EXPLAINABLE ARTIFICIAL INTELLIGENCE IN ADDITIVE MANUFACTURING: A SYSTEMATIC REVIEW ON METHOD CONVERGENCE AND ASSESSMENT OF STANDARDIZATION GAPS

QUOC THANH NGUYEN[1], THANH-CONG TRUONG[1*], THANH PHONG NGUYEN[2], THANH-NAM TRAN[3], JAKUB MESICEK[4], MAREK PAGAC[4]

[1]Faculty of Data Science, University of Finance - Marketing, No. 778 Nguyen Kiem, Ho Chi Minh, 70000, Viet Nam

[2] Ho Chi Minh City Technical and Economic College, No 215 Nguyen Van Luong, Ho Chi Minh, 70000, Viet Nam

[3] Data Science Laboratory, Faculty of Information Technology, Ton Duc Thang University, No. 19 Nguyen Huu Tho, Ho Chi Minh, 70000, Viet Nam

[4] Department of Machining, Assembly and Engineering Metrology, Faculty of Mechanical Engineering, VSB-Technical University of Ostrava, 70833, Czech Republic

DOI: 10.17973/MMSJ.2025_10_2025095

e-mail to corresponding author: ttcong@ufm.edu.vn

ABSTRACT

Additive manufacturing has revolutionized production capabilities across industries, yet quality assurance and process optimization remain significant challenges due to complex, multi-parameter interactions. Explainable Artificial Intelligence (XAI) offers potential solutions by providing interpretable insights into manufacturing processes, though its systematic application remains fragmented. Nevertheless, a comprehensive understanding of integration patterns, effectiveness metrics, and implementation barriers remains limited. Following PRISMA guidelines, this systematic review searched Scopus, Web of Science, IEEE Xplore, and ScienceDirect databases for studies published from inception to 2025. From 211 initial records, 38 peer-reviewed studies met inclusion criteria after screening and quality assessment. Results reveal that while XAI achieves high predictive performance, critical interpretability standardization gaps hinder industrial deployment. SHAP dominates applications (58% adoption), with quality control representing 39% of studies. Regression tasks achieve $R^2 > 0.90$ in 76% of cases, and classification tasks report >95% accuracy in 71% of cases. However, only 21% of studies provide quantitative interpretability assessment. These findings establish a foundation for developing standardized XAI evaluation frameworks in manufacturing contexts. Ensemble methods and physics-informed approaches offer the most promising pathways for achieving both high performance and mechanistic interpretability in safety-critical manufacturing environments.

KEYWORDS

Explainable Artificial Intelligence, XAI, Additive Manufacturing, 3D Printing, Interpretable Machine Learning, Transparency in AI

1 INTRODUCTION

Explainable Artificial Intelligence (XAI) has become a critical enabler for trustworthy AI deployment across high-stakes domains, providing transparency and interpretability to complex machine learning systems. In manufacturing contexts, where safety, quality, and regulatory compliance are essential, understanding and validating AI decision-making processes has become necessary for practical implementation [Ma 2024; Truong 2025]. Additive Manufacturing (AM), one of the most data-intensive and process-sensitive manufacturing paradigms, generates vast quantities of sensor data that can benefit from AI-driven analysis, yet requires interpretable insights for operational acceptance and regulatory approval [Ukwaththa 2024].

Recent comprehensive reviews have identified three critical systematic gaps in XAI-AM integration that fundamentally constrain industrial adoption: (1) the absence of standardized interpretability assessment frameworks across diverse AM applications, (2) insufficient integration between real-time manufacturing constraints and XAI computational overhead, and (3) limited human-centered evaluation methodologies for manufacturing practitioners [Ukwaththa 2024] methods in additive manufacturing. These gaps explicitly represent the primary barriers preventing the transition from technically feasible XAI implementations to widespread industrial deployment, despite demonstrated accuracy improvements exceeding 95% in controlled environments.

The convergence of XAI and AM has gained unprecedented momentum since 2023, driven by what recent literature characterizes as *the explainability revolution* in smart manufacturing systems [Abhilash 2024]. This paradigm shift represents a fundamental transition from Industry 4.0's focus on automation toward Industry 5.0's emphasis on human-AI collaboration and trustworthy manufacturing systems. The year 2024 marked a critical inflection point, with Kharate et al. demonstrating successful integration of SHapley Additive exPlanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME), and Partial Dependence Plot (PDP) techniques in FDM-based biocomposite manufacturing, achieving R^2 values exceeding 0.95 for mechanical property prediction [Kharate 2024], while systematic reviews of XAI in process engineering revealed that current applications remain predominantly exploratory due to restricted access to large, reliable datasets [Di Bonito 2024].

Recent advances in model-agnostic explanation methods have transcended the traditional SHAP-LIME paradigm, with comparative analyses revealing critical limitations related to model-dependency and feature collinearity that affect interpretation reliability [Salih 2025]. The emergence of physics-informed XAI approaches [Arinez 2020] and ontology-based explainable AI frameworks [Dolgui 2024] represents a paradigm shift toward mechanistically grounded interpretability that addresses the causality limitations observed in 39% of current implementations.

Despite growing academic interest, recent surveys indicate that XAI integration in manufacturing cyber-physical systems remains predominantly in proof-of-concept stages, with limited industrial deployment due to computational overhead constraints and unclear ROI justification [Moosavi 2024]. The most significant barrier lies in the absence of comprehensive evidence synthesis that could bridge the gap between theoretical XAI capabilities and practical manufacturing requirements. Recent semiconductor manufacturing case studies demonstrate that XAI-enhanced quality control can achieve substantial process improvements [Senoner 2022], yet systematic understanding of scalability, generalizability, and cross-domain applicability remains fragmented.

This systematic review addresses these critical gaps by providing the first comprehensive analysis of XAI integration in additive manufacturing, examining 38 peer-reviewed studies published between 2021 and 2025. Our analysis reveals that 92.1% of publications occurred during 2023-2025, indicating rapid field maturation, while simultaneously exposing systematic methodological deficiencies where only 21% of studies provide quantitative interpretability assessment. The convergence toward SHAP-based approaches (58% adoption rate) reflects practical consensus, yet masks underlying diversity in application contexts and evaluation methodologies that compromise cross-study comparability.

Our primary objectives are to systematically map the current landscape of XAI methods applied to AM applications, evaluate the effectiveness of different XAI approaches across diverse manufacturing contexts, identify methodological patterns and performance benchmarks, and determine critical research gaps and future research priorities that address the identified systematic barriers to industrial XAI deployment.

