Classification of Degree of Degradation around Scribe for Coil-Coated Metallic Samples Using Convolutional Neural Models

1st Pavel Rozsival*, 2nd Petr Dolezel†, 3rd Bruno Baruque Zanon‡, 4th Dominik Stursa§
Faculty of Electrical Engineering and Informatics
University of Pardubice
Pardubice, Czech Republic
Email: *pavel.rozsival@upce.cz, †petr.dolezel@upce.cz,
§dominik.stursa@upce.cz
Departamento de Ingeniería Civil
Universidad de Burgos
Burgos, Spain
Email: ‡bbaruque@ubu.es

***Abstract*—Coil coating, a technique for applying organic coatings to rolled metal strip substrates, plays a critical role in achieving consistent, high-quality surface finishes. However, these protective coatings are vulnerable to mechanical damage, which can lead to irreversible alterations when exposed to environmental elements. Traditionally, assessing degradation resistance in coil-coated materials involves manual determination of degraded areas. In this study, a classification-based approach to automate this assessment is proposed. Additionally, the selected classification models are compared with semantic segmentation, highlighting their performance and computational efficiency. The results demonstrate that both approaches (classification and semantic segmentation) can assess degradation, with semantic segmentation providing highly accurate results and classification models offering efficient practical deployment alternatives.**

***Index Terms*—coil coating, delamination, degradation, deep learning, classification**

I. Introduction

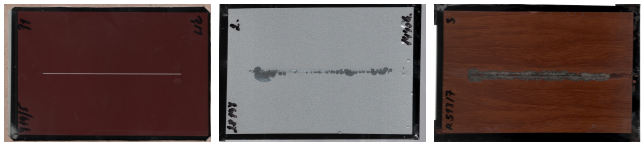
Coil coating stands as a pivotal technique in the continuous, automated application of organic coatings onto rolled metal strip substrates within industrial settings. The primary objective is to achieve a consistent, high-quality surface finish that endures over time [1], [2]. This method plays a critical role in providing long-lasting protection for various applications demanding resilient finishes, including building exteriors, metal roofing, wall panels, garage doors, office furniture, vending machines, food service equipment, among others. Furthermore, it extends its utility to specialized coatings like cool metal roofing materials, pollution-reducing building panels, antimicrobial products, corrosion-resistant metal components, and solar panels.

In essence, coil coating provides a thin yet robust and flexible protective layer, effectively shielding the substrate from corrosion. However, this protective layer remains vulnerable to mechanical damage such as scratches or abrasions, which could lead to irreversible alterations when exposed to environmental elements like moisture, UV radiation, salt, or corrosive gases [3]. Such damages may manifest in diverse forms, ranging from chalking and blistering to flaking or rusting of the coated surface. Consequently, it becomes imperative to assess the performance of coated surfaces under conditions simulating real-world outdoor exposure.

The evaluation of degradation resistance in coil-coated materials aligns with the guidelines outlined in the European Standard EN 13523-8 titled "Coil coated metals. Test methods. Resistance to salt spray (fog)." This standard entails subjecting a coated specimen to a salt fog under predefined temperature and duration conditions. Subsequently, the specimen undergoes assessment following the criteria set forth in the International Organization for Standardization (ISO) 4628 standard titled "Paints and varnishes. Evaluation of degradation of coatings. Designation of the quantity and size of defects and intensity of uniform changes to appearance." In simpler terms, the objective is to gauge the extent of degradation on the specimen, as illustrated in Fig. 1. The degree of degradation is then quantified by calculating the ratio of the degraded area to the total specimen area. The degree of degradation is determined on a scale of integers from 1 to 5.

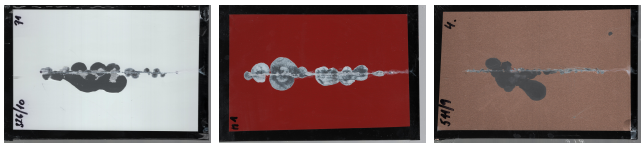
Traditionally, the determination of the degraded area is performed manually, as indicated in prior studies [3], [4]. Surprisingly, few studies have explored automatic image processing-based approaches. Some have proposed specific methods for degradation evaluation using image processing techniques [5], [6]. Others have suggested utilizing office

The work has been supported by the SGS grant at the Faculty of Electrical Engineering and Informatics, University of Pardubice, Czech Republic. This work was supported by the European Regional Development Fund (ERDF)/European Social Fund (ESF) "Cooperation in Applied Research between the University of Pardubice and companies, in the Field of Positioning, Detection and Simulation Technology for Transport Systems (PosiTrans)" under Grant No. CZ.02.1.01/0.0/0.0/17_049/0008394. This support is very gratefully acknowledged.



