Materials Science Forum
ISSN: 1662-9752, Vol. 1187, pp 17-27
doi:10.4028/p-4P6a64

Submitted: 2026-01-19
Revised: 2026-02-26
Accepted: 2026-02-26
Online: 2026-04-17

Deep Learning-Based Quality Classification of Cold Spray Coatings with Explainable Feature Analysis

Perna Alessia Serena[1,a], Rossi Domenico[2,b], Auriemma Citarella Alessia[2,c], De Fazio Dario[1,d], De Marco Fabiola[2,e*], Di Biasi Luigi[2,f], Durante Massimo[1,g],Tortora Genoveffa[2,h] and Viscusi Antonio[3,i]

[1]Department of Chemical, Materials and Production Engineering, University of Naples Federico II, Piazzale V. Tecchio 80, 80125 Napoli, Italy

[2]Department of Computer Science, University of Salerno, Fisciano, SA, Italy

[3]Department of Engineering and Sciences, Faculty of Technology and Innovation Sciences, Universitas Mercatorum, Piazza Mattei 10, 00186 Rome, Italy

[a]alessiaserena.perna@unina.it, [b]dorossi@unisa.it, [c]aauriemmacitarella@unisa.it, [d]dario.defazio@unina.it, [e*]fdemarco@unisa.it, [f]ldibiasi@unisa.it, [g]mdurante@unina.it, [h]tortora@unisa.it, [i]antonio.viscusi@unimercatorum.it

*corresponding author

Keywords: deep learning, cold spray, eXplainable artificial intelligence.

Abstract. This study investigates the applicability of deep learning models for automated quality classification of cold spray coatings, focusing on three deposition categories: good, degraded, and poor deposition. Three state-of-the-art convolutional architectures, ResNet-50, EfficientNet-B0, and ConvNeXt-Tiny, were evaluated across two training phases designed to assess the impact of dataset balancing, data augmentation, and higher input resolution. In the first phase, models were trained on an imbalanced dataset using only class weighting; EfficientNet-B0 achieved the best performance (ACC 80%, F1 77%), while ResNet-50 showed notable instability (ACC 60%, F1 56%). In the refined second phase, oversampling, advanced augmentation, 380×380 resolution, and early stopping led to substantial performance gains for all models. ConvNeXt-Tiny achieved the most robust and balanced results (ACC 93.3%, F1 90.3%), outperforming EfficientNet-B0 and ResNet-50 particularly in sensitivity and specificity for minority classes. Grad-CAM analysis provided qualitative insights into the decision-making process: poor samples elicited strong, spatially extended activations corresponding to defective regions, degraded samples produced more localized responses aligned with mid-scale irregularities, and good samples yielded diffuse, low-intensity activation patterns associated with surface uniformity. These interpretable attention maps validated the physical relevance of the learned features and confirmed the suitability of ConvNeXt-Tiny for reliable and explainable cold spray quality assessment.

Introduction

Cold spray is an advanced solid-state deposition technique that enables the formation of metallic coatings by propelling micron-sized powders at supersonic velocities onto a target substrate, without reaching the melting point of the material. This unique feature significantly limits oxidation, phase transformation, and thermal degradation, making the process particularly advantageous for the deposition of temperature-sensitive alloys and multi-material systems. Owing to these capabilities, cold spray is emerging as a key enabling technology in high-performance sectors such as aerospace, automotive, energy, and defense industries, where lightweight design, corrosion resistance, and durability are essential requirements [1, 2]. The resulting coating quality, characterized in terms of thickness uniformity, porosity, hardness, adhesion strength, and microstructural integrity, is strongly influenced by a complex interplay of factors. These include powder characteristics (e.g., morphology, grain size distribution, material type), process conditions (such as gas temperature, pressure, and nozzle geometry), and substrate mechanical properties [3]. Establishing accurate relationships between these variables and the final coating performance represents one of the most critical challenges in cold spray process optimization. Traditionally, process development and property prediction rely on costly and iterative experimental campaigns and computationally demanding simulation approaches based on the Finite Element Method (FEM). While these techniques enable detailed physical insight, their scalability is limited when exploring large design spaces or multi-objective optimization scenarios [4, 5]. As industries increasingly seek flexible and rapid design-to-manufacture workflows, the need for efficient and data-driven predictive tools has become more pressing.

