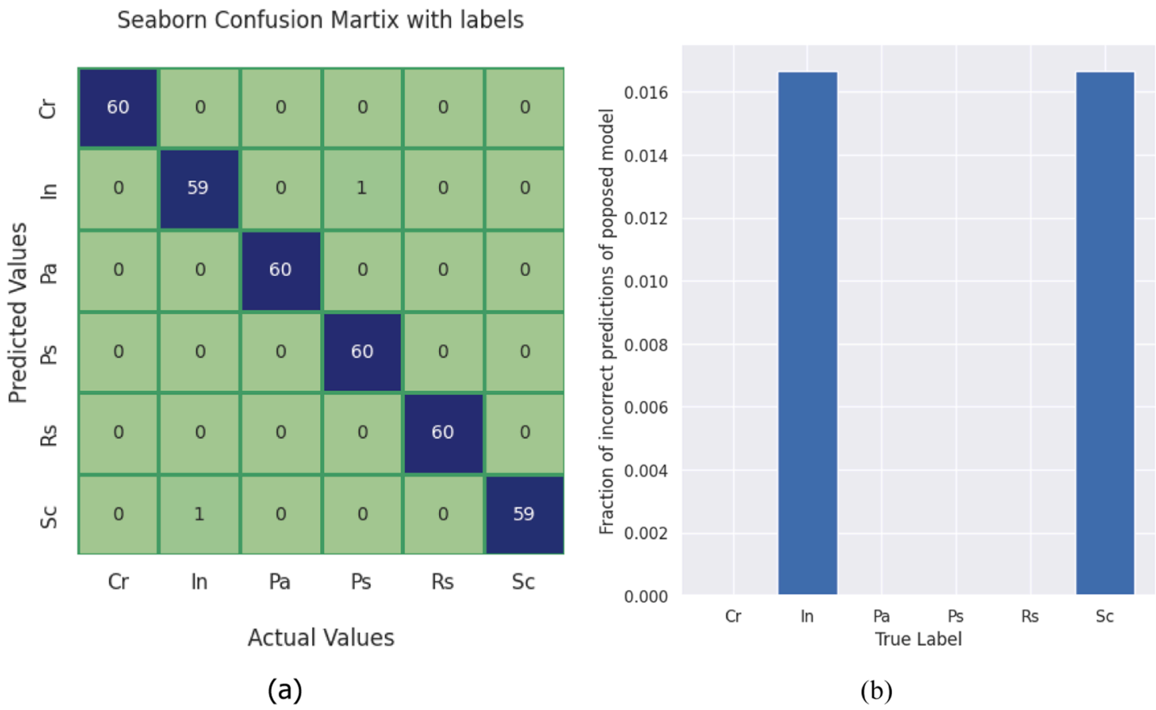

Fig. 7. Confusion matrix and fraction of incorrect predictions of the test dataset.

Table 3
Comparison of defect classification results on NEU dataset using machine learning approaches (the data are based on Ref. [38]).

Feature extract Method	Classifier	Accuracy (%)
GLRM	KNN	91.94
GLRM	SVM	94.72
GLRM	RF	93.89
GLCM	KNN	92.78
GLCM	SVM	95.56
GLCM	RF	94.72
VGG16(Ours)	CNN	**99.44**

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

The experimental results are shown in Table 2. The test dataset's confusion matrix and the fraction of inaccurate predictions are shown in Fig. 7. The performance of our model in classifying defects across many categories was consistently strong, with only a small number of misclassifications found in a few photos. Based on the observation of Fig. 7(a), our model can accurately classify several types of surface defects, including crazing, patches, pitted surfaces, and rolled-in scales. According to Fig. 7(b), our model made an error in identifying one image, classifying it incorrectly as a pitted surface in the defect category of inclusions. Furthermore, our algorithm misidentified a single image in the scratches category as an inclusion. The plot for accuracy and loss value curves over 100 epochs is shown in Fig. 6.

Discussion

This section presents and discusses a comparative analysis of our model results with existing machine learning and deep learning methodologies. In order to assess the efficacy of the CNN model proposed, first, we conduct a comparative analysis using several machine learning models. In Table 3, we compared our classification accuracy with the recent paper proposed by Rao et al. [38] to detect and classify steel surface defects by combining feature extraction and classification techniques. This study utilized the feature extraction approaches of GLCM (Gray-Level Cooccurrence Matrix) and GLRM (Gray-Level Run-Length Matrix). Three separate algorithms have been employed to construct three different machine-learning models. The Support Vector Machine (SVM) method, the Random Forest (RF) technique, and the K-Nearest Neighbors (KNN) algorithm have been used. As we can see from Table 3 our proposed method outperforms the three constructed machine learning methods. Additionally, five combination classification methods

Table 4
Comparison of defect classification accuracy on test set images using various machine learning models.

