BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 74(5), 2026, Article number: 1141
DOI: 10.24425/bpasts.2026.1141

Weld pool analysis from images using deep learning methods

Przemysław FRANKIEWICZ*, **Jan PAWLIK**, and **Michał BEMBENEK**

AGH University of Krakow, Faculty of Mechanical Engineering and Robotics, Department of Manufacturing Systems, Krakow, Poland

Abstract. This paper presents a method for weld quality assessment based on static images of weld beads produced during the cladding process. The dataset consisted of 400 images from real industrial cladding operations, equally divided into acceptable and unacceptable samples. Three convolutional neural network models were evaluated: a custom CNN, ResNet50, and MobileNetV2. MobileNetV2 achieved the best overall performance, with accuracy = 0.7625, precision = 0.8000, recall = 0.7000, F1-score = 0.7467, and AUC = 0.8375, while ResNet50 achieved the highest AUC = 0.8719. Grad-CAM analysis showed that the models focused on relevant weld regions, including fusion lines and structural irregularities. In addition, manually isolated weld pool regions were analyzed using six geometric features: area, ellipticity, symmetry_x,symmetry_y, offset_x, and offset_y. Statistical analysis showed that symmetry_x and offset_x were the most significant features related to weld quality. Among classical machine learning models trained on these features, Random Forest achieved the best results with accuracy = 0.603,F1-score = 0.582, and AUC = 0.634. The results confirm that transfer learning is effective under low-data conditions, while weld pool geometry provides useful supplementary information but is not sufficient as a standalone input for reliable weld quality assessment.

Keywords: weld pool; WAAM additive cladding; transfer learning; weld quality; geometric feature analysis; Grad-CAM; CNNs.

1. INTRODUCTION

The rise of Industry 4.0, growing concern for welders' health, and the increasing availability of collaborative robots (cobots) have accelerated the automation of both welding and additive manufacturing methods, such as Wire Arc Additive Manufacturing (WAAM) [1, 2]. In this context, maintaining high weld quality is critical to the structural integrity of components in industries such as automotive, aerospace, energy, and construction [3]. Defective welds can result in failures, financial loss, or even safety hazards [4]. Traditional weld quality assessment methods, such as visual inspection, ultrasonic testing, radiography, and dye penetrant testing, are effective but time-consuming, costly, and reliant on skilled personnel [5]. As a result, there is growing interest in automating not only welding but also post-process inspection using computer vision and machine learning [6, 7]. Convolutional neural networks (CNNs) have proven effective in image classification and object detection, making them suitable for weld quality evaluation [8,9]. However, training deep learning models typically requires large labeled datasets, often unavailable in niche welding scenarios, such as cladding with titanium-based or wear-resistant alloys [10, 11]. This limitation can be addressed through transfer learning, where pretrained models (e.g., ResNet50, MobileNetV2) are fine-tuned for specific tasks using smaller datasets [12].

In this study, we investigate weld quality classification based on static images captured after cladding. Beyond comparing the performance of standard CNNs and transfer learning models [13, 14], we also explore the relationship between weld pool geometry and final weld quality [15, 16]. Selected weld pool regions were manually extracted and analyzed in terms of area, ellipticity, symmetry, and positional offset [17]. Additionally, Grad-CAM visualization was used to highlight image regions most influential to the model predictions [18, 19]. This hybrid approach aims to improve interpretability and supports automated quality assessment in robotic cladding processes, including WAAM [20, 21].

2. METHODOLOGY

The dataset used in this study consisted of 400 images of weld beads obtained from real-world cladding processes. Each image captured a single weld pass and was categorized by certified welding inspectors, Visual Testing (VT1+VT2), into one of two quality classes: acceptable or unacceptable, resulting in 200 samples per class. The labeling process was guided by standards such as PN-EN ISO 5817 and PN-EN ISO 6520-1, ensuring that the ground truth was both reliable and aligned with industrial practices [22, 23]. All images were saved in JPG format and resized to 224×224 pixels to match the input requirements of the convolutional neural networks used in the study [24]. No additional preprocessing or filtering was applied beyond normalization to the 0 to 1 range. Example images from both classes are shown in Fig. 1 to illustrate the visual diversity of the dataset.

To evaluate model performance, the dataset was initially divided into a training set (80%) and a test set (20%), using stratified sampling to preserve the class balance. To ensure the robustness and generalization of the models, a five-fold stratified

*e-mail: pfrankie@agh.edu.pl

Manuscript submitted 2026-02-14, revised 2026-04-17, initially accepted for publication 2026-04-27, published in September 2026.

(a)

(b)

Fig. 1. Weld images from the database: (a) acceptable welds and (b) unacceptable welds

cross-validation was applied within the training data. In each fold, four subsets were used for training and one for validation. This approach allowed for a balanced assessment of each model predictive capabilities while reducing the risk of overfitting, which is a particularly important consideration when working with a relatively small dataset.

Three different convolutional neural network architectures were implemented and compared:

- A custom lightweight CNN, built from scratch
- A ResNet50 model using transfer learning
- A MobileNetV2, also using transfer learning

The custom CNN was designed to be simple and lightweight, making it suitable for small datasets [25]. It consisted of three convolutional layers with 3×3 filters and ReLU activation functions, interleaved with max pooling layers to reduce spatial dimensions and control overfitting [26, 27]. After feature extraction, the output was flattened and passed through a fully connected (Dense) layer, followed by a final output neuron with a sigmoid activation for binary classification [28, 29]. The ResNet50 and MobileNetV2 architectures were imported from keras.applications, pre-trained on the ImageNet dataset [30, 31]. These models leverage deep, hierarchical feature extraction capabilities learned on a large and diverse set of natural images. For our task, the convolutional base of each model was frozen, and a custom classifier head was added: a GlobalAveragePooling2D layer, followed by a Dense(128) layer with ReLU activation, and a final sigmoid output layer. This transfer learning approach enables efficient model adaptation while significantly reducing training time and required data volume, which is a key advantage when working with specialized welding images. All models were trained using Adam optimizer with a learning rate of 0.0001, Binary cross-entropy loss, Batch size of 16 and up to 20 training epochs, depending on convergence.

Training was conducted in Google Colaboratory using a Tesla T4 GPU. Importantly, no data augmentation techniques were applied to assess raw model performance in realistic low-data conditions [32]. The key characteristics of the three evaluated architectures, including model complexity, number of trainable parameters, batch size, number of training epochs, and average training time per fold, are summarized in Table 1.