In this context, Deep Learning (DL) techniques are gaining prominence as powerful surrogate modeling approaches to map high-dimensional process input parameters to coating outcomes. By using experimental or synthetic training datasets, DL models can learn nonlinear correlations and predict coating quality metrics with high accuracy, drastically reducing development time and resource consumption. Recent studies demonstrate the potential of Artificial Intelligence (AI)-based techniques in supporting real-time monitoring, automated parameter tuning, and defect detection, thus paving the way toward fully digitalized and intelligent cold spray manufacturing systems [6, 7, 8]. As the technique enables solid-state bonding without melting, even subtle surface defects or non-uniform deposition patterns can significantly influence the functional performance of the resulting coatings. For this reason, the ability to reliably distinguish between good, degraded, and poor deposition states is fundamental for quality assurance, process optimization, and the prevention of premature component failure. However, manual inspection is time-consuming, subjective, and often unable to capture the fine-grained texture variations that characterize borderline quality conditions.

In this work, we address this challenge by developing an automated deep learning–based pipeline for classifying cold-spray surfaces into three quality classes: good, degraded, and poor. We evaluate three modern convolutional architectures— ResNet-50, EfficientNet-B0, and ConvNeXt-Tiny— using a two-phase training strategy designed to explore the impact of dataset imbalance, data augmentation, and image resolution. The first phase uses class weighting to partially address imbalance, while the second integrates oversampling, extensive augmentation, higher-resolution inputs, and dynamic early stopping to enhance robustness and generalization. Performance is assessed through accuracy, macro-F1, macro-sensitivity, and macro-specificity to ensure an unbiased evaluation across quality classes. Model interpretability is then investigated through Grad-CAM visualizations, which highlight the image regions most relevant to the predictions and help verify the physical consistency of the learned features.

Materials and Methods

Dataset.

The dataset employed in this study consists of optical images of cold spray depositions obtained under different processing conditions, resulting in a range of surface morphologies and deposition outcomes. Cold spray coatings were produced using a low pressure cold spray system (Dymet 423), with air employed as the carrier gas. The feedstock material consisted of spherical copper powders supplied by Carpenter Additive. The substrates were composed of different thermoplastic polymers, namely polyamide (PA), polycarbonate (PC), and polypropylene (PP).

The images were acquired employing a Hirox digital microscope, using the same magnification for all the samples. Each image represents the surface state of a cold-sprayed region and captures variations in texture, uniformity, and coating distribution associated with changes in particle impact behavior and bonding efficiency. All images were manually annotated by domain experts and assigned to one of three qualitative deposition classes: good, degraded, and poor. The labeling procedure was based on visual assessment of surface continuity, homogeneity, and the presence or absence of irregular deposition features, such as localized discontinuities, heterogeneous particle clustering, or extended defective regions. The original dataset comprises 78 images, with 35 samples labeled as good, 23 as degraded, and 20 as poor. This class distribution reflects the inherent imbalance commonly observed in experimental cold spray campaigns, where optimal or near-optimal deposition conditions are more frequently achieved than severely defective ones. Image acquisition was performed under consistent optical and illumination conditions in order to minimize variability unrelated to deposition quality while preserving the intrinsic texture characteristics of each class. Validation and test sets were composed solely of original, non-augmented experimental images to ensure an unbiased evaluation of model performance under realistic inspection conditions. This dataset configuration enabled a systematic comparison between minimally processed data and a data-centric training strategy, allowing the impact of dataset composition and preprocessing choices on classification performance to be rigorously assessed.

Consequently, although the original dataset consists of 78 images, several precautions were adopted to mitigate overfitting to synthetic samples. First, validation and test sets were composed exclusively of original, non-augmented images, ensuring that performance metrics reflect generalization to real data only. Second, early stopping and macro-averaged metrics were used to monitor class-wise stability rather than raw accuracy.

Models.

To assess the capability of deep learning models in recognizing coating quality, we tested multiple neural network architectures on a dataset of cold spray deposition images. In particular, we trained and evaluated three networks: ResNet, EfficientNet-B0 and ConvNeXt-Tiny.

The ResNet-50 is a deep convolutional neural network (CNN) architecture designed to address the vanishing gradient problem commonly encountered when training very deep models. Introduced by He et al. [9], ResNet employs a residual learning framework that allows the network to learn identity mappings through the use of shortcut (or skip) connections between non-adjacent layers. Each residual block typically consists of two or three convolutional layers followed by batch normalization and ReLU activation [10], with the input added directly to the block's output via an identity or projection shortcut. This architecture allows the model to maintain representational efficiency while avoiding the overfitting and training instability often observed in traditional deep CNNs. In image classification tasks, ResNet has demonstrated superior performance compared to other earlier architectures by achieving state-of-the-art accuracy with significantly deeper models. Its modular design also makes it a robust backbone for various computer vision applications, including feature extraction, segmentation, and defect detection [11].