Ref	Method	Accuracy (%)
Zaghdoudi et al. [39]	HOG+GLCM-SVM	90.16
Boudiaf et al. [40]	HOG+PCA-KNN	91.12
Ibrahim and Tapamo [35]	BRISK+SVM	95.00
Song and Yan [42]	AECLBP+NNC	98.3
Mohamed and Yampolskiy [41]	AELTP+MLR	98.6
Ours	VGG16+CNN	**99.44**

Table 5
Comparison of defect classification accuracy on test set images using various deep learning models.

Ref	Method	Accuracy (%)
Yeung et al. [43]	Fused attention CNN model	89.30
Tian et al. [44]	SegNet + CNN	89.60
Singla [46]	CNN	95.63
Fadli et al. [31]	VGG19	97.20
Piwal et al. [45]	VGG19	97.20
Alkapov [7]	AVDDCS	98.16
Yi et al. [47]	Deep CNN	99.05
Ours	VGG16+CNN	**99.44**

based on machine learning have been compared. These methods include four traditional machine learning approaches as classifiers, including Support Vector Machine (SVM), Nearest Neighbor Clustering (NNC), K-Nearest Neighbors Algorithm (KNN), and Multiple Linear Regression (MLR). Five feature extraction methods, including Gray Level Co-occurrence Matrix (GLCM) [39], Histogram of Oriented Gradients (HOG) [40], Adaptive Extended Local Ternary Pattern (AELTP) [41], Binary Robust invariant scalable keypoints (BRISK) [35], and Adjacent Evaluation Completed Local Binary Patterns (AECLBP) [42]. Comparison results of classification accuracy of six surface defects on testing set images using various combination machine learning models are shown in Table 4.

Moreover, some recent deep-learning-based surface defect classification approaches were also considered, including fused attention CNN model proposed by Yeung et al. [43], SegNet + CNN proposed by Tian et al. [44], VGG-16 and VGG 19 proposed by Piwal et al. [45], convolutional neural network (CNN) proposed by Singla et al. [46], model-based deep learning using pre-trained VGG19 proposed by Fadli et al. [31], automatic visible defect detection and classification system (AVDDCS) proposed by Alkapov et al. [7] and Deep CNN proposed by Yi et al. [47]. Comparison results of classification accuracy of six surface defects on test set images using various deep learning models are shown in Table 5. From Table 5, you can see that the accuracy of our proposed method surpasses other deep learning systems, ranging from 89.30% to 99.05%. It's important to note that all the techniques undergo training and testing using the same size of training and testing datasets, respectively. As it is shown in Tables 4 and 5 our method outperforms the traditional methods and has higher accuracy than other deep learning methods.

Conclusion

This research article addressed the challenge of accurately detecting and classifying surface defects on steel strips. We have proposed an innovative method that utilizes deep learning techniques, explicitly employing a part of the pre-trained VGG16 model as a feature extractor and a newly developed convolutional neural network (CNN) as a classifier. Our experimental findings on the Northeastern University dataset (NEU) proved our technique's efficacy, including six surface defects. The resulting classification accuracy was 99.44%, surpassing other existing methods. The success of our method can be attributed to the combination of a powerful feature extractor and a well-designed CNN classifier. This combination enables us to overcome the challenges posed by diversity and similarity in some categories of defects. However, it is worth noting that the dataset used in this study is relatively small, containing a limited number of classes. The ability of the model to effectively generalize may be impeded by several challenges faced in real world scenarios, including fluctuations in lighting conditions, interference from noise, and variations in defect appearances. Additionally, the process of collecting labeled data can be costly or time consuming. Moreover, our future work will investigate semi-supervised learning techniques that combine a small labeled dataset with a larger unlabeled dataset to detect and classify steel strip surface defects.

CRediT authorship contribution statement

Alaa Aldein M.S. Ibrahim: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Visualization, Investigation. **Jules R. Tapamo:** Supervision, Validation, Formal analysis, Resources, Writing – review & editing, Project administration, Funding acquisition.

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.

