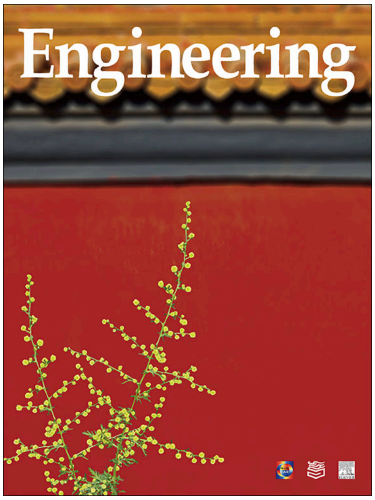


Article

LLM-Inspired Vision Transformer Framework for Intelligent Quality Recognition in Aerosol Jet Printing

Haining Zhang, Joon Phil Choi, Xingchen Liu, Nannan Liang, Yangwen Yu,
Yongrae Kim, Dongwoon Shin, Seung Ki Moon, Yong-Jin Yoon

PII: S2095-8099(26)00248-1
DOI: https://doi.org/10.1016/j.eng.2026.03.021
Reference: ENG 2327

To appear in: *Engineering*

Received Date: 15 October 2025
Revised Date: 20 March 2026
Accepted Date: 20 March 2026

Please cite this article as: H. Zhang, J.P. Choi, X. Liu, N. Liang, Y. Yu, Y. Kim, D. Shin, S.K. Moon, Y-J. Yoon, LLM-Inspired Vision Transformer Framework for Intelligent Quality Recognition in Aerosol Jet Printing, *Engineering* (2026), doi: https://doi.org/10.1016/j.eng.2026.03.021

Research

Frontiers of Extreme Manufacturing—Article

LLM-Inspired Vision Transformer Framework for Intelligent Quality Recognition in Aerosol Jet Printing

Haining Zhang [a,#], Joon Phil Choi [b,c,#,*], Xingchen Liu [d], Nannan Liang [e], Yangwen Yu [f], Yongrae Kim [b], Dongwoon Shin [b], Seung Ki Moon [g], Yong-Jin Yoon [a,g,h,*]

[a] *KAIST InnoCORE PRISM-AI Center, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea*

[b] *Department of 3D Printing, Korea Institute of Machinery & Materials, Daejeon 34103, Republic of Korea*

[c] *School of Mechanical Engineering, Chung-Ang University, Seoul 06974, Republic of Korea*

[d] *Department of Industrial and Systems Engineering, Hong Kong Polytechnic University, Hong Kong 999077, China*

[e] *College of Mechanics and Engineering Science, Hohai University, Nanjing 211100, China*

[f] *Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong 999077, China*

[g] *School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798, Singapore*

[h] *Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea*

* Corresponding authors.

E-mail addresses: jpchoi@kimm.re.kr (J.P. Choi), yongjiny@kaist.ac.kr (Y.J. Yoon).

[#] These authors contributed equally to this manuscript.

Abstract: Aerosol jet printing (AJP) is an advanced direct-ink writing technique within additive manufacturing (AM) that enables the high-resolution fabrication of micro-scale electronic components through the controlled deposition of aerosolized functional inks. Despite its potential, AJP faces significant challenges in maintaining consistent printing quality due to dynamic process variability and material heterogeneity. Existing approaches are often insufficient to address these challenges. In particular, traditional data-driven models relying on numerical inputs inherently omit critical visual morphology features, while conventional supervised vision methods struggle to generalize under practical variability. To address this, we propose an intelligent quality recognition framework that adapts large language model (LLM)-inspired Transformer architectures to manufacturing vision tasks. Specifically, Vision Transformers (ViT), which translate the self-attention mechanisms of LLMs into computer vision, are employed to achieve robust and generalizable defect detection in AJP. The proposed approach leverages a dataset composition of approximately 46 000 unlabeled images for self-supervised pre-training to enhance feature representation, followed by fine-tuning on 9369 labeled images generated through randomized printing experiments. Afterwards, comparative studies demonstrate the effectiveness of the proposed system, achieving classification accuracy of 98.1% with high robustness. By leveraging LLM-derived architectures in the context of smart manufacturing, this work introduces a scalable, adaptable, and unified machine vision framework for intelligent quality control across diverse material deposition systems. The integration of self-attention-based ViT models constructs a synergistic system that not only advances the precision of defect recognition in AJP but also contributes to the broader realization of an autonomous, data-driven manufacturing ecosystem.

Keywords: Large language models; Vision transformer; Machine learning; Self-supervised representation learning; Anomaly detection; Aerosol jet printing; Smart manufacturing

1. Introduction

Conventional printed electronics techniques, such as screen printing [1,2] and inkjet printing (IJP) [3–5], often fail to achieve the high-resolution and precise multilayer structures required for advanced electronics. Aerosol jet printing (AJP), an emerging additive manufacturing (AM) technology, overcomes these limitations by enabling the direct deposition of fine-featured (~10 μm) multi-material patterns [6–8]. This unique combination of a fine resolution, broad material compatibility, and 3D conformal printing capability positions AJP as a powerful technology for producing next-generation electronics—including flexible and wearable devices [9–11], functional sensors [12–14], and customized circuit designs [15–17].

Despite exhibiting distinct advantages, AJP faces significant challenges with regard to maintaining consistent printing quality due to dynamic process variability and material heterogeneity—factors that hinder its scalability for high-reliability applications [18–20]. To illustrate, suboptimal parameters in AJP—particularly the print speed, carrier gas flow rate (CGFR), and sheath gas flow rate (SHGFR)—can significantly degrade printed line quality outcomes, thereby compromising the electrical characteristics of the affected microelectronics [21–23]. Specifically, defects such as line discontinuities may result in open circuits, while excessive overspray increases the risk of shorting between adjacent traces [24,25]. On the other hand, insufficient line density levels can lead to concave features or voids, severely reducing line conductivity [26,27]. Furthermore, high line edge roughness degrades electrical uniformity, and excessive line spreading impairs both feature resolutions and print precision outcomes [28,29]. As such, accurate recognition of the printing quality is critical for ensuring the reliable electrical performance of advanced microelectronics [30,31].

In recent years, the recognition of AJP quality has emerged as a critical research area, driven by the increasing demand for consistent and reliable outcomes in manufacturing processes. Given that CGFR and SHGFR have been experimentally identified as critical process parameters [19,25], more researchers are now attempting to define their effects on print outcomes on both global and local scales, ultimately enabling more robust and accurate predictive quality assessments. For instance, multiple computational fluid dynamics (CFD) models have been proposed to investigate the effects of the gas flow rates on the transport of aerosols [32–34], with results indicating that the aerodynamic interaction between different gas flow rates is critical for the effective collimation ink aerosols, as it helps mitigate flow instabilities and enhance the printed line quality [35,36]. However, the intricate design of the aerosol jet printhead makes aerodynamic collimation of this component highly sensitive to gas flow fluctuations, which can quickly disrupt the velocity and pressure balance, resulting in immediate reductions in the printing quality [37,38]. Therefore, data-driven modeling is preferred for efficient quality recognition, as it identifies line patterns and anomalies without extensive domain expertise [39,40]. However, such models are not always suitable for in-process data processing, require extensive preprocessing, and are generally limited to numerical inputs—a format that frequently omits the visual details needed for rapid and precise quality evaluations [41,42].

Under such circumstances, deep learning-based printing quality evaluation methods have garnered significant recognition as they facilitate the on-line monitoring of visual variations during the process. For example, Choi et al. [43,44] introduced a convolutional neural network (CNN)-driven technique to assess jetting conditions and classify jetting states in real time for IJP. Yao et al. [45] designed a CNN-based method for monitoring defects and predicting resistance throughout the IJP process. Similarly, Zhang et al. [46] proposed a lightweight CNN-based framework for in-process printing status monitoring, facilitating immediate corrections in the AJP process. Despite these advances, several critical gaps remain unaddressed. First, obtaining large-scale labeled datasets in specialized manufacturing domains remains a resource-intensive task due to the extensive expert knowledge required. Furthermore, the inherent variability of practical environments often limits the adaptability of standard recognition techniques under fluctuating conditions. Finally, the limited interpretability of conventional deep learning models may complicate practical diagnostics, as the reasoning behind predictions remains somewhat unclear.

In parallel with these developments in deep learning, the emergence of large language models (LLMs) has led to transformative advances in artificial intelligence, particularly through their use of transformer-based architectures that capture global dependencies through self-attention mechanisms [47–49]. Originally developed for natural language processing (NPL), these architectures have since evolved to address tasks in other domains, including computer vision and manufacturing [50,51]. One of the most notable extensions is Vision Transformer (ViT), which applies LLM-inspired self-attention to visual inputs. ViT models offer strong generalization, scalability, and interpretability, making them especially appealing for complex inspection and recognition tasks in advanced manufacturing industries. Several recent studies have demonstrated the advantages of ViTs in quality inspection applications, such as the automated detection of welding penetration depths [52] and surface flaws in metal printing processes [53]. Furthermore, ViTs have consistently demonstrated exceptional anomaly detection capabilities in industrial inspections, outperforming traditional methods on benchmark datasets such as MVTec AD by precisely identifying subtle defects in high-resolution images of manufactured products [54,55]. These findings collectively highlight the growing potential of LLM-inspired transformer architectures as foundational tools for intelligent and data-efficient quality recognition in smart manufacturing systems, offering strong interpretability, reliability, and scalability. However, their reliance on extensive labeled datasets remains a challenge that hinders their real-world deployment, particularly in resource-constrained manufacturing environments. As such, developing techniques that either improve the efficiency of compiling a labeled dataset or enhance modeling performance with a limited dataset is essential, representing a significant step for future advancements in printing quality recognition for AM.

To bridge these critical gaps, this research proposes a novel machine learning-enhanced ViT framework for quality recognition in AJP. Specifically, by integrating upstream machine learning-driven data generation with downstream interpretable process analysis into computer vision, the developed framework effectively adapts the self-attention mechanisms of LLMs to achieve robust and generalizable defect detection in AJP. There are several distinct contributions of this work. First, it offers an integrated data automation strategy combining Latin hypercube sampling (LHS) and Gaussian mixture modeling (GMM) to systematically explore the process parameter space and automatically generate labeled datasets, significantly reducing the reliance on manual annotation. Second, it pioneers the application of a self-supervised "Self-Distillation with No Labels v2" (DINOv2)-powered ViT model is pioneered for quality recognition in AJP, leveraging its superior global context ability to achieve robust accuracy even with limited data. Finally, it incorporates an attention-based mechanism to visualize morphological signatures, confirming that predictions are based on physically meaningful line characteristics and assisting users in diagnosing the root causes of process instabilities, thereby surpassing mere defect detection and moving toward informative quality assessments. By adapting LLM-derived transformer architectures to visual inspection in AM, this work contributes to the broader goal of applying LLM-based intelligence to smart manufacturing environments—highlighting the role of ViT-based approaches as scalable, interpretable, and data-efficient solutions for in-process quality recognition in complex production settings.

The structure of the paper is as follows: Section 2 introduces the developed machine learning-enhanced ViT approach. Section 3 details the experimental results. Finally, Section 4 concludes the paper and explores potential further research directions.

2. Developed machine learning-enhanced ViT approach

2.1. Overview of the proposed approach

Fig. 1 illustrates a machine learning-enhanced ViT framework designed for AJP. In this research, an aerosol jet printer is equipped with two imaging systems, as depicted in Fig 1(a): a coaxial camera (Jingtuo Youcheng Technology, China) designed for the automatic analysis of printed line quality levels within the design space, and a process camera (AnMo Electronics Corporation, China) for in-process image acquisition and decision support during printing. Specifically, as illustrated in Fig. 1(b), an integrated machine learning approach directs the coaxial camera to analyze critical process parameters and determine the distribution of line patterns, facilitating the automated generation of a labeled training dataset through randomized printing experiments; Fig. 1(c) illustrates the development of a ViT model, in this case trained on a large annotated dataset, enabling continuous *in-situ* monitoring throughout the AJP process; and Figs. 1(d-i)–(d-vi) demonstrates the capability of the proposed approach to perform in-process quality recognition, thereby improving the printing stability and ensuring consistent product quality.

