2026 IEEE International Conference on Signal Processing, Information, Communication and Systems (SPICSCON)
13-14 August 2026, Faculty of ECE, Bangladesh Army University of Engineering & Technology (BAUET), Qadirabad, Natore-6431, Bangladesh

CBAM-Net: An Attention-Enhanced Lightweight Convolutional Network for Steel Surface Defect Classification with Ablation and Robustness Analysis

1st Mugdha Farheen Kibria
Dept. of Computer Science and Engineering
BRAC University
Dhaka-1212, Bangladesh
mugdhafarheenkibria@gmail.com

2nd Md. Abdus Sium
Dept. of Electrical and Electronic Engineering
BRAC University
Dhaka-1212, Bangladesh
siumprof9005@gmail.com

3rd Mohammad Musgfiqur Rahman
Dept. of Computer Science and Engineering
BRAC University
Dhaka-1212, Bangladesh
mohammad.mushfiqur.rahman17@gmail.com

4th S M Mushfiq Reza
Faculty of Science
Queensland University of Technology
Brisbane QLD-4000, Australia
mushfiqreza25@gmail.com

5th Al Hasan Sakib
International Engineering and Science
Jeonbuk National University
Jeonju, Republic of Korea
sakibalhasan@jbnu.ac.kr

6th Md Jobayer Sazzad
Dept. of Mechanical and Production Engineering
Ahsanullah University of Science and Technology
Dhaka-1208, Bangladesh
jobayersazzad@gmail.com

7th Tauhidur Rahman Sakib*
Dept. of Computer Science and Engineering
BRAC University
Dhaka-1212, Bangladesh
tauhidurrahmansakib@gmail.com

***Abstract*—Automated steel surface inspection is challenged by irregular defect shapes, complex textures, and illumination variation. This paper presents CBAM-Net, a lightweight CNN applying CBAM after each feature-extraction block. Rather than introducing a new attention mechanism, the study characterises a compact configuration through six architecture comparisons, controlled ablation, computational analysis, and synthetic robustness testing. On the held-out NEU test set ($n = 300$), CBAM-Net achieved 96.33% accuracy and 0.963 Micro-F1 with 2.03 M parameters, outperforming EfficientNetB0, MobileNetV3-Large, ResNet-18, and the no-attention baseline. Ablation showed that SE channel attention outperformed CBAM channel attention alone, indicating that the main benefit comes from combining channel and spatial recalibration. Grad-CAM highlighted defect-relevant regions, while robustness tests measured degradation under brightness shifts, Gaussian noise, and motion blur. Only the improvement over the no-attention baseline was significant under the unpaired proportion test; paired McNemar testing is required for stronger baseline comparisons.**

***Index Terms*—Steel Surface Defect Detection, Deep Learning, Attention Mechanism, CBAM, Lightweight CNN, Industrial Inspection, Grad-CAM, Robustness Analysis, Ablation Study.**

I. Introduction

Surface quality control is critical in steel manufacturing because undetected defects such as crazing, inclusions, and rolled-in scale can cause downstream failures. Manual inspection is slow and operator-dependent, while classical image-processing methods rely on hand-tuned thresholds that generalise poorly across production lines. CNNs reduce dependence on handcrafted features, and attention mechanisms improve discrimination by emphasising defect regions while suppressing irrelevant surface texture.

Recent studies have examined acquisition quality and sensor effects [1], compared classical and deep-learning approaches [2], and introduced learned preprocessing for steel defect analysis [3]. Reviews have also documented the growing use of CNNs in industrial inspection [4]–[6], [11]. Further developments include hybrid convolutional-classical methods [7], autoencoder-CNN segmentation [8], lightweight real-time networks [9], transformer-based detectors [10], deep segmenta-

979-8-3195-1270-3/26/$31.00 ©2026 IEEE

tion models [12], [13], and channel-shuffled architectures [14]. Attention-based approaches include spatial-attention bilinear networks [15], magnetic-tile defect models [16], trainable guided attention [17], and multitask attention networks [18].