The EfficientNet architecture represents a family of convolutional neural networks designed through a compound scaling strategy that uniformly scales network depth, width, and input resolution in a balanced manner. Unlike traditional approaches that arbitrarily increase one of these dimensions to improve accuracy, often at the cost of computational efficiency, EfficientNet employs a compound coefficient that optimally allocates model capacity across all three dimensions, achieving a superior trade-off between performance and efficiency [12]. EfficientNet-B0 is the baseline model of the EfficientNet family, which employs mobile inverted bottleneck convolution (MBConv) blocks combined with squeeze-and-excitation (SE) modules and the Swish activation function, enabling effective feature extraction with reduced computational cost. The architecture follows a compound scaling strategy, which uniformly balances network depth, width, and input resolution to maximize performance while minimizing parameters [13].

ConvNeXt-Tiny is a modern convolutional neural network architecture that reinterprets traditional convolutional networks through the design principles of Vision Transformers (ViTs). It retains the hierarchical structure of classic CNNs but incorporates architectural elements inspired by transformer-based models to enhance representational capacity and training stability. The Tiny variant represents the smallest configuration of the ConvNeXt family, optimized for computational efficiency while maintaining high accuracy. It consists of approximately 28 million parameters and is particularly suitable for image classification and transfer learning tasks where a balance between performance and inference cost is required [15].

Gradient-Weighted Class Activation Mapping (Grad-CAM) is a widely used interpretability technique designed to provide visual explanations for the decisions of convolutional neural networks. The method highlights the image regions that contribute most strongly to a model's prediction, allowing researchers to inspect whether the network focuses on semantically meaningful features or on spurious artifacts [19]. By identifying the spatial regions most influential to a prediction, Grad-CAM provides qualitative insights into model behavior, helps detect biases or incorrect reasoning, and supports model validation in sensitive domains such as medical imaging, materials science, and quality control. Its architecture-agnostic design and compatibility with a broad range of CNN-based models make it a standard tool for explainability in deep learning.

Experimental design and Training Protocol.

Two training phases were conducted to evaluate the performance of deep learning models in classifying cold-spray deposition quality into the three categories. The following subsections describe in detail the datasets, preprocessing strategies, training procedures, and model-specific results. In the initial training setup, the dataset exhibited a clear class imbalance, with 35 good, 23 degraded, and 20 poor samples. No resampling strategies or data augmentation techniques were applied during this first phase. Instead, class weighting was introduced during training, assigning higher weights to under-represented classes to partially mitigate imbalance effects. A fixed training schedule of 15 epochs was used for all models, without early-stopping [16].

In the second phase of experimentation, the training pipeline was substantially refined to enhance model robustness and to better address the intrinsic challenges posed by the dataset. First, the class imbalance observed in the original dataset was mitigated through oversampling of minority classes, ensuring a more uniform representation of degraded and poor samples during training. This adjustment was essential to prevent the model from developing a bias toward the majority class and to promote more equitable learning across all coating-quality categories [17]. In addition to oversampling, a comprehensive data augmentation strategy was introduced to simulate realistic variations in the visual appearance of cold-spray depositions. Techniques such as geometric transformations, intensity perturbations, and localized texture modifications were applied to artificially expand the dataset and improve generalization. These operations help mimic real-world variability in coating morphology, enabling the models to better recognize subtle texture differences associated with each quality class [18]. A further improvement involved increasing the input image resolution to 380×380 pixels. The cold-spray process generates fine-grained surface features and microtexture patterns whose discriminative details may be lost at lower resolutions. By adopting a higher spatial resolution, the models were provided with richer visual information, facilitating more accurate extraction of texture-based cues that differentiate good, degraded, and poor deposition states.

Performance Metrics.