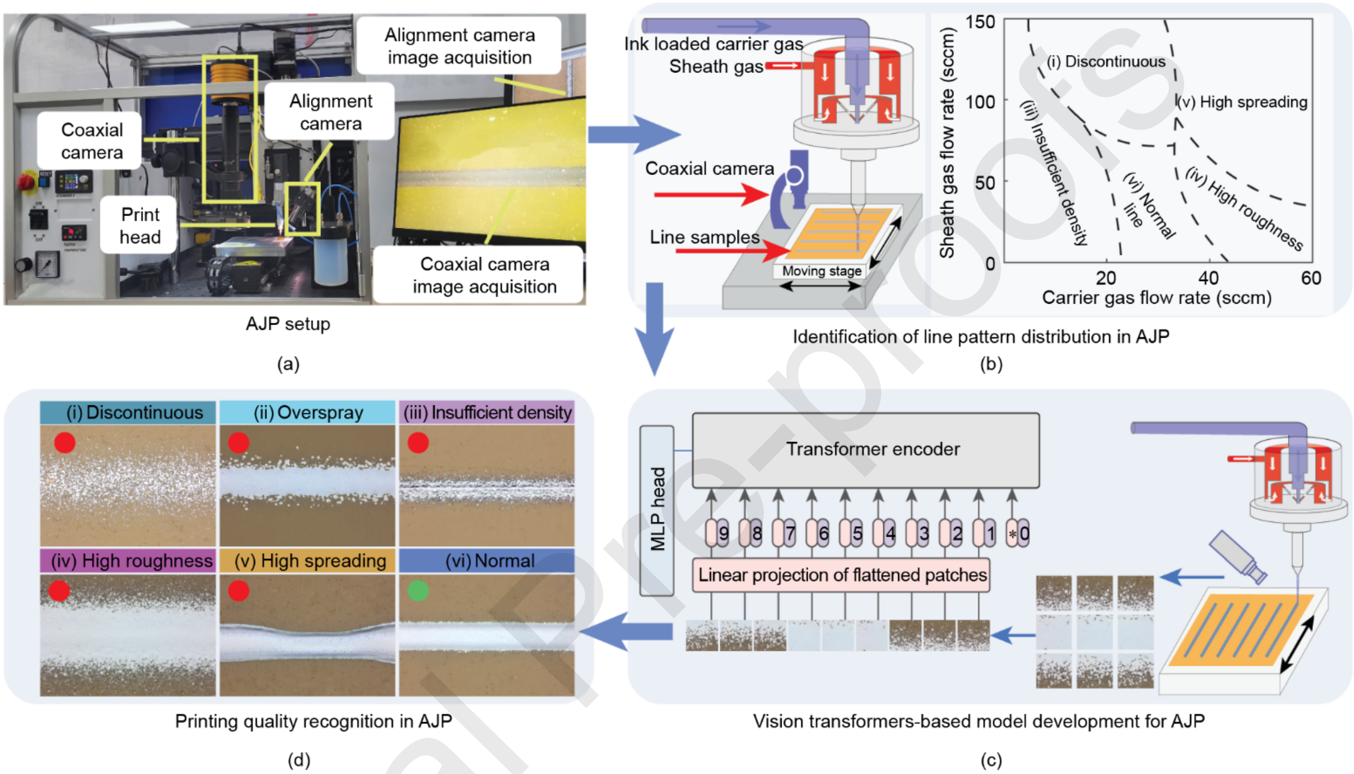


Fig. 1. Machine learning-enhanced ViT framework for printing quality recognition in AJP: (a) architecture of the dual-camera AJP system; (b) identification of the line pattern distribution in AJP based on an integrated machine learning approach; (c) ViT-based model development; and (d-i)–(d-vi) quality recognition based on the developed ViT model.

2.2. Line pattern distribution analysis driven by machine learning

Fig. 2(a) presents the integration of an aerosol jet printer with a coaxial camera system, where the imaging device is aligned with the nozzle and positioned perpendicularly to the substrate. An 8-megapixel charge-coupled device camera with a resolution of 3840 pixel × 2160 pixel was utilized to capture in-process images. After deposition, the substrate moves beneath the coaxial camera, enabling immediate image acquisition to evaluate dynamic processes. This setup allows for a rapid analysis of printed line patterns and aids in AJP process modeling, as shown in Figs. 2(b)–(g), by providing high-resolution visual feedback during the deposition process.

Specifically, the integrated machine learning approach for AJP is structured into the following six key stages: ① Defining deposited printed line features, including roughness and overspray, to evaluate the overall quality of the print; ② a central composite design (CCD) approach, supported by a coaxial camera and image-processing algorithms, is utilized to quantify the effect of the main process variables on key printed line characteristics within a defined design space; ③ analysis of variance (ANOVA) is adopted to confirm the significance of the key process variables on the printing quality; ④ a LHS experimental design is then utilized to investigate the design space efficiently in relation to the key process variables with a limited dataset; ⑤ a GMM is subsequently applied to determine the distribution of line patterns, aiding in the generation of a labeled training dataset for the subsequent development the ViT model; ⑥ Bayesian neural networks (BNNs) are incorporated to explore the relationships among process inputs and outputs, offering valuable guidance related to the fine-tuning of the printing process. Furthermore, a comprehensive workflow, as detailed in Table 1, clarifies the procedural steps for the identification of the line patterns distribution, highlighting the inputs, methods, and expected outputs at each stage.

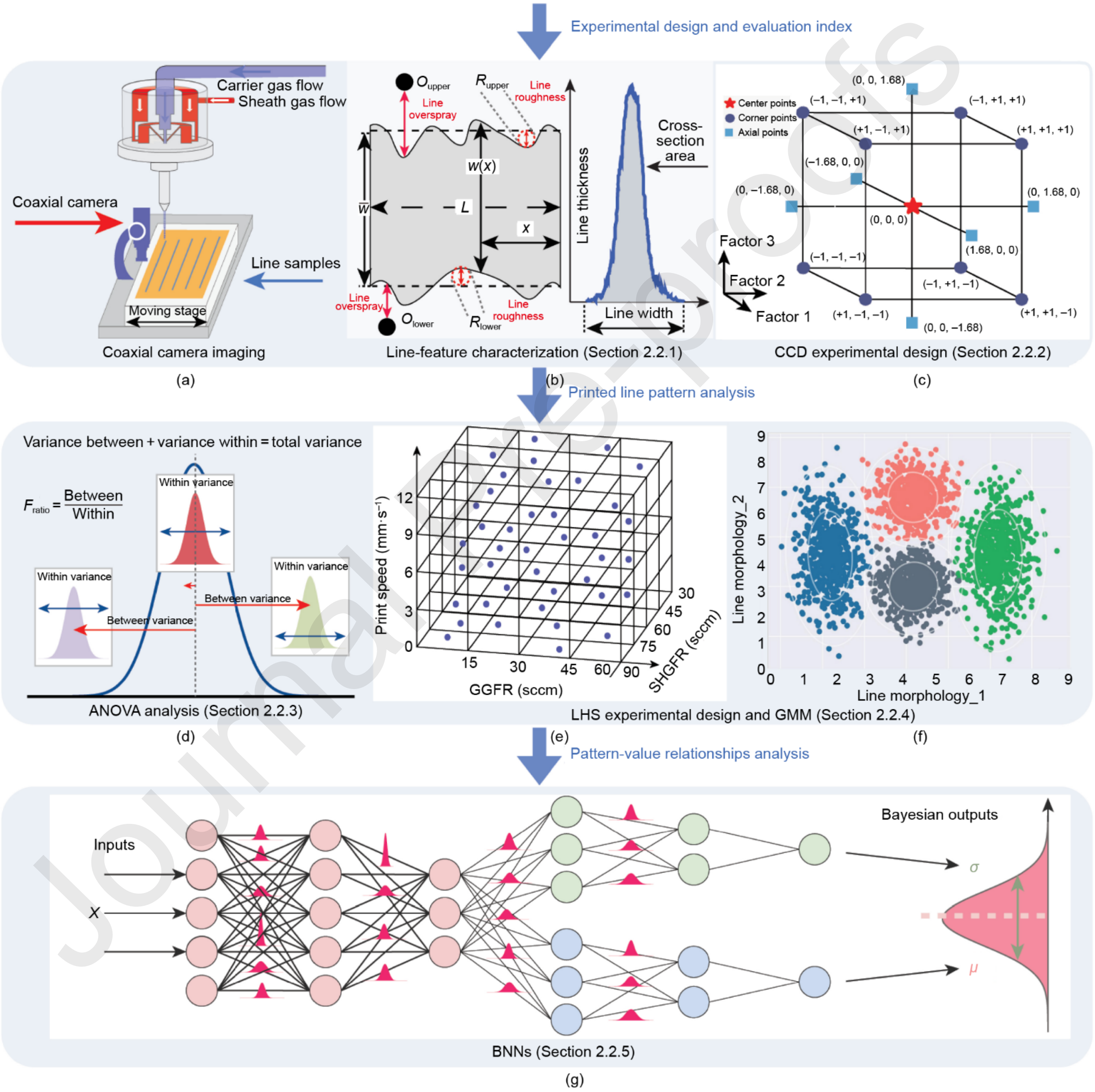


Fig. 2. Analysis of main process parameters and determination of the line pattern distribution via the integrated machine learning approach. (a) Illustration of a coaxial camera system for capturing printed line samples; (b) definition of printed line characteristics, where O_{upper} and O_{lower} represent the upper and lower overspray widths,

R_{upper} and R_{lower} denote the upper and lower line roughness, $w(x)$ is the actual line width at position x, and L is the reference line length; (c) CCD-based experimental design; (d) ANOVA analysis of AJP main process parameters, where F_{ratio} is the variance ratio; (e) LHS experimental design within the space of the main process parameters, where sccm stands for standard cubic centimeter per minute; (f) GMM-based analysis of the line pattern distribution; and (g) BNN modeling of the AJP process, where X represents the input features, μ is the predicted mean, and σ denotes the standard deviation.

Table 1

Detailed process for the determination of the line pattern distribution as illustrated in Fig. 2.

Step	Inputs	Approaches	Responses
1	Initial design space: {Print speed, SHGFR, CGFR}	CCD experimental design	Designed experiments: {Print speed, SHGFR, CGFR}
2	Designed experiments: {Print speed, SHGFR, CGFR}	Coaxial camera and laser microscope imaging	Mean line thickness ($\overline{T}$); mean line width ($\overline{w}$)
3	($\overline{w}$, $\overline{T}$) {Print speed, SHGFR, CGFR}	ANOVA analysis	Significance analysis; correlation analysis
4	Confirmed design space: {Print speed, SHGFR, CGFR}	LHS experimental design and coaxial camera imaging	Line discontinuity (L_{disc}); mean line edge roughness (R_{m}); mean line density ($\overline{L_{\rho}}$)
5	($\overline{L_{\rho}}$, R_{m}, L_{disc})	GMM	Line pattern distribution analysis
6	($\overline{w}$, $\overline{T}$)	BNNs	Developed process model

{Print speed, SHGFR,
CGFR}

2.2.1. Definition of printed line characteristics

An effective analysis of the line pattern distribution in AJP requires quantitative definitions of the pertinent printing quality indicators. This study introduces five fundamental quality metrics for evaluating the printing quality. In particular, as illustrated in Fig. 2(b), the average line width $\overline{w}$ is calculated by spatially averaging pixel intensity values across *N* columns in the discretized dataset,

$$\overline{w} = \frac{1}{N}\sum_{i=1}^{N} w_i \tag{1}$$

where w_i denotes the quantized line width in the *i*th column. Additionally, the 3D surface topography of each line sample is analyzed using a laser microscope. For each line sample, the ratio of the cross-sectional area (c_i) to the line width (d_i) was averaged across *M* localized measurements to derive the mean line thickness (T_{m}).

$$T_{\mathrm{m}} = \frac{1}{M}\sum_{i=1}^{M} \frac{c_i}{d_i} \tag{2}$$

The line discontinuity (L_{disc}) and average line edge roughness (R_{m}) and are then defined relative to the determined reference lines (mean lines):

$$L_{\mathrm{disc}} = \frac{\mathrm{F}_u + F_{\mathrm{l}}}{2N} \tag{3}$$

$$R_{\mathrm{m}} = \sqrt{\frac{1}{2N}\sum_{i=1}^{N}\left(R_{\mathrm{upper},i}^2 + R_{\mathrm{lower},i}^2\right)} \tag{4}$$

Subsequently, the line density ($\overline{L_\rho}$) is measured by computing the mean grayscale intensity of the pixels enclosed by the identified line edges

$$\overline{L_\rho} = \frac{1}{N}\sum_{i=1}^{N} I_i - I_b, \tag{5}$$

where I_i denotes the average intensity in each column, while I_b is the background intensity used to correct for illumination effects. R_i reflects the deviations of the actual line edge from the mean edge, while F_{u} and F_{l} represent the number of failures during the detection of the upper and lower edges, respectively.

In this study, the printed line width (PLW) and printed line thickness (PLT) are not utilized as input features for the ViT image classifier. Instead, these parameters provide quantitative physical descriptors of the printed line morphology to facilitate the analysis of process morphology variations. Furthermore, they serve as the target variables for the BNN-based predictive modeling, linking the process parameters to measurable geometric outcomes. While the image classifier operates on captured images for visual defect recognition, these width and thickness descriptors are employed to interpret morphology variations and assist with the decision-support strategy.

2.2.2. CCD for statistical experimentation

The AJP process exhibits significant nonlinearity due to complex aerodynamic interactions within the printhead, requiring an experimental design capable of effectively analyzing the nonlinearity of response surfaces [26]. In this study, CCD is employed to address curvature effects and nonlinear interactions systematically. As illustrated in Fig. 2(c), the CCD approach incorporates factorial points to examine the interaction and linear effects of key process factors, center points to measure process variability through iterative testing, and axial points to evaluate quadratic terms along experimental axes. This method allows for a more efficient study of important quality metrics with fewer experimental runs, substantially lowering resource consumption levels compared to exhaustive trial designs while maintaining high accuracy in the second-order model.

2.2.3. ANOVA analysis of main process parameters

CGFR, SHGFR, and the print speed have been experimentally determined as critical process variables. However, a thorough quantitative analysis is required to assess both the independent effects and interactions of these parameters on the printing quality. To achieve this, ANOVA, as shown in Fig. 2(d), is employed to analyze the results of the CCD experiments, providing a systematic statistical framework by which to quantify the significance and relative contributions of process parameters (print speed, SHGFR, CGFR) key to the printed line geometry [56]. This facilitates the identification of the dominant factors driving morphology variations, enhancing the overall understanding of their impact on the printing quality.