Table 1
Summary of model architectures and training parameters

Model	Custom CNN	ResNet50 (TL)	MobileNetV2 (TL)
Type	From scratch	Transfer learning	Transfer learning
Depth	Low	Very deep	Moderate
Layers	8	> 175	> 85
Trainable parameters	~ 28 000	~ 2 million	~ 500 000
Total parameters	~ 30 000	~ 23 million	~ 2.2 million
Batch size	16	16	16
Epochs	20	10–15	10–15
Training time / Fold	~ 1 min	~ 4–5 min	~ 3 min

To quantitatively assess the classification performance of each model, we used five standard evaluation metrics. Accuracy measures the overall proportion of correctly classified samples, while precision refers to the proportion of correctly identified defective welds among all predictions labeled as defective. Recall indicates the proportion of truly defective welds that were correctly identified by the model. The F1-score, being the harmonic mean of precision and recall, provides a balanced measure that accounts for both false positives and false negatives. Finally, AUC (Area Under the ROC Curve) serves as a threshold-independent metric, capturing the overall ability of the model to distinguish between acceptable and defective welds. In this study, recall and AUC were considered the most important indicators, given that the main industrial objective is to minimize the risk of missing defective welds. Precision and F1-score were used as supporting metrics, while accuracy served as a secondary reference due to the balanced dataset. To enhance model interpretability, we applied Gradient-weighted Class Activation Mapping to selected predictions from the test set. Grad-CAM visualizations were generated for correctly and incorrectly classified samples, highlighting the image regions that most influenced the decision-making of the model. This interpretability layer provides valuable insights into whether the model relies on relevant visual cues such as defects, irregular geometries, or surface irregularities [33, 34].

In addition to classifying full weld bead images, we extracted weld pool regions from all 400 samples (200 acceptable, 200 unacceptable). These cropped images are hypothesized to contain geometrical cues that may correlate with the final weld quality [35, 36]. Example weld pool crops for both classes are shown in Fig. 2.

Each weld pool image was analyzed based on four geometric features: area (total number of pixels in the segmented region), ellipticity (ratio of major to minor axis of a fitted ellipse), symmetry (calculated as pixel distribution across horizontal and vertical axes), and centroid offset (distance between the pool center and the image center) [37, 38]. These features were standardized and subjected to multiple levels of analysis to investigate their relevance.

(a) (b)

Fig. 2. Weld pool images from the database: (a) weld pool from acceptable welds and (b) weld pool from unacceptable welds

To analyze the relationship between weld pool geometry and weld quality, we applied statistical tests, such as t-test and Mann–Whitney, to compare feature distributions, and calculated Pearson correlations to examine dependencies between features [39]. Principal Component Analysis (PCA) was used for dimensionality reduction and visualization. Finally, we trained three classical classifiers: Logistic Regression, Random Forest, and SVM to evaluate whether weld quality could be predicted solely from geometric features [40, 41].

3. RESULTS

In order to evaluate the effectiveness of various deep learning approaches in weld bead quality classification, three convolutional neural networks were trained and compared: a custom CNN developed from scratch, and two pretrained models, ResNet50 and MobileNetV2, adapted through transfer learning. This comparison aimed to assess whether more complex, pretrained architectures would outperform a simpler baseline network, especially when trained on a relatively small dataset of weld images. Training for all three models was conducted under the same conditions, using the same dataset split and preprocessing techniques. Their learning behavior was monitored through validation curves, as shown in the combined training and validation plots in Fig. 3 [42].

The custom CNN struggled to converge effectively and displayed considerable fluctuations in both training and validation accuracy, as well as high loss values. This suggests underfitting and a lack of representational capacity, likely due to the limited depth of the network. In contrast, both ResNet50 and MobileNetV2 exhibited more stable and consistent training curves, with lower validation loss and higher accuracy, indicating better learning dynamics and generalization potential.

After training, the performance of each model was evaluated on the validation set using standard classification metrics such as accuracy, precision, recall, F1-score, and area under the curve. The results are summarized in Table 2 and visualized in a radar plot in Fig. 4.

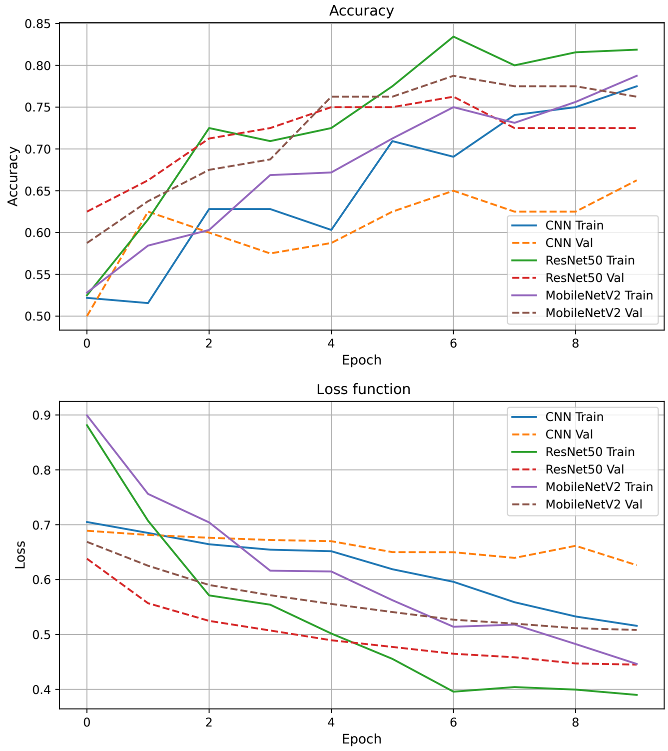

Fig. 3. Learning curves and loss function curves for the simple CNN model, ResNet50, and MobileNetV2