However, many attention-based inspection models lack controlled ablations separating channel and spatial attention or comparisons with lightweight alternatives such as Squeeze-and-Excitation. Robustness is also often inferred from benchmark accuracy, although real deployment may involve illumination changes, sensor noise, and motion blur.

This work does not introduce a new attention operator; it uses the published CBAM module [19]. Instead, it characterises a lightweight three-block CNN with CBAM after each feature-extraction block. The model is evaluated against six architectures under a common protocol, analysed through a five-variant ablation including SE attention, assessed for computational cost and statistical significance, and tested under synthetic acquisition perturbations. The main contributions are:

- **Lightweight CBAM design** – A 2.03 M-parameter, three-block CNN with CBAM after each block.
- **Attention ablation** – Comparison of channel-only, spatial-only, SE, and full CBAM variants.
- **Architecture comparison** – Evaluation against six CNN baselines using accuracy, parameters, FLOPs, and latency.
- **Robustness and interpretability** – Testing under brightness, noise, and blur, supported by Grad-CAM analysis.

The remainder of the paper is organised as follows. Section II describes the methodology and architecture, Section III reports results, and Section IV concludes with limitations and future work.

II. Methodology

CBAM-Net follows four stages: data preparation, augmentation, attention-based feature learning, and evaluation. Experiments used TensorFlow 2.13 and Python 3.10 on an RTX 3060 workstation with 32 GB RAM and fixed random seeds.

A. Dataset Overview

Experiments used the publicly available NEU Surface Defect Database [20], a standard benchmark collected from hot-rolled steel strips. It contains 1,800 grayscale images of 200×200 pixels, evenly distributed across six classes: crazing, inclusion, patches, pitted surface, rolled-in scale, and scratches, with 300 images per class. Despite its limited size, the dataset is challenging because defect appearance varies within classes, while some classes share similar textures. For example, crazing and rolled-in scale show diffuse surface disruption with changing scale and orientation, whereas pitted surface and rolled-in scale exhibit similar granular patterns. Non-uniform illumination further reduces the reliability of grey-level intensity, requiring the model to learn local texture features. These characteristics make NEU suitable for evaluating attention mechanisms that suppress background texture and emphasise defect regions. The dataset was divided into disjoint, class-stratified partitions of 1,200 training, 300 validation, and 300 test images, corresponding to 200, 50, and 50 images per class. Validation data were used only to monitor training, while all reported results were obtained from the unseen test set. Because each partition was balanced, the reported accuracy was not influenced by class imbalance.

B. Data Preprocessing and Augmentation

Images were resized to 200×200 pixels for CBAM-Net, the no-attention baseline, ResNet-18, and the MobileNetV3 and EfficientNetB0 models, and to 224×224 for the VGG16 transfer baseline. Grayscale inputs were replicated across three channels to match the pretrained backbones. Intensities were scaled to $[0, 1]$:

$$I_{\text{norm}}(x, y, c) = \frac{I(x, y, c)}{255} \tag{1}$$

where $I(x, y, c)$ is the raw intensity at location (x, y) in channel c. Augmentation was applied to the training partition only, using random rotation, width and height shifting, and horizontal flipping. These transformations were chosen because they correspond to plausible variation in strip orientation and camera alignment on a production line, whereas vertical flipping and heavy colour jitter do not correspond to physically realisable acquisition changes for this modality. Evaluation images received resizing and normalisation only.