2 METHODOLOGY

2.1 Search Strategy

This systematic review traced XAI development in AM from the early stages through 2025, capturing the critical transition from proof-of-concepts to practical implementation. We conducted comprehensive searches across four major academic databases: Scopus, Web of Science, IEEE Xplore, and ScienceDirect. The standardized search string applied across all databases was: ("explainable AI" or "interpretable AI" or “XAI" or "explainable artificial intelligence" or "transparent AI") and ("additive manufacturing" or "3D printing" or "rapid prototyping" or "digital fabrication"). Initial retrieval yielded 211 records: Scopus (99), Web of Science (58), IEEE Xplore (28), ScienceDirect (26).

2.2 Selection Criteria and Screening

The inclusion criteria required studies to: (i) be peer-reviewed and published in English, (ii) present a clear implementation of explainable artificial intelligence (XAI) techniques applied specifically to additive manufacturing (AM), and (iii) include quantitative results derived from empirical experimentation or simulation, with sufficient methodological detail to support comparative analysis.

The exclusion criteria were defined as follows: (i) non-peer-reviewed or grey literature (e.g., preprints, theses, whitepapers); (ii) studies addressing general artificial intelligence in AM without any XAI component; (iii) papers lacking empirical validation (e.g., conceptual frameworks or position papers); and (iv) duplicate entries across databases.

A two-stage screening process was employed. In the first stage, two independent reviewers screened titles and abstracts based on the defined criteria. In the second stage, full-text reviews were conducted for shortlisted articles. Inter-rater reliability was assessed using Cohen’s κ, achieving a value of 0.87, indicating substantial agreement. Discrepancies were resolved through consensus discussion. The final corpus comprised 38 studies that met all eligibility conditions.

.2.3 Data Extraction and Quality Assessment

We conducted quality assessment through a differentiated evaluation approach that recognized the distinct characteristics and contributions of journal articles and conference papers to the field. Journal articles were evaluated using six criteria, each scored from 0-5 points for a total of 30 points: research methodology, theoretical foundation, data analysis and validation, result interpretation, impact and implications, and overall quality of presentation. Conference papers were evaluated using five criteria totaling 30 points: technological novelty (0-7 points), implementation rigor (0-7 points), experimental validation (0-6 points), clarity of results presentation (0-5 points), and prospects for future research (0-5 points).

To reduce potential subjective bias and enhance reliability, multiple safeguards were employed. Each paper was independently assessed by two reviewers following a standardized evaluation framework. Inter-rater reliability was quantified using Cohen’s Kappa coefficient, yielding a value of 0.82. Any discrepancies were addressed through structured discussions aimed at reaching consensus. Building upon the defined evaluation criteria, the adapted framework accounted for the specific characteristics of conference publications. While such works often present more concise methodological sections, they frequently deliver substantial technological innovations and detailed implementation insights, elements of particular relevance to the advancement of XAI applications in AM. To ensure consistency, the maximum attainable score was standardized at 30 points across both conference and journal publications, reflecting their respective strengths. Based on the total score, publications were classified into three quality tiers: superior (24-30 points), moderate (16-23 points), and inferior (below 16 points), irrespective of format. Nonetheless, certain methodological limitations were acknowledged. The expertise and professional background of reviewers could influence the weighting of specific quality dimensions, and the rapid evolution of XAI technologies may, over time, reshape perceptions of quality.

3 RESULTS

3.1 Study Selection Process

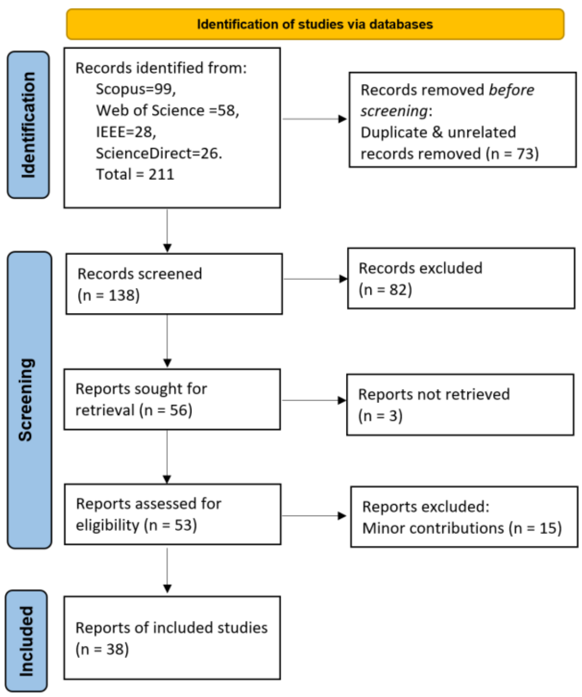

Figure 1. PRISMA flow diagram of study selection process

A total of 211 records were retrieved from four electronic databases during our initial search: Scopus (n=99), Web of Science (n=58), IEEE Xplore (n=28), and ScienceDirect (n=26). This systematic search and screening methodology was thoroughly documented following PRISMA guidelines (Fig. 1).

The screening process proceeded through several stages following systematic review protocols to ensure comprehensive and transparent study selection:

- Identification: Initial database searches across multiple platforms yielded records that underwent preliminary filtering. After removing 73 duplicates and clearly unrelated records through automated deduplication tools and manual verification, 138 unique records remained for detailed screening assessment.
- Screening & Eligibility: From the remaining 138 records, systematic title and abstract review was conducted by independent reviewers applying predetermined inclusion and exclusion criteria. This process resulted in the exclusion of 82 records based on established criteria such as lack of empirical XAI implementation, absence of additive manufacturing application focus, and insufficient methodological detail for systematic analysis. Inter-reviewer disagreements were resolved through discussion and consensus.
- Full-text assessment: Of the 56 articles identified for full-text retrieval and detailed evaluation, three were inaccessible despite direct author contact attempts and institutional library searches, leaving 53 articles for comprehensive assessment.
- Final inclusion: After evaluating 53 full-text articles against the complete inclusion criteria, 15 were excluded due to insufficient XAI validation methodology, lack of quantitative results suitable for systematic synthesis, or peripheral relevance to the specific research questions. This rigorous selection process resulted in 38 studies meeting all inclusion criteria and providing adequate detail for comprehensive systematic analysis.

3.2 Overview

A total of 38 peer-reviewed studies published between 2021 and 2025 were included in the review. Among them, 87% were journal articles and 13% were conference proceedings. The dominant AI application type was regression (68%), followed by classification (24%) and other tasks such as inverse design or anomaly detection (8%). Publication frequency peaked in 2023 and 2024, each accounting for 39% of the total studies, while 2025 contributed 18%, indicating increased interest in XAI within AM in recent years. Publication trends show accelerating growth, with 2023 and 2024 marking peak year (39% of total publications each year), followed by continued momentum in 2025 (18%). This pattern reflects the maturation of both XAI methodologies and their practical implementation in manufacturing contexts, with 2023-2025 accounting for 92% of all publications in this domain.