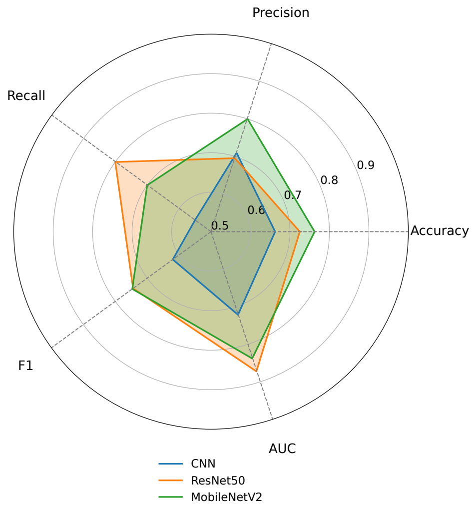

Fig. 4. Radar chart comparing the classification of the three evaluated models: CNN, ResNet50, and MobileNetV2

Table 2
Classification performance of the evaluated CNN models

Model	CNN	ResNet50	MobileNetV2
Accuracy	0.6625	0.7250	0.7625
Precision	0.7097	0.6957	0.8000
Recall	0.5500	0.8000	0.7000
F1-score	0.6197	0.7442	0.7467
AUC	0.7213	0.8719	0.8375

Among the three, MobileNetV2 achieved the best overall performance with an accuracy of 0.76 and an F1-score of 0.75, indicating a strong balance between precision and recall. ResNet50 followed closely, with a slightly lower F1-score but the highest AUC (0.87), suggesting that it had the best capacity for distinguishing between the two classes across all thresholds. Meanwhile, the custom CNN underperformed on all metrics, with an accuracy of just 0.66 and an F1-score of 0.62, further confirming its limited suitability for this task. To gain deeper insight into the nature of classification errors, confusion matrices for each model were generated, shown in Fig. 5.

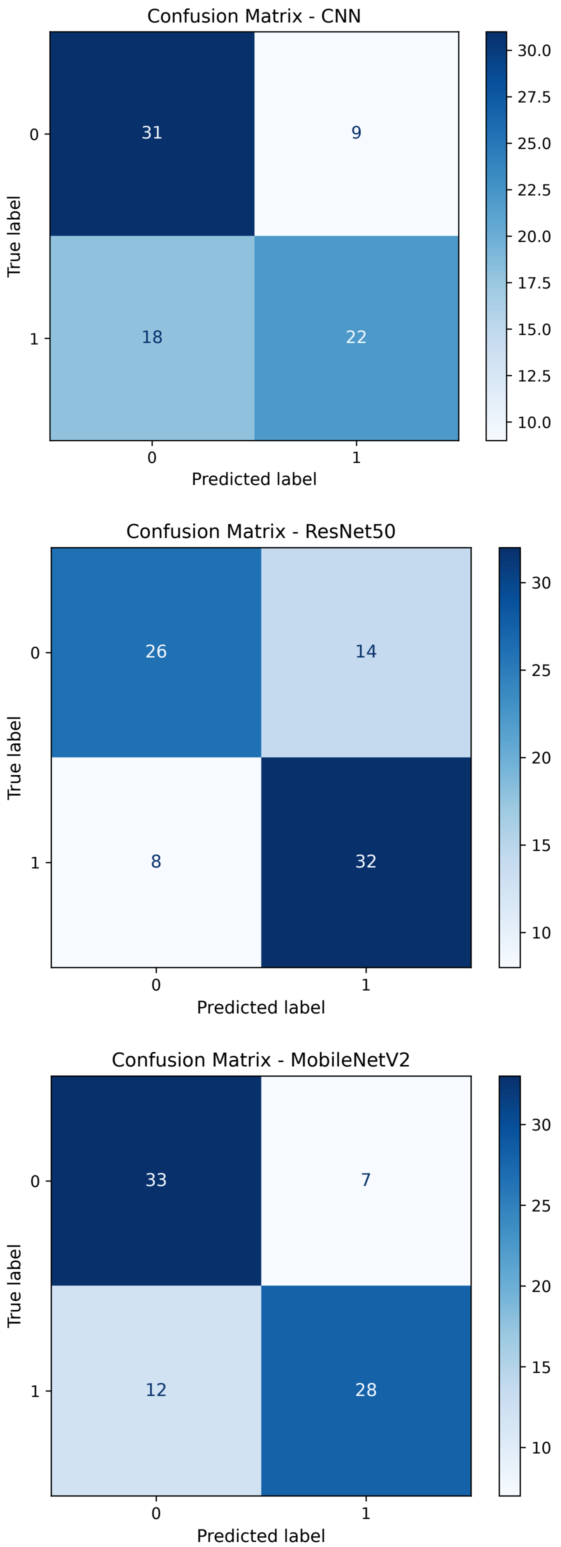

Fig. 5. Confusion matrices for: CNN, ResNet50, and MobileNetV2

The CNN was prone to misclassifying good welds as defective (false negatives), suggesting poor feature extraction and a lack of robustness. ResNet50 and MobileNetV2, on the other hand, produced more balanced predictions overall, with MobileNetV2 showing particularly low false-positive and false-negative rates, while ResNet50 favored higher sensitivity (lower false negatives) at the cost of more false positives. These results were further supported by the ROC curves in Fig. 6, which provide a threshold-independent view of classifier performance.

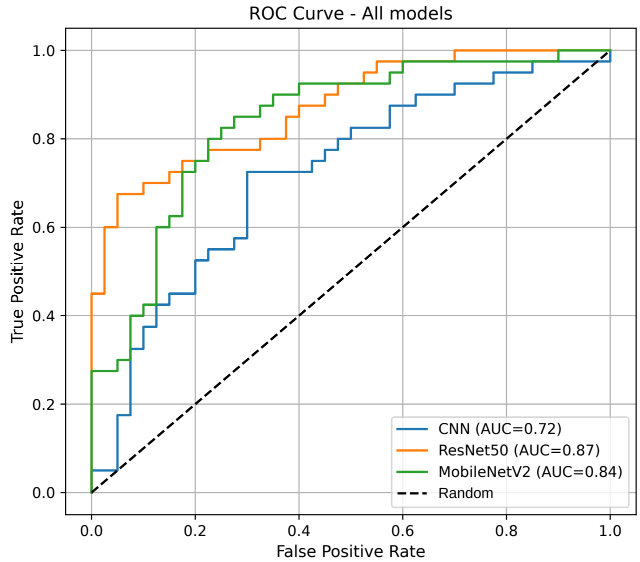

Fig. 6. ROC curves for CNN, ResNet50, and MobileNetV2

ResNet50 achieved the highest curve, closely followed by MobileNetV2, while the CNN curve remained below both, reflecting its inferior discriminatory power. Because the dataset was relatively small performance across data splits was a concern, five-fold stratified cross-validation was conducted for all three models. This approach allowed us to estimate the variance in model performance and ensure that the results were not overly dependent on a specific validation subset. The cross-validation results, shown in the heatmap in Fig. 7, confirm the stability of the pretrained models, particularly MobileNetV2, across all folds, with consistently high accuracy and F1-scores. The custom CNN again demonstrated the weakest overall performance, with CNN folds clustering around an accuracy of 0.65–0.69. Although fold-wise performance was recorded, no additional retraining was performed on selected folds; rather, the results were used strictly for comparative and analytical purposes.