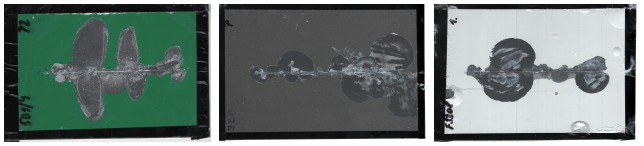
Degree of degradation 1

Degree of degradation 2



Degree of degradation 3

Degree of degradation 4



Degree of degradation 5

Fig. 1: Examples of degraded area determination. Each sample is evaluated according to ISO 4628 and the final degree is shown under each three samples.

scanners in tandem with commercial software like Adobe Photoshop [7]. However, these approaches commonly suffer from limitations tied to specific material types, surface characteristics, or sample colors. In contrast, industrial demands necessitate methods that exhibit versatility across a broad spectrum of materials. Addressing this need, the authors of this study propose a semantic segmentation-based approach employing fully-convolutional neural networks [8]. See Fig. 2 for the pipeline of the discussed method.

The aforementioned method based on semantic segmentation provides highly accurate results; however, a drawback lies in the complexity and time-intensive nature of designing the semantic segmentation model. Conversely, the task at hand, as defined in ISO 4628 standard, essentially involves classifying a coated specimen into five categories, thus constituting an image classification task. To address this, highly reliable convolutional neural networks have been developed. These networks excel in extracting features from images and making predictions based on learned patterns.

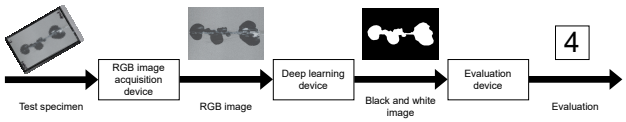


Fig. 2: A pipeline of the autonomous method for the assessment of the degree of degradation around a scribe presented in [8]. First, the specimen is scanned and the image is preprocessed. Then, an image segmentation task is performed. Ultimately, the degree of degradation is then quantified by calculating the ratio of the degraded area to the total specimen area.

In this study, our aim is to verify the efficacy of solving this task through classification and compare the results with those obtained through semantic segmentation. Therefore, a series of convolutional neural models for image classification will be designed in order to classify coated specimens, as illustrated in Fig. 3.

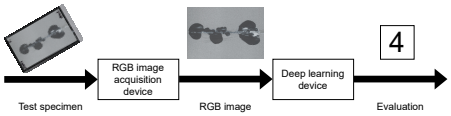


Fig. 3: A pipeline of the autonomous method for the assessment of the degree of degradation around a scribe based on image classification convolutional neural networks.

II. Materials and Methods

In this section, the methodologies employed to investigate the performance of classification-based approaches in contrast to semantic segmentation for evaluating the degradation resistance of coil-coated materials are delineated. The study primarily focuses on the task of categorizing metal specimens into distinct classes, thereby necessitating the utilization of convolutional neural networks tailored for image classification. The experimental setup, data collection procedures, network architectures, training methodologies, and evaluation metrics utilized to assess the efficacy of classification-based techniques in comparison to semantic segmentation are outlined below.

A. Standard ISO 4628-8

Various methodologies for examining coil-coated metals include assessing their resistance to degradation induced by salt fog, as specified by the EN 13523-8 standard. A crucial aspect of this evaluation involves quantifying the extent of degradation surrounding a scribe, a parameter meticulously defined by the ISO 4628-8 standard. The testing procedure entails horizontally scribing the test specimen with a sharp edge and subjecting it to a corrosive salt fog environment. After a predefined exposure duration, the specimen is cleansed with tap water and residual water is removed using compressed air. Any loosely adhered coating is then removed with a blade held at a precise angle.