2.2.4. Identification of line pattern distribution using LHS and GMM

To analyze the distribution of line patterns within the design space, a thorough examination of the effects of the key process variables on the AJP quality is necessary. In this regard, LHS design is utilized here, as it offers the ability to investigate the design space comprehensively while minimizing the required number of trials. As depicted in Fig. 2(e), LHS achieves space-filling coverage by dividing the parameter space into intervals, randomly sampling one point from each, and ensuring uniform representation across the entire domain [57]. This strategy enables an effective exploration with well-distributed and statistically representative samples.

Based on the designed LHS experiments, GMM is employed to study the effects of the primary process variables on the distribution of the printed line patterns [58]. As depicted in Fig. 2(f), GMM offers strong adaptability to data variability and heterogeneity, making it particularly well suited for extracting nuanced insights into parameter effects. For the clustering analysis, key printed line attributes ($\overline{L_\rho}$, R_m, L_disc) are used as input features, while the parameters ($\overline{w}$, T_m) are reserved for a correlation analysis in Section 2.2.5. The resulting distribution of distinct line patterns across the design space provides valuable prior knowledge, significantly enhancing the efficiency of generating a labeled training dataset for subsequent ViT model development.

2.2.5. BNNs for process modeling

The selection of BNNs for process modeling is motivated by the requirement to address the high inherent variability and practical constraints of limited experimental data within AJP [59]. Unlike deterministic models, BNNs establish a probabilistic framework that quantifies the prediction uncertainty through confidence intervals (CIs), providing risk-aware guidance for robust process adjustments in real-world manufacturing settings. This approach serves as an uncertainty-aware process model that transforms the predictive tool into a transparent framework for process assessments, offering confidence levels that enhance the reliability of the overall printing process while aligning with the framework's goal of enabling adaptable quality monitoring.

2.3. ViT model development for quality recognition in AJP

The development of the ViT-based quality recognition model represents a key innovation within this framework, characterized by its seamless integration of upstream automated data generation and downstream interpretable process analysis. Departing from conventional methodologies that compartmentalize data collection and model training, the proposed integrated system utilizes statistically representative datasets derived via LHS and GMM, which are further enriched with semantic features through DINOv2 to ensure superior generalization and robustness. To minimize manual annotation burdens, the framework incorporates a multi-stage efficiency strategy. This approach initially leverages automated data acquisition aligned with identified line patterns to transform defect scarcity into a controllable data generation task, followed by the utilization of self-supervised representation learning via DINOv2 to capture invariant features from extensive unlabeled data.

Ultimately, transfer-learning is utilized to capitalize on pre-trained visual features through fine-tuning, thereby reducing the reliance on expert labels for reliable domain-specific classification.

Specifically, Fig. 3 presents an overall flowchart of the development of the ViT-based quality recognition system in AJP. As shown in Fig. 3(a), a process camera is installed next to the printhead, utilizing the identified distribution of line patterns to automate the generation of a labeled training dataset through randomized experiments. For this study, a 1.3 MP charge-coupled device camera is used for image collection during printing. In the next step, as depicted in Fig. 3(b), the DINOv2 framework—designed to learn invariant representations without labeled data—performs representation learning, extracting unsupervised and robust visual features from the collected images. Subsequently, as shown in Fig. 3(c), the ViT architecture is then paired with the self-supervised DINOv2 framework and fine-tuned for the recognition of the AJP printing quality using the collected and labeled dataset. As ViT excels at detecting localized anomalies and interpreting their temporal and spatial relationships, it effectively eliminates the need for extensive manual adjustments to materials or equipment through trial and error while also enhancing the precision of classification under challenging conditions. Therefore, as illustrated in Fig. 3(d), the developed ViT model is implemented for in-process quality assessments in AJP. The derived process correlations between the input parameters and output quality responses can serve as a systematic guide for dynamically adjusting critical process variables when anomalies are detected, ensuring consistent printing quality through continuous corrective actions.

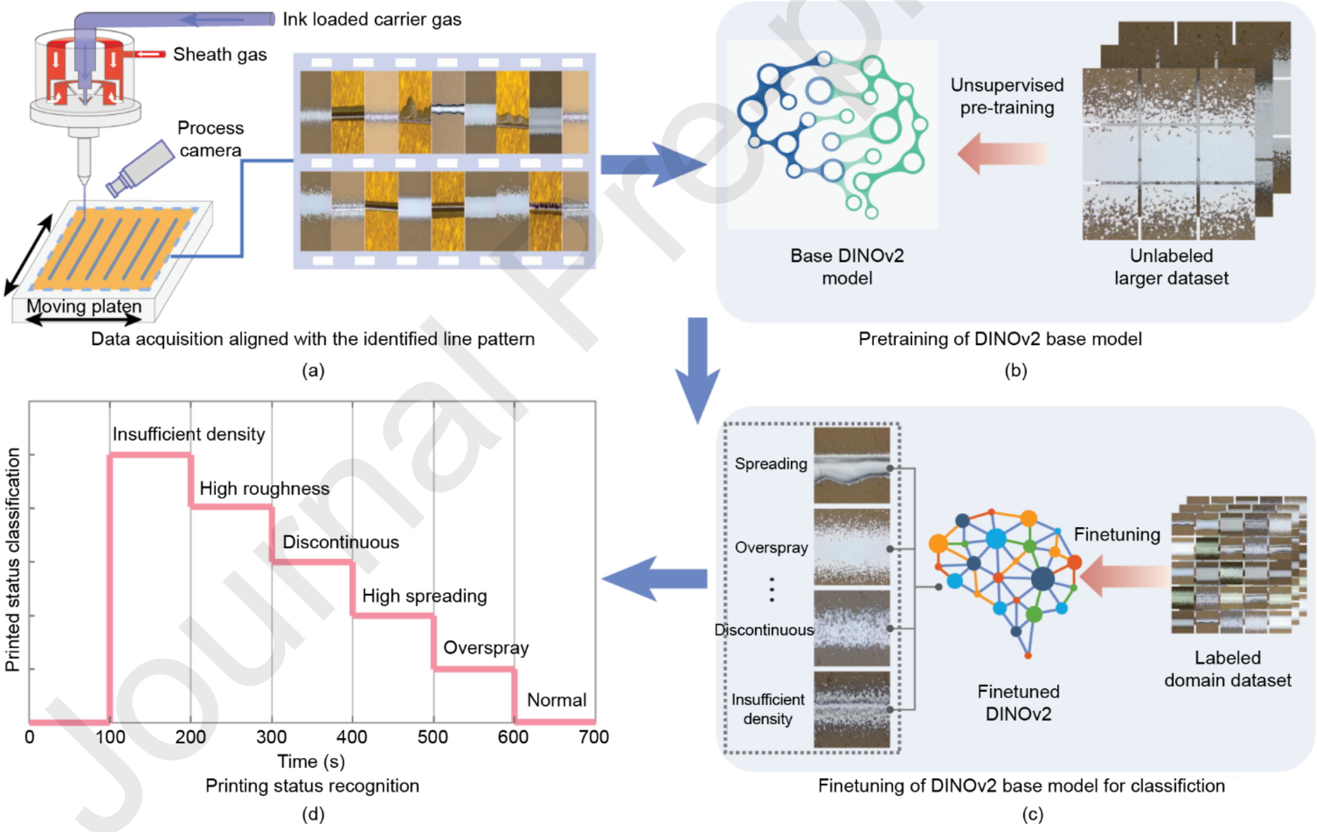


Fig. 3. Overall flowchart of the development of the ViT-based quality recognition system in AJP: (a) Illustration of the process camera system that automatically collects line images during printing; (b) DINOv2-based representation learning of unsupervised visual features; (c) fine-tuning of the ViT model with the collected labeled datasets; and (d) *in-situ* quality recognition based on the developed ViT model.

2.3.1. DINOv2 for unsupervised representation learning

DINOv2, an advanced self-supervised learning approach, is utilized in this study to extract high-quality visual features without relying on labeled data. Renowned for its strong generalization capabilities across various computer vision tasks, DINOv2 proves particularly effective in contexts where annotated datasets are scarce or costly to obtain [60]. Specifically, quality defects in the AJP process, caused by variations in materials, process parameters, or equipment conditions, lead to diverse and intricate defect patterns that significantly hinder traditional supervised learning methods. To overcome this, the present study utilizes pretrained DINOv2 representations and applies fine-tuning to address the task of print quality classification more effectively in AJP. By leveraging the robustness of DINOv2's self-supervised features, the proposed method can significantly reduce dependence on extensive labeled datasets, thereby improving both the generalization performance and reliability of defect detection under complex and variable production conditions.

Specifically, DINOv2 employs a discriminative self-supervised learning framework based on the Teacher–Student architecture. It utilizes a cross-view training strategy where the Student network learns to predict the representation of the Teacher network from different augmented views (e.g., global and local crops) of the same image. By minimizing the cross-entropy loss between the outputs, the model forces the student to learn invariant and robust visual features (such as the object geometry and global structure) rather than superficial low-level details. Consequently, unlike traditional supervised learning methods that often overfit to local textural details, this self-distillation mechanism enables the model to generalize more effectively using domain-specific data. By capturing robust semantic features, the framework contributes to mitigating relevant challenges regarding sensitivity to environmental noise in AM. In this research, the unlabeled dataset constructed for the self-supervised pre-training phase consists of approximately 46 000 images. These images were collected from historical experimental records and random sampling during various printing runs. This separate dataset was utilized to perform domain specific pre-training. Given that the original DINOv2 backbone is pre-trained on general domain images, this step assists in bridging the domain gap and allows the model to adapt more effectively to the visual characteristics of aerosol jet printed lines.

2.3.2. ViT for printing quality classification

ViT is a pioneering deep learning model tailored for image classification that works by adapting the transformer architecture, originally designed for NPL, to the domain of visual data [61]. In contrast to CNNs, ViT splits an image into a grid of patches with a fixed size, considering each patch as a separate token within a sequence. Once these patches are flattened, they are linearly embedded, with positional encodings incorporated to maintain the spatial dependencies among them. Additionally, a unique token designated for classification is introduced into the sequence, which ultimately collects and synthesizes universal information after passing through the transformer encoder. The encoder, composed of feedforward networks and multiple layers of multi-head self-attention mechanisms, enables ViT to extract both local and long-range dependencies in the images [62]. This ability of global attention to capture complex patterns makes ViT particularly well suited for discerning subtle and distributed patterns in images, which is essential for achieving high-quality recognition in applications such as AJP.

In this research, the DINOv2 model is pretrained on an unlabeled dataset, after which the paired ViT model undergoes fine-tuning for AJP quality classification using an automatically collected and labeled dataset. Fine-tuning consists of modifying the pretrained feature extractor for a targeted implementation by incorporating a layer for classification above the pretrained encoder. This layer is initially assigned random weights and is then optimized by means of supervised learning. During the fine-tuning process, the parameters of the DINOv2 model can either be frozen or updated, based on factors such as the size and quality of the task-specific dataset. Keeping the backbone weights frozen helps maintain the pretrained features and mitigates the risk of overfitting, especially when the dataset is limited. On the other hand, permitting updates to the backbone during training allows the model to better adjust to the specific task, although this approach demands more data and greater computational resources.

3. Experimental results and discussion

3.1. Experimental setup

In this research, an atomizer-based aerosol jet printer was employed to conduct the experiments, utilizing a CCD design with actual factor levels corresponding to the center (0), axial (±1.68) points, and factorial (±1), as detailed in Table 2. The technical configuration features a 150 µm tip diameter and a 4 mm working height, with operations conducted at an atomization current of 0.4 mA and an ink temperature of 20 °C while maintaining the plate at ambient temperature. This configuration aligns with typical AJP setups by focusing on universal process variables fundamental to aerosol collimation. The key independent factors under investigation were the print speed (0.23 to 4.02 mm·s^{-1}), SHGFR (29.92 to 70.03 sccm, where sccm stands for standard cubic centimeter per minute;), and CGFR (30 to 44.95 sccm), while the target responses consisted of the PLW and PLT. Prior to printing, a functional nanoparticle ink was prepared by diluting silver ink at 1:1 with deionized water, while the substrate was treated in a 3 min isopropyl alcohol (IPA) bath under sonication at room temperature to improve its wettability. During printing, experimental conditions were adjusted based on the CCD design, with every condition replicated three times to produce single-pass line samples characterized by unique morphometric properties. Immediately following the printing step, the morphology of the printed lines was examined using a microscope and the profiles were analyzed through an image-processing algorithm to confirm the reliability of the characterization.

Meanwhile, to construct the dataset with which to train the ViT model, a charge-coupled device camera was utilized to collect 9369 images of printed lines by modifying adjustable process variables based on the determined line pattern distribution. To enhance the generalization capability of the model, a training strategy was implemented that treats the multi-view setup as a form of data augmentation. Specifically, a subset of coaxial camera images was incorporated alongside the charge-coupled device camera images during the ViT training phase. By exposing the model to defect features from both perspectives, the network is encouraged to learn intrinsic view-invariant representations of the defects as opposed to overfitting to the specific angle or lighting of a single sensor. In this research, these images were categorized into the following six distinct morphology classes: “normal line,” “insufficient density,” “high spreading,” “discontinuities,” “high roughness,” and “overspray,” with data collection guided by the morphology distribution to maintain a balanced dataset. Consequently, the collected data exhibit a relatively uniform distribution with class proportions ranging from 14.7% to 18.0%, specifically comprising 1689 samples for normal line, 1491 for insufficient density, 1641 for high spreading, 1509 for discontinuities, 1381 for high edge roughness, and 1658 for overspray. Moreover, various data-augmentation techniques, in this case lighting adjustments, flipping, resizing, and rotation were applied to increase the diversity, with this followed by an 80-20 train-validation split. Both the baseline models and the proposed framework are evaluated on this specific dataset to maintain consistency in the cross-method comparison, thus facilitating a more direct interpretation of the results and supporting the precision of the evaluation procedure.