Finally, to better understand the decision-making process of the deep learning models and improve model explainability, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to visualize the regions of interest identified by the networks. This analysis was performed for the ResNet50 and

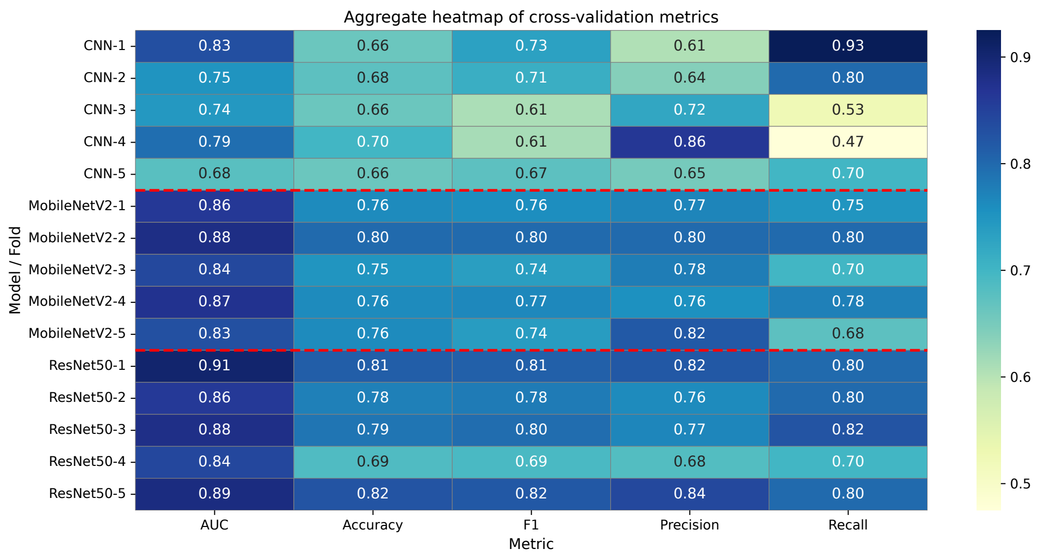

Fig. 7. Cross-validation metrics for CNN, ResNet50, and MobileNetV2

MobileNetV2 models only, since the CNN limited performance rendered interpretation less meaningful, as illustrated in Fig. 8 and Fig. 9.

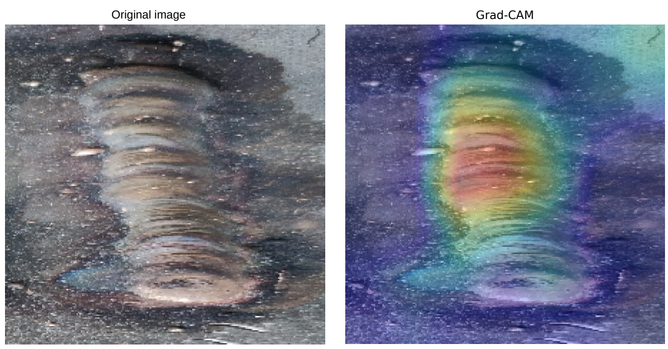

Fig. 8. Grad-CAM visualizations for MobileNetV2 correct prediction

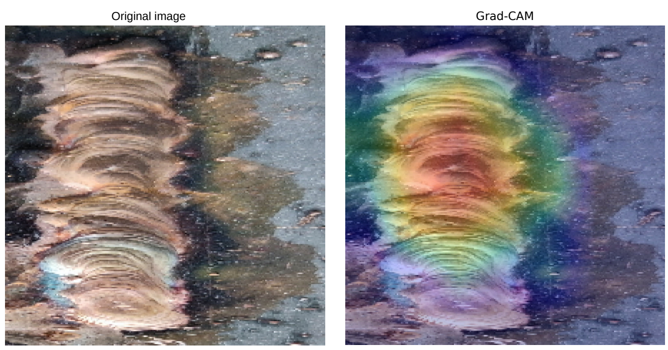

Fig. 9. Grad-CAM visualizations for ResNet50 correct prediction

Both models concentrated their attention on the central part of the weld bead, especially around the fusion lines and regions with structural irregularities, which aligns with the areas typically assessed during expert visual inspection (VT1 + VT2). Interestingly, ResNet50 showed more sharply localized activation patterns, while MobileNetV2 produced slightly broader attention maps, suggesting differences in how each network processes texture and shape features. These Grad-CAM visualizations reinforce the reliability of the model predictions.

On the next stage of the research, we proceeded to investigate whether the geometry of the weld pool alone could provide useful information for weld quality assessment. This secondary analysis focused on six geometric features extracted from cropped weld pool images: area, ellipticity, horizontal and vertical, symmetry_x, symmetry_y, centroid offset_x, and offset_y. At the beginning, we segmented the weld pool region in each image and computed the relevant shape descriptors. Outlier values were removed using the interquartile range (IQR) method to ensure robustness. To evaluate which geometric features differed significantly between acceptable and unacceptable welds, we conducted statistical significance tests (Student's t-test and Mann–Whitney U-test) [43]. Results are presented in Table 3.

Table 3
P-values of statistical tests for geometric features

Feature	t-test (p-value)	Mann–Whitney U (p-value)
area	0.8336	0.3906
ellipticity	0.0437	0.0934
symmetry_x	0.0196	0.0165
symmetry_y	0.0997	0.0783
offset_x	0.0016	0.0020
offset_y	0.3099	0.1555

Symmetry_x and offset_x were the most statistically significant, $p < 0.05$ in both tests, indicating that horizontal asymmetry and lateral misalignment are associated with reduced weld quality. Ellipticity also reached borderline significance in the t-test $p \approx 0.044$, while area and offset_y did not show meaningful differences. This analysis was supported by a correlation heatmap shown in Fig. 10, which shows the strongest correlations between class label and both symmetry_x and offset_x. This suggests that geometrical irregularities in the weld pool may be a useful proxy for underlying process instabilities.