TABLE I: Training and test set

Degree of degradation	1	2	3	4	5
Occurrence in training set	210	84	47	51	64
Occurrence in test set	50	18	19	18	23

In accordance with the guidelines outlined in the ISO 4628-8 standard, quantifying the degree of degradation around a scribe follows a specific protocol: the actual area of degradation is measured. The standard recommends placing transparent millimeter-grid paper over the plate and counting the squares corresponding to the degraded region. Subsequently, the degree of degradation (d) in millimeters can be calculated using the following equation.

$$d = \text{round}\left(\frac{A_d - A_l}{2l}\right), \tag{1}$$

where A_d is the area of degradation, including the area of the scribe (in square millimeters), A_l is the area of the scribe in the evaluated area in square millimeters, and l is the length of the scribe in the evaluated area (in millimeters).

B. Dataset

In order to devise a robust data-driven approach for the automated detection of coating degradation, it is essential to accumulate a comprehensive dataset comprising diverse annotated samples exhibiting variations in color, texture, and reflectivity, while also encompassing various levels of degradation. In the seminal work by Rozsivalova et al. [8], a meticulously curated dataset consisting of 584 coated samples measuring 150×100 mm was assembled. These samples featured coatings in a spectrum of colors, spanning black, white, green, gray, orange, red, brown, blue, dark blue, and yellow, encompassing both fine and coarse textures.

To expose the uncoated substrate, a small horizontal scratch, measuring 0.5 mm in width, was deliberately introduced through the coating of each sample using an iron nail. Subsequently, the samples underwent exposure in a salt fog chamber for defined duration of 120h, 240h, 480h, 720h, and 1440 h. Following the exposure period, the samples underwent thorough cleaning and were scanned using an office scanner. Subsequently, each sample was manually annotated to accurately determine the degree of degradation in accordance with Standard ISO 4628-8.

The collected dataset was divided into classes according to the degree of degradation and divided into training and test sets. The occurrences of each class are summarized in Table I.

It is evident that the frequency distribution of individual classes within the training dataset is uneven; however, this imbalance was unavoidable due to the strictly experimental and time-intensive nature of dataset creation. This phenomenon will be taken into account during the training of neural models.

C. Convolutional neural models for classification

In this chapter, the utilization of convolutional neural networks for the classification of coil-coated metal samples is dealt with. Convolutional neural networks are recognized as a cornerstone in contemporary image analysis methodologies, primarily owing to their inherent ability to autonomously discern and extract salient features from raw image data, facilitating precise predictions. The process is initiated by presenting input images to the neural network architecture, where a series of convolutional layers, augmented by pooling and activation functions, progressively analyze the spatial characteristics of the input. Within these convolutional layers, intricate patterns and features within the images are adeptly detected, capturing both local and global information crucial for accurate classification. Subsequently, the extracted features undergo flattening and are fed into fully connected layers, where they are further refined, and output predictions corresponding to the distinct classes of metal samples based on their degradation levels are generated. Throughout this process, subtle nuances and discriminative features inherent in the images are learned by the convolutional network, enabling robust and reliable classification performance [9].

Following a comprehensive survey of the literature, architectures such as ResNet [10], EfficientNet [11], and DenseNet [12] were selected to address our problem. These architectures were chosen based on their proven effectiveness in a wide array of image classification tasks. ResNet, renowned for its residual learning framework, enables the training of exceptionally deep networks while mitigating the vanishing gradient problem. EfficientNet, characterized by its compound scaling method, achieves superior performance by balancing model depth, width, and resolution. DenseNet, distinguished by its dense connectivity pattern, fosters feature reuse and enhances gradient flow throughout the network. The rationale behind selecting these architectures lies in their robustness and adaptability to diverse datasets, ensuring optimal performance in our classification task. Some of the important features of the selected architectures are shown in Table II.

1) ResNet: ResNet is characterized by its residual learning framework, allowing the training of exceptionally deep neural networks. This architecture has been widely recognized for its ability to alleviate the vanishing gradient problem, which often hinders the training of deep networks. ResNet152V2, which is utilized in this study, offers a deeper and more sophisticated network structure compared to its predecessors, thus enabling enhanced feature extraction and representation learning capabilities [13].