Table 2

Designed experimental points based on CCD.

NO.	SHGFR (sccm)	CGFR (sccm)	Print speed (mm·s^{-1})
1	49.97	35.58	2.12
2	70.03	35.58	2.12
3	49.97	35.58	2.12
4	49.97	35.58	4.02

5	61.90	30.00	3.25
6	38.05	30.00	3.25
7	49.97	35.58	2.12
8	38.05	41.15	3.25
9	49.97	26.20	2.12
10	61.90	41.15	3.25
11	49.97	35.58	2.12
12	61.90	30.00	1.00
13	29.92	35.58	2.12
14	61.90	41.15	1.00
15	49.97	35.58	2.12
16	49.97	35.58	2.12
17	38.05	30.00	1.00
18	49.97	44.95	2.12
19	38.05	41.15	1.00
20	49.97	35.58	0.23

3.2. Analysis of the distribution of printed line patterns

The quantitative analysis in this section follows a sequential procedure comprising ANOVA-based significance testing, GMM clustering, and BNN-based process modeling. To ensure physical accuracy and maintain fidelity of the data, these evaluations rely exclusively on high-fidelity measurements from a coaxial metrology system. This approach ensures that the statistical inferences and predictive modeling remain grounded in precise quantitative data, providing a robust foundation for interpreting the printed line morphology.

3.2.1. ANOVA analysis of main process parameters

The target responses derived from the CCD design (Table 3) were analyzed using quadratic regression models to quantify the effects of key process parameters on the printed line characteristics. The ANOVA results confirmed the statistical significance of both the overall models and individual factors, confirming their reliability in capturing process-response relationships. Detailed ANOVA outcomes for responses R1 (line width) and R2 (line thickness) are presented in Table 4, involving main factors A (SHGFR), B (CGFR), and C (print speed), along with their interactions (e.g., AB, AC). Significant factors were identified by low *P* values, while high *F* values indicated that the model variation exceeded the inherent process variability. The interactions AB and AC for R2, deemed statistically insignificant, were removed to enhance the model robustness. A lack-of-fit test, with *P* values below 0.05, showed that residual errors were minor compared to replicate variability, further confirming the adequacy of the model. Strong coefficient of determination (R^2) values highlighted the models' explanatory strength, and precision ratios above 15 indicated sufficient signal-to-noise ratios for reliable predictive use. Overall, the CCD model successfully identified the statistically significant impacts of the selected main process parameters and revealed the critical interactions between them, thereby providing valuable insights for further quality-driven process optimization.

Table 3

CCD-derived target responses of AJP.

NO.	PLW (μm)	PLT (μm)
1	43.97	1.63
2	42.35	1.78
3	43.98	1.61
4	40.18	1.03
5	35.76	0.82
6	38.1	0.61
7	43.97	1.79
8	45.6	0.89
9	34.6	0.85
10	38.1	1.36

11	43.06	1.55
12	41.11	1.81
13	46.15	1.19
14	49.91	2.61
15	43.96	1.56
16	41.53	1.59
17	38.51	1.21
18	52.16	1.82
19	53.01	2.15
20	46.82	2.11

Table 4

ANOVA results of the CCD study.

Factors	R1		R2	
	F value	*P* value	*F* value	*P* value
Model	50.29	< 0.0001	49.79	< 0.0001
A	13.63	0.0027	29.45	< 0.0001
B	191.89	< 0.0001	69.25	< 0.0001
C	63.90	< 0.0001	136.90	< 0.0001
AB	9.88	0.0078	—	—

AC	7.28	0.0182	—	—
BC	15.17	0.0018	5.66	0.0321
B^2	—	—	7.71	0.0148
R^2	0.9587	—	0.9468	—
Adjusted R^2	0.9396	—	0.9277	—
Adequate precision	24.93	—	25.145	—

Specifically, the modeling accuracy was assessed by analyzing studentized residuals (standardized in units of standard deviations). For instance, Figs. 4(a) and (b) display normal probability plots of the target responses, where the residuals closely follow a straight-line pattern with limited scatter, thus supporting the statistical soundness of the models. Figs. 4(c) and (d) show that residuals are randomly distributed around zero with no discernible pattern, indicating their independence from the test order and ruling out any confounding effect on the CCD models. Moreover, Figs. 4(e) and (f) compare the predicted printed line features with the actual values against a reference line ($y = x$). The elevated coefficients of determination and acceptable precision levels demonstrate robust model performance, effectively capturing the causal relationships between the key process variables and the produced line features. Overall, the ANOVA results confirm the reliability of the developed CCD model for subsequent process optimization.

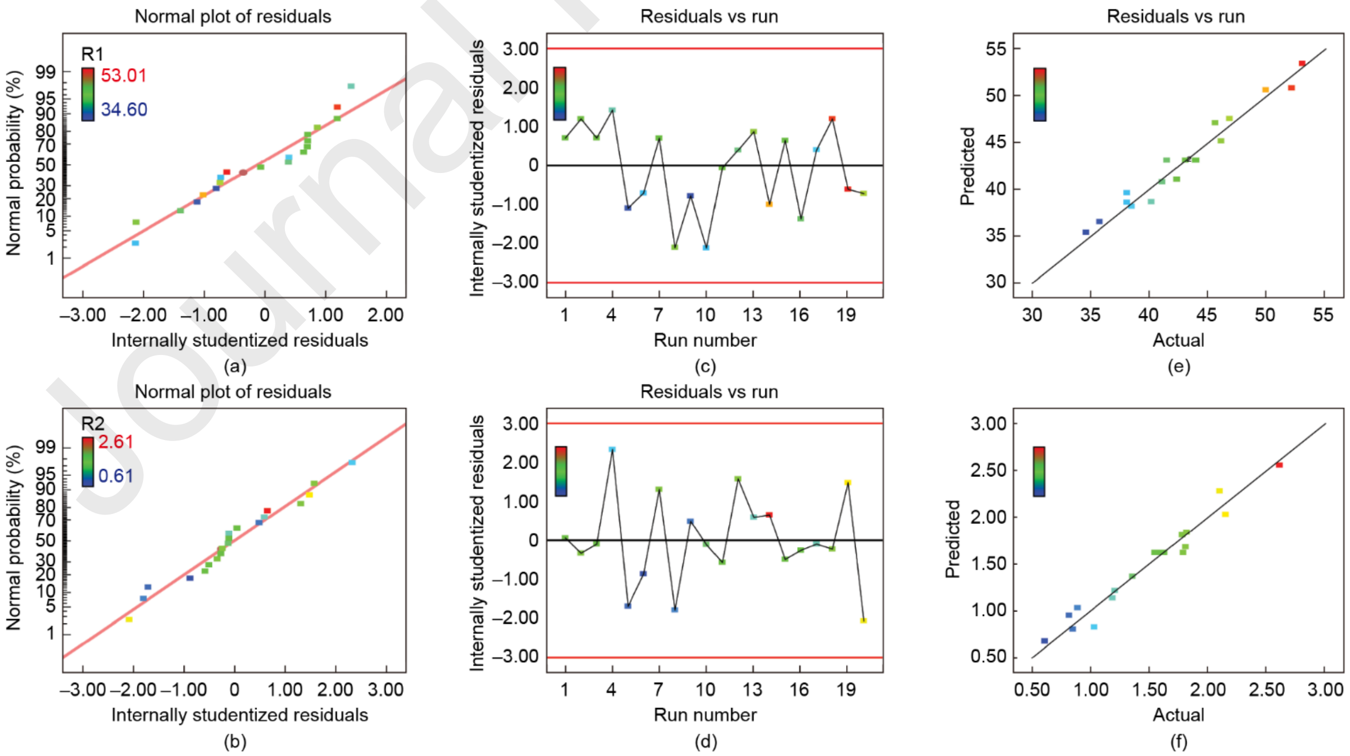


Fig. 4. Analysis of residuals and model verification of the PLW and PLT: (a, b) Assessing residual normality using probability plots of (a) PLW and (b) PLT; (c, d) residuals plotted against test orders of (c) PLW and (d) PLT; and (e, f) predicted vs observed correlations of (e) PLW and (f) PLT.

3.2.2. GMM-based identification of the line pattern distribution

Following the ANOVA analysis, the 2D design space corresponding to SHGFR and CGFR was explored further using LHS at a range of print speeds. During this process, line sample images were captured, revealing anomalies related to example discontinuities, overspray, low density levels, significant edge roughness, and excessive spreading. These images were then processed as shown in Fig. 5. The deposited line patterns were subsequently assessed according to the predetermined metrics, in this case the mean line density ($\overline{L_\rho}$), line discontinuity (L_{disc}), and line overspray (O_{sp}), using GMM.

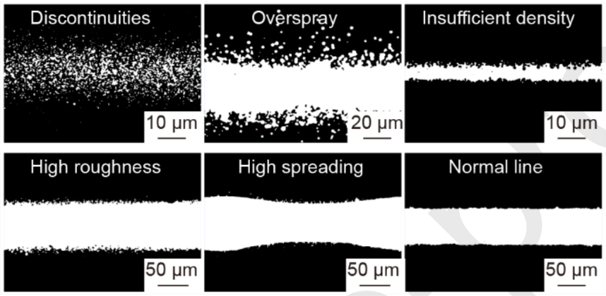


Fig. 5. Examples of six distinct printed line morphologies observed under different operating conditions.

The GMM clustering results shown in Fig. 6 systematically illustrate the evolving distribution of the printed line patterns across the CGFR and SHGFR parameter space at six different print speeds. Each graph highlights regions representing distinct line quality characteristics, marked by dashed boundaries in this case. As the print speed increases from Fig. 6(a) 0.5 to Fig. 6(f) 8 mm·s^{-1}, the distribution of the printed line patterns shifts noticeably: the optimal "normal line" region contracts and moves diagonally toward the bottom right. Simultaneously, the defect regions "discontinuous" and "high overspray" expand, particularly dominating the parameter space at 8 mm·s^{-1}. This shift underscores the stricter gas flow tolerances required to maintain line quality levels at higher print speeds. Overall, the GMM clustering method effectively captures these trends, revealing the complex interaction between key process parameters and the printed line quality. The GMM method helps identify separate clusters in the design space, offering a clear depiction of how the printed line patterns are distributed. This approach not only enhances our understanding of the underlying structure but also aids in the efficient generation of a labeled training dataset, crucial for model development and improving the predictive accuracy.

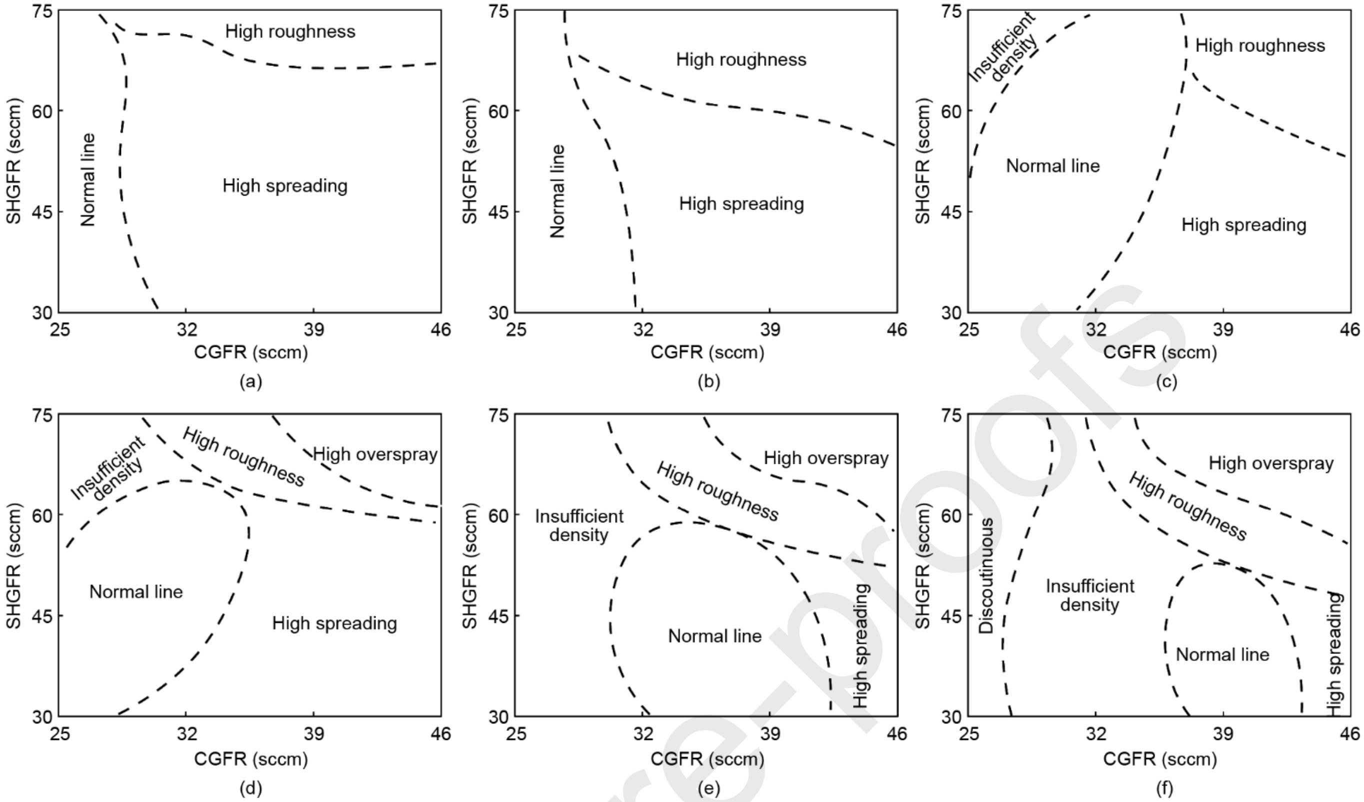


Fig. 6. Distributions of the printed line morphology at print speed of (a) 0.5, (b) 1, (c) 2, (d) 4, (e) 6, and (f) 8 mm·s^{-1}.