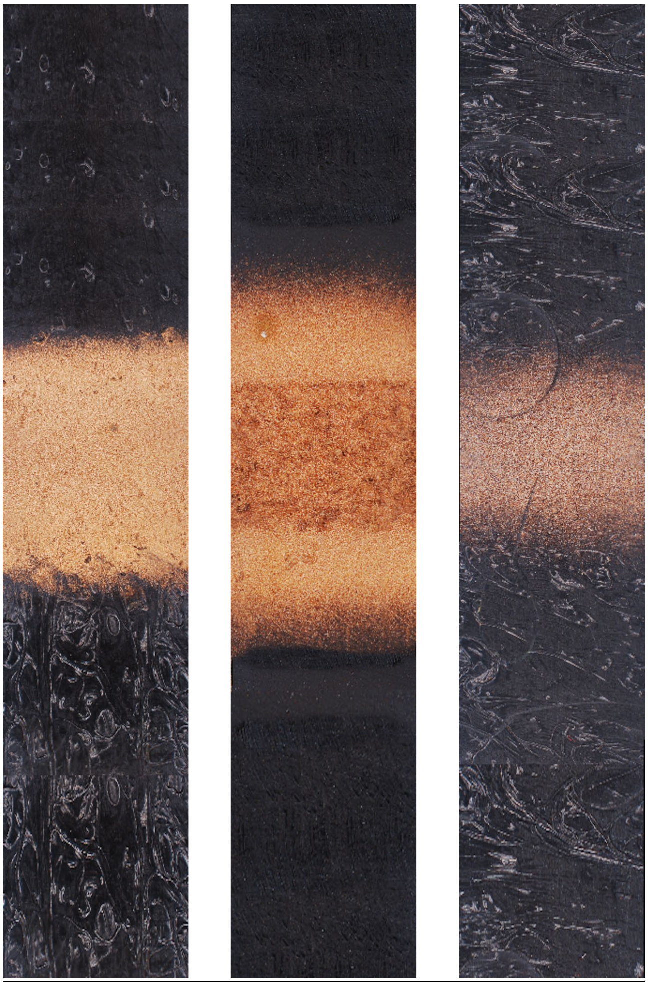
The performance of the classification models was assessed using several statistical indicators, including Accuracy, macro-averaged F1-score, macro-averaged Sensitivity, and macro-averaged Specificity. These metrics collectively provide a comprehensive evaluation of the ability of the models to correctly classify samples across all classes, regardless of potential class imbalance. Accuracy (ACC) quantifies the overall proportion of correctly classified samples among the total number of observations. It provides a global measure of performance but may not fully reflect class-wise behavior when class distributions are unbalanced. F1-score (macro) is the harmonic mean of precision and recall, computed independently for each class and then averaged. This macro-averaging ensures that each class contributes equally to the final score, regardless of its frequency. Sensitivity (macro) (SN), also referred to as macro recall, measures the model's ability to correctly identify positive samples across all classes. It is obtained by averaging the sensitivity values computed per class. Specificity (macro) (SP) evaluates the model's capacity to correctly recognize negative samples for each class. The class-wise specificities are averaged to provide a balanced estimate across the entire dataset. The use of macro-averaged metrics is particularly appropriate for multi-class classification tasks, as it treats all classes equally, preventing dominant classes from biasing the overall performance evaluation [14].

Results

This section presents a comprehensive analysis of the experimental results obtained from the two training phases. The first phase employed a minimally processed dataset and a static training protocol, while the second phase introduced several methodological improvements, especially oversampling, data augmentation, higher-resolution inputs, and dynamic early stopping.

Labelling of the dataset.

The acquired images were manually labeled according to the three predefined categories. The labeling procedure was based on the traditional classification adopted within the cold spray deposition framework, which defines the deposition window as comprising three distinct regimes: the rebound region, the degradation region, and the good adhesion region. These regimes reflect the process parameter ranges that respectively prevent deposition, lead to degraded coating formation, or ensure effective deposition and adhesion of the coating to the substrate. Representative examples of coatings corresponding to the three categories are shown in Figure 1.



(a) (b) (c)

Fig. 1. a) Example of a copper coating deposited on a polyamide substrate and classified as good; b) example of a copper coating deposited on a polyamide substrate and classified as degraded; c) example of a copper coating deposited on a polyamide substrate and classified as poor.

Results of the first training phase.

In the initial phase, models were trained with class weighting only, without any augmentation or resampling strategies. This configuration exposed the ability of each architecture to directly cope with class imbalance and limited sample variability. Table 1 summarizes the obtained results of this first phase.

Table 1. Performance Metrics (First training phase.

Model	ACC (%)	F1 macro (%)	SN macro (%)	SP macro (%)
ResNet-50	60	56	56.8	80.3
EfficientNet-B0	**80**	**77**	**79.4**	**91.1**
ConvNeXt-Tiny	67	62	63.5	85

EfficientNet-B0 clearly dominates the first phase, indicating stronger generalization under limited augmentation. ConvNeXt-Tiny shows moderate performance, while ResNet-50 struggles the most, reflecting a higher sensitivity to intra-class variability.

The behavior of the three architectures during the first training phase reveals important differences in how each model interprets the visual properties of cold-spray depositions.