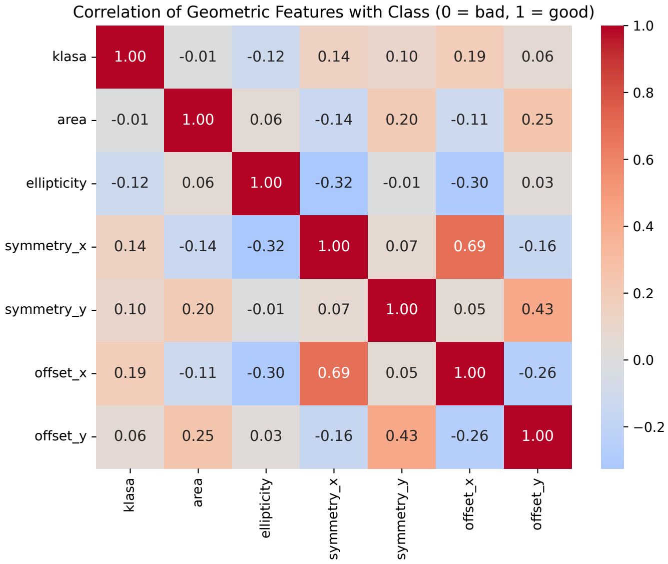

Fig. 10. Correlation of geometric features describing weld beads

To better understand the relationships between the geometric features and to reduce dimensionality while preserving variance, Principal Component Analysis (PCA) was applied to the standardized dataset. The amount of variance explained by each successive component is presented in Fig. 11, which shows the cumulative explained variance curve. As can be seen, the first three components account for approximately 73% of the total variance, while five components already capture over 95%, indicating moderate feature redundancy. These values are summarized in Table 4.

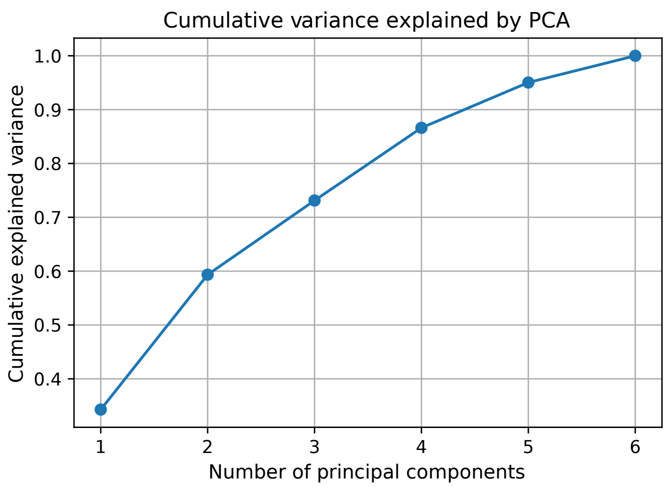

Fig. 11. Cumulative variance explained by PCA

Table 4
P-values of statistical tests for geometric features

Principal component	Cumulative explained variance
PC1	34.29%
PC2	59.34%
PC3	73.12%
PC4	86.62%
PC5	95.04%
PC6	100.00%

To visualize potential separability between classes, the dataset was projected onto the first three principal components and plotted in a 3D scatter plot, shown in Fig. 12. This visualization revealed a partial but noticeable separation between acceptable and unacceptable welds, supporting the hypothesis that certain geometric traits systematically differ across classes. Furthermore, a pairplot of the three principal components, shown in Fig. 13, was generated to highlight how combinations of principal components (PCs) distribute across classes. Although overlap still exists, particularly along PC2 and PC3, the most significant differences appear along PC1, which is largely influenced by symmetry_x and offset_x. Together, these results suggest that PCA is able to capture structure in the geometric feature space related to weld quality. While not fully linearly separable, the distribution indicates that some degree of discrimination is achievable using these features alone.

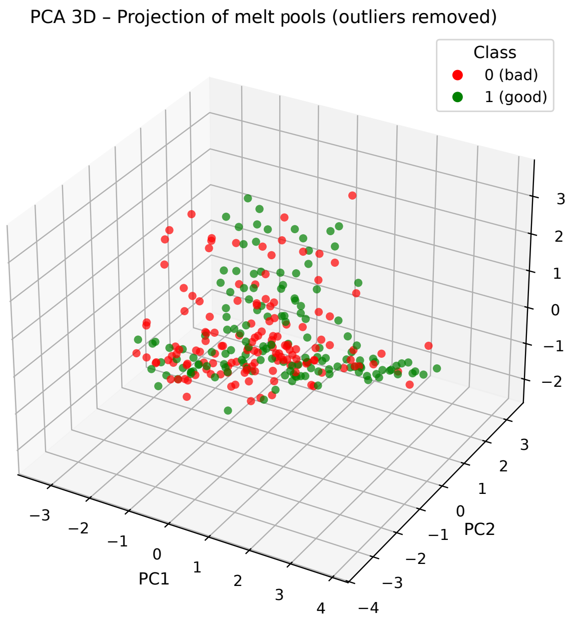

Fig. 12. PCA projection for the first three principal components

Finally, to evaluate whether weld quality could be predicted based solely on weld pool geometry, three classical machine learning classifiers were trained on the extracted features: Logistic Regression, Random Forest, and Support Vector Machine

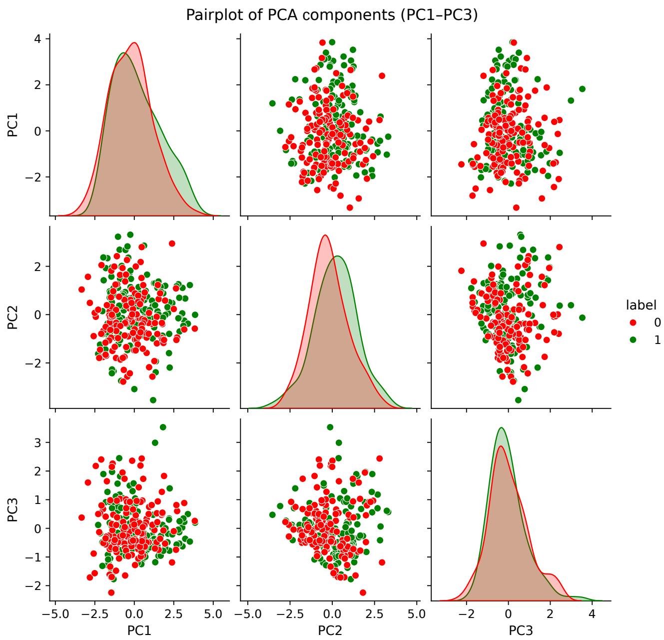

Fig. 13. Pairplot of PCA components

(SVM) [44, 45]. Models were evaluated using standard metrics on a stratified 80/20 train-test split. The results of this classification are shown in Table 5 and Fig. 14.