3.2.3. BNN-based process modeling

To adapt to system variations effectively and enable prompt responses, it is crucial to establish robust correlations between input parameters and printed line characteristics within the AJP process. This research employs BNNs for advanced process modeling. Specifically, the BNN provides posterior mean predictions of key geometric responses, in this case the line width and thickness, alongside credible intervals that quantify the associated uncertainty of these predictions. In each section of Fig. 7, the relationships between these features and specific process variables are depicted, with the remaining variables held constant at their mean values. The red and blue grid surfaces in Fig. 7 indicate the lower and upper bounds of the 95% CIs, respectively. In deployment, these estimates are utilized to identify regimes with elevated predictive uncertainty and to recommend additional inspection or data-collection steps. Furthermore, the uncertainty quantification enables the system to prioritize conservative parameter changes toward regions with lower predicted variance. This probabilistic output can also be integrated into multi-objective optimization tasks to support robust decision-making, such as achieving a customized line width while maximizing the aspect ratio, by weighing performance targets against process reliability.

In this research, the BNN modeling results demonstrate that achieving optimal printing quality requires a careful balance of the main process parameters, as extremes lead to greater uncertainty and reduced process stability. Specifically, Fig. 7 shows that the line width is mainly influenced by the SHGFR-to-CGFR ratio, where higher ratios narrow the line due to material constriction, though the effects diminish at extremes. The line thickness, however, increases with higher CGFR values due to more rapid faster material delivery. Both printed line features decrease with faster higher print speeds as the reduced deposition time hinders material accumulation. This balance supports informed process parameter adjustments to rectify inconsistencies in the printed line characteristics. For instance, if the "insufficient density" status is observed, the CGFR can be adjusted upwards to boost the aerosol mass flow rate. Subsequently, the SHGFR should be increased to enhance the focusing effect, while the print speed should be reduced to amplify the volume of aerosol deposited on the substrate.

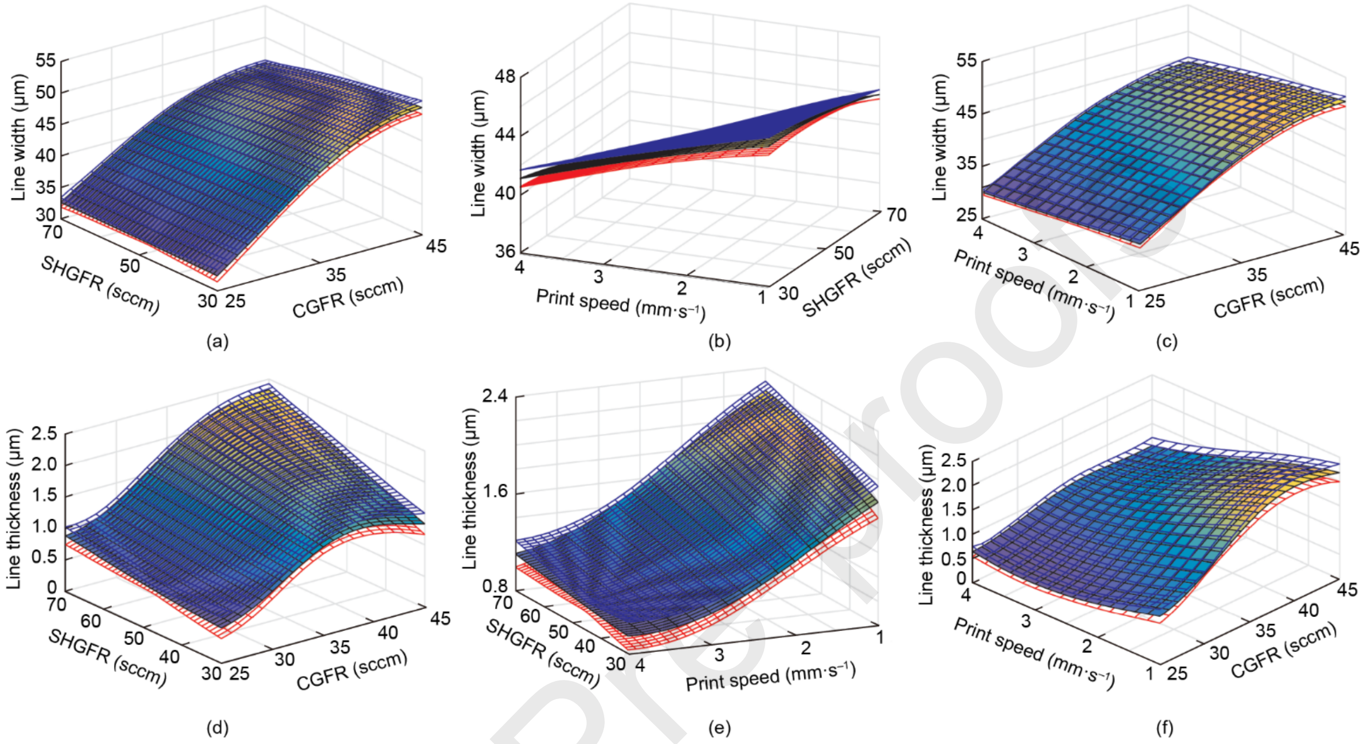


Fig. 7. Surfaces representing the posterior mean and 95% CIs of the printed line characteristics relative to the main process variables: (a–c) PLW as the response variable, and (d–f) PLT as the response variable. The red and blue grid surfaces indicate the lower and upper bounds of the 95% CIs, respectively.

3.3. ViT-based quality recognition in AJP

3.3.1. ViT development and comparative analysis

In this research, ViT along with two representative deep learning models (ResNet [63,64] and YOLOv8 [65,66]) were evaluated and compared for quality recognition in AJP. The input image size used for all models was 224 pixels by 224 pixels. The classification head was trained using the cross-entropy loss, which aims to minimize the discrepancy between predicted classes and the corresponding true labels. The fine-tuning process employed stochastic gradient descent (SGD) to enhance the modeling performance and accelerate convergence. Critical hyperparameters, in this case the batch size, learning rate, and weight decay, were carefully fine-tuned. Table 5 summarizes the comparative performance outcomes of these models. ViT, ResNet, and YOLOv8 attained test accuracy rates of 98.1%, 92.6%, and 93.9%, respectively. While direct comparisons between architecturally distinct models such as ResNet, YOLOv8, and ViT are often challenging, a meaningful comparison has been made possible in this study through a uniform dataset generated though GMM and LHS. This strategy ensures fully consistent training conditions, thereby guaranteeing that the performance differences primarily reflect architectural distinctions rather than variability in the data.

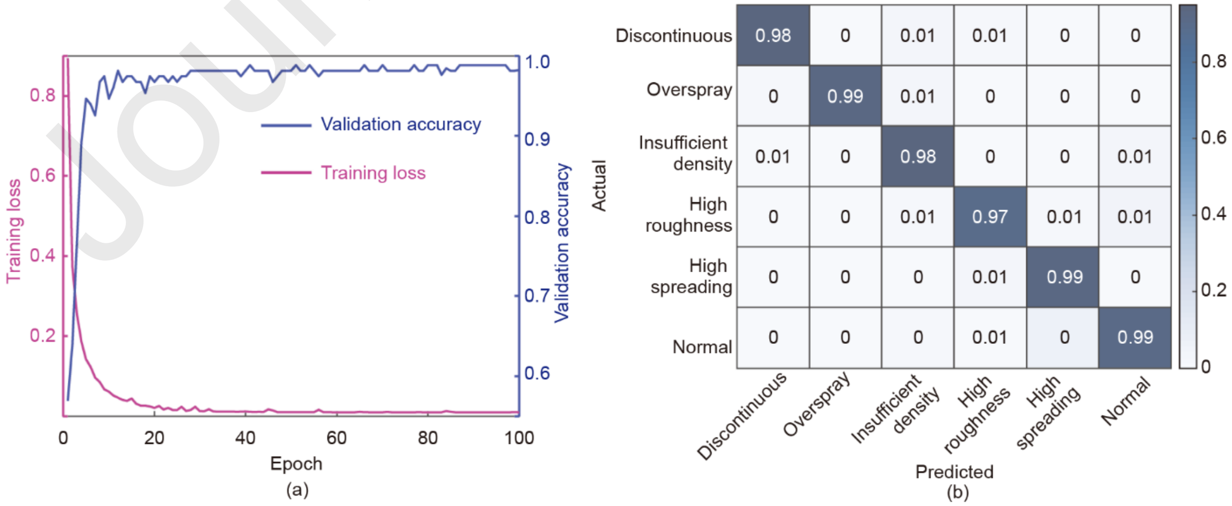
Specifically, ViT achieves superior classification accuracy through its use of self-attention mechanisms, which enable the model to capture global dependencies and long-range contextual relationships across the image. This contrasts with ResNet, which addresses the vanishing gradient problem through deep residual learning, allowing for the effective training of very deep networks. YOLOv8, on the other hand, is specifically designed for object detection, leveraging efficient convolutional layers, anchor boxes, and advanced upsampling techniques to provide in-process performance assessments and high accuracy in localizing objects within images. In contrast to CNN-based models that are constrained by local receptive fields, ViT is exceptionally proficient at recognizing intricate patterns and fine-grained details. Additionally, ViT reduces reliance on the inductive biases that are often inherent in CNN-based models such as ResNet and YOLO, allowing it to generalize better across diverse datasets. Given these advantages, ViT has demonstrated superior classification performance in this study compared to both the ResNet and YOLOv8 models.

Table 5

Comparison of modeling performance outcomes.

Model	Training		Testing	
	Class loss	Accuracy	Class loss	Accuracy
ResNet	0.259	0.931	0.291	0.926
YOLOv8	0.043	0.946	0.056	0.939
ViT	0.011	0.986	0.026	0.981

Building on the developed ViT model, Fig. 8(a) illustrates the detailed curves of the training loss and testing accuracy based on finely tuned hyperparameters, whereas Fig. 8(b) presents an extensive analysis of the testing accuracy across a range of printed line patterns. Generally, leveraging its innovative self-attention mechanisms, the ViT model exhibits strong overall performance, characterized by rapid training convergence in Fig. 8(a) and high-precision quality recognition in Fig. 8(b).



(a)

(b)

Fig. 8. ViT model performance evaluation. (a) Training loss and validation accuracy evolution across 100 epochs, and (b) confusion matrix for quality recognition in AJP.

To evaluate printing quality recognition across varying dataset sizes in AJP, the performance capabilities of ViT and the representative deep learning models using datasets of 100, 200, 600, and 800 samples were analyzed further, covering scenarios ranging from limited to moderate data availability. As shown in Table 6, with smaller datasets (100 and 200 samples), the ViT model—enhanced with DINOv2-extracted representations—significantly outperforms the YOLO and CNN models by leveraging robust unsupervised feature representations that effectively capture both local and universal patterns, thus generalizing effectively even with scant labeled data. In contrast, YOLOv8 and ResNet, which rely on localized feature extraction and require larger amounts of data to build hierarchical representations, struggle under these constraints. Although the performance gap narrows with moderate datasets (600 and 800 samples), the pre-trained DINOv2 continues to give ViT a competitive edge by enabling it to learn nuanced global relationships, while YOLO and CNN remain limited in their capacity to capture long-range dependencies without deeper architectures or more extensive data.

To assess the robustness of the developed model under realistic conditions, the impact of data noise on printing quality recognition was examined. Given that real-world data can be affected by lighting changes, sensor issues, environmental interference, and preprocessing errors, the dataset as shown in Fig. 9 was intentionally perturbed using random adjustments to brightness, contrast, saturation, hue, image mirroring, and rotations. The experimental results (Table 6) show that ViT consistently outperformed the ResNet and YOLOv8 models across these noisy scenarios. This performance advantage is largely due to ViT's global self-attention mechanism, which adeptly captures contextual cues and long-range dependencies by computing query-key-value representations for each image patch, thereby allowing the model to downplay noisy regions and focus on informative ones. Furthermore, ViT leverages DINOv2's strategy of learning invariant representations through diverse augmentations during training, which further reinforces its robustness against perturbations and enhances its reliability in unpredictable real-world scenarios.

Table 6

Performance comparison across varying dataset sizes and noise conditions.