3.3 Evolution and Distribution of XAI Techniques

Among the 38 reviewed studies, SHAP emerged as the most frequently applied explainability method, reported in 22 studies (58%). LIME and PDP followed, each appearing in six studies (16%). Visual interpretability techniques, including Layer-wise Relevance Propagation (LRP) and Grad-CAM, were used in a smaller number of studies (3 and 2, respectively). A subset of works employed domain-specific and physics-informed approaches, such as Mahalanobis distance metrics, governing equation integration, and handcrafted feature attribution schemes. Tab. 1 summarizes the frequency and representative examples of each XAI method identified in the corpus.

Table 1. Summary of XAI Techniques Used in Reviewed Studies

XAI Technique	No. of Studies	Representative Studies
SHAP	22	[Uddin 2023; Akbari 2024; [Ackermann 2023; Ghasemi 2023; Maitra 2024]
LIME	6	[Bordekar 2025; Ryan 2024; Xie 2025]
PDP	6	[Kharate 2024; Mishra 2023; Ryan 2024]
LRP	3	[Kiran 2025; Weeks 2025]
Grad-CAM	2	[Yoo 2024]
Mahalanobis Distance	3	[Kumar 2025; Kumar 2024a; Kumar 2024b]
Physics-Informed Explanations	4	[Du 2021; Zhu 2024]
Feature Importance (ML-based)	4	[Gawade 2025; Akbari 2024]

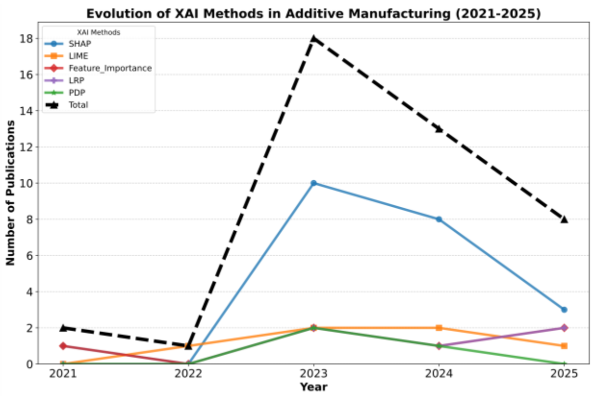

Figure 2 Evolution of XAI Methods Over Time

The temporal adoption of XAI methods from 2021 to 2025 is illustrated in Fig. 2, highlighting the growing dominance of SHAP and the gradual emergence of hybrid and domain-specific techniques.

3.4 Application-Specific XAI Implementation

3.4.1 Quality Control and Defect Detection

Quality control represents the largest application area with 15 studies (39%), reflecting the critical importance of defect prevention and process validation in manufacturing environments. The XAI implementations in this domain utilize diverse methodological approaches including Zero-bias Deep Neural Networks combined with Mahalanobis distance metrics, SHAP analysis for feature attribution, and visual interpretability techniques such as Grad-CAM and Layer-wise Relevance Propagation (LRP) for spatial explanations.

These methods demonstrate exceptional technical performance across multiple manufacturing contexts. Zero-bias Deep Neural Networks achieve remarkable accuracy (>99%) in detecting unknown defects without requiring defect-specific training data [Kumar 2024b], while SHAP analysis consistently identifies key process parameters in eight quality control studies, enabling systematic understanding of defect causation factors. Furthermore, Grad-CAM and LRP visualizations provide spatial explanations for image-based defect detection systems [Yoo 2024], allowing operators to understand both the location and characteristics of detected anomalies. Additional quality control implementations include CT scan analysis with SVM achieving AUC = 0.94 and F1-score = 94.4% explainable CT-based defect detection [Bordekar 2025], textile defect detection addressing LIME limitations, privacy-preserving anomaly detection achieving 94.43% accuracy [Piran 2025], cross-modal defect detection with 98.6% accuracy [Xie 2025], WAAM anomaly detection with LightGBM achieving F1 = 0.945 [Vozza 2024], injection molding quality control [Muaz 2025] and physics-informed defect prediction with 90% accuracy [Du 2021].

The applications span multiple additive manufacturing technologies and focus primarily on real-time anomaly detection and defect characterization processes. These implementations support automated inspection workflows, enable predictive quality management strategies, and provide interpretable feedback for process adjustment decisions. However, the computational overhead and real-time processing requirements present ongoing challenges for deployment in high-throughput manufacturing environments.

3.4.2 Process Parameter Optimization

Process parameter optimization comprises 11 studies (29%) that leverage XAI methodologies to reveal complex, nonlinear relationships between process variables and material properties in additive manufacturing systems. The primary XAI approaches include SHAP feature importance analysis, Partial Dependence Plots (PDP) for individual parameter visualization, and interaction analysis techniques for identifying synergistic effects between multiple process variables.

SHAP feature importance analysis appears in eight optimization studies, consistently identifying critical parameters such as layer height, infill density, and thermal conditions across different AM technologies [Kharate 2024; Uddin 2023]. Additionally, Partial Dependence Plots enable visualization of individual parameter effects in three studies, while advanced interaction analysis reveals synergistic effects between multiple variables that cannot be captured through individual parameter assessment alone [Wang 2024]. These methodological combinations provide a comprehensive understanding of parameter-property relationships that guide systematic optimization strategies.

The applications focus on mechanical property prediction, surface quality optimization, and process efficiency improvement across diverse additive manufacturing processes and material systems. The studies demonstrate how XAI-guided parameter optimization can replace traditional trial-and-error approaches with systematic, knowledge-based optimization strategies. Nevertheless, the complexity of parameter interactions and the need for extensive validation across different manufacturing contexts remain significant implementation challenges.

3.4.3 Design Optimization and Inverse Design

Design optimization applications represent seven studies (18%) that utilize XAI techniques to guide topology optimization, lattice structure design, and material discovery processes in additive manufacturing contexts. The methodological approaches include SHAP-based analysis for design principle extraction, physics-informed XAI methods for mechanistic understanding, and interpretable surrogate modeling techniques for complex design space exploration.

SHAP-based analysis appears in three design studies, revealing fundamental design principles for composite structures and lattice optimization [Chiu 2023; Thawon 2025], while physics-informed approaches enable interpretable surrogate modeling for complex design spaces where traditional computational methods become prohibitively expensive. These implementations combine data-driven pattern recognition with mechanistic understanding to create design tools that respect physical constraints while optimizing multiple performance objectives simultaneously.