2) EfficientNet: Similarly, the EfficientNet architecture was utilized, specifically employing the EfficientNetB3 variant [14]. EfficientNet is distinguished by its compound scaling method, which optimally balances model depth, width, and resolution. This unique scaling approach enables EfficientNet to achieve superior performance by efficiently allocating computational resources across different network dimensions.

3) DenseNet: The DenseNet architecture was also employed, explicitly selecting the DenseNet201 variant [15]. DenseNet is characterized by its dense connectivity pattern, which facilitates feature reuse and gradient flow enhancement throughout the network. This connectivity scheme allows for the seamless integration of features across different layers, pro-

TABLE II: Selected neural models for classification

Architecture	Input size	Parameters	Model size
ResNet152V2	$(288 \times 288 \times 3)$	58×10^6	685 MB
EfficientNetB3	$(288 \times 288 \times 3)$	11×10^6	127 MB
DenseNet201	$(288 \times 288 \times 3)$	18×10^6	216 MB

moting effective information propagation and gradient signal preservation. DenseNet201, chosen for its deeper and more densely connected network structure, demonstrates enhanced feature extraction and representation learning capabilities, thereby bolstering the efficacy of defined classification task.

D. Semantic segmentation-based approach employing fully-convolutional neural networks

This approach leverages a computer-driven, deep learning-based methodology, which is composed of two sequential U-shaped convolutional networks. This approach is implemented as "Deep learning device" in Fig. 2 and serves as a competitor to the classification models described above.

The first network in the sequence is designed for preliminary detection. It scans the input data and identifies potential areas of degradation around the scribe. This initial detection phase serves as a rough estimate, flagging areas of interest for further analysis.

The second network in the sequence is designed for refined detection of the degradation area around a scribe. It takes the output from the first network and applies a more detailed and rigorous analysis to accurately identify and quantify the degradation.

The detailed description is available in [8]. A simplified pipeline of this approach is depicted in Fig. 4.

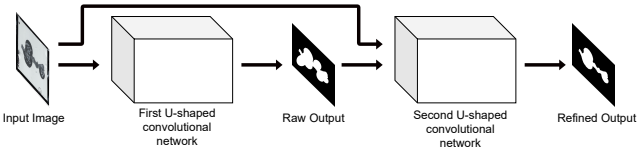


Fig. 4: A pipeline of the Deep learning device module in Fig. 2.

E. Training details

All classification models are trained using the Adam optimization algorithm, widely recognized for its generally satisfactory performance in various scenarios [16], [17]. The model weights are initialized randomly with a Gaussian distribution. During training, categorical cross-entropy is employed as the loss function. A validation set comprising 15% of the training dataset is utilized for model validation. To address the stochastic nature of training, the experiments are conducted ten times. The models demonstrating the best performance, determined by the loss function on the validation set, are subsequently chosen for further evaluation. Detailed parameters regarding the training process are summarized in Table III.

Due to the uneven distribution of individual classes within the training dataset, class weighting was employed during training to ensure that each class in the training batch was included with equal probability. This approach aimed to address the imbalance in class representation and prevent the model from being biased towards the majority class during training.

TABLE III: Parameters of the training

Input shape	288 x 288 x 3
Training algorithm	Adam optimizer
Number of training experiments	10
Number of samples	456
Validation split	0.15
Initialization	Normal distrib. (mean = 0, std = 0.05)
Number of epochs	500
Batch size	16
Criterion for resultant model	Loss function value over validation set
Learning rate α	0.001
Exponential decay rate 1 β_1	0.9
Exponential decay rate 2 β_2	0.999

Furthermore, data augmentation was implemented to ensure greater diversity within the training dataset, given the relatively low number of samples available. Specifically, methods such as rotation, horizontal and vertical shifting, shear, zoom, and horizontal flipping were employed, each applied at a maximum rate of 10 percent relative to the original image. This augmentation strategy aimed to introduce variations in the training data, thereby enhancing the model's ability to generalize to unseen samples and improve overall performance.

