A Systematic Review of AI-Enabled Predictive Quality Control in Advanced Metal Manufacturing Systems

Tahmina Akter Rainy[1]; Md. Ashfaq Siddiquee[3];

[1]. *Wright State University, OH, USA;*
Email: tahminarainy332@gmail.com

[2]. *Louisiana Tech University, Louisiana, USA*
Email: ashfaqsiddiquee01@gmail.com

Abstract

This study addresses the persistent problem that quality control in advanced metal manufacturing remains largely reactive, while defect formation is nonlinear, high-dimensional, and costly to detect late, creating scrap, rework, warranty risk, and inspection bottlenecks. The purpose was to quantify and synthesize how AI-enabled predictive quality control (PQC) performs and how deployment-ready it is across manufacturing settings, using a quantitative cross-sectional, case-based design in which each eligible publication is treated as a comparable "case." The sample comprised 52 peer-reviewed cases (2005–2023) spanning industrial and laboratory environments, including enterprise production cases (real factory datasets) and digitally integrated settings where analytics are typically deployed through cloud and enterprise workflows. Key variables included (1) model family (deep learning, classical ML, hybrid), (2) data modality (machine vision, sensor time-series, NDT-based data), (3) validation rigor (internal vs external), (4) performance metrics (F1, AUC, precision, recall), and (5) implementation readiness (1–5 rubric). The analysis plan used structured coding, frequency distributions, cross-tabulation by context and modality, and comparative summary statistics (means, SD). Headline findings show that deep learning dominated 61.5% (32/52) of cases versus 34.6% (18/52) classical ML and 3.8% (2/52) hybrid approaches; machine vision was the most common input (44.2%, 23/52), followed by sensor time-series (36.5%, 19/52) and NDT-driven data (19.2%, 10/52). Multi-modal fusion appeared in 30.8% (16/52) of cases and 75.0% (12/16) reported performance gains over single-modality baselines. Reporting rigor was uneven: while 86.5% (45/52) reported standard predictive metrics, only 21.2% (11/52) reported external validation. Industrial enterprise datasets showed lower average performance (mean F1 = 0.84, SD = 0.07) than lab datasets (mean F1 = 0.91, SD = 0.05), yet more often reported operational impact (63.2% vs 21.7%). Implications are that PQC benefits most when embedded in enterprise-quality data ecosystems, prioritized for drift-aware validation, and paired with explainability and monitoring to support auditability, risk-based inspection, and scalable deployment.

Keywords

*Predictive quality control; Deep learning; Metal manufacturing; Sensor fusion; F1-score and AUC.*

INTRODUCTION

Artificial intelligence (AI) refers to computational methods that approximate human cognitive functions such as learning, pattern recognition, and decision-making, typically operationalized through machine learning (ML) and deep learning (DL) algorithms that infer relationships from data rather than relying exclusively on fixed rules. In manufacturing scholarship, AI is often discussed as an enabling layer inside cyber-physical systems (CPS) where sensing, connectivity, analytics, and decision logic are integrated across shop-floor assets and enterprise workflows (Bai et al., 2018). Predictive quality control (PQC) is defined here as a quality-management approach that anticipates product nonconformance and process drift before final inspection by using data-driven models trained on historical and real-time signals, thereby shifting quality assurance from reactive detection toward anticipatory classification and prediction (Everton et al., 2016). In advanced metal manufacturing systems—covering machining, forming, casting, welding, and metal additive manufacturing—quality states are shaped by multiscale phenomena (thermal gradients, metallurgical phase evolution, residual stress, tool–work interactions) that are observable only indirectly through sensors and inspection technologies, including machine vision, ultrasonic testing, and digital radiography. Internationally, metal manufacturing remains central to infrastructure, energy, transportation, and defense supply chains, which makes quality failures economically systemic rather than local, extending from safety risk to warranty cost, scrap generation, delivery disruption, and reputational loss across global networks (Harchaoui & Meziou, 2018). Industry 4.0 scholarship frames these realities through digitization of production and the coupling of analytics with connected equipment; in that framing, data pipelines and computational intelligence become quality infrastructures rather than optional tools. Research also formalizes how manufacturing intelligence is organized architecturally: the 5C CPS model articulates progression from connection and conversion to cyber, cognition, and configuration, clarifying where PQC functions reside (Du et al., 2019). Alongside CPS, conceptualizations of Quality 4.0 emphasize the combination of digital technologies with quality principles, connecting organizational quality practices to analytics and integrated data ecosystems. In metal manufacturing, these definitions translate to a practical research problem: quality is high-dimensional, data are heterogeneous, defect classes are imbalanced, and decisions must align with process physics and inspection standards, which positions AI-enabled PQC as a contemporary research axis spanning materials processing, industrial informatics, and quality engineering (Bengio, 2009).