EfficientNet-B0 demonstrates the most reliable performance, particularly in its perfect discrimination of the degraded class. This suggests that the visual signatures associated with degraded coatings, typically characterized by clear non-uniformities or texture irregularities, are highly separable and well captured by the network's feature hierarchy. However, the model exhibits some confusion between good and poor, misclassifying a subset of good samples as poor. This behavior reflects the intrinsic visual overlap that may arise in borderline samples, where subtle irregularities in an otherwise good deposition resemble the more pronounced defects of the poor category. In contrast, ConvNeXt-Tiny shows a different error distribution. The model frequently misclassifies good → poor and poor → degraded, indicating a pessimistic bias in its decision-making process. In other words, when uncertainty is present, ConvNeXt tends to err by assigning samples to a lower-quality class. This pattern suggests that the model may be particularly sensitive to small deviations in surface texture or illumination that it interprets as defects. Despite these misclassifications, ConvNeXt-Tiny exhibits consistently strong performance on the degraded class, implying that degraded coatings contain the most distinctive and easily recognizable visual patterns among the three quality categories.

ResNet-50, on the other hand, displays the most heterogeneous and inconsistent error pattern. The model misclassifies good samples as both degraded and poor, and occasionally produces degraded → good errors, suggesting that it struggles to establish stable class boundaries without additional data regularization. Furthermore, poor samples are most often confused with degraded, highlighting difficulties in capturing the fine-grained distinctions that separate the two lower-quality classes. This error structure indicates that ResNet-50's feature extraction is less suited to the subtle and texture-rich characteristics of cold-spray surfaces, particularly when trained without augmentation. The model appears unable to reliably disentangle the nuanced patterns that differentiate visually similar deposition states, resulting in lower overall performance and class-wise stability.

Results of the second training phase.

The second training phase integrates oversampling, image augmentation, higher-resolution inputs (380×380) and dynamic early stopping. These improvements substantially enhance model performance and stability across all architectures, as we can see in Table 2.

Table 2. Performance Metrics (Second training phase).

Model	ACC (%)	F1 macro (%)	SN macro (%)	SP macro (%)
ResNet-50	93.3	86.7	83.3	95.8
EfficientNet-B0	86.7	83.3	89.7	94.9
ConvNeXt-Tiny	**93.3**	**90.3**	**94.4**	**97.4**

The improvement between training phases is substantial, particularly for ConvNeXt-Tiny and ResNet-50, showing that the architectural capacity was previously underutilized due to training constraints.

EfficientNet-B0 demonstrates a generally strong and reliable classification behavior. The model achieves near-perfect recognition of the good class. This indicates that the features extracted by the network are well aligned with the morphological characteristics typically associated with high-quality coatings, such as uniform deposition and consistent surface texture. The model also performs perfectly on the poor class, successfully identifying all corresponding samples without any false positives. This suggests that the network is highly sensitive to the pronounced irregularities and defects that define poor surface quality. However, EfficientNet-B0 exhibits a single misclassification in the degraded category, assigning one degraded sample to the poor class. This particular error suggests that the model may be especially responsive to marked texture irregularities within borderline degraded cases, interpreting them as more severe than they actually are. Despite the overall performance improvement relative to the first training phase, the model remains slightly less stable on minority classes, where the number of available samples is low. This sensitivity results in minor fluctuations in how intermediate deposition defects are interpreted.