The applications encompass topology optimization for structural components, lattice structure design for lightweight applications, and automated material discovery for specific property targets. The studies demonstrate XAI's potential to transform design from empirical iteration toward knowledge-based synthesis, enabling extraction of generalizable design rules and principles. However, the relatively smaller number of design optimization applications compared to other domains indicates field immaturity and suggests significant potential for future development as methodologies become more sophisticated.

3.4.4 Thermal Modeling and Process Monitoring

Thermal modeling and process monitoring applications comprise 4 studies (11%) that employ sophisticated hybrid architectures combining Physics-Informed Neural Networks (PINNs) with XAI methods to address the critical importance of thermal management in additive manufacturing processes. The primary methodological approaches include SHAP analysis of temporal neural networks, Layer-wise Relevance Propagation for spatial feature importance, and physics-informed interpretability techniques that ground explanations in fundamental thermal dynamics principles.

These hybrid implementations achieve remarkable predictive accuracy with R^2 values exceeding 0.97 while maintaining real-time processing capabilities essential for manufacturing deployment. SHAP analysis of LSTM networks reveals temporal dependencies in thermal gradient evolution [Kiran 2025], enabling understanding of how thermal history influences current and future thermal states. Bayesian learning-enabled XAI demonstrates improved prediction accuracy with MAE reduced from 5.31% to 3.43% [Zhu 2024], while emission prediction models achieve high accuracy with RMSE around 0.66 mV [Bock 2024; Guo 2023]. Concurrently, LRP provides insights into spatial feature importance in thermal field prediction, identifying which spatial regions most strongly influence thermal predictions and enabling both predictive control and diagnostic analysis.

The applications focus on temperature prediction, thermal gradient monitoring, and real-time process control across multiple additive manufacturing technologies including polymer extrusion and metal powder bed fusion processes. These implementations enable predictive thermal management strategies that prevent thermal-related defects and optimize process parameters for improved quality outcomes. However, the computational complexity of hybrid architectures and the need for real-time processing create technical challenges that limit widespread deployment in industrial manufacturing environments.

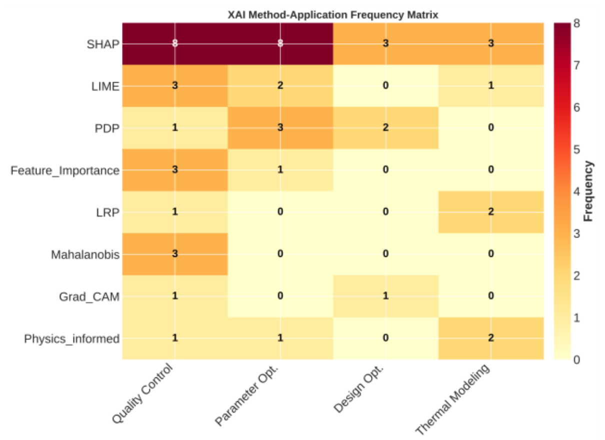

Figure 3. XAI Method vs. Application Area

Fig. 3 shows the frequency distribution of XAI methods across different application domains; SHAP clearly dominates in Quality Control and Parameter Optimization, whereas methods like LIME, Feature Importance, and Mahalanobis are also prevalent in Quality Control, and PDP is frequently used for Parameter and Design Optimization.

3.5 AI Model Architectures and XAI Integration

The distribution of AI model architectures across the 38 reviewed studies reveals distinct patterns that reflect the balance between predictive performance requirements, interpretability needs, and computational constraints in manufacturing applications. Ensemble methods dominate with 16 studies (42%), followed by deep learning approaches in 14 studies (37%), while physics-informed models and Bayesian approaches each appear in 4 studies (11%). This distribution indicates strategic preference for interpretable-by-design approaches that provide inherent explainability while maintaining competitive predictive accuracy.

Ensemble methods, including Random Forest, XGBoost, and LightGBM implementations, offer inherent interpretability through feature importance metrics while maintaining high predictive accuracy [Akbari 2024; Mishra 2023]. These approaches provide transparent decision-making processes that manufacturing engineers can interpret without specialized XAI training, thereby addressing critical adoption barriers in industrial environments. Moreover, the computational efficiency and robustness of ensemble methods make them particularly suitable for manufacturing deployment where reliability and interpretability are paramount considerations.

Deep learning approaches require sophisticated XAI integration but demonstrate superior performance in complex pattern recognition tasks including defect morphology classification, process signature analysis, and multi-modal sensor data fusion. SHAP provides model-agnostic explanations in six deep learning studies, enabling standardized interpretability frameworks across different neural network architectures, while attention mechanisms offer architecture-specific interpretability that reveals which temporal or spatial features drive manufacturing predictions. However, the computational overhead and complexity of deep learning XAI integration present challenges for real-time manufacturing deployment.

Physics-informed models represent an emerging paradigm that combines mechanistic understanding with data-driven learning capabilities [Du 2021; Zhu 2024]. These approaches achieve interpretability through physical constraint satisfaction and governing equation residuals, providing explanations grounded in fundamental manufacturing physics rather than purely statistical associations. Similarly, Bayesian approaches offer uncertainty quantification alongside interpretability, particularly valuable in data-scarce scenarios common in AM research [Drakoulas 2024], where confidence bounds on predictions become essential for risk management in manufacturing decision-making.

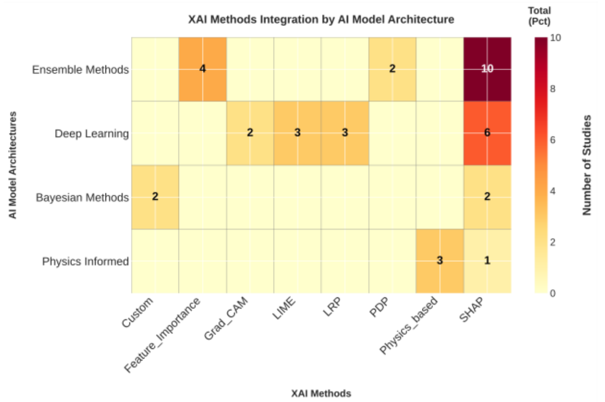

Figure 4. XAI Methods Integration by AI Model Architecture

Figure 4 illustrates the distribution of XAI methods across different AI model architectures, revealing that SHAP demonstrates universal applicability across ensemble and deep learning models due to its robust model-agnostic capabilities. Conversely, physics-informed and Bayesian models tend to rely on custom or domain-specific interpretability tools tailored to their mathematical frameworks, indicating areas where cross-compatibility and standardized XAI integration require further methodological development to achieve comprehensive coverage across all architectural approaches.