Quality control in metal manufacturing historically relies on statistical process control, end-of-line inspection, and non-destructive testing (NDT) regimes, yet advanced metal processes introduce nonlinearities and coupled variables that complicate purely rule-based monitoring. Casting generates internal porosity, inclusions, and shrinkage defects whose detectability depends on imaging modality and sampling strategy, while welding introduces discontinuities shaped by heat input, joint geometry, shielding conditions, and operator or automation variability (Cantero-Chinchilla et al., 2022). Machining quality is influenced by tool wear, cutting forces, chatter, thermal expansion, and material heterogeneity, with defects emerging as dimensional error, poor surface integrity, or microstructural damage. These process realities connect directly to AI-enabled PQC because algorithms can map complex signal–quality relationships when the data are sufficiently informative and labeled, and when evaluation aligns with accepted inspection metrics (Çaydaş & Ekici, 2010). In NDT-centric workflows, digital radiography and ultrasonic inspection produce image or waveform representations where defect characterization becomes a recognition task under noise, attenuation, and geometry constraints, encouraging ML/DL methods that learn discriminative representations rather than relying on handcrafted features alone. A concrete example is deep learning used to improve X-ray defect detection for casting aluminum parts by integrating detection architectures and augmentation strategies to raise mean average precision in industrial datasets. Similarly, studies on radiographic inspection stress that performance reporting benefits from NDT-standard evaluation logic such as probability of detection, reflecting the industrial expectation that automation complements inspector judgment with measurable reliability. Parallel to NDT, machine vision has become a dominant pathway for surface-quality monitoring in rolled, formed, and finished products, where metallic surface flaws and texture anomalies are detected and classified via CNN-based segmentation and classification pipelines (Wang et al., 2020). These developments build an empirical backdrop for a literature review: AI-enabled PQC

in metal manufacturing is not one technique but an ecosystem of sensing modalities, data representations, model families, and validation practices, each conditioned by process physics and operational constraints. Industry 4.0 literature reinforces that quality analytics is most meaningful when embedded in connected architectures that sustain data acquisition, labeling, feedback, and traceability at scale. The academic record therefore positions AI-enabled PQC as a multidisciplinary convergence where "quality" is operationalized through measurable defect metrics, prediction error, detection reliability, and cost-linked performance indicators rather than a single universal definition (Ventura-Carvalho et al., 2021).

Figure 1: AI–ML–DL Framework for Predictive Quality Control in Metal Manufacturing
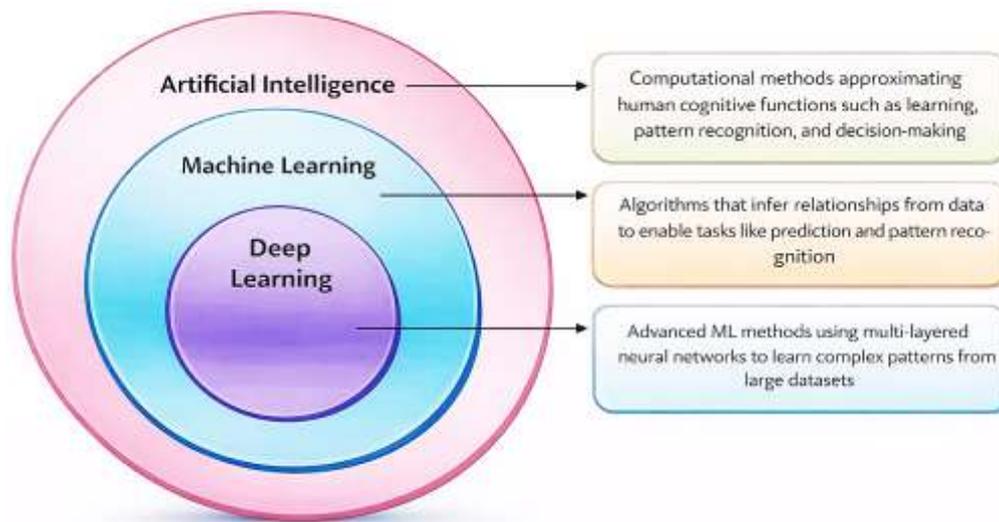
This study is designed to systematically examine, organize, and critically synthesize the body of scholarly evidence on artificial intelligence–enabled predictive quality control (PQC) within advanced metal manufacturing systems. The primary objective is to construct a structured evidence map that clarifies how predictive quality approaches are conceptualized, implemented, validated, and evaluated across machining, casting, forming, welding, and metal additive manufacturing environments. The study aims to identify the dominant artificial intelligence methodologies applied to quality prediction tasks, including classical machine learning, deep learning architectures, hybrid models, and data-fusion strategies, and to categorize their usage according to process context and defect typology. A second objective is to analyze the data ecosystems that underpin predictive quality applications, focusing on the types of sensors, non-destructive testing methods, machine vision systems, and process signals that serve as input variables, as well as the ways these heterogeneous data sources are integrated within modeling pipelines. A further objective is to compare reported performance metrics, validation strategies, and evaluation designs to determine how predictive accuracy, reliability, and operational effectiveness are measured and communicated in the literature. This includes examining whether studies rely on laboratory datasets or real industrial production data and how those contexts influence reported outcomes. The study also seeks to assess implementation readiness by identifying recurring technical, organizational, and operational factors that shape successful deployment, including system integration, data governance, interpretability, computational constraints, and alignment with existing quality management systems. In addition, the research aims to develop a coherent conceptual synthesis that links process characteristics, data modalities, modeling strategies, and operational impact within a unified analytical structure tailored to metal manufacturing environments. By organizing existing findings into clearly defined categories and thematic dimensions, this review intends to clarify the current state of knowledge, highlight methodological patterns, and establish a structured foundation for evaluating predictive quality control as a system-level capability within advanced metal manufacturing systems.