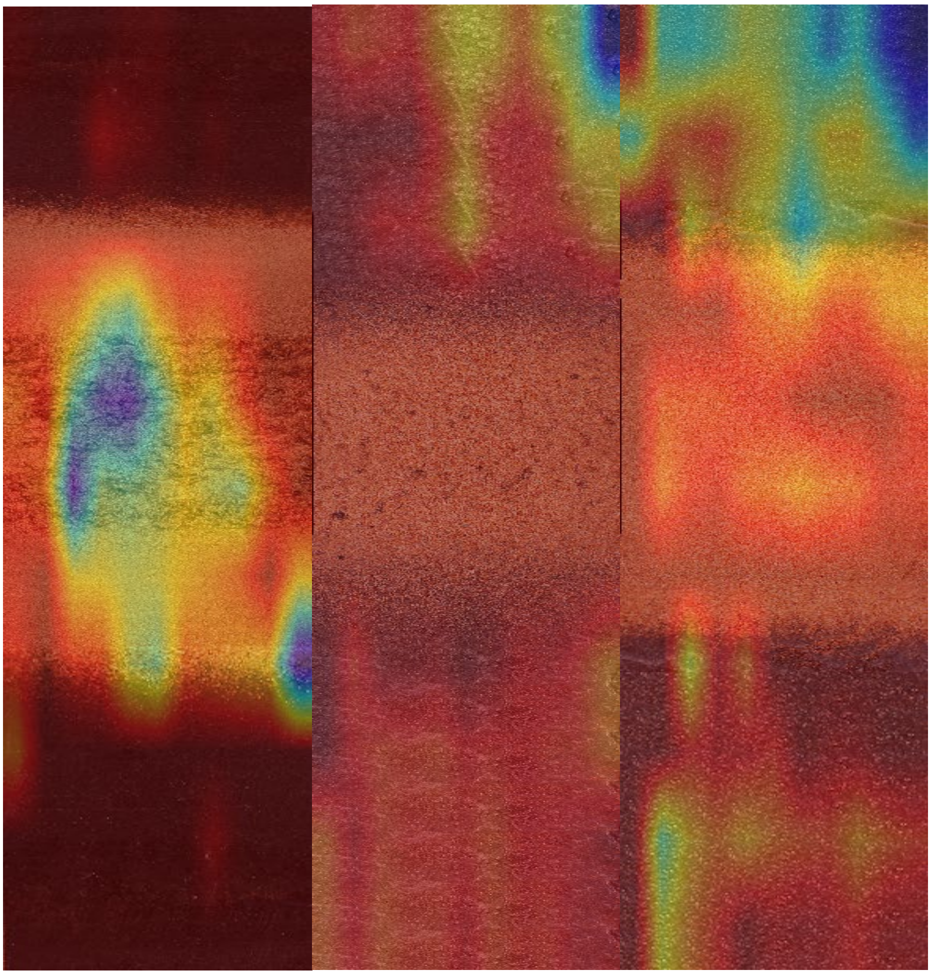
ConvNeXt-Tiny exhibits the most consistent and robust behavior across all classes. The model achieves perfect classification for both the good and poor categories, recognizing all samples correctly and without ambiguity. This is particularly relevant for industrial deployment, as accurately detecting good samples while never misclassifying poor samples as acceptable is essential for maintaining reliable quality control standards. The only error made by ConvNeXt-Tiny occurs within the degraded class, where one sample is incorrectly labeled as poor. Importantly, this misclassification reflects a conservative bias: rather than overestimating the quality of a potentially flawed deposition, the model errs on the side of caution by assigning it to a lower-quality class. In quality-critical manufacturing contexts, this behavior is generally preferable, as it minimizes the risk of approving defective components. ConvNeXt-Tiny also achieves the highest macro-F1, macro-sensitivity, and macro-specificity among all tested models, confirming that it maintains balanced performance across classes. Its ability to generalize effectively despite class imbalance and subtle inter-class differences underscores its suitability for texture-rich classification tasks like cold-spray surface assessment.

ResNet-50 also performs well overall, achieving perfect classification of both the good and degraded classes, which indicates that the network can reliably distinguish high-quality coatings and moderate deposition defects. However, performance declines for the poor class, where ResNet-50 misclassifies one sample as degraded. This poor → degraded misclassification is more critical than the error observed in ConvNeXt-Tiny. While ConvNeXt's error tends to underestimate the quality of a marginal sample, ResNet-50's error has the opposite effect: it reduces the perceived severity of an actual defect, potentially allowing faulty coated components to pass inspection. In safety-critical or performance-critical settings, this type of error is far less acceptable because it compromises the reliability of the quality assurance process. Despite its strong performance on the majority classes, this limitation indicates that ResNet-50 is less reliable for identifying the most defective samples, making it less suitable for deployment in industrial inspection environments where strict defect detection is mandatory.

The results demonstrate that data quality and training strategies have a stronger influence than architecture alone. When trained with minimal preprocessing (Phase 1), models show high variance in performance and substantial confusion, particularly for the poor class. However, after implementing a more sophisticated training pipeline (Phase 2), all architectures improve, with ConvNeXt-Tiny emerging as the most reliable for industrial cold-spray quality assessment. The model's stability, superior class-wise performance, and conservative error dynamics make it highly suitable for deployment in automated inspection systems where misclassifying a poor-quality deposition as acceptable must be strictly avoided.

eXplainable results with Grad-CAM.

To investigate whether ConvNeXt-Tiny makes decisions based on semantically meaningful features of the cold-spray depositions, we applied Grad-CAM to a representative subset of validation images from each class. Grad-CAM provides class-discriminative saliency maps by computing the gradients of the target class score with respect to the activations of a late convolutional block, averaging these gradients to obtain importance weights, and then forming a weighted sum of the feature maps followed by a ReLU. For ConvNeXt-Tiny we extracted activation maps from the final convolutional stage immediately prior to any classification head (i.e., the deepest stage that retains spatial layout), upsampled the resulting heatmaps to the original image resolution, and overlaid them on the input images. The Grad-CAM maps shown in Figure 2 highlight how the ConvNeXt-Tiny model focuses on different regions of the coating surface when predicting each class. Across all three examples, the model's attention is concentrated on texture-rich areas of the deposition.



(a) (b) (c)

Fig. 2. Grad-CAM on a) degraded b) poor c) good images.