C. CBAM-Net Architecture

CBAM-Net comprises three convolutional feature-extraction blocks followed by a classifier head. Each block applies a 3×3 convolution with ReLU activation, then a CBAM module, then 2×2 max pooling. Table I gives the layer configuration and output shapes. Three design choices warrant justification. Three blocks were selected for progressive receptive-field growth, hierarchical feature extraction, and repeated attention refinement rather than to concentrate capacity in the backbone: as Table I shows, the convolutional and attention layers hold 107,491 parameters (5.3% of the model) against 94.6% in the dense head, so the network is a compact attention-guided extractor followed by a high-capacity classifier. The channel progression $80 \rightarrow 120 \rightarrow 12$ expands the feature space across the first two blocks and then compresses; restricting the final map to 12 channels holds the flattened vector to 7,500 elements, without which the dense layer would grow by an order of magnitude. CBAM is applied after every block rather than only before the classifier, because defect evidence spans scales – scratches are thin and elongated, crazing a coarse regional disruption – so recalibrating only the deepest map forgoes suppression of background texture where the two remain separable. A uniform reduction ratio $r = 8$ was used throughout to avoid block-specific tuning and keep the configuration identical across ablation variants; in block 3 this yields a one-unit hidden layer, so that module performs strongly compressed channel recalibration and its spatial branch carries more of the refinement.

The CBAM module operates in two sequential stages. Channel attention aggregates the input feature map F by parallel

global average and global max pooling, passes both descriptors through a shared multilayer perceptron, and combines them with a sigmoid to produce channel weights:

$$F' = M_c(F) \otimes F \tag{2}$$

where $M_c(F)$ is the channel attention map and $\otimes$ denotes element-wise multiplication. Spatial attention then computes channel-wise average and max pooling over F', concatenates the two maps, and applies a 7×7 convolution with sigmoid activation:

$$M_s(F') = \sigma\big(f^{7\times7}([\text{AvgPool}(F'); \text{MaxPool}(F')])\big) \tag{3}$$

where $f^{7\times7}$ is a convolution with a 7×7 kernel and $\sigma(\cdot)$ is the sigmoid function. The resulting map is multiplied with F' to give the refined representation. The classifier head flattens the final feature map into a dense layer of 256 units with ReLU activation, applies dropout at 0.35, and produces a six-class distribution through softmax.

TABLE I
CBAM-NET LAYER CONFIGURATION

Stage	Output	Params
Input	$200 \times 200 \times 3$	0
Block 1: Conv $3 \times 3/80$ + ReLU, CBAM (r=8), MaxPool 2×2	$100 \times 100 \times 80$	4,029
Block 2: Conv $3 \times 3/120$ + ReLU, CBAM (r=8), MaxPool 2×2	$50 \times 50 \times 120$	90,354
Block 3: Conv $3 \times 3/12$ + ReLU, CBAM (r=8), MaxPool 2×2	$25 \times 25 \times 12$	13,108
Flatten	7,500	0
Dense + ReLU, Dropout (0.35)	256	1,920,256
Dense + Softmax	6	1,542
Total	—	**2,029,289**

D. Training Strategy and Experimental Setup

All models were trained with the Adam optimiser under categorical cross-entropy loss:

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{4}$$

where y_i is the ground-truth label, $\hat{y}_i$ the predicted probability, and $C = 6$. Batch size was 32 and training ran for 20 epochs with convergence monitored on validation data. Every comparison model – the no-attention baseline, MobileNetV3-Small, MobileNetV3-Large, EfficientNetB0, ResNet-18, and VGG16 – used the same partition, augmentation, optimiser, schedule, and evaluation script, so differences in reported performance are attributable to architecture rather than to training conditions. The ablation variants differed from full CBAM-Net only in which attention branch was active.

E. Evaluation Metrics and Statistical Analysis

Accuracy was computed as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives. For the multi-class setting, Micro-F1 was used for global performance:

$$\text{Micro-F1} = \frac{2\sum TP}{2\sum TP + \sum FP + \sum FN} \tag{6}$$

Grad-CAM visualised defect-relevant regions, while baseline margins were assessed using Pearson's Chi-square test at $\alpha = 0.05$; its limitations are discussed in Section III.J.