Modeling accuracy	Impact of data scale				Impact of data noise			
	100	200	600	800	100	200	600	800
ViT	65.6%	69.3%	80.2%	85.1%	56.9%	62.3%	71.1%	79.6%
YOLOv8	52.9%	58.5%	73.9%	79.3%	39.6%	46.6%	59.6%	68.9%
ResNet	51.3%	57.3%	72.8%	78.6%	40.3%	46.1%	58.8%	68.3%

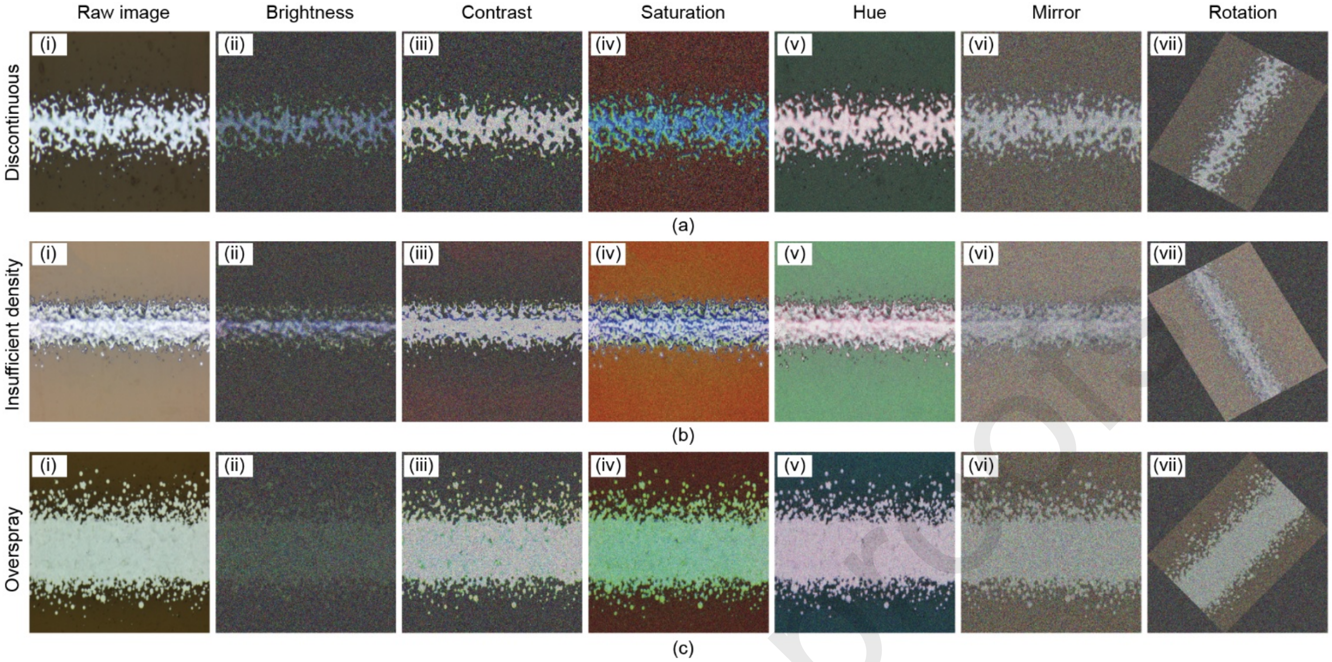


Fig. 9. Examples of operations applied to raw images to introduce artificial noise for a comparison of the robustness of ViT, ResNet and YOLOv8: (a) discontinuity, (b) insufficient density, and (c) overspray.

3.3.2. ViT feature extraction and comparative analysis

To explore the underlying factors contributing to the superior modeling performance, generalizability, and robustness of ViT over CNN-based models, the extracted features from both architectures were compared and analyzed. Specifically, Fig. 10 (ViT/DINOv2) illustrates a clear hierarchical evolution of feature representations—from the initial block that starts to capture abstract, high-level patterns, through the intermediate block that refines these into structured, class-discriminative features, to a final block where the features are highly globally integrated and semantically distilled, effectively representing each image as an invariant, high-level conceptual identifier (e.g., distinguishing various line types) with minimal spatial noise. In contrast, Fig. 11 (CNNs) demonstrates a progression more focused on local pattern extraction: the early layers capture fragmented edges and textures with significant noise, the intermediate layers enhance the spatial clarity and structure, and the final layers retain pronounced local details such as specific line segments or junctions while still preserving pixel-level fidelity, resulting in a representation that is less abstract and more sensitive to geometric variations.

Fundamentally, the self-attention mechanism of ViT enables early and efficient global context integration, allowing it to develop high-level, semantically distilled representations that are invariant to spatial noise and capture abstract patterns from the outset. This results in robust, transferable representations ideal for high-level semantic tasks. In contrast, CNNs rely on a convolutional strategy that gradually expands local receptive fields, building representations through a strategy that involves localized pattern extraction. While CNNs are adept at capturing precise spatial structures, they are more sensitive to geometric variations and less capable of abstracting pure semantic concepts. As a result, CNNs excel at capturing spatial details, but their representations are less abstract and robust, whereas ViT's approach leads to more refined, transferable features that enhance both the modeling precision and overall robustness in high-level semantic tasks.

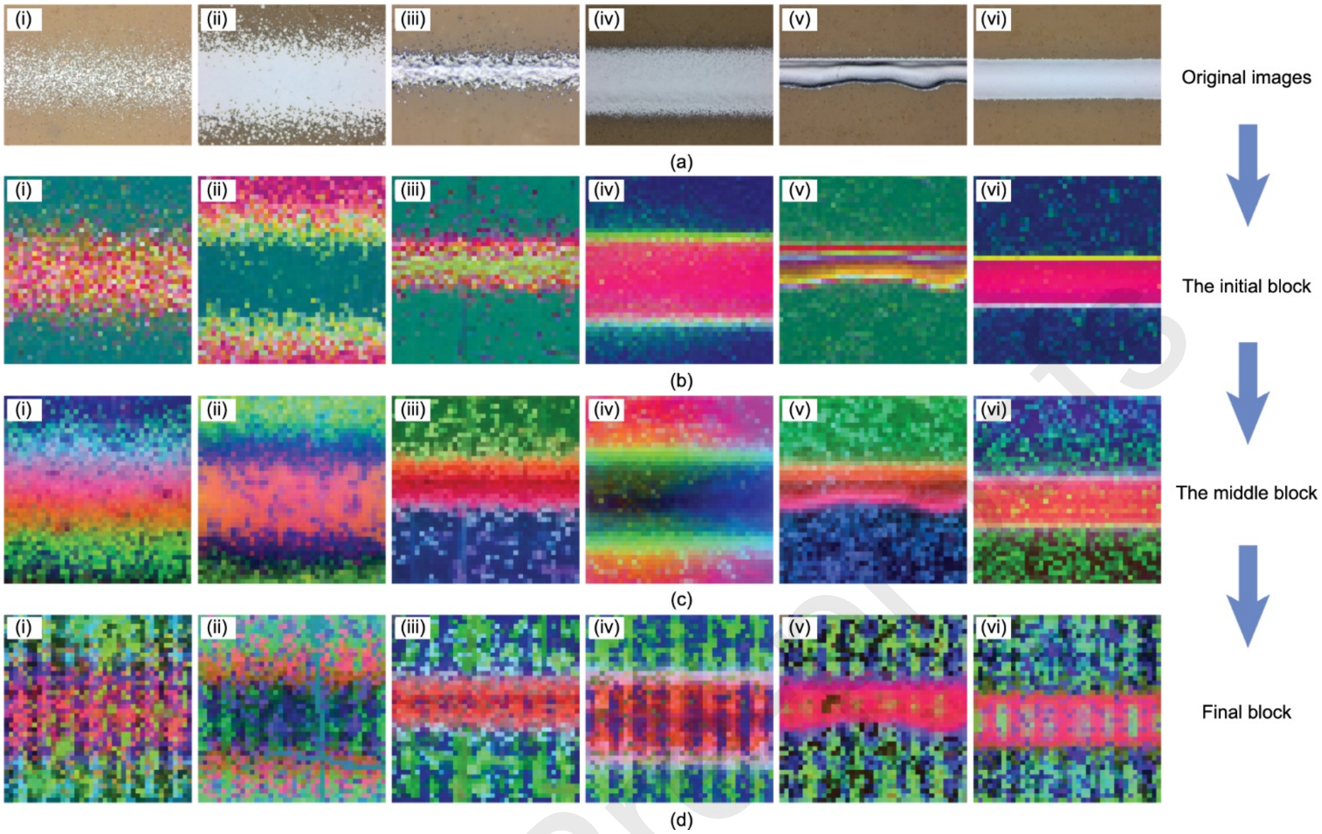


Fig. 10. Visualization of line feature extraction throughout various DINOv2 layers. The sequence includes (a-i)–(a-vi) the initial input images displaying printed lines, followed by the extracted features after (b-i)–(b-vi) the initial transformer block, (c-i)–(c-vi) the middle transformer block, and finally (d-i)–(d-vi) the last transformer block.

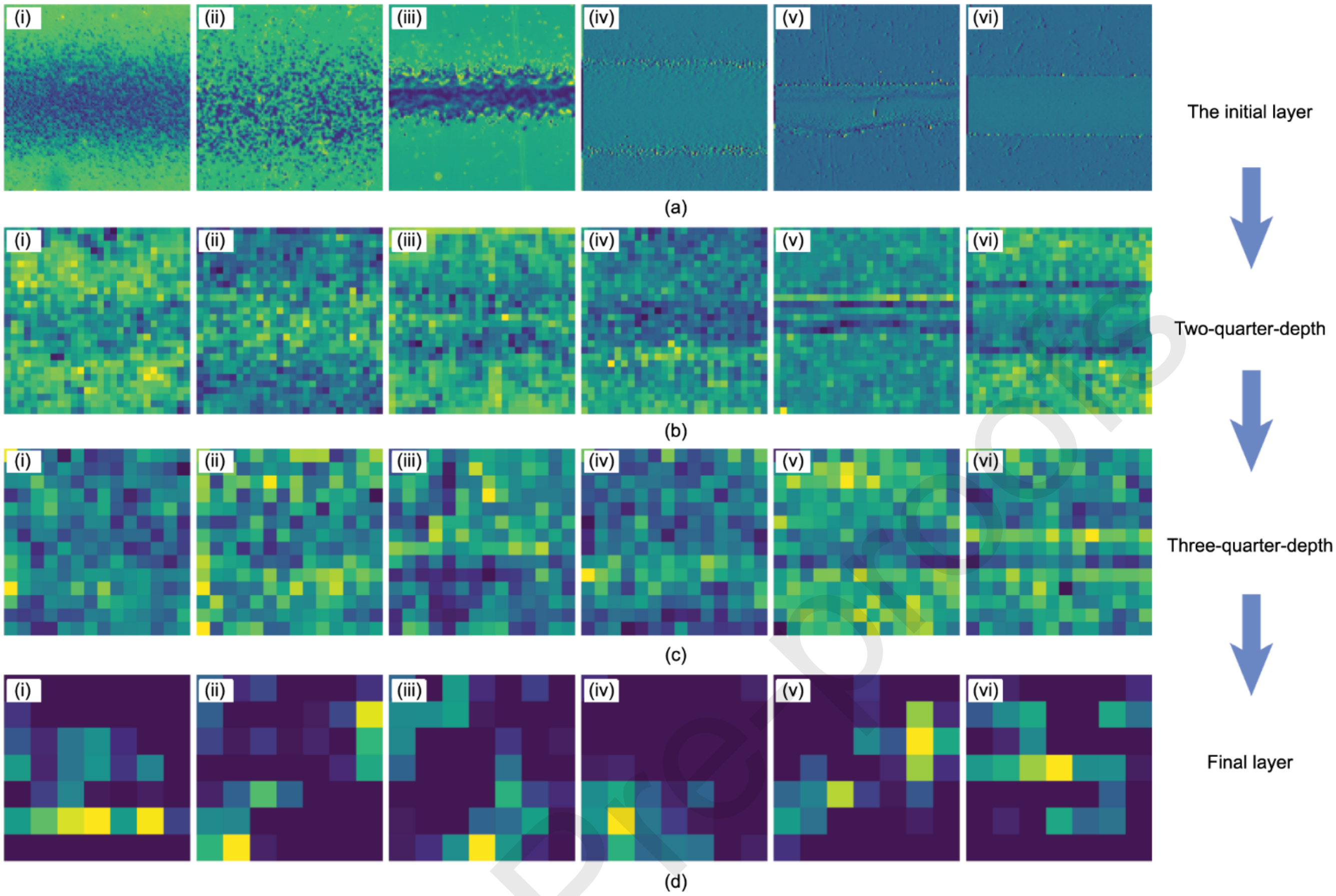


Fig. 11. Visualization of line feature extraction throughout various CNN layers. The sequence includes the extracted features after (a-i)–(a-vi) the initial convolutional layer, (b-i)–(b-vi) the two-quarter-depth convolutional layer, (c-i)–(c-vi) the three-quarter-depth convolutional layer, and (d-i)–(d-vi) the final convolutional layer.

3.3.3. Attention map analysis of ViT

Beyond achieving high classification accuracy, the significance of interpretability is essential when deploying learning-based quality recognition in practical manufacturing environments. In AJP, quality-related decisions often necessitate costly process interventions such as parameter adjustments, maintenance actions, or part qualification. Consequently, it is vital to verify that predictions are grounded in physically meaningful morphological features rather than being affected by spurious background patterns. Ensuring that the model's logic aligns with physical reality is fundamental to building the integrity required for industrial deployment.