3.6. Performance Metrics and Interpretability Assessment

Performance analysis across the 38 reviewed studies reveals consistently high predictive accuracy that demonstrates technical readiness for industrial deployment, while simultaneously highlighting a critical gap in interpretability evaluation methodology that constrains systematic validation of explainability claims. Regression tasks achieve R^2 values exceeding 0.90 in 29 out of 38 studies (76%), with notable examples including $R^2 = 0.998$ for lattice structure prediction [Thawon 2025], $R^2 = 0.976$ for thermal modeling [Kiran 2025], and $R^2 = 0.958$ for mechanical property prediction [Akbari 2024]. These performance levels demonstrate that XAI integration maintains or enhances predictive accuracy, contradicting traditional assumptions about interpretability-performance trade-offs.

Classification tasks report equally impressive results with accuracy above 95% in 27 studies (71%), including exceptional performance examples such as 99.75% for fault detection [Chowdhury 2023], 99.72% for defect detection [Kumar 2024b] and 99.61% for thermal state classification [Yoo 2024]. These accuracy levels significantly exceed typical manufacturing AI requirements and suggest robust algorithm performance across diverse application contexts, manufacturing environments, and material systems. The consistency of high performance across different XAI methods and applications indicates that explainability enhancement does not compromise predictive capabilities.

However, a fundamental methodological deficiency emerges in interpretability assessment, where only eight studies (21%) provide quantitative XAI evaluation metrics. This critical gap prevents systematic comparison of explanation quality, validation of explanation accuracy, and evidence-based improvement of interpretability methods. The absence of standardized interpretability evaluation creates a situation where explainability claims cannot be objectively validated or systematically compared across different studies and applications.

Recent developments suggest growing recognition of this evaluation gap, as novel interpretability metrics emerge in contemporary studies. These include relevancy, stability, and discernability measures for SHAP explanations [Gawade 2025], which represent promising advances toward rigorous XAI evaluation frameworks. However, the limited adoption of these metrics indicates that the field requires standardized interpretability assessment protocols to enable systematic validation and improvement of explanation methods.

Fig. 5 provides a comparative analysis of three prominent XAI techniques across five key performance dimensions: interpretability level, computational overhead, accuracy impact, user comprehension, and generalizability. The scores on each dimension are assigned by the authors based on a qualitative synthesis of the evidence presented in the reviewed literature. The radar chart comparison reveals that SHAP excels in interpretability and generalizability while imposing moderate computational costs, LIME demonstrates strong performance in overhead efficiency and user comprehension with slightly lower interpretability, and physics-informed methods offer maximal interpretability grounded in domain knowledge though with higher specificity and lower accuracy impact. This multi-dimensional analysis indicates that optimal XAI method selection depends on specific application requirements rather than universal method superiority, suggesting the need for application-specific evaluation criteria that balance multiple performance considerations.

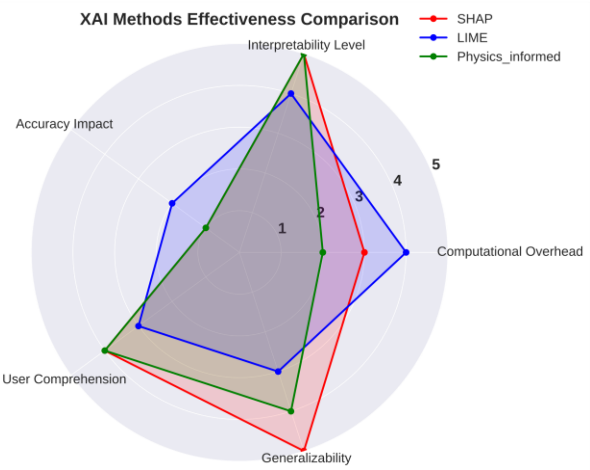

Figure 5. Radar chart comparing SHAP, LIME, and Physics-Informed XAI methods on five performance and interpretability dimensions

4 DISCUSSION

4.1 Key Findings and Evidence Quality

This systematic analysis of 38 peer-reviewed studies represents the first comprehensive evidence synthesis of XAI integration in additive manufacturing, revealing substantial growth and methodological convergence in this interdisciplinary domain. Our findings demonstrate that SHAP-based explanations have achieved dominance (58% adoption) not merely due to technical superiority, but through proven model-agnostic capabilities that address the diverse algorithmic landscape in AM applications, from ensemble methods (42% of studies) to deep learning architectures (37%).

The concentration of publications in 2023-2025 (92.1% of total output) reflects the maturation of both XAI methodologies and their practical implementation in manufacturing contexts, contrasting sharply with earlier isolated attempts at AI interpretability in manufacturing. This temporal clustering suggests a critical mass threshold where XAI techniques became sufficiently robust for industrial applications, supported by our observed high predictive performance ($R^2 > 0.90$ in 76% of regression tasks).

The predominance of quality control applications (39% of studies) aligns with manufacturing industry priorities for risk mitigation and regulatory compliance, while the substantial representation of process optimization (29%) indicates growing confidence in XAI-guided parameter tuning. However, the limited adoption in design optimization (18%) suggests barriers to integrating interpretability with complex design space exploration.

4.2 Interpretability Assessment Gap and Methodological Implications

A striking limitation of the current XAI-AM landscape is the paucity of quantitative interpretability assessment. Our finding that only 21% of reviewed studies conduct any formal evaluation of explanation quality reveals a methodological blind spot, where explanations are often presumed valid without empirical grounding. This gap suggests that the field, while rapidly adopting XAI techniques, has not yet established a commensurate culture of rigorous validation. It is likely attributable to a lack of standardized evaluation benchmarks and the inherent complexity of designing experiments that measure abstract qualities such as explanation fidelity or human utility.

The implication of this oversight is profound: it poses a direct challenge to the scientific legitimacy of XAI applications in manufacturing. In the absence of rigorous empirical validation, the epistemic reliability of generated explanations remains indeterminate. This indeterminacy introduces substantial deployment risks in safety-critical industrial contexts, where erroneous interpretative outputs may precipitate economically detrimental or operationally hazardous decision-making. This assessment gap has a direct and tangible consequence: it fosters an environment where method selection may be driven by popularity rather than contextual suitability.