F. Robustness Protocol

Because NEU was acquired under controlled conditions, robustness was evaluated using fixed model weights under increasing brightness shifts, Gaussian noise, and motion blur. This synthetic perturbation analysis does not constitute external validation on industrial imagery.

III. Results and Discussion

All results were computed on the held-out test partition ($n = 300$) under the shared protocol of Section II.

A. Training Convergence

Fig. 1 shows the accuracy and loss curves. Training accuracy rose sharply over the first epochs before flattening, and validation accuracy tracked it without sustained divergence. Both loss curves decreased monotonically in trend, indicating stable optimisation and no substantial overfitting within the 20-epoch budget.

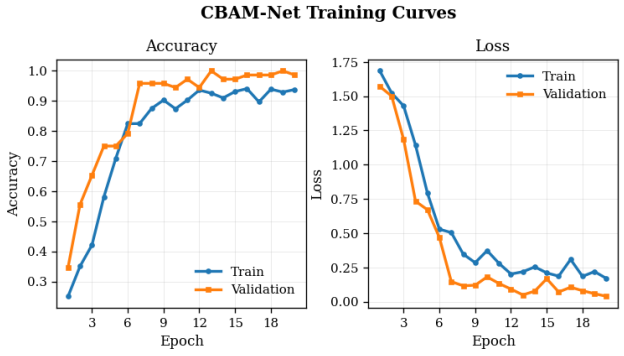


Fig. 1. Training and validation accuracy and loss curves of CBAM-Net.

B. Class-wise Performance

Fig. 2 shows class-wise precision, recall, and F1-score. Patches achieved the highest F1 (0.990) with perfect precision, followed by crazing and inclusion (0.970 each). Scratches had the lowest F1 (0.939) due to 0.920 recall, while rolled-in scale had the lowest precision (0.941) because of confusion with pitted surface. The macro-F1 of 0.963 matched the Micro-F1, consistent with the balanced test set.

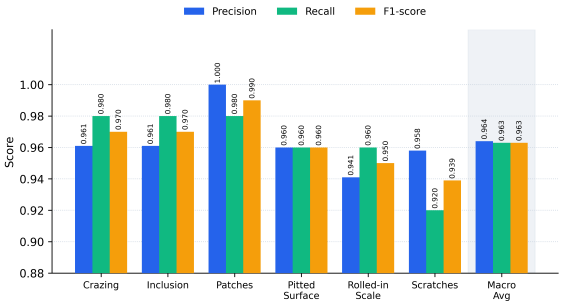


Fig. 2. Class-wise precision, recall, and F1-score.

C. Attention Ablation

Table II compares five attention variants under identical settings. Channel-only CBAM (93.33%) slightly outperformed spatial-only CBAM (92.67%), while SE achieved 94.33%. Full CBAM reached 96.33%, showing that its gain mainly comes from combining channel and spatial recalibration.

TABLE II
ABLATION OF ATTENTION COMPONENTS

Variant	Chan. Att.	Spat. Att.	Acc. (%)	Micro-F1
CNN, no attention	—	—	90.67	0.907
CNN + CBAM channel only	CBAM	—	93.33	0.933
CNN + CBAM spatial only	—	CBAM	92.67	0.927
CNN + SE channel only	SE	—	94.33	0.943
CBAM-Net (full)	CBAM	CBAM	96.33	0.963

D. ROC Analysis

Fig. 3 shows strong class separation, with AUCs of 0.996 for inclusion, 0.995 for scratches, and approximately 0.990 overall.

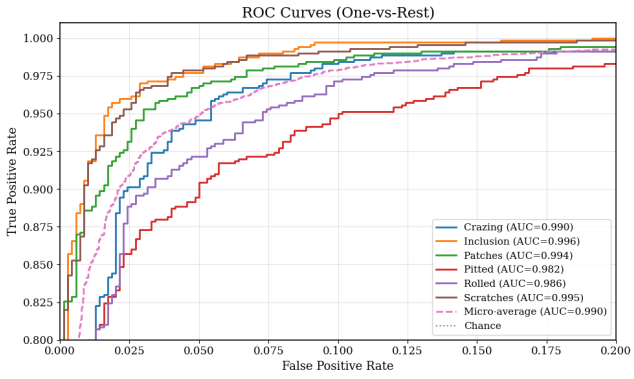


Fig. 3. One-vs-rest ROC curves for six defect classes, zoomed to distinguish high-performing classes.