LITERATURE REVIEW

The literature on artificial intelligence–enabled predictive quality control (PQC) in advanced metal manufacturing systems emerges at the intersection of manufacturing engineering, industrial informatics, materials science, and quality management research. Over the past two decades, the evolution from traditional inspection-based quality assurance toward data-driven predictive frameworks has been closely aligned with the broader digital transformation of manufacturing systems. Within this body of work, predictive quality control is generally conceptualized as the anticipatory identification of product defects or process deviations through the use of machine learning and deep learning models trained on historical and real-time production data. In metal manufacturing contexts—such as machining, casting, forming, welding, and additive manufacturing—quality challenges are characterized by nonlinear process behavior, multiscale defect formation mechanisms, high-dimensional sensor signals, and strict performance tolerances tied to safety and regulatory requirements. As a result, the literature reflects a strong emphasis on integrating sensing technologies, non-destructive testing methods, machine vision systems, and process-parameter monitoring into cohesive data ecosystems capable of supporting predictive analytics. Studies vary in scope, ranging from narrowly focused algorithmic performance comparisons to system-level investigations of cyber-physical integration and Quality 4.0 architectures. A recurring theme across the literature is the need to align predictive modeling techniques with process physics, data availability, and operational constraints, recognizing that quality prediction is not merely a classification task but a component of a broader production and governance framework. The research corpus also reveals substantial heterogeneity in experimental design, dataset composition, validation procedures, and performance reporting metrics, which complicates cross-study comparison and synthesis. Consequently, a structured review is necessary to categorize methods, contextualize applications, and identify consistent patterns across diverse manufacturing scenarios. By synthesizing findings across defect types, data modalities, modeling strategies, and implementation contexts, the literature review establishes the intellectual foundation for understanding how AI-enabled predictive quality control functions as a system capability within advanced metal manufacturing environments.

Predictive Quality Control in Manufacturing

Predictive quality control (PQC) in manufacturing can be described as the set of methods that estimate conformance outcomes (e.g., dimensional compliance, defect occurrence, functional test pass/fail) early enough in the production flow to support timely quality decisions such as parameter adjustment, inspection targeting, rework routing, or disposition. PQC extends classical quality control—centered on control charts and end-of-line inspection—by introducing data-driven inference over upstream process states and intermediate product characteristics (Faysal & Shamsunnahar, 2022; Habibullah & Zaheda, 2022). Early PQC research emphasized how production data could be mined to uncover workflow patterns and quality-relevant states across operations, treating "quality" as an evolving attribute of a product as it moves through manufacturing steps. Workflow-mining approaches formalized the idea that quality improvement can be supported by extracting knowledge from production-event data and using that knowledge to identify conditions associated with nonconformance in near real time (Ho et al., 2005; Jahangir & Shahab, 2022; Ratul, 2022). A second foundational line of work framed PQC as a manufacturing-data mining problem, where the central challenge is selecting and structuring feature sets that represent production conditions while avoiding overfitting and spurious correlations. Feature-set decomposition strategies were proposed to manage high-dimensional manufacturing datasets and to improve the stability of learned quality rules or classifiers when quality is measured at batch or lot level rather than per individual unit (Ratul & Subrato, 2022; Rokach & Maimon, 2006; Bhuya & Rebeka, 2022). These perspectives position PQC as a bridge between quality engineering and manufacturing informatics: it assumes measurable traces of the process and product state, transforms them into predictive features, and evaluates prediction quality in ways compatible with manufacturing decision cycles. In multi-stage production, the motivation is amplified because variation introduced early can propagate and compound downstream, making late inspection expensive and information-poor. PQC therefore emphasizes prediction horizons aligned to stage gates, inspection intervals, and takt time, so the predicted signal arrives while corrective actions remain feasible. This orientation fits data-rich factories.