Fig. 12 illustrates the progressive refinement of the model's attention across encoder layers, demonstrating its evolution from broad, generalized pattern recognition to focused, class-discriminative feature extraction. In the early layer (b-i)–(b-vi), the model exhibits broad attention patterns, focusing on large areas of the image that capture basic line structures, including potentially less relevant features. As the model progresses to the intermediate layer (c-i)–(c-vi), there is a significant shift towards more localized attention, with the model learning to suppress irrelevant background areas and less important line segments, while concentrating on smaller, more distinct patterns that are more discriminative for class identification. In the final layer (d-i)–(d-vi), the attention maps show highly precise localization, with the model focusing intensely on class-characteristic strokes, junctions, and micro-patterns. Notably, the attention maps at layers 9 and 16 are visually similar, indicating that by the intermediate layer, the model has already identified the key discriminative features, with the final layer further refining this focus. This progression illustrates the model's gradual refinement, starting with the recognition of broad patterns and progressively zeroing in on the most class-defining features, ultimately focusing on the critical regions essential for accurate classification.

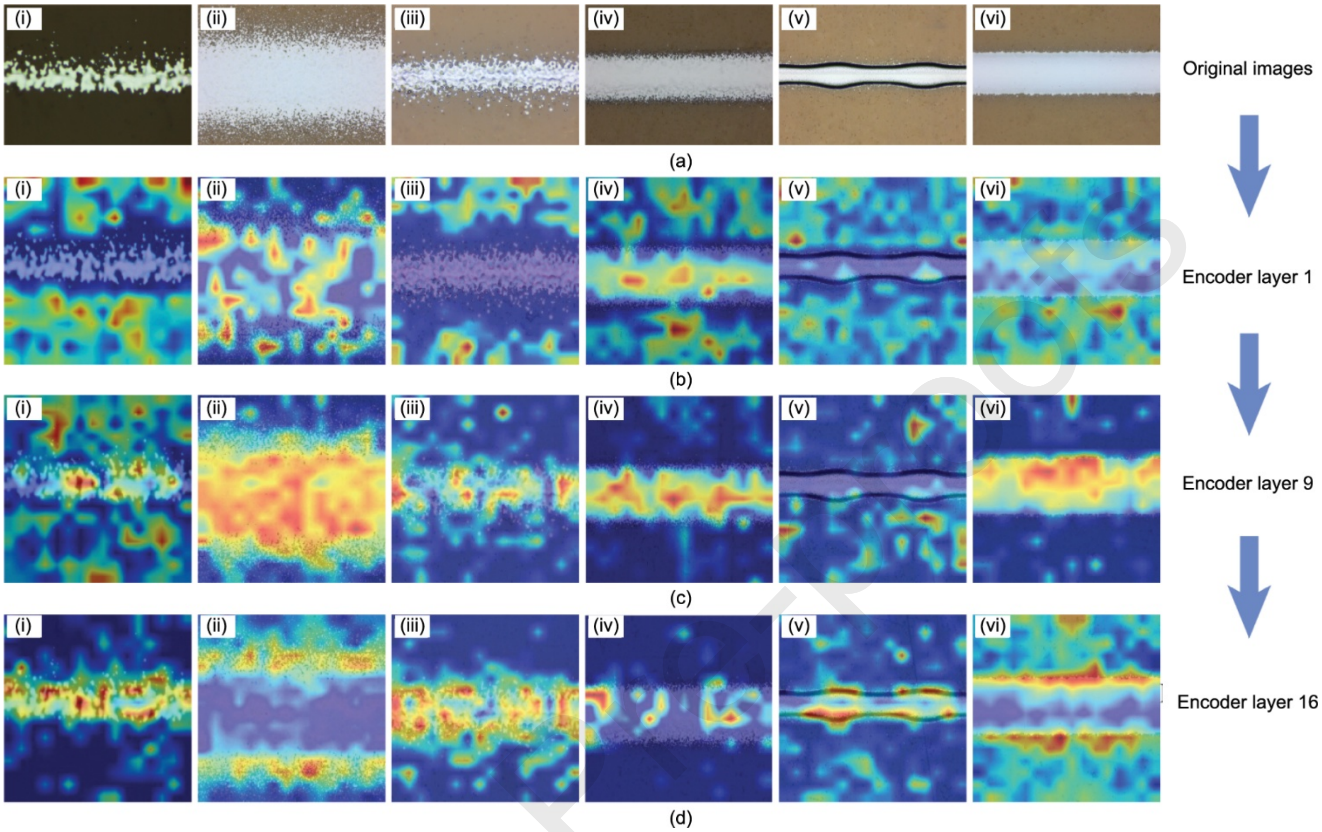


Fig. 12. Visualization of class attention maps across different encoder layers. (a-i)–(a-vi) Original input printed line images. Attention maps generated after (b-i)–(b-vi) the encoder layer 1, (c-i)–(c-vi) the encoder layer 9, and (d-i)–(d-vi) the encoder layer 16.

The functional feasibility and computational performance of the proposed framework were evaluated within a dynamic environment on a high-performance workstation equipped with an NVIDIA RTX 4090 GPU (NVIDIA Corporation, USA), utilizing in-process images acquired from operational AJP runs. Figs. 13(a)–(i) present representative pairs of captured images and their corresponding interpretable attention maps across various morphological states. The inclusion of multiple samples for representative states serves to illustrate the diagnostic consistency of the attention mechanism across diverse printing instances. These visualizations suggest that the framework primarily aligns with characteristic morphological signatures, thereby mitigating the potential influence of incidental image noise.

Fig. 13(j) presents the sequential classification results, illustrating the framework's capacity to track morphological variations over a continuous period. The general agreement between the predicted and true classes indicates the functional stability of the framework for in-process assessment scenarios. These observations are further supported by a performance evaluation indicating that the framework yields an average per-frame processing time of 16.8 ms, which is expected to support a throughput rate of 30 FPS with a model memory footprint of 652 MB. Such performance is facilitated by a streamlined workflow incorporating ROI extraction and resizing to the 224 × 224 input resolution, allowing the framework to capture key printing signatures while meeting the operational requirements for in-process quality assessments.

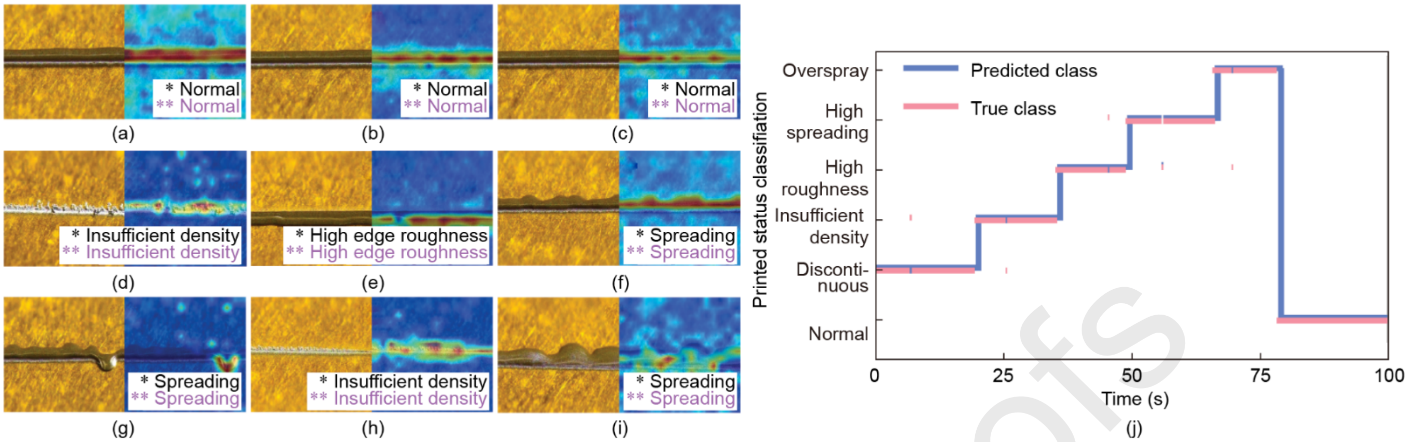


Fig. 13. In-process quality assessment and sequential classification results. (a–i) Representative pairs of captured images and their corresponding interpretable attention maps across various morphological states; asterisks (*) and double asterisks (**) denote the true and predicted classes within each image, respectively. (j) Sequential classification results comparing the predicted class (blue line) against the true class (red line) over a continuous period.

4. Conclusion

This study presents a machine learning-enhanced ViT framework, leveraging attention-based architectures inspired by LLMs to address critical challenges in AJP quality assessments. By integrating a coaxial camera system with a GMM for automated dataset generation and a BNN for uncertainty-aware modeling, the proposed system establishes a cohesive pipeline for in-process quality assessments. The results indicate that the ViT model, fine-tuned with self-supervised DINOv2 representations, exhibits enhanced accuracy and robustness compared to conventional CNN-based architectures, in this case ResNet and YOLOv8, offering a scalable and efficient solution for real-world manufacturing environments.

Despite these advancements, the current work serves as an algorithmic validation, identifying several directions for subsequent research. Future efforts can prioritize systematic ablation studies to isolate the individual contributions of representation learning, automated data generation, and morphology-based feature capturing. Moreover, to strengthen the cross-domain robustness of the architecture, the framework will be evaluated on representative public defect datasets and compared against an expanded range of baseline model families. These evaluations are intended to verify that the recognition capabilities remain robust across various imaging-perspectives and computational constraints.

Another priority involves broadening the material and physical scope of the framework to accommodate a wider variety of ink formulations and compositions. Future investigations should aim to quantify the sensitivity of aerosol stability to solids loading and correlate particle-size statistics with the electrical performance and sintering behavior of the printed lines. Furthermore, incorporating quantitative contact-angle analysis and addressing clogging-related operational states will provide a more comprehensive understanding of ink-substrate interactions. These enhancements are expected to improve the practical applicability of the framework under significant material variability and complex fluid dynamics.

Finally, the transition toward full hardware-integrated closed-loop deployment remains the primary objective for industrial implementation. This will require characterizing end-to-end latency under continuous data acquisition and validating the BNN-driven decision-making strategies with industrial controllers. While the current study successfully validates the functional potential of the integrated machine vision and ViT-based framework, extending this system to additional nozzle diameters and diverse AM platforms will facilitate the development of a unified, intelligent quality assurance solution for next-generation printing technologies.

Acknowledgments

This work was supported by the InnoCORE program of the Ministry of Science and ICT (N10260002). This research was also supported by the Technology Innovation Program (No. RS-2024-00441774) and the Industrial Technology Alchemist Project (No. 20025702) funded by the Ministry of Trade, Industry, and Energy (MOTIE, Korea). Additionally, this research was partially supported by the Start-up Fund for RAPs under the Strategic Hiring Scheme of PolyU (No. P0054859) and the National Natural Science Foundation of China (No. 72501245).

Author Contributions

Conceptualization, H.Z.; methodology, H.Z., Y.Y. and J.P.C.; software, D.S., Y.Y. and H.Z.; validation, Y.K. D.S.; formal analysis, X.L.; investigation, X.L., Y.K. and D.S.; resources, X.L.; data curation, D.S., N.L. and Y.K; writing—original draft preparation, H.Z. and N.L.; writing—review and editing, J.P.C. ,Y.J.Y. and S.K.M.; visualization, H.Z. and Y.K.; supervision, J.P.C., Y.J.Y. and S.K.M.; project administration, J.P.C.; funding acquisition, J.P.C. and H.Z. All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability statement

Data will be made available on request.

[1] Zeng P, Tian B, Tian Q, Yao W, Li M, Wang H, et al. Screen-printed, low-cost, and patterned flexible heater based on Ag fractal dendrites for human wearable application. Adv Mater Technol 2019;4(3):1800453.

[2] He P, Cao J, Ding H, Liu C, Neilson J, Li Z, et al. Screen-printing of a highly conductive graphene ink for flexible printed electronics. ACS Appl Mater Interfaces 2019;11(35):32225–34.

[3] Xiong J, Chen J, Chen W, Yue X, Zhao Z, Yin Z. Intelligent path planning algorithm system for printed display manufacturing using graph convolutional neural network and reinforcement learning. J Manuf Syst 2025;79:73–85.

[4] Zhang Z, Yang H, Chen J, Yin Z. Multi-scale conditional diffusion model for deposited droplet volume measurement in inkjet printing manufacturing. J Manuf Syst 2023;71:595–608.

[5] Yue X, Chen J, Li Y, Li X, Zhu H, Yin Z. Intelligent control system for droplet volume in inkjet printing based on stochastic state transition soft actor–critic DRL algorithm. J Manuf Syst 2023;68:455–64.

[6] Elsersawy R, Rahman A, Sakib-Uz-Zaman C, Khondoker MAH. Multifunctional inks in aerosol jet printing: performance, challenges, and applications. Front Manuf Technol. 2025;5:1558209.

[7] Fisher C, Skolrood LN, Li K, Joshi PC, Aytug T. Aerosol-jet printed sensors for environmental, safety, and health monitoring: a review. Adv Mater Technol 2023;8(15):2300030.

[8] Hines DR, Gu Y, Martin AA, Li P, Fleischer J, Clough-Paez A, et al. Considerations of aerosol-jet printing for the fabrication of printed hybrid electronic circuits. Addit Manuf 2021;47:102325.

[9] Guo Z, Yu P, Li B, Xuan F, Zhao J. Direct wire writing technique benefitting the flexible electronics. Virtual Phys Prototyp 2024;19(1):e2286514.

[10] Lim HR, Kim HS, Qazi R, Kwon YT, Jeong JW, Yeo WH. Advanced soft materials, sensor integrations, and applications of wearable flexible hybrid electronics in healthcare, energy, and environment. Adv Mater 2020;32(15):1901924.

[11] Agarwala S, Goh GL, Dinh Le TS, An J, Peh ZK, Yeong WY, et al. Wearable bandage-based strain sensor for home healthcare: combining 3D aerosol jet printing and laser sintering. ACS Sens 2019;4(1):218–26.