F. Evaluation metrics

A set of evaluation metrics must be considered to clearly demonstrate the benefits or shortcomings of the proposed method. For the purposes of this study, the metrics include accuracy, which measures the proportion of correctly classified samples among the total dataset, providing an overall indication of model performance. The confusion matrix is employed to analyze classification results in detail, depicting the distribution of true positive, true negative, false positive, and false negative predictions for each class. From the confusion matrix, metrics such as precision, recall, and F1-score can be derived, offering insights into the model's performance across different classes. Additionally, relative inference time is utilized to gauge the computational efficiency of the models during the inference phase. This metric quantifies the time taken by each model to process a single input sample relative to a baseline model, providing crucial information for real-world deployment scenarios where inference speed is essential. Together, these evaluation metrics offer a comprehensive assessment of the classification models' performance in terms of accuracy, class-wise performance, and computational efficiency.

III. Experiments and results

A. Setup of experiments

All experimental models were implemented using Python 3.9 with TensorFlow 2.0 and Keras framework. Experiments were performed using the following hardware specification: Intel Core i5-8600K (3.6 GHz) processor, 16 GB DDR4 (2666 MHz) internal memory, NVIDIA PNY Quadro P5000

16 GB GDDR5 PCIe 3.0 (2560 CUDA cores) video card, SATA M.2 512 GB SSD.

B. ResNet152V2

The overall accuracy over the test set for this model is 0.796. The Confusion Matrix is shown in Table IV.

TABLE IV: Confusion Matrix for Resnet152V2

		Actual values				
		1	**2**	**3**	**4**	**5**
Predicted values	**1**	45	2	0	0	0
	2	3	13	3	0	0
	3	0	7	6	3	1
	4	0	0	3	15	2
	5	0	0	0	3	22

C. EfficientNetB3

The overall accuracy over the test set for this model is 0.719. The Confusion Matrix is shown in Table V.

TABLE V: Confusion Matrix for EfficientNetB3

		Actual values				
		1	**2**	**3**	**4**	**5**
Predicted values	**1**	46	1	0	0	0
	2	4	10	5	0	0
	3	0	6	8	2	1
	4	0	0	9	4	7
	5	0	0	1	5	19

D. DenseNet201

The overall accuracy over the test set for this model is 0.797. The Confusion Matrix is shown in Table VI.

TABLE VI: Confusion Matrix for DenseNet201

		Actual values				
		1	**2**	**3**	**4**	**5**
Predicted values	**1**	45	2	0	0	0
	2	3	12	3	1	0
	3	0	7	7	2	1
	4	0	0	0	16	4
	5	0	0	0	2	23

E. Semantic segmentation-based approach

The overall accuracy over the test set for this approach is 0.891. The Confusion Matrix is shown in Table VII.

TABLE VII: Confusion Matrix for semantic segmentation-based approach

		Actual values				
		1	**2**	**3**	**4**	**5**
Predicted values	**1**	45	1	0	0	1
	2	0	18	1	0	0
	3	0	3	13	1	0
	4	0	0	4	16	0
	5	0	0	0	3	22

TABLE VIII: Relative inference time

Architecture	**Relative inference time**
ResNet152V2	0.7981
EfficientNetB3	0.4235
DenseNet201	0.6817
Semantic segmentation-based approach	1.000

F. Relative inference time

In order to determine this metric, the responses of all three models together with the competitive approach were measured over the test set. The inference time of the competitive approach was set as a baseline time and the inference times of the classification models are shown in Table VIII relative to this baseline time.

G. Discussion

In this study, the efficacy of classification-based approaches in assessing the degree of degradation around scribes on coil-coated metallic samples was investigated. The goal was to compare these approaches with a semantic segmentation-based method employing fully-convolutional neural networks.

From the performance point of view, ResNet152V2 achieved an overall accuracy of 0.796, EfficientNetB3 achieved 0.719, and DenseNet201 achieved 0.797. While all models provided reasonable accuracy, their class-wise performance varied. For instance, EfficientNetB3 excelled in classifying degradation level 1, while DenseNet201 performed well in classes 4 and 5. Nevertheless, the semantic segmentation-based approach achieved an overall accuracy of 0.891 and it outperformed the accuracy of other models in most classes. It is worth noting that designing the semantic segmentation model is complex and time-intensive.