Within complex production systems, PQC research frequently focuses on the "where" and "when" of prediction: which upstream variables best explain downstream quality, and at what point in the routing a reliable estimate becomes available.

Figure 2: Concept And Evolution of Predictive Quality Control in Manufacturing
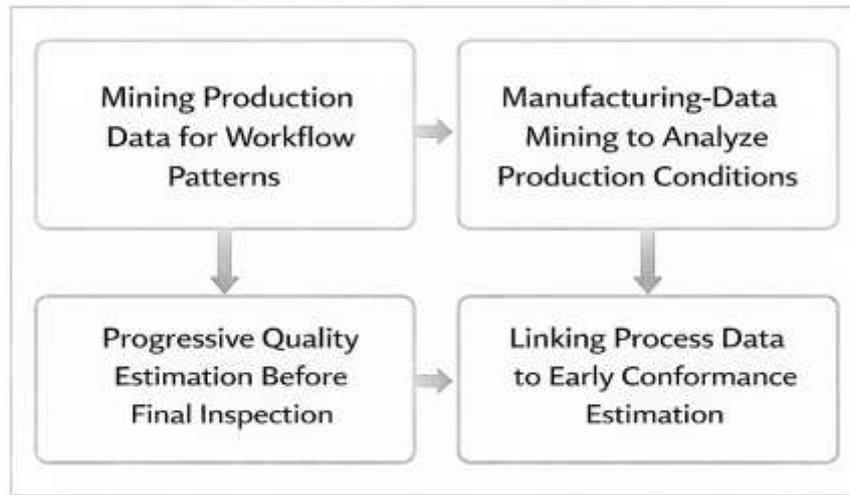
One influential approach is to treat downstream quality as a performance variable that can be predicted progressively as additional process information is accumulated, allowing manufacturing teams to intervene before final test or assembly completion. In a manufacturing setting guided by Six Sigma logic, neural-network models predicted downstream performance and linked accuracy to reduced process variation, illustrating how predictive modeling can be embedded within structured quality-improvement programs rather than operating as a standalone analytics exercise (Jahangir & Mohiul, 2023; Jinnat & Molla Al Rakib, 2023; Johnston et al., 2009). This framing is especially relevant to advanced metal manufacturing, where product quality is only partially observable at intermediate steps and where rework windows are constrained by heat history, surface finish requirements, or subsequent joining operations. A complementary stream of work addresses the data structure of interlinked or multi-stage processes, in which intermediate quality characteristics and process parameters interact across stations (Khaled & Mosheur, 2023; Shahab & Aditya, 2023). In a rolling-mill case study, an inline quality prediction framework combined data preprocessing, feature extraction, and both supervised and unsupervised learning to identify operational patterns associated with intermediate physical quality, emphasizing design around coupled stages and material flow (Lieber et al., 2013; Mostafa, 2023; Mostafa & Tahmina Akter Bhuya, 2023). Together, these studies highlight that PQC is not limited to algorithm selection; it is a system-design problem involving measurement placement, traceability of units through stages, synchronization of process and inspection data, and governance of feedback actions (Ratul & Aditya, 2023; Rifat & Rebeka, 2023). For metal manufacturing lines that include machining, forming, heat treatment, and inspection, the literature therefore treats prediction as an evolving estimate updated as the product accrues evidence, with the practical aim of shifting quality effort upstream toward prevention-oriented decisions. This logic supports alarms and a focused search for root causes.