[12] Han T, Kundu S, Nag A, Xu Y. 3D printed sensors for biomedical applications: a review. Sensors 2019;19(7):1706.

[13] Huang B, Wu S, Liu J, Liu J, Peng B, Zhou Z. Aerosol jet printing of polyelectrolyte-modified MXene ink for a multifunctional humidity and temperature flexible sensor. Chem Eng J 2025;519:165403.

[14] Karipoth P, Chandler JH, Lee J, Taccola S, Macdonald J, Valdastri P, et al. Aerosol jet printing of strain sensors for soft robotics. Adv Eng Mater 2024;26(1):2301275.

[15] Khan Y, Garg M, Gui Q, Schadt M, Gaikwad A, Han D, et al. Flexible hybrid electronics: direct interfacing of soft and hard electronics for wearable health monitoring. Adv Funct Mater 2016;26(47):8764–75.

[16] Serpelloni M, Cantù E, Borghetti M, Sardini E. Printed smart devices on cellulose-based materials by means of aerosol-jet printing and photonic curing. Sensors 2020;20(3):841.

[17] Min SH, Lee GY, Ahn SH. Direct printing of highly sensitive, stretchable, and durable strain sensor based on silver nanoparticles/multi-walled carbon nanotubes composites. Compos Part B Eng 2019;161:395–401.

[18] Seifert T, Sowade E, Roscher F, Wiemer M, Gessner T, Baumann RR. Additive manufacturing technologies compared: morphology of deposits of silver ink using inkjet and aerosol jet printing. Ind Eng Chem Res 2015;54(2):769–79.

[19] Mahajan A, Frisbie CD, Francis LF. Optimization of aerosol jet printing for high-resolution, high-aspect ratio silver lines. ACS Appl Mater Interfaces 2013;5(11):4856–64.

[20] Zhang H, Moon SK, Ngo TH. 3D printed electronics of Non-contact ink writing techniques: status and promise. Int J Precis Eng Manuf-Green Technol 2020;7(2):511–24.

[21] Chen G, Gu Y, Tsang H, Hines DR, Das S. The effect of droplet sizes on overspray in aerosol-jet printing. Adv Eng Mater 2018;20(8):1701084.

[22] Zhang H, Huang J, Zhang X, Wong CN. Uncertainty quantification of aerosol jet 3D printing process using non-intrusive polynomial chaos and stochastic collocation. Adv Eng Inform 2025;65(Part A):103175.

[23] Liu Z, Liu Y, He L, Cui L, Liang N, Choi JP, et al. A comprehensive investigation of process parameters and material properties effects on printed line quality of aerosol jet printing based on coupled three-dimensional numerical models. Int J Precis Eng Manuf-Green Technol 2024;11:727–42.

[24] Yoo D, Mahoney CM, Deneault JR, Grabowski C, Austin D, Berrigan JD, et al. Mapping drift in morphology and electrical performance in aerosol jet printing. Prog Addit Manuf 2021;6(2):257–68.

[25] Smith M, Choi YS, Boughey C, Kar-Narayan S. Controlling and assessing the quality of aerosol jet printed features for large area and flexible electronics. Flex Print Electron 2017;2(1):015004.

[26] Zhang H, Choi JP, Moon SK, Ngo TH. A hybrid multi-objective optimization of aerosol jet printing process via response surface methodology. Addit Manuf 2020;33:101096.

[27] Secor EB. Guided ink and process design for aerosol jet printing based on annular drying effects. Flex Print Electron 2018;3(3):035007.

[28] Zhang H, Choi JP, Moon SK, Ngo TH. A multi-objective optimization framework for aerosol jet customized line width printing via small data set and prediction uncertainty. J Mater Process Technol 2020;285:116779.

[29] Secor EB. Light scattering measurements to support real-time monitoring and closed-loop control of aerosol jet printing. Addit Manuf 2021;44:102028.

[30] Divakaran N, Das JP, PV AK, Mohanty S, Ramadoss A, Nayak SK. Comprehensive review on various additive manufacturing techniques and its implementation in electronic devices. J Manuf Syst 2022;62:477–502.

[31] Kim S, Kim EH, Lee W, Sim M, Kim I, Noh J, et al. Real-time in-process control methods of process parameters for additive manufacturing. J Manuf Syst 2024;74:1067–90.

[32] Feng JQ. Multiphase flow analysis of mist transport behavior in aerosol jet® system. Int J Comput Methods Exp Meas 2017;6(1):23–34.

[33] Salary R, Lombardi JP, Samie Tootooni M, Donovan R, Rao PK, Borgesen P, et al. Computational fluid dynamics modeling and online monitoring of aerosol jet printing process. J Manuf Sci Eng 2017;139(2):021015.

[34] Salary R, Lombardi JP, Weerawarne DL, Rao P, Poliks MD. A computational fluid dynamics investigation of pneumatic atomization, aerosol transport, and deposition in aerosol jet printing process. J Micro Nano-Manuf 2021;9(1):010903.

[35] Zhang H, Xu H, Cui L, Pan Z, Lee PH, Jung MK, et al. An extensive study of the influence of key flow variables on printed line quality outcomes during aerosol jet printing using coupled three-dimensional numerical models. Materials 2024;17(13):3179.

[36] Ramesh S, Mahajan C, Gerdes S, Gaikwad A, Rao P, Cormier DR, et al. Numerical and experimental investigation of aerosol jet printing. Addit Manuf 2022;59(Part A):103090.

[37] Zhang H, Liu Z, Yin S, Xu H. A hybrid multi-objective optimization of functional ink composition for aerosol jet 3D printing via mixture design and response surface methodology. Sci Rep 2023;13(1):2513.

[38] Akhatov IS, Hoey JM, Swenson OF, Schulz DL. Aerosol focusing in micro-capillaries: theory and experiment. J Aerosol Sci 2008;39(8):691–709.

[39] Li M, Liu Z, Yin S, Choi JP, Zhang H. Comparison and identification of optimal machine learning model for rapid optimization of printed line characteristics of aerosol jet printing technology. Int J Precis Eng Manuf-Green Technol 2024;11(1):71–87.

[40] Zhang H, Choi JP, Moon SK, Ngo TH. A knowledge transfer framework to support rapid process modeling in aerosol jet printing. Adv Eng Inform 2021;48:101264.

[41] Zhang H, Hong E, Chen X, Liu Z. Machine learning enables process optimization of aerosol jet 3D printing based on the droplet morphology. ACS Appl Mater Interfaces 2023;15(11):14532–45.

[42] Zhang H, Moon SK. Reviews on machine learning approaches for process optimization in noncontact direct ink writing. ACS Appl Mater Interfaces 2021;13(45):53323–45.

[43] Choi E, An K, Kang KT. Deep-learning-based microfluidic droplet classification for multijet monitoring. ACS Appl Mater Interfaces 2022;14(13):15576–86.

[44] Choi E, Choi S, An K, Kang KT. Deep learning-based inkjet droplet detection for jetting characterizations and multijet synchronization. ACS Appl Mater Interfaces 2024;16(14):18040–51.

[45] Yao C, Wang L, Wang Q, Liu Z, Liu G, Zhang M. Deep-learning-guided evaluation method for the high-volume preparation of flexible sensors based on inkjet printing. ACS Appl Mater Interfaces 2024;16(10):13326–34.

[46] Zhang H, Huang J, Zhang X, Wong CN. Autonomous optimization of process parameters and *in-situ* anomaly detection in aerosol jet printing by an integrated machine learning approach. Addit Manuf 2024;86:104208.

[47] Liang X, Wang Z, Liu J. A survey of large language model-augmented knowledge graphs for advanced complex product design. J Manuf Syst 2025;80:883–901.

[48] Jeon J, Sim Y, Lee H, Han C, Yun D, Kim E, et al. ChatCNC: conversational machine monitoring via large language model and real-time data retrieval augmented generation. J Manuf Syst 2025;79:504–14.

[49] Shi D, Liedl P, Bauernhansl T. Interoperable information modelling leveraging asset administration shell and large language model for quality control toward zero defect manufacturing. J Manuf Syst 2024;77:678–96.

[50] Wu T, Li J, Bao J, Liu Q. ProcessCarbonAgent: a large language models-empowered autonomous agent for decision-making in manufacturing carbon emission management. J Manuf Syst 2024;76:429–42.

[51] Fan H, Liu C, Janvisloo NE, Bian S, Fuh JYH, Lu WF, et al. MaViLa: unlocking new potentials in smart manufacturing through vision language models. J Manuf Syst 2025;80:258–71.

[52] Wang Z, Chen H, Zhong Q, Lin S, Wu J, Xu M, et al. Recognition of penetration state in GTAW based on vision transformer using weld pool image. Int J Adv Manuf Technol 2022;119(7–8):5439–52.

[53] Wei H, Zhao L, Li R, Zhang M. RFAConv-CBM-ViT: enhanced vision transformer for metal surface defect detection. J Supercomput 2025;81(1):155.

[51] Zhang J, Chen X, Wang Y, Wang C, Liu Y, Li X, et al. Exploring plain ViT reconstruction for multi-class unsupervised anomaly detection. 2024. arXiv:2312.07495.

[55] Yang Q, Guo R. An unsupervised method for industrial image anomaly detection with vision transformer-based autoencoder. Sensors 2024;24(8):2440.

[56] Liu Z, Li M, Weng Y, Wong TN, Tan MJ. Mixture design approach to optimize the rheological properties of the material used in 3D cementitious material printing. Constr Build Mater 2019;198:245–55.

[57] Choi Y, Song D, Yoon S, Koo J. Comparison of factorial and latin hypercube sampling designs for meta-models of building heating and cooling loads. Energies 2021;14(2):512.

[58] Yang MS, Lai CY, Lin CY. A robust EM clustering algorithm for gaussian mixture models. Pattern Recognit 2012;45(11):3950–61.

[59] Franchi G, Bursuc A, Aldea E, Dubuisson S, Bloch I. Encoding the latent posterior of bayesian neural networks for uncertainty quantification. IEEE Trans Pattern Anal Mach Intell 2024;46(4):2027–40.

[60] Oquab M, Darcet T, Moutakanni T, Vo H, Szafraniec M, Khalidov V, et al. DINOv2: learning robust visual features without supervision. 2024. arXiv:2304.07193.

[61] Khan S, Naseer M, Hayat M, Zamir SW, Khan FS, Shah M. Transformers in vision: a survey. ACM Comput Surv 2022;54(10s):1–41.

[62] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. 2023. arXiv:1706.03762.

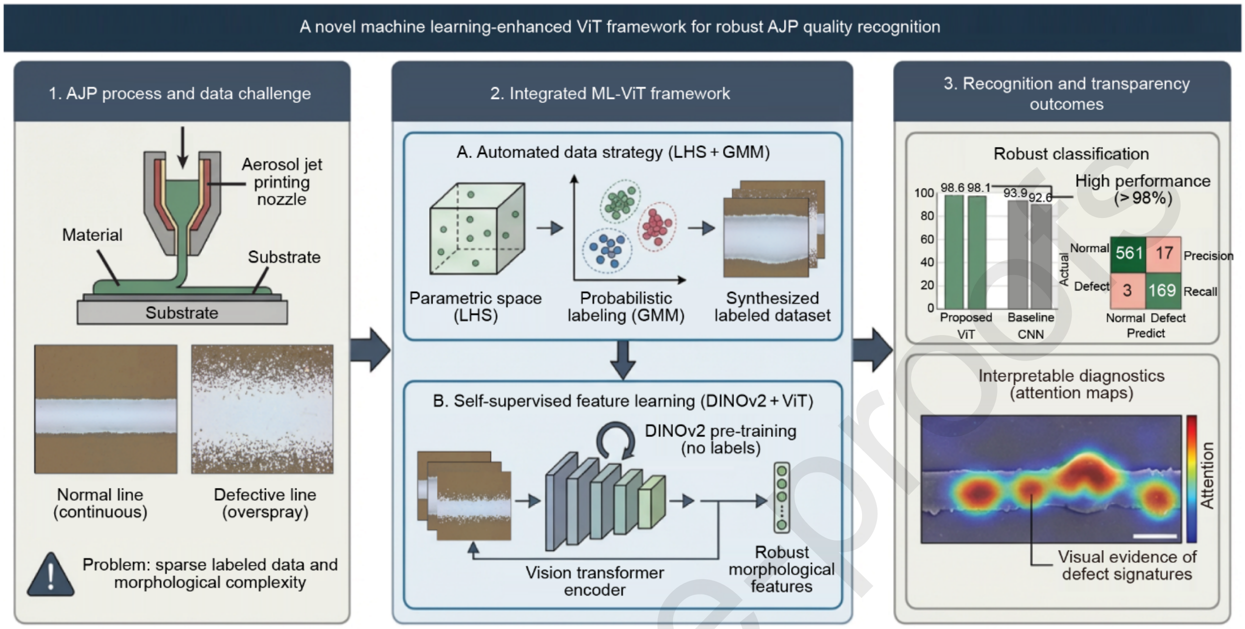
[63] Wu Z, Shen C, Van Den Hengel A. Wider or deeper: revisiting the ResNet model for visual recognition. Pattern Recognit 2019;90:119–33.

[64] Krichen M. Convolutional neural networks: a survey. Computers 2023;12(8):151.

[65] Hussain M. YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. Machines 2023;11(7):677.

[66] Terven J, Córdova-Esparza DM, Romero-González JA. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS. Mach Learn Knowl Extr 2023;5(4):1680–716.

Graphical abstract



Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.