E. Architecture Comparison and Computational Cost

Table III compares accuracy, parameters, FLOPs, and inference time across seven architectures. CBAM-Net achieved the highest accuracy with 2.03 M parameters, fewer than EfficientNetB0, ResNet-18, and VGG16. MobileNetV3-Small was the lightest model at 0.95 M parameters and 0.06 G FLOPs but trailed CBAM-Net by 3.66 percentage points. EfficientNetB0 was the strongest baseline at 95.00%, outperforming ResNet-18 with fewer parameters.

TABLE III
ARCHITECTURE COMPARISON UNDER A MATCHED TRAINING PROTOCOL

Model	Par. (M)	FLOPs (G)	Time (ms)	Acc. (%)	Micro-F1
CNN, no attention	1.24	0.42	4.8	90.67	0.907
MobileNetV3-Small	0.95	0.06	3.7	92.67	0.927
VGG16 (transfer)	138.36	15.50	28.4	93.33	0.933
ResNet-18	11.18	1.82	11.6	94.00	0.940
MobileNetV3-Large	3.00	0.22	5.2	94.33	0.943
EfficientNetB0	4.06	0.39	7.4	95.00	0.950
CBAM-Net	2.03	0.67	6.1	96.33	0.963

F. Confusion Matrix Analysis

Fig. 4 presents the confusion matrix. Predictions concentrate on the diagonal across all six categories. The residual errors are informative: confusion is heaviest between rolled-in scale and pitted surface, the two classes whose granular texture statistics overlap most closely, and scratches attract a small number of misassignments from several classes, consistent with its lower class-wise F1.

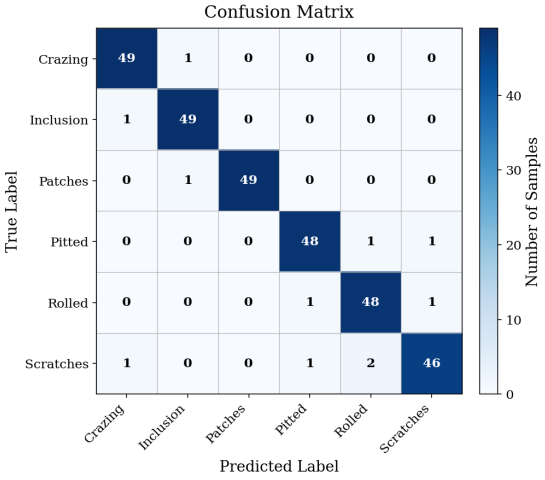


Fig. 4. Confusion matrix for six defect categories.

G. Robustness Under Acquisition Perturbations

Table IV shows that CBAM-Net is relatively robust to moderate brightness changes, with 15% darkening reducing accuracy by 1.67 percentage points. Performance drops more

sharply under Gaussian noise and motion blur, reaching 15.67 points at $\sigma = 0.10$ and 13.33 points with a 7×7 kernel, as both disrupt high-frequency defect textures. Thus, deployment requires effective noise control, suitable exposure, and adequate shutter speed.