Table 5
Performance of classical machine learning models trained on the geometric features of weld pool beads

Model	Logistic regression	Random forest	SVM
Accuracy	0.534	0.603	0.603
Precision	0.538	0.615	0.615
Recall	0.483	0.552	0.552
F1-score	0.509	0.582	0.582
AUC	0.483	0.634	0.592

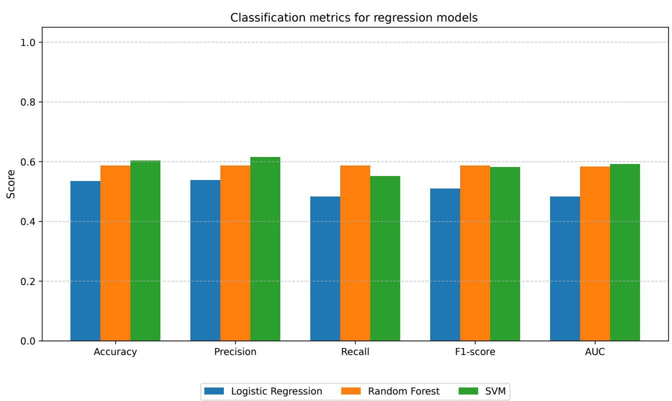

Fig. 14. Classification metrics for classification models

The Random Forest classifier achieved the best performance among the evaluated models, with an accuracy and F1-score of around 0.58–0.60 and an AUC of 0.63. While this result is not high in absolute terms for binary classification, it outperformed Logistic Regression (F1 ≈ 0.51) and matched SVM in F1-score (≈ 0.58), while achieving a higher AUC than both. The ROC curves, shown in Fig. 15, confirmed that Random Forest offered the most favorable trade-off between sensitivity and specificity.

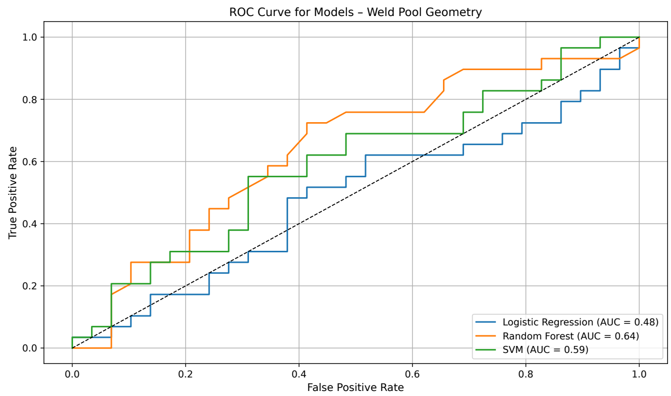

Fig. 15. ROC curves for all classification models

Although the geometric features alone did not enable high-confidence classification, the moderate performance indicates that these features do carry some information about weld quality. Rather than serving as standalone predictors, the weld pool shape descriptors, especially horizontal asymmetry and lateral offset, show promise as complementary features. These results also suggest that the local shape of the weld pool does not always reflect the overall quality of the entire bead. A weld pool can appear symmetrical and well-formed at a specific moment, yet the full weld bead may contain defects such as undercuts, lack of fusion, or porosity that render it unacceptable. Therefore, weld pool geometry alone may not be sufficient for reliable quality assessment, but it can still provide valuable supplementary information, especially when integrated into multi-modal inspection systems or interpretable machine learning models.

4. CONCLUSIONS

This study demonstrated the feasibility of using deep learning and geometric analysis to support weld quality assessment in cladding processes. Among the three evaluated convolutional neural networks, MobileNetV2 achieved the highest overall classification performance, closely followed by ResNet50, both significantly outperforming the custom lightweight CNN. The application of transfer learning proved effective in compensating for the small dataset size, enabling pretrained models to generalize well to the task of weld quality classification [46]. Grad-CAM visualizations verified that the models focused on visually relevant regions, such as fusion lines and surface irregularities, offering insights into their decision-making processes. This increases trust in model predictions and provides a useful interpretability layer, essential for deployment in industrial environments. The analysis of geometric features extracted from weld pool images revealed that certain traits, most notably horizontal symmetry and lateral offset, are statistically associated with final weld quality. Classical machine learning models trained on these features achieved moderate classification performance, with Random Forest and SVM reaching identical F1-scores of 0.58, and Random Forest achieving a slightly higher AUC. Although geometric features alone were not sufficient for high-confidence classification, they do capture aspects of the weld pool that relate to process quality and can complement image-based deep learning models [47, 48]. However, the geometry of the weld pool at a single point may not fully reflect the overall quality of the entire weld bead, which can contain defects such as undercuts, porosity, or lack of fusion beyond the pool region [49]. While the presented results demonstrate the potential of combining deep learning and geometric analysis for post-process weld quality assessment, several limitations remain. The study relied on a relatively small dataset of static images, with manually extracted weld pools and no temporal or sensor-based data. As a result, the models may not capture the full complexity of the cladding process or its dynamic aspects. Future research will focus on expanding the dataset, automating weld pool segmentation, and i ncorporating multi-modal process data, which could enhance the reliability and applicability of the method, especially in robotic and WAAM-based cladding systems.

ACKNOWLEDGEMENTS

The article was financed from the Research Subsidy of the Faculty of Mechanical Engineering and Robotics, AGH University of Krakow.

REFERENCES

[1] A. Adamczak, "Deep Recurrent and Convolutional Networks in the Detection of Welding Defects for Systems with an Industrial Robot," *Pomiary Autom. Robot.*, vol. 25, no. 2, pp. 17–22, 2021, doi: 10.14313/par_240/17.