Our radar chart comparison (Fig. 5) provides clear evidence for this challenge, revealing that different XAI methods possess distinct, complementary strengths, SHAP excels in interpretability and generalizability, LIME in computational efficiency, and physics-informed approaches in domain grounding. This finding directly challenges the current SHAP dominance, indicating that the *one-size-fits-all* application of a single popular method is suboptimal. Therefore, a more strategic, application-specific deployment of XAI is required, where the choice of method is tailored to the specific goals of the manufacturing task, whether it is real-time process control, post-hoc failure analysis, or materials discovery.

4.3 Technical Performance Validation and Deployment Readiness

A key finding from our synthesis is that the integration of XAI in Additive Manufacturing does not necessitate a performance trade-off. On the contrary, model utility is often enhanced, with 76% of regression tasks reporting R^2 values exceeding 0.90 and classification tasks frequently surpassing 95% accuracy. While these figures strongly support the technical readiness of XAI, they primarily reflect performance under controlled, post hoc evaluation settings. A critical gap remains between this laboratory validated performance and industrial grade robustness, as there is limited reporting on generalization to unseen operational conditions, multi machine environments, or process drift. This laboratory to factory gap is a primary challenge for deployment on a scale.

In response to this challenge, our analysis reveals a strategic trend toward architectural interpretability over post hoc explanation. The prevalence of inherently interpretable ensemble methods (42% of studies) suggests a growing recognition that transparency should be integrated by design rather than retrofitted. This movement is further reinforced by the rise of physics informed approaches (11% of studies). By grounding explanations in fundamental manufacturing principles, these hybrid models not only enhance accuracy but also directly address the causality limitations observed in 39% of all reviewed studies. Together, these trends indicate a strategic pivot toward more scientifically grounded and intrinsically trustworthy XAI implementations as the most promising path to bridge the laboratory to factory gap.

4.4 Methodological Limitations and Evidence Quality Constraints

Several methodological limitations constrain the generalizability of our findings. The heavy concentration of publications in recent years (2023-2025) may reflect publication bias toward positive results as XAI gained prominence, potentially overestimating technique effectiveness. The predominance of journal articles (87%) over conference proceedings may bias toward more mature, validated approaches while underrepresenting emerging methodologies.

The interdisciplinary nature of XAI-AM research creates evaluation challenges where traditional systematic review quality assessment tools prove inadequate. The absence of standardized benchmarks across studies limits cross-study comparison and meta-analytic synthesis. Additionally, the rapid pace of methodological development means some included studies may represent outdated approaches by current standards.

The limited geographic and institutional diversity in the reviewed literature, with concentration in specific research centers, may reflect accessibility bias and limit the generalizability of findings across different manufacturing contexts and organizational cultures. The underrepresentation of negative results and failed XAI implementations likely overestimates success rates and underestimates deployment challenges.

4.5 Implications for Manufacturing Practice and Policy

This comprehensive evidence synthesis demonstrates that XAI integration in additive manufacturing has moved beyond proof-of-concept to practical implementation readiness, with significant implications for industry adoption and regulatory frameworks. The demonstrated high performance across diverse applications provides confidence for industrial deployment, while the identified methodological gaps highlight areas requiring standardization before widespread adoption.

Manufacturing organizations should prioritize SHAP-based approaches for broad applicability while considering specialized methods for specific applications, develop internal capabilities for interpretability assessment, and establish frameworks for human-AI collaboration that leverage XAI transparency. The evidence supports strategic investment in XAI infrastructure as a competitive advantage in increasingly automated manufacturing environments.

Regulatory bodies should consider the maturity of XAI techniques in developing compliance frameworks for AI-assisted manufacturing, while recognizing the current limitations in interpretability assessment. The convergence toward standardized evaluation metrics represents an opportunity for proactive policy development that could accelerate safe, transparent AI deployment in manufacturing contexts.

The transition toward XAI-enabled AM systems that are simultaneously autonomous and transparent addresses fundamental tensions between efficiency and oversight in AI-driven manufacturing. This systematic evidence synthesis provides the foundation for evidence-based decision-making in XAI adoption, supporting the evolution toward Industry 5.0 paradigms of trustworthy human-AI collaboration in advanced manufacturing systems.

4.6 Future Research

The systematic gaps identified throughout this review point toward several critical research priorities that are essential to advance the field. A foundational requirement for the scientific legitimacy of XAI in manufacturing is the establishment of standardized interpretability evaluation frameworks. Future research must transcend model-centric metrics to prioritize multi-dimensional protocols that encompass human-centered usability, domain-relevance, and context-specific fidelity. To address the mechanistic understanding limitations affecting 39% of current research, the expansion of Causal XAI frameworks is equally pressing. The exploration of hybrid models that bridge data driven predictions with physically grounded insights is essential to transition from proof-of-concept prototypes to trustworthy, auditable AI systems.

Emerging trends observed in the literature indicate promising avenues for future work. Multi-modal interpretability approaches, identified in four studies, present a significant opportunity to develop unified explanation frameworks capable of transforming complex manufacturing decision-making. Concurrently, the nascent appearance of privacy-preserving XAI (one study) and human-centered design (two studies) signals a growing recognition of deployment constraints that extend beyond mere technical performance. Furthermore, real-time adaptive systems that combine XAI with active learning, noted in three studies, point toward a new paradigm of intelligent manufacturing that could revolutionize human-AI collaboration in the industry 5.0 era.

To ensure long-term impact, research should also prioritize longitudinal studies to assess the deployment effectiveness of XAI in real world settings, alongside comparative analyses across diverse manufacturing domains. The integration of XAI with emerging technologies such as digital twins and edge computing offers a strategic path to create more holistic and responsive systems. Finally, the development of domain-specific XAI methods tailored to the unique physics and materials science of AM represents a particularly fruitful direction for advancing both fundamental scientific understanding and practical, high-value applications.

5 CONCLUSIONS

This systematic review of 38 studies reveals growing adoption of XAI in additive manufacturing, with SHAP-based approaches showing preference across diverse applications. However, critical gaps in interpretability assessment methodology (present in only eight studies, 21%) indicate the field requires standardized evaluation frameworks before claims of technical maturity can be supported. The field's rapid crystallization during 2023-2025 (92.1% of publications) and preference for ensemble methods (42%) signals implicit recognition that transparency should be architectural rather than retrofitted; while emerging causal frameworks (3 studies) and multi-modal approaches (4 studies) point toward the next evolutionary phase, inherently interpretable manufacturing systems that align with physical principles rather than requiring post-hoc explanation. The evidence demonstrates technical readiness for industrial deployment while highlighting an urgent need for standardized interpretability evaluation frameworks; success in advancing trustworthy manufacturing AI depends not on refining explanation methods, but on fundamentally rethinking system architecture to achieve transparency by design rather than explanation by afterthought.