Regarding computational efficiency, the semantic segmentation-based approach served as the baseline (relative inference time = 1.000). ResNet152V2 had a relative inference time of 0.7981, EfficientNetB3 was faster at 0.4235, and DenseNet201 fell in between at 0.6817. Thus, it is important to emphasize that all classification models are significantly more computationally efficient than the semantic segmentation-based approach.

Hence, the findings suggest that both classification and semantic segmentation approaches can assess degradation in coil-coated materials. Researchers and practitioners can choose the most suitable method based on their specific requirements. While semantic segmentation provides highly accurate results, classification models offer a simpler alternative for practical deployment.

IV. Conclusions

This study demonstrated the effectiveness of a classification-based approach in automating the assessment of degradation in coil-coated materials. By utilizing convolutional neural networks, the approach was able to classify coated specimens into distinct categories, providing a practical alternative to manual determination. While semantic segmentation method, considered as a competitive approach in this study, achieves

highly accurate results, it is computationally intensive and difficult to design. In contrast, regular classification models offer efficient deployment alternatives without sacrificing performance. Researchers and practitioners can choose the most suitable method based on their specific requirements.

For future research, exploring hybrid approaches that combine classification and semantic segmentation could potentially lead to even better performance. Additionally, investigating other neural network architectures and fine-tuning hyperparameters may enhance the robustness of degradation assessment.

REFERENCES

[1] A.-S. Jandel, "Sustainability in the coil coatings sector [nachhaltigkeit in der coil-coating-branche]," *JOT, Journal fuer Oberflaechentechnik*, vol. 59, no. 8, pp. 10–13, 2019.

[2] The National Coil Coating Association, "NCCA promotes growth of coil coating industry," *Light Metal Age*, vol. 78, no. 5, pp. 66–67, 2020.

[3] A. Bastos and A. Simões, "Effect of deep drawing on the performance of coil-coatings assessed by electrochemical techniques," *Progress in Organic Coatings*, vol. 65, no. 2, pp. 295–303, 2009.

[4] J. Hanus, "Selection and evaluation of singlelayer coating compositions in corrosive environments," *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, vol. 59, no. 5, pp. 53–64, 2011.

[5] P. Kapsalas, M. Zervakis, and P. Maravelaki-Kalaitzaki, "Evaluation of image segmentation approaches for non-destructive detection and quantification of corrosion damage on stonework," *Corrosion Science*, vol. 49, no. 12, pp. 4415–4442, 2007.

[6] E. C. Cringasu, A. Dragomirescu, and C. A. Safta, "Image processing approach for estimating the degree of surface degradation by corrosion," in *2017 International Conference on ENERGY and ENVIRONMENT (CIEM)*, Oct 2017, pp. 275–278.

[7] P. Blanchard, D. Hill, G. Bretz, and R. McCune, *Evaluation of Corrosion Protection Methods for Magnesium Alloys in Automotive Applications*, 2014, vol. 9781118858943.

[8] V. Rozsivalova, P. Dolezel, D. Stursa, and P. Rozsival, "Sequence of u-shaped convolutional networks for assessment of degree of delamination around scribe," *International Journal of Computational Intelligence Systems*, vol. 15, no. 1, 2022.

[9] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Computation*, vol. 29, no. 9, p. 2352 – 2449, 2017.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," vol. 2016-December, 2016, p. 770 – 778.

[11] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," vol. 2019-June, 2019, p. 10691 – 10700.

[12] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," vol. 2017-January, 2017, p. 2261 – 2269.

[13] P. Shourie, V. Anand, and S. Gupta, "Automatic classification of alzheimer's using brain mri data and the resnet152v2 architecture," 2023.

[14] D. Mehta, A. Mohite, V. Shinde, R. Khatri, and I. Dokare, "Mri-based alzheimer's disease detection using efficientnetb3," 2023.

[15] S. Aluvala, A. A. Issa, K. Saranya, Z. A. Almoussawi, and A. M. Shakir, "Densenet201 based feature extraction for remote sensing image classification in land cover monitoring," 2023.

[16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980

[17] E. M. Dogo, O. J. Afolabi, N. I. Nwulu, B. Twala, and C. O. Aigbavboa, "A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks," in *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, 2018, pp. 92–99.