TABLE IV
ROBUSTNESS UNDER SYNTHETIC ACQUISITION PERTURBATIONS

Perturbation	**Severity**	**Acc. (%)**	**Macro-F1**	**Drop (pp)**
None (clean)	—	96.33	0.963	0.00
Brightness down	-15%	94.67	0.946	1.67
Brightness down	-30%	91.33	0.913	5.00
Brightness up	$+15\%$	95.00	0.949	1.33
Brightness up	$+30\%$	92.33	0.922	4.00
Gaussian noise	$\sigma = 0.02$	94.00	0.939	2.33
Gaussian noise	$\sigma = 0.05$	89.67	0.895	6.67
Gaussian noise	$\sigma = 0.10$	80.67	0.803	15.67
Motion blur	3×3	93.67	0.935	2.67
Motion blur	5×5	89.33	0.891	7.00
Motion blur	7×7	83.00	0.827	13.33

H. Comparison with Reported Methods

Table V compares CBAM-Net with published CNN, transformer, and attention-based methods for steel and metal defect classification. Because datasets, preprocessing, and evaluation protocols differ, this comparison is indirect; the controlled benchmark is Table III. Recent YOLO, Swin, ViT, ECA, and GAM methods are not included and will be retrained under the same protocol in future work.

TABLE V
COMPARISON WITH REPORTED METHODS

Study	**Method**	**Acc. (%)**
Fu *et al.* [9]	Deep CNN classification	93.0
Wang and Xie [10]	Transformer (MeDERT)	94.5
Sampath *et al.* [18]	Attention-guided multitask	95.1
CBAM-Net (proposed)	Attention-enhanced CNN	96.33

I. Grad-CAM Interpretability

A representative Grad-CAM activation map is shown in Fig. 5. The highlighted region aligns with the visible defect area rather than with the surrounding mill texture, indicating that the attention-refined features drive the prediction from defect evidence. Grad-CAM is a qualitative saliency method and does not establish causal attribution; it is reported as supporting evidence that the model attends to plausible regions.

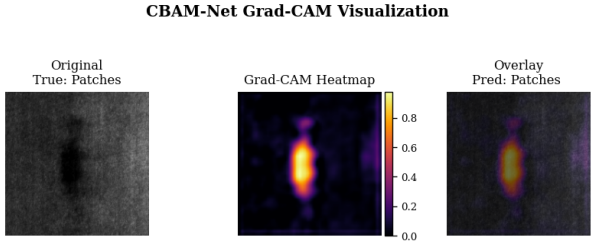


Fig. 5. Grad-CAM visualisation of defect regions.

J. Statistical Comparison

Because every model was evaluated on the same 300 test images, the outcomes are paired and McNemar's test is the appropriate choice. That test requires the per-image agreement counts rather than the marginal accuracies, and those counts were not retained for the runs reported here, so McNemar statistics are not presented. As an approximation, treating the outcomes as independent proportions gives Pearson $\chi^2 = 1.78$ ($p = 0.183$) against ResNet-18, $\chi^2 = 2.76$ ($p = 0.097$) against VGG16, and $\chi^2 = 7.93$ ($p = 0.0049$) against the no-attention baseline. On this basis only the margin over the no-attention baseline reaches significance at $\alpha = 0.05$; the 2.33 and 3.00 percentage-point margins over ResNet-18 and VGG16 do not. This is the expected outcome at $n = 300$, where a three-point difference corresponds to nine images. The independence assumption is violated by the shared test partition, so these values are indicative rather than definitive, and a paired analysis on retained predictions is needed to resolve the smaller margins in either direction.

K. Qualitative Predictions

Fig. 6 presents correctly classified examples of rolled-in scale, pitted surface, scratches, and patches. The visual similarity between rolled-in scale and pitted surface also explains their confusion in Fig. 4.

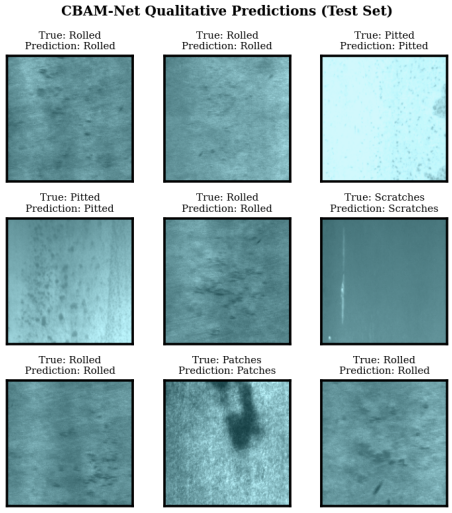


Fig. 6. Qualitative defect prediction examples.