References

[1] D. Zhang, X. Hao, L. Liang, W. Liu, C. Qin, A novel deep convolutional neural network algorithm for surface defect detection, J. Comput. Des. Eng. 9 (5) (2022) 1616–1632.
[2] D. Zhang, X. Hao, D. Wang, C. Qin, B. Zhao, L. Liang, W. Liu, An efficient lightweight convolutional neural network for industrial surface defect detection, Artif. Intell. Rev. (2023) 1–27.
[3] H. Zhao, T. Zhang, Fabric surface defect detection using SESSDNet, Symmetry 14 (11) (2022) 2373.
[4] G. Zhao, S. Qin, High-precision detection of defects of tire texture through X-ray imaging based on local inverse difference moment features, Sensors 18 (8) (2018) 2524.
[5] Y. Xian, Y. Yu, Y. Lian, J. Fan, Z. Wang, An EA-based pruning on improved YOLOv3 for rapid copper elbow surface defect detection, Eng. Appl. Artif. Intell. 123 (2023) 106412.
[6] H. Hu, Z. Zhu, Sim-YOLOv5s: a method for detecting defects on the end face of lithium battery steel shells, Adv. Eng. Inform. 55 (2023) 101824.
[7] R. Alkapov, A. Konyshev, N. Vetoshkin, N. Valkevich, P. Kostenetskiy, Automatic visible defect detection and classification system proto- type development for iron-and-steel works, in: Proceedings of the Global Smart Indus- try Conference (GloSIC), IEEE, 2018, pp. 1–8.
[8] W. Chen, Y. Gao, L. Gao, X. Li, A new ensemble approach based on deep convolutional neural networks for steel surface defect classification, Procedia CIRP 72 (2018) 1069–1072.
[9] N. Neogi, D. Mohanta-K, P.K. Dutta, Review of vision-based steel surface inspection systems, EURASIP J. Image Video Process. 2014 (1) (2014) 1–19.
[10] Q. Luo, X. Fang, L. Liu, C. Yang, Y. Sun, Automated visual defect detection for flat steel surface: a survey, IEEE Trans. Instrum. Meas. 69 (3) (2020) 626–644.
[11] X. Zheng, S. Zheng, Y. Kong, J. Chen, Recent advances in surface defect inspection of industrial products using deep learning techniques, Int. J. Adv. Manuf. Technol. 113 (1) (2021) 35–58.
[12] Q. Jin and L. Chen, A survey of surface defect detection of industrial products based on a small number of labeled data. arXiv preprint arXiv (2022) 2203.05733.
[13] R. Mordia, A.K. Verma, Visual techniques for defects detection in steel products: a comparative study, Eng. Fail. Anal. (2022) 106047.
[14] M. Chu, J. Zhao, X. Liu, R. Gong, Multi-class classification for steel surface defects based on machine learning with quantile hyper-spheres, Chemom. Intell. Lab. Syst. 168 (2017) 15–27.
[15] P. Menesatti, A. Zanella, S. D'Andrea, C. Costa, G. Paglia, F. Pallottino, Supervised multivariate analysis of hyper-spectral NIR images to evaluate the starch index of apples, Food Bioprocess Technol. 2 (3) (2009) 308–314.
[16] Q. Wang, G.M. Garrity, J.M. Tiedje, J.R. Cole, Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy, Appl. Environ. Microbiol. 73 (16) (2007) 5261–5267.
[17] Q. Huangpeng, W. Huang, H. Shi, Automated paper impurities evaluation using feature representations based on ADMM sparse codes, J. Intell. Fuzzy Syst. 34 (2) (2018) 797–805.
[18] R. Gong, C. Wu, M. Chu, Steel surface defect classification using multiple hyper-spheres support vector machine with additional information, Chemom. Intell. Lab. Syst. 172 (2018) 109–117.
[19] H. Wang, J. Zhang, Y. Tian, H. Chen, H. Sun, K. Liu, A simple guidance template-based defect detection method for strip steel surfaces, IEEE Trans. Ind. Inform. 15 (5) (2018) 2798–2809.
[20] H. Hu, Y. Li, M. Liu, W. Liang, Classification of defects in steel strip surface based on multiclass support vector machine, Multimed. Tools Appl. 69 (2014) 199–216.
[21] R. Gong, M. Chu, Y. Yang, Y. Feng, A multi-class classifier based on support vector hyper-spheres for steel plate surface defects, Chemom. Intell. Lab. Syst. 188 (2019) 70–78.
[22] H. Zhao, Z. Yang, J. Li, Detection of metal surface defects based on YOLOv4 algorithm, J. Phys. Conf. Ser. 1907 (2021) 012043.
[23] Y. He, K. Song, Q. Meng, Y. Yan, An end-to-end steel surface defect detection approach via fusing multiple hierarchical features, IEEE Trans. Instrum. Meas. 69 (4) (2019) 1493–1504.