[2] A. Adamczak, "Detection Method of Welding Defects in a Robotic Station Using the Deep Neural Network," *Pomiary Autom. Robot.*, vol. 25, no. 1, pp. 67–72, 2021, doi: 10.14313/par_239/67.

[3] R. Tsuzuki, "Development of automation and artificial intelligence technology for welding and inspection process in aircraft industry," *Weld. World*, vol. 66, no. 1, pp. 105–116, 2022, doi: 10.1007/s40194-021-01210-3.

[4] E.A. Gyasi, H. Handroos, and P. Kah, "Survey on artificial intelligence (AI) applied in welding: A future scenario of the influence of AI on technological, economic, educational and social changes," *Procedia Manuf.*, vol. 38, pp. 702–714, 2019, doi: 10.1016/j.promfg.2020.01.095.

[5] Y. Li, M. Hu, T. Wang, A. Maseleno, X. Yuan, and V.E. Balas, "Visual inspection of weld surface quality," *J. Intell. Fuzzy Syst.*, vol. 39, no. 4, pp. 5075–5084, 2020, doi: https://journals.sagepub.com/doi/abs/10.3233/JIFS-179993.

[6] G. Senthil Kumar, U. Natarajan, and S.S. Ananthan, "Vision inspection system for the identification and classification of defects in MIG welding joints," *Int. J. Adv. Manuf. Technol.*, vol. 61, no. 9, pp. 923–933, 2012, doi: 10.1007/s00170-011-3770-z.

[7] S. Kajan, M. Trebul'a, F. Duchoň, Z. Kovaríková, M. Švolík, and D. Švec, "Robotic Vision Inspection of Weld Quality Using Convolutional Neural Networks," in *2024 25th International Carpathian Control Conference* (*ICCC*), 22–24 May 2024, pp. 01–05, doi: 10.1109/ICCC62069.2024.10569157.

[8] R. Ranjan *et al.*, "Classification and identification of surface defects in friction stir welding: An image processing approach," *J. Manuf. Process.*, vol. 22, pp. 237–253, 2016, doi: 10.1016/j.jmapro.2016.03.009.

[9] R. Miao *et al.*, "Online inspection of narrow overlap weld quality using two-stage convolution neural network image recognition," *Mach. Vis. Appl.*, vol. 32, no. 1, p. 27, 2021, doi: 10.1007/s00138-020-01158-2.

[10] M. Ha, Y. Byun, J. Kim, J. Lee, Y. Lee, and S. Lee, "Selective Deep Convolutional Neural Network for Low Cost Distorted Image Classification," *IEEE Access*, vol. 7, pp. 133030–133042, 2019, doi: 10.1109/ACCESS.2019.2939781.

[11] C. Ajmi, J. Zapata, J.J. Martínez-Álvarez, G. Doménech, and R. Ruiz, "Using Deep Learning for Defect Classification on a Small Weld X-ray Image Dataset," *J. Nondestruct. Eval.*, vol. 39, no. 3, p. 68, 2020, doi: 10.1007/s10921-020-00719-9.

[12] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), 2018, pp. 4510–4520, doi: 10.48550/arXiv.1801.04381.

[13] S. Kumaresan, K.S.J. Aultrin, S.S. Kumar, and M.D. Anand, "Transfer Learning With CNN for Classification of Weld Defect," *IEEE Access*, vol. 9, pp. 95097–95108, 2021, doi: 10.1109/ACCESS.2021.3093487.

[14] M. Madhav, S.S. Ambekar, and M. Hudnurkar, "Weld defect detection with convolutional neural network: an application of deep learning," *Ann. Oper. Res.*, 2023, doi: 10.1007/s10479-023-05405-3.

[15] H. Pan, Z. Pang, Y. Wang, Y. Wang, and L. Chen, "A New Image Recognition and Classification Method Combining Transfer Learning Algorithm and MobileNet Model for Welding Defects," *IEEE Access*, vol. 8, pp. 119951–119960, 2020, doi: 10.1109/ACCESS.2020.3005450.

[16] M. Xiao, B. Yang, S. Wang, Y. Chang, S. Li, and G. Yi, "Research on recognition methods of spot-welding surface appearances based on transfer learning and a lightweight high-precision convolutional neural network," *J. Intell. Manuf.*, vol. 34, no. 5, pp. 2153–2170, 2023, doi: 10.1007/s10845-022-01909-0.

[17] L. Yang, H. Wang, W. Meng, and H. Pan, "CvT-UNet: A weld pool segmentation method integrating a CNN and a transformer," *Heliyon*, vol. 10, no. 15, 2024, doi: 10.1016/j.heliyon.2024.e34738.

[18] R.R. Selvaraju, M. Cogswell, and A. Das, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, 2020, doi: 10.48550/arXiv.1610.02391.

[19] F. Zhou, X. Liu, X. Zhang, Y. Liu, C. Jia, and C. Wu, "Keyhole status prediction based on voting ensemble convolutional neural networks and visualization by Grad-CAM in PAW," *J. Manuf. Process.*, vol. 80, pp. 805–815, 2022, doi: 10.1016/j.jmapro.2022.06.034.

[20] J. Pawlik *et al.*, "On the Influence of Heat Input on Ni-WC GMAW Hardfaced Coating Properties," *Materials*, vol. 16, no. 11, p. 3960, 2023, doi: 10.3390/ma16113960.

[21] G. Bruno, G. Antal, E. Traini, and M. De Maddis, "Convolutional Neural Network for Quality Monitoring and Predictive Maintenance in Resistance Spot Welding," in *Real-World Applications of AI Innovation. IGI Global*, 2024, pp. 307–330.

[22] PN-EN ISO 6520-1:2007, "Spawalnictwo – Klasyfikacja geometrycznych niezgodności spawalniczych – Część 1: Spawanie metali," 2007.

[23] PN-EN ISO 5817:2014-05, "Spawanie – Połączenia spawane metali – Nieciągłości spawalnicze i poziomy jakości," 2014.

[24] M.S. Elmahdy, S.S. Abdeldayem, and I.A. Yassine, "Low quality dermal image classification using transfer learning," in *2017 IEEE EMBS International Conference on Biomedical & Health Informatics* (*BHI*), 16–19 Feb. 2017, pp. 373–376, doi: 10.1109/bhi.2017.7897283.