ACKNOWLEDGMENTS

This paper was funded by the University of Finance – Marketing. Additionally, it was completed in connection with the project “Innovative and additive technologies for sustainable energy industry”, registration no. CZ.02.01.01/00/23_021/0010117 financed by the Structural Funds of European Union project. It was also realized in connection with project Students Grant Competition SP2025/062 “Specific research on progressive and sustainable production technologies”, and SP2025/063 “Specific research into innovative and progressive production technologies” financed by the Ministry of Education, Youth and Sports and Faculty of Mechanical Engineering VSB-TUO.

REFERENCES

[Abhilash 2024] Abhilash, P.M., et al. Towards next-gen smart manufacturing systems: the explainability revolution. *npj Advanced Manufacturing*, November 2024, Vol.1, pp 8, ISSN 3004-8621. https://doi.org/10.1016/j.jmsy.2023.10.015

[Ackermann 2023] Ackermann, M., & C. Haase. Machine learning-based identification of interpretable process-structure linkages in metal additive manufacturing. *Additive Manufacturing*, June 2023, Vol.71, pp 103585, ISSN 2214-8604. https://doi.org/10.1016/j.addma.2023.103585

[Akbari 2024] Akbari, P., M. Zamani & A. Mostafaei. Machine learning prediction of mechanical properties in metal additive manufacturing. *Additive Manufacturing*, July 2024, Vol.91, pp 104320, ISSN 2214-8604. https://doi.org/10.1016/j.addma.2024.104320

[Arinez 2020] Arinez, J.F., et al. Artificial Intelligence in Advanced Manufacturing: Current Status and Future Outlook. *Journal of Manufacturing Science and Engineering*, November 2020, Vol.142, pp 110804, ISSN 1087-1357, 1528-8935. https://doi.org/10.1115/1.4047855

[Bock 2024] Bock, F.E., et al. Data-driven and physics-based modelling of process behaviour and deposit geometry for friction surfacing. *Computer Methods in Applied Mechanics and Engineering*, January 2024, Vol.418, pp 116453, ISSN 0045-7825. https://doi.org/10.1016/j.cma.2023.116453

[Bordekar 2025] Bordekar, H., et al. eXplainable artificial intelligence for automatic defect detection in additively manufactured parts using CT scan analysis. *Journal of Intelligent Manufacturing*, February 2025, Vol.36, pp 957–974, ISSN 0956-5515, 1572-8145. https://doi.org/10.1007/s10845-023-02272-4

[Chiu 2023] Chiu, Y.-H., Y.-H. Liao, & Juang, J.-Y. Designing Bioinspired Composite Structures via Genetic Algorithm and Conditional Variational Autoencoder. *Polymers*, January 2023, Vol.15, pp 281, ISSN 2073-4360. https://doi.org/10.3390/polym15020281

[Chowdhury 2023] Chowdhury, D., A. Sinha, & Das, D. XAI-3DP: Diagnosis and Understanding Faults of 3-D Printer With Explainable Ensemble AI. *IEEE Sensors Letters*, January 2023, Vol.7, pp 1–4, ISSN 2475-1472. https://doi.org/10.1109/LSENS.2022.3228327

[Di Bonito 2024] Di Bonito, L.P., et al. eXplainable Artificial Intelligence in Process Engineering: Promises, Facts, and Current Limitations. *Applied System Innovation*, November 2024, Vol.7, pp 121, ISSN 2571-5577. https://doi.org/10.3390/asi7060121

[Dolgui 2024] Dolgui, A., et al. Editorial for the special issue: AI and data-driven decisions in manufacturing. *Journal of Intelligent Manufacturing*, December 2024, Vol.35, pp 3599–3604, ISSN 0956-5515, 1572-8145. https://doi.org/10.1007/s10845-024-02499-9

[Drakoulas 2024] Drakoulas, G., et al. An explainable machine learning-based probabilistic framework for the design of scaffolds in bone tissue engineering. *Biomechanics and Modeling in Mechanobiology*, June 2024, Vol.23, pp 987–1012, ISSN 1617-7959, 1617-7940. https://doi.org/10.1007/s10237-024-01817-7

[Du 2021] Du, Y., T. Mukherjee, & T. DebRoy. Physics-informed machine learning and mechanistic modeling of additive manufacturing to reduce defects. *Applied Materials Today*, September 2021, Vol.24, pp 101123, ISSN 2352-9407. https://doi.org/10.1016/j.apmt.2021.101123

[Gawade 2025] Gawade, V., & M. Chen. Explaining multimodal CNN-DNN model predictions for quality monitoring of porosity in laser metal deposition. *Knowledge-Based Systems*, February 2025, Vol.311, pp 113095, ISSN 0950-7051. https://doi.org/10.1016/j.knosys.2025.113095

[Ghasemi 2023] Ghasemi, A., & M.Z. Naser. Tailoring 3D printed concrete through explainable artificial intelligence. *Structures*, October 2023, Vol.56, pp 104850, ISSN 2352-0124. https://doi.org/10.1016/j.istruc.2023.07.040

[Guo 2023] Guo, W. “Grace”, et al. Explainable AI for layer-wise emission prediction in laser fusion. *CIRP Annals*, 2023, Vol.72, pp 437–440, ISSN 0007-8506. https://doi.org/10.1016/j.cirp.2023.03.009

[Kharate 2024] Kharate, N., et al. Explainable AI Techniques for Comprehensive Analysis of the Relationship between Process Parameters and Material Properties in FDM-Based 3D-Printed Biocomposites. *Journal of Manufacturing and Materials Processing*, August 2024, Vol.8, pp 171, ISSN 2504-4494. https://doi.org/10.3390/jmmp8040171

[Kiran 2025] Kiran, A., et al. Explainable AI in thermal modelling enhancing precision in thermal gradient monitoring for additive manufacturing using LSTM networks. *Thermal Science and Engineering Progress*, April 2025, Vol.60, pp 103465, ISSN 2451-9049. https://doi.org/10.1016/j.tsep.2025.103465

[Kumar 2024a] Kumar, D., et al. Explainable deep neural network for in-plain defect detection during additive manufacturing. *Rapid Prototyping Journal*, January 2024, Vol.30, pp 49–59, ISSN 1355-2546, 1355-2546. https://doi.org/10.1108/RPJ-05-2023-0157

[Kumar 2024b] Kumar, D., et al. Multisource data fusion for defect detection in composite additive manufacturing using explainable deep neural network. *Aerospace Science and Technology*, December 2024, Vol.155, pp 109729, ISSN 1270-9638. https://doi.org/10.1016/j.ast.2024.109729