L. Deployment Considerations

At 2.03 M parameters, CBAM-Net requires about 8.1 MB in FP32 and 2.0 MB with INT8 quantisation, making it roughly two orders smaller than VGG16. Its 6.1 ms RTX 3060 latency gives about 164 images/s at batch size one, although embedded devices may be slower due to memory-bandwidth limits. Despite a modest 0.67 G FLOP cost, 94.6% of parameters lie in the dense head, making the model memory-bound; therefore, pruning this layer is more promising than pruning the 107,491-parameter convolutional backbone. Quantisation and pruning were not evaluated and remain future work.

IV. Conclusion

This paper characterised CBAM-Net, a 2.03 M-parameter CNN applying CBAM after each of three feature-extraction blocks. Under a matched protocol, it achieved 96.33% accuracy and 0.963 Micro-F1 on the held-out NEU test set, outperforming six alternatives, including EfficientNetB0 and MobileNetV3-Large, while using fewer parameters than all except MobileNetV3-Small. A five-variant ablation showed that the gain arises from combining channel and spatial recalibration, although SE was the stronger single-channel mechanism. Robustness was relatively high to illumination shifts but lower under sensor noise and motion blur. Among the baseline comparisons, only the improvement over the no-attention model was statistically supported. Several limitations remain. Evaluation used only 1,800 laboratory-acquired NEU images, while robustness was tested through synthetic perturbations rather than production data, so industrial generalisation is unverified. Published comparisons were obtained under different protocols and omit recent YOLO-based detectors, transformer models, and alternative attention mechanisms. The novelty lies in characterising an existing CBAM configuration, whose advantage depends on combining channel and spatial attention. Latency was measured only on a workstation GPU, and the absence of retained per-image predictions prevented paired McNemar testing; therefore, the margins over ResNet-18 and VGG16 are not statistically established. Future work will compare YOLO, Swin Transformer, ViT, ECA, GAM, and coordinate attention under the same protocol; evaluate transfer on Severstal and GC10-DET; test quantised and pruned models on Jetson-class hardware; and use production-line imagery for genuine external validation.

Acknowledgement

AI-assisted language tools were used solely to improve linguistic clarity; all research design, implementation, analysis, and interpretation were undertaken by the authors.

References

[1] H.-I. Lin and F. S. Wibowo, "Image data assessment approach for deep learning-based metal surface defect-detection systems," *IEEE Access*, vol. 9, pp. 47621–47638, Jan. 2021, doi: 10.1109/access.2021.3068256.

[2] R. Usamentiaga, D. G. Lema, O. D. Pedrayes, and D. F. Garcia, "Automated surface defect detection in metals: a comparative review of object detection and semantic segmentation using deep learning," in *Proc. IEEE Industry Applications Society Annual Meeting (IAS)*, Oct. 2021, pp. 1–8, doi: 10.1109/ias48185.2021.9677231.

[3] K. B. Chigateri and A. M. Hebbale, "A steel surface defect detection model using machine learning," *Materials Today Proceedings*, vol. 100, pp. 51–58, May 2023, doi: 10.1016/j.matpr.2023.04.646.

[4] P. M. Bhatt *et al.*, "Image-based surface defect detection using deep learning: a review," *Journal of Computing and Information Science in Engineering*, vol. 21, no. 4, Jan. 2021, doi: 10.1115/1.4049535.

[5] W. Li *et al.*, "Deep learning based online metallic surface defect detection method for wire and arc additive manufacturing," *Robotics and Computer-Integrated Manufacturing*, vol. 80, p. 102470, Oct. 2022, doi: 10.1016/j.rcim.2022.102470.