In Figure 2a (degraded sample), the heatmap exhibits strong, localized activations, primarily in the central band, corresponding to irregular patches and mid-scale texture disruptions typical of degraded coatings. The model concentrates on transitions in roughness and subtle clustering of particles, which serve as reliable indicators of partial deposition defects. This focused activation pattern aligns well with the model's ability to correctly isolate the degraded class and also explains the occasional conservative misclassification of borderline degraded regions as poor, as the highlighted zones often coincide with visually severe irregularities. In Figure 2b (poor sample), the activations are more widespread and intense, extending across the upper and lower regions of the image. This reflects the extensive heterogeneity, abrupt texture discontinuities, and defective local structures that define poor-quality deposits. The model captures these broad defective areas, leading to a coherent and intuitive activation map. The large, contiguous red–yellow regions in the heatmap indicate high confidence in identifying severe surface defects, consistent with the model's perfect classification of the poor class in the second phase of the experiment. In Figure 2c (good sample), the heatmap reveals a markedly different behavior: activations are diffuse and of lower intensity, often aligned with subtle texture transitions rather than clear defect zones. This suggests that, for good samples, the model relies on the absence of strong irregularities rather than on specific localized cues. The uniformity of the underlying texture corresponds to low-saliency regions in the Grad-CAM output, reflecting the model's correct identification of high-quality deposition. The sparse, low-intensity activations highlight the network's focus on subtle surface features while avoiding false attention to noise or benign variations, an essential property for robust industrial inspection. Overall, Figure 2 demonstrates that Grad-CAM provides consistent and interpretable insights into the model's internal reasoning. The network attends to defect-related features for low-quality samples while exhibiting diffuse and low-saliency focus for good-quality coatings. This behavior is both physically plausible and aligned with expert interpretation, reinforcing the trustworthiness of the ConvNeXt-Tiny model for automated cold-spray quality assessment.

Conclusion

The results of this study highlight the combined importance of model architecture and data-centric strategies in achieving reliable classification of cold spray deposition quality. While the first training phase emphasized the limitations of relying solely on class weighting to address imbalance, the second phase clearly demonstrated that oversampling, domain-appropriate augmentation, and high-resolution inputs are essential for extracting discriminative texture features characteristic of cold spray surfaces. These findings align with recent trends in automated inspection, where data quality and representativeness increasingly drive model performance beyond architectural complexity. Among the evaluated models, ConvNeXt-Tiny consistently exhibited the most stable behavior across all classes, achieving both high accuracy and balanced class-wise metrics. Its conservative error tendency, misclassifying ambiguous degraded samples as poor rather than overestimating quality, is especially valuable in industrial contexts where false negatives (i.e., poor samples classified as acceptable) pose a greater risk than false positives. In contrast, ResNet-50, although benefiting significantly from the refined training pipeline, remained less reliable for identifying the poorest-quality samples. EfficientNet-B0 maintained strong performance but showed slightly higher sensitivity to borderline cases, particularly within the degraded class, confirming that architectural efficiency alone is not sufficient to ensure optimal discrimination of subtle material defect patterns. Grad-CAM analysis provided critical evidence supporting the interpretability and trustworthiness of ConvNeXt-Tiny's predictions. The model consistently focused on physically meaningful regions: broad defective zones in poor samples, localized irregularities in degraded samples, and diffuse non-salient regions in good samples. This alignment between model attention and expert-understood morphological features reinforces confidence in the system's internal representations and supports its integration into automated quality assurance pipelines. Importantly, the absence of spurious activations indicates that the model does not rely on artefacts such as illumination gradients or background noise, a frequent concern in vision-based industrial inspection. This preliminary study demonstrates that deep convolutional models-when trained with appropriate data balancing and augmentation strategies-can achieve not only high classification accuracy but also explainable and physically grounded decision-making. ConvNeXt-Tiny, in particular, emerges as a strong candidate for deployment in real-world cold spray inspection systems due to its robustness, conservative error profile, and interpretability. Future work will focus on expanding the dataset, incorporating multi-modal inputs (e.g., height maps or thermal data), and exploring transformer-based or hybrid architectures to further enhance generalization across varying deposition conditions and material systems. Since this is a preliminary study, another relevant direction for future work concerns the systematic evaluation of the robustness of Grad-CAM explanations under varying acquisition conditions. Although the current results indicate that ConvNeXt-Tiny focuses on physically meaningful defect-related regions, further experiments will be conducted to explicitly assess its ability to distinguish real surface defects from acquisition-related artefacts. In particular, future studies will include images acquired under different lighting setups, magnifications, and controlled perturbations, as well as synthetic noise injection protocols, to verify the stability and consistency of the attention maps and to further validate the physical reliability of the model's interpretability.