Recent scholarship consolidates predictive quality within smart manufacturing, where the technical pipeline typically links process data (machine signals, sensor streams, and operational settings) with quality measurements (metrology results, inspection labels, or functional test outcomes) to enable early conformance estimation (Faysal & Bhuya, 2024; Zaheda & Farabe, 2023). Systematic synthesis work characterizes predictive quality solutions as spanning multiple use cases: quality prediction during production from sensor data, early defect risk classification to support routing decisions, and automated interpretation of measurement data to standardize inspection outcomes. Maturity is often judged by how well methods address constraints such as limited labels, data imbalance, changing process conditions, and the need for interpretable outputs that can be translated into actionable quality

levers. A key finding across the predictive quality literature is that performance claims are sensitive to study design choices, including the definition of the target quality variable, the temporal alignment of inputs and outputs, the granularity of prediction (part, batch, or lot), and the validation regime used to simulate deployment conditions. A systematic review of machine learning and deep learning approaches to predictive quality in manufacturing emphasizes this diversity and organizes the field around data sources, modeling strategies, and evaluation practices, showing that predictive quality studies often differ in how they encode process context, integrate heterogeneous variables, and report generalization to new products or operating windows (Towhidul & Uddin, 2024; Sazzadul & Rebeka, 2024; Tercan & Meisen, 2022). For advanced metal manufacturing systems, these observations imply that the literature must be interpreted with explicit attention to process-specific quality mechanisms and to measurement technologies that define ground truth (Tasnim & Anick, 2024; Zaheda & Md Hamidur, 2024). Metal production also requires robust handling of scrap, rework, and traceability. Accordingly, this subsection positions PQC as a system capability realized when data, models, and quality governance align so that predictions arrive at decision points with sufficient reliability to influence manufacturing outcomes. In practice today.

Advanced Metal Manufacturing Systems

Advanced metal manufacturing systems include high-throughput, high-precision production routes such as machining, casting, forming, welding, heat treatment, and metal additive manufacturing, each characterized by tightly coupled thermal–mechanical–metallurgical interactions that shape final part integrity (Amena Begum, 2025; Faysal & Aditya, 2025). Within these environments, "quality" is not limited to visible defects; it spans dimensional accuracy, surface integrity, internal soundness, microstructural stability, and mechanical performance under service conditions (Jahangir, 2025; Md Shahab, 2025). Metal production frequently involves multistage transformations in which variation introduced early can propagate into downstream nonconformance, making defect formation both cumulative and context-dependent. Casting, for example, is susceptible to shrinkage porosity, gas porosity, hot tearing, inclusions, and segregation, where the severity and performance consequences depend on defect morphology, location, and loading mode (Syeedur, 2025; Amin, 2025). Quality challenges become more complex when inspection evidence is incomplete or when defects are sub-surface and only partially observable through sampling-based non-destructive testing. A core quality difficulty in metal manufacturing lies in the gap between an indication found during inspection and its functional impact, which complicates disposition decisions and strengthens the need for process-linked prediction and defect-risk characterization (Blair & Beckermann, 2005; Towhidul & Rebeka, 2025; Mostafa, 2025). At the same time, modern metal supply chains demand consistent quality under compressed lead times, multiple alloys, varied geometries, and changing customer specifications, which intensifies the value of reliable upstream quality indicators. These realities place quality control at the intersection of process physics and measurement capability, where the success of predictive quality strategies depends on the stability of process windows, the traceability of units, and the ability to convert sensor signals and inspection outputs into actionable representations of part conformance (Ratul, 2025; Rifat, 2025).

Machining and forming operations add another layer of quality complexity through surface integrity and geometric accuracy constraints. In machining, even when dimensional tolerances are met, surface integrity outcomes such as residual stress profiles, microhardness variation, phase transformations, and microcracking can determine fatigue life, corrosion behavior, and tribological performance. These quality characteristics are influenced by tool geometry, wear state, cutting speed, feed, lubrication conditions, and thermal loading, which together create a high-dimensional quality landscape. Quality challenges in machining are therefore often expressed as a mixture of measurable geometry and latent integrity variables that are costly to inspect comprehensively on every part (Sazzadul, 2025; Shamsunnahar, 2025). The concept of surface integrity synthesizes these concerns by linking process conditions to subsurface damage and functional performance, making it a central quality lens for advanced metal manufacturing and a major driver for predictive quality monitoring that uses process signals as proxies for integrity risk (M'Saoubi et al., 2008; Yousuf et al., 2025; Azam, 2025). Forming systems likewise face defects such as thinning, wrinkling, tearing, and springback-driven geometric deviation, where small shifts in material properties or tool conditions may trigger out-of-spec

outcomes. Across both machining and forming, measurement limitations, sampling constraints, and the cost of destructive testing create incentives for predictive quality control frameworks that infer quality risk from process signatures. This need becomes even more salient in high-mix, low-volume industrial contexts where statistical stability is weaker and historical quality baselines are less transferable across product variants (Akter & Aditya, 2025; Tasnim, 2025).

**Figure 3: Advanced Metal Manufacturing Systems and Quality Challenges**