[6] S. Qi, J. Yang, and Z. Zhong, "A review on industrial surface defect detection based on deep learning technology," in *Proc. 3rd Int. Conf. Machine Learning and Machine Intelligence (MLMI)*, New York, NY, USA, 2020, pp. 24–30, doi: 10.1145/3426826.3426832.

[7] K. Demir, M. Ay, M. Cavas *et al.*, "Automated steel surface defect detection and classification using a new deep learning-based approach," *Neural Computing and Applications*, vol. 35, pp. 8389–8406, 2023, doi: 10.1007/s00521-022-08112-5.

[8] X. Tao, D. Zhang, W. Ma, X. Liu, and D. Xu, "Automatic metallic surface defect detection and recognition with convolutional neural networks," *Applied Sciences*, vol. 8, no. 9, p. 1575, Sep. 2018, doi: 10.3390/app8091575.

[9] G. Fu *et al.*, "A deep-learning-based approach for fast and robust steel surface defects classification," *Optics and Lasers in Engineering*, vol. 121, pp. 397–405, May 2019, doi: 10.1016/j.optlaseng.2019.05.005.

[10] C. Wang and H. Xie, "MeDERT: a metal surface defect detection model," *IEEE Access*, vol. 11, pp. 35469–35478, Jan. 2023, doi: 10.1109/access.2023.3262264.

[11] E. Cumbajin *et al.*, "A systematic review on deep learning with CNNs applied to surface defect detection," *Journal of Imaging*, vol. 9, no. 10, p. 193, Sep. 2023, doi: 10.3390/jimaging9100193.

[12] S. Ashrafi, S. Teymouri, S. Etaati, J. Khoramdel, Y. Borhani, and E. Najafi, "Steel surface defect detection and segmentation using deep neural networks," *Results in Engineering*, vol. 25, p. 103972, Jan. 2025, doi: 10.1016/j.rineng.2025.103972.

[13] W. Zhao, F. Chen, H. Huang, D. Li, and W. Cheng, "A new steel defect detection algorithm based on deep learning," *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, p. 5592878, Jan. 2021, doi: 10.1155/2021/5592878.

[14] S. M. Yasir and H. Ahn, "Faster metallic surface defect detection using deep learning with channel shuffling," *Computers, Materials & Continua*, vol. 75, no. 1, pp. 1847–1861, Jan. 2023, doi: 10.32604/cmc.2023.035698.

[15] Z. Tang, E. Tian, Y. Wang, L. Wang, and T. Yang, "Nondestructive defect detection in castings by using spatial attention bilinear convolutional neural network," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 82–89, Apr. 2020, doi: 10.1109/tii.2020.2985159.

[16] J. Li, K. Wang, M. He, L. Ke, and H. Wang, "Attention-based convolution neural network for magnetic tile surface defect classification and detection," *Applied Soft Computing*, vol. 159, p. 111631, Apr. 2024, doi: 10.1016/j.asoc.2024.111631.

[17] M. Aslam, S. S. Naqvi, T. M. Khan, G. Holmes, and R. Naffa, "Trainable guided attention based robust leather defect detection," *Engineering Applications of Artificial Intelligence*, vol. 124, p. 106438, Jun. 2023, doi: 10.1016/j.engappai.2023.106438.

[18] V. Sampath, I. Maurtua, J. J. A. Martín, A. Rivera, J. Molina, and A. Gutierrez, "Attention-guided multitask learning for surface defect identification," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 9, pp. 9713–9721, Jan. 2023, doi: 10.1109/tii.2023.3234034.

[19] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. European Conf. Computer Vision (ECCV)*, Munich, Germany, 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1.

[20] K. Dixit, "NEU surface defect database," Kaggle, 2019. [Online]. Available: https://www.kaggle.com/datasets/kaustubhdikshit/neu-surface-defect-database