[25] S. Perri, F. Spagnolo, F. Frustaci, and P. Corsonello, "Welding defects classification through a Convolutional Neural Network," *Manuf. Lett.*, vol. 35, pp. 29–32, 2023, doi: 10.1016/j.mfglet.2022.11.006.

[26] N. Thi Hoa, T. Ha Minh Quan, and Q.B. Diep, ' 'Weld-CNN: Advancing non-destructive testing with a hybrid deep learning model for weld defect detection," *Adv. Mech. Eng.*, vol. 17, no. 5, p. 16878132251341615, 2025, doi: 10.1177/16878132251341615.

[27] G.A. Elhendawy and Y. El-Taybany, "Machine Vision-Assisted Welding Defect Detection System with Convolutional Neural Networks," *Int. J. Precis. Eng. Manuf.*, 2025, doi: 10.1007/s12541-025-01281-y.

[28] M. Abadi, A. Agarwal, and P. Barham, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," in *Proc. 12th USENIX Symposium on Operating Systems Design and Implementation* (*OSDI*), 2016, doi: 10.48550/arXiv.1603.04467.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), 2016, pp. 770–778, doi: 10.48550/arXiv.1512.03385.

[30] F. Chollet, "Keras," 2015. [Online]. Available: https://github.com/fchollet/keras.

[31] J. Deng, W. Dong, and R. Socher, "ImageNet: A Large-Scale Hierarchical Image Database," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.

[32] H. Huang, B. Wu, and J. Fan, "Analysis to the relationship of classification accuracy, segmentation scale, image resolution," in *IGARSS 2003–2003 IEEE International Geoscience and Remote Sensing Symposium. Proc.*, 21–25 July 2003, vol. 6, pp. 3671–3673 vol.6, doi: 10.1109/IGARSS.2003.1295233.

[33] S.Q. Moinuddin, S.S. Hameed, A.K. Dewangan, K. Ramesh Kumar, and A. Shanta Kumari, "A study on weld defects classification in gas metal arc welding process using machine learning techniques," *Mater. Today Proc.*, vol. 43, pp. 623–628, 2021, doi: 10.1016/j.matpr.2020.12.159.

[34] R.M. Nazarov, Z.M. Gizatullin, and E.S. Konstantinov, "Classification of Defects in Welds Using a Convolution Neural Network," in *2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering* (*ElConRus*), 26–29 Jan. 2021, pp. 1641–1644, doi: 10.1109/ElConRus51938.2021.9396301.

[35] H. Dong, M. Cong, Y. Zhang, Y. Liu, and H. Chen, "Modeling and real-time prediction for complex welding process based on weld pool," *Int. J. Adv. Manuf. Technol.*, vol. 96, no. 5, pp. 2495–2508, 2018, doi: 10.1007/s00170-018-1685-7.

[36] Y. Cheng, Q. Wang, W. Jiao, J. Xiao, S. Chen, and Y. Zhang, "Automated recognition of weld pool characteristics from active vision sensing," *Weld. J.*, vol. 100, no. 5, pp. 183S–192S, 2021, doi: 10.29391/2021.100.015.

[37] J. Hassan, A.M. Awan, and A. Jalil, "Welding Defect Detection and Classification Using Geometric Features," in *2012 10th International Conference on Frontiers of Information Technology*, 17-19 Dec. 2012, pp. 139–144, doi: 10.1109/FIT.2012.33.

[38] Z. Jiao *et al.*, "Image Processing and Feature Extraction for Hull Structure GMAW Based on Weld Pool Visual Sensing," *J. Sens.*, vol. 2023, no. 1, pp. 1–23, 2023, doi: 10.1155/2023/6317992.

[39] R. Kuppuswamy, K. Calo, and J. Ramakumar, "Use of ResNet modelling for TIG weld feature digitization and correlation : a technique for AI based welding system," *Manuf. Technol. Today*, vol. 22, no. 1, pp. 25–32, 2023, doi: 10.58368/mtt.22.1.2023.25-32.

[40] Q. Martin, V. Ahedo, J.I. Santos, and J.M. Galan, "Comparative Study of Classification Algorithms for Quality Assessment of Resistance Spot Welding Joints From Pre- and Post-Welding Inputs," *IEEE Access*, vol. 10, pp. 6518–6527, 2022, doi: 10.1109/ACCESS.2022.3142515.

[41] R. Ghimire and R. Selvam, "Machine Learning-Based Weld Classification for Quality Monitoring," *Eng. Proc.*, vol. 59, no. 1, p. 241, 2023, doi: 10.3390/engproc2023059241.

[42] J.D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007.

[43] P. Virtanen, R. Gommers, and T.E. Oliphant, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nat. Methods*, vol. 17, no. 3, pp. 261–272, 2020, doi: 10.1038/s41592-019-0686-2.

[44] F. Pedregosa, G. Varoquaux, and A. Gramfort, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011. [Online]. Available: https://jmlr.org/papers/v12/pedregosa11a.html.

[45] R. Liang, R. Yu, Y. Luo, and Y. Zhang, "Machine learning of weld joint penetration from weld pool surface using support vector regression," *J. Manuf. Process.*, vol. 41, pp. 23–28, 2019, doi: 10.1016/j.jmapro.2019.01.039.

[46] W. Jiao, Q. Wang, Y. Cheng, and Y. Zhang, "End-to-end prediction of weld penetration: A deep learning and transfer learning based method," *J. Manuf. Process.*, vol. 63, pp. 191–197, 2021, doi: 10.1016/j.jmapro.2020.01.044.

[47] B.D.R. Veitía *et al.*, "Deep learning for quality prediction in dissimilar spot welding dp600-aisi304 using a convolutional neural network and infrared image processing," in *32nd European Modeling and Simulation Symposium*, EMSS 2020, 2020, pp. 393–399.

[48] R. Verma and A.K. Verma, "TIG weld defect prediction from weld pool images using deep convolutional neural network and transfer learning," *Int. J. Manuf. Res.*, vol. 19, no. 2, pp. 181–210, 2024, doi: 10.1504/IJMR.2024.140279.

[49] J.-K. Park, W.-H. An, and D.-J. Kang, "Convolutional Neural Network Based Surface Inspection System for Non-patterned Welding Defects," *Int. J. Precis. Eng. Manuf.*, vol. 20, no. 3, pp. 363–374, 2019, doi: 10.1007/s12541-019-00074-4.