Acknowledgment

This study was carried out within the "OPTIMA: depOsition of cold sPray in the realm of green addiTIve manufacturing through construction of MAchine learning models" and received funding from the Research Projects of Significant National Interest (PRIN) 2022 PNRR, project n. D53D23017410001.

References

[1] Falco, R., & Bagherifard, S. (2025). Cold spray additive manufacturing: A review of shape control challenges and solutions. Journal of Thermal Spray Technology, 1-19.

[2] Rezzoug, A., Gholenji, R. B., & Yandouzi, M. (2025). Innovative Thermal Spray Deposition Techniques for Polymer and Polymeric Matrix Composite Substrates: Methodologies, Characteristics, and Real-World Applications. Journal of Thermal Spray Technology, 1-44.

[3] Zhang, C., Molla, T., Brandl, C., Watts, J., Mccully, R., Tang, C., & Schaffer, G. (2025). Critical velocity and deposition efficiency in cold spray: A reduced-order model and experimental validation. Journal of Manufacturing Processes, 134, 547-557.

[4] Li, W. Y., Zhang, C., Li, C. J., & Liao, H. (2009). Modeling aspects of high velocity impact of particles in cold spraying by explicit finite element analysis. Journal of Thermal Spray Technology, 18(5), 921-933.

[5] Rahmati, S., Mostaghimi, J., Coyle, T., & Dolatabadi, A. (2024). Jetting phenomenon in cold spray: a critical review on finite element simulations. Journal of Thermal Spray Technology, 33(5), 1233-1250.

[6] Eberle, M., Pinches, S., King, H., Guzman, P., Qin, K., & Ang, A. (2025). Predicting Deposition Efficiency Across Diverse Cold Spray Process Parameters Using Machine Learning. Journal of Thermal Spray Technology, 1-24.

[7] Citarella, A. A., Carrino, L., De Marco, F., Di Biasi, L., Perna, A. S., Viscusi, A., & Tortora, G. (2025). AI Data-Driven Optimization of Cold Spray Coating Manufacturing. IEEE Transactions on Industrial Informatics.

[8] Savangouder, R. V., Patra, J. C., & Palanisamy, S. (2024). A machine learning technique for prediction of cold spray additive manufacturing input process parameters to achieve a desired spray deposit profile. IEEE Transactions on Industrial Informatics, 20(10), 12275-12283.

[9] He, F., Liu, T., & Tao, D. (2020). Why resnet works? residuals generalize. IEEE transactions on neural networks and learning systems, 31(12), 5349-5362.

[10] Banerjee, C., Mukherjee, T., & Pasiliao Jr, E. (2019, April). An empirical study on generalizations of the ReLU activation function. In Proceedings of the 2019 ACM southeast conference (pp. 164-167).

[11] Mascarenhas, S., & Agarwal, M. (2021, November). A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification. In 2021 International conference on disruptive technologies for multi-disciplinary research and applications (CENTCON) (Vol. 1, pp. 96-99). IEEE.

[12] Koonce, B. (2021). EfficientNet. In Convolutional neural networks with swift for Tensorflow: image recognition and dataset categorization (pp. 109-123). Berkeley, CA: Apress.

[13] Mistry, J., Koshti, N., & Gautam, G. K. (2024, December). EfficientNetB0 for AI-Generated and Real Image Classification. In International Conference on Information and Communication Technology for Competitive Strategies (pp. 307-316). Singapore: Springer Nature Singapore.

[14] Mortaz, E. (2020). Imbalance accuracy metric for model selection in multi-class imbalance classification problems. Knowledge-Based Systems, 210, 106490.

[15] Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I. S., & Xie, S. (2023). Convnext v2: Co-designing and scaling convnets with masked autoencoders. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 16133-16142).

[16] Saini, M., & Susan, S. (2023). Tackling class imbalance in computer vision: a contemporary review. Artificial Intelligence Review, 56(Suppl 1), 1279-1335.

[17] Zhu, T., Liu, X., & Zhu, E. (2022). Oversampling with reliably expanding minority class regions for imbalanced data learning. IEEE Transactions on Knowledge and Data Engineering, 35(6), 6167-6181.

[18] Wang, Z., Wang, P., Liu, K., Wang, P., Fu, Y., Lu, C. T., ... & Zhou, Y. (2025). A comprehensive survey on data augmentation. IEEE Transactions on Knowledge and Data Engineering.

[19] Fayyaz, A. M., Abdulkadir, S. J., Talpur, N., Al-Selwi, S. M., Hassan, S. U., & Sumiea, E. H. (2025). Grad-CAM (Gradient-weighted Class Activation Mapping): A systematic literature review. Computers in Biology and Medicine, 198, 111200.