[Kumar 2025] Kumar, D., Y. Liu, & S. Namilae. Evaluating AI Algorithms for Identifying Anomalies in Composite Additive Manufacturing. *Applied Composite Materials*, May 2025, ISSN 0929-189X, 1573-4897. https://doi.org/10.1007/s10443-025-10340-6

[Ma 2024] Ma, Q.P., et al. A bibliometric review on application of machine learning in additive manufacturing and practical justification. *Applied Materials Today*, October 2024, Vol.40, pp 102371, ISSN 23529407. https://doi.org/10.1016/j.apmt.2024.102371

[Maitra 2024] Maitra, V., C. Arrasmith, & J. Shi. Introducing explainable artificial intelligence to property prediction in metal additive manufacturing. *Manufacturing Letters*, October 2024, Vol.41, pp 1125–1135, ISSN 2213-8463. https://doi.org/10.1016/j.mfglet.2024.09.138

[Mishra 2023] Mishra, A., et al. Explainable Artificial Intelligence (XAI) and Supervised Machine Learning-based Algorithms for Prediction of Surface Roughness of Additively Manufactured Polylactic Acid (PLA) Specimens. *Applied Mechanics*, May 2023, Vol.4, pp 668–698, ISSN 2673-3161. https://doi.org/10.3390/applmech4020034

[Moosavi 2024] Moosavi, S., et al. Explainable AI in Manufacturing and Industrial Cyber–Physical Systems: A Survey. *Electronics*, September 2024, Vol.13, pp 3497, ISSN 2079-9292. https://doi.org/10.3390/electronics13173497

[Muaz 2025] Muaz, M., et al. Explainable AI for correct root cause analysis of product quality in injection moulding. *Journal of Manufacturing Processes*, July 2025, Vol.145, pp 371–380, ISSN 1526-6125. https://doi.org/10.1016/j.jmapro.2025.03.114

[Piran 2025] Piran, F.J., et al. Explainable Differential Privacy-Hyperdimensional Computing for Balancing Privacy and Transparency in Additive Manufacturing Monitoring. *Engineering Applications of Artificial Intelligence*, May 2025, Vol.147, pp 110282, ISSN 0952-1976. https://doi.org/10.1016/j.engappai.2025.110282

[Ryan 2024] Ryan, D., E. Harris, & G.M. O'Connor. Explainable machine learning for the regulatory environment: A case study in micro-droplet printing. *Additive Manufacturing*, May 2024, Vol.88, pp 104237, ISSN 2214-8604. https://doi.org/10.1016/j.addma.2024.104237

[Salih 2025] Salih, A.M., et al. A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. *Advanced Intelligent Systems*, January 2025, Vol.7, pp 2400304, ISSN 2640-4567, 2640-4567. https://doi.org/10.1002/aisy.202400304

[Senoner 2022] Senoner, J., T. Netland, & S. Feuerriegel. Using Explainable Artificial Intelligence to Improve Process Quality: Evidence from Semiconductor Manufacturing. *Management Science*, August 2022, Vol.68, pp 5704–5723, ISSN 0025-1909, 1526-5501. https://doi.org/10.1287/mnsc.2021.4190

[Thawon 2025] Thawon, I., et al. Integrating explainable artificial intelligence in machine learning models to enhance the interpretation of elastic behaviors in three-dimensional-printed triangular lattice plates. *Engineering Applications of Artificial Intelligence*, March 2025, Vol.144, pp 110148, ISSN 0952-1976. https://doi.org/10.1016/j.engappai.2025.110148

[Truong 2025] Truong, T.C., et al. A SYSTEMATIC LITERATURE REVIEW ON THE IMPACT OF GENERATIVE AI ON ADDITIVE MANUFACTURING: CURRENT STATE AND FUTURE DIRECTIONS. *MM Science Journal*, March 2025, Vol.2025, ISSN 1803-1269, 1805-0476. https://doi.org/10.17973/MMSJ.2025_03_2025010

[Uddin 2023] Uddin, M.N., et al. Interpretable machine learning for predicting the strength of 3D printed fiber-reinforced concrete (3DP-FRC). *Journal of Building Engineering*, August 2023, Vol.72, pp 106648, ISSN 2352-7102. https://doi.org/10.1016/j.jobe.2023.106648

[Ukwaththa 2024] Ukwaththa, J., S. Herath, & D.P.P. Meddage. A review of machine learning (ML) and explainable artificial intelligence (XAI) methods in additive manufacturing (3D Printing). *Materials Today Communications*, December 2024, Vol.41, pp 110294, ISSN 23524928. https://doi.org/10.1016/j.mtcomm.2024.110294

[Vozza 2024] Vozza, M., et al. Explaining the Anomaly Detection in Additive Manufacturing via Boosting Models and Frequency Analysis. *Mathematics*, October 2024, Vol.12, pp 3414, ISSN 2227-7390. https://doi.org/10.3390/math12213414

[Wang 2024] Wang, J., et al. A machine learning approach for predicting evaporation-induced composition variability in directed energy deposition in-situ alloying. *Additive Manufacturing*, July 2024, Vol.92, pp 104384, ISSN 2214-8604. https://doi.org/10.1016/j.addma.2024.104384

[Weeks 2025] Weeks, R.D., et al. In-Situ Rheology Measurements via Machine-Learning Enhanced Direct-Ink-Writing. *Advanced Intelligent Systems*, January 2025, Vol.7, ISSN 2640-4567, 2640-4567. https://doi.org/10.1002/aisy.202400293

[Xie 2025] Xie, J., et al. Audio-visual cross-modality knowledge transfer for machine learning-based in-situ monitoring in laser additive manufacturing. *Additive Manufacturing*, March 2025, Vol.101, pp 104692, ISSN 2214-8604. https://doi.org/10.1016/j.addma.2025.104692

[Yoo 2024] Yoo, Y.-J. Thermal imaging-based diagnostic process using explainable artificial intelligence for 3D printing system. *Soft Computing*, May 2024, Vol.28, pp 6741–6752, ISSN 1432-7643, 1433-7479. https://doi.org/10.1007/s00500-023-09530-w

[Zhu 2024] Zhu, Y., et al. Thermal prediction of additive friction stir deposition through Bayesian learning- enabled explainable artificial intelligence. *Journal of Manufacturing Systems*, February 2024, Vol.72, pp 1–

15, ISSN 0278-6125. https://doi.org/10.1016/j.jmsy.2023.10.015

CONTACTS:

Thanh-Cong Truong, Phd.
Faculty of Data Science, University of Finance - Marketing,
No. 778 Nguyen Kiem, Ho Chi Minh, 70000, Viet Nam
ttcong@ufm.edu.vn