[24] R. Singari, K. Singla, G. Chawla, Deep learning framework for steel surface defects classification, Int. J. Adv. Prod. Ind. Eng. 1 (2019) 25–32.
[25] Y. He, K. Song, H. Dong, Y. Yan, Semi-supervised defect classification of steel surface based on multi-training and generative adversarial network, Opt. Lasers Eng. 122 (2019) 294–302.
[26] Y. Liu, K. Xu, J. Xu, Periodic surface defect detection in steel plates based on deep learning, Appl. Sci. 9 (15) (2019) 3127.
[27] J. Li, Z. Su, J. Geng, Y. Yin, Real-time detection of steel strip surface defects based on improved yolo detection network, IFAC Papers OnLine 51 (21) (2018) 76–81.
[28] F. Jingwen, Z. Xiaoyan, L. Yingbin, Recognition of Surface Defects on Steel Sheet Using Transfer Learning, 5, Xian Jiaotong University, 2020, pp. 25–34.
[29] S. Wang, X. Xia, L. Ye, B. Yang, Automatic detection and classification of steel surface defect using deep convolutional neural networks, Metals 11 (3) (2021) 388.
[30] G. Fu, P. Sun, W. Zhu, J. Cao, Y, M.Y. Yang, Y. Cao, A deep- learning-based approach for fast and robust steel surface defects classification, Opt. Lasers Eng. 121 (2019) 397–405.
[31] V.F. Fadli, I.O. Herlistiono, Steel surface defect detection using deep learning, Int. J. Innov. Sci. Res. Technol 5 (2020) 244–250.
[32] A.M Nagy, L. Czúni, Classification and fast few-shot learning of steel surface defects with randomized network, Appl. Sci. 12 (8) (2022) 3967.
[33] M. Abu, A. Amir, Y.H. Lean, N.A.H. Zahri, S.A. Azemi, The performance analysis of transfer learning for steel defect detection by using deep learning, J. Phys. Conf. Ser. 1755 (2021) 012041.
[34] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv preprint (2014), arXiv:1409.1556.
[35] A.A.M. Ibrahim, J.R. Tapamo, Steel surface defect detection and classification using bag of visual words with BRISK, in: Proceedings of the Congress on Smart Computing Technologies, Singapore, Springer Nature Singapor, 2022, pp. 235–246.
[36] X. Wan, X. Zhang, L. Liu, An improved VGG19 transfer learning strip steel surface defect recognition deep neural network based on few samples and imbalanced datasets, Appl. Sci. 11 (6) (2021) 2606, 15.
[37] D.M. Powers, Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint (2020), arXiv:2010.16061.
[38] K.B. Rao, G.N.R. Prasad, S. Amudha, A. Kumar, D.V.S.S.S.V. Prasad, V. Mohanavel, A significant feature selection to improve the accuracy of a classification algorithm for steel defect, in: Proceedings of the 4th International Conference on Inventive Research in Computing Applications (ICIRCA), IEEE, 2022, pp. 200–206.
[39] R. Zaghdoudi, H. Seridi, A. Boudiaf, Detection and classification of steel defects using machine vision and SVM classifier, in: Proceedings of the 2nd International Conference on Automatic Control, Telecommunication and Signals (ICATS'17, 2017.
[40] A. Boudiaf, A. Moussaoui, R. Zaghdoudi, Z. Mentouri and A., Saadoune, An automatic system to surface defect classification of hot rolled steel (2018).
[41] A.A. Mohamed, R.V. Yampolskiy, Adaptive extended local ternary pattern (aeltp) for recognizing avatar faces, in: Proceedings of the 11th International Conference on Machine Learning and Applications 1, IEEE, 2012, pp. 57–62.
[42] K. Song, Y. Yan, A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects, Appl. Surf. Sci. 285 (2013) 858–864.
[43] C.C. Yeung, K.M. Lam, Efficient fused-attention model for steel sur- face defect detection, IEEE Trans. Instrum. Meas. 71 (2022) 1–11.

[44] S. Tian, P. Huang, H. Ma, J. Wang, X. Zhou, S. Zhang, J. Zhou, R. Huang, Y. Li, CASDD: automatic surface defect detection using a complementary adversarial network, IEEE Sens. J. 22 (20) (2022) 19583–19595.
[45] H. Piwal, M. Dhokale, R. Biswas, S. Raut, A. Malge, Surface defect detection using deep learning, in: Proceedings of the AIP Conference 2717, AIP Publishing, 2023.
[46] K. Singla, G. Chawla, R.M. Singari, Deep learning framework for steel surface defects classification, Int. J. Adv. Prod. Ind. Eng. 4 (1) (2019).
[47] L. Yi, G. Li, M. Jiang, An end-to-end steel strip surface defects recognition system based on convolutional neural networks, Steel Res. Int. 88 (2) (2017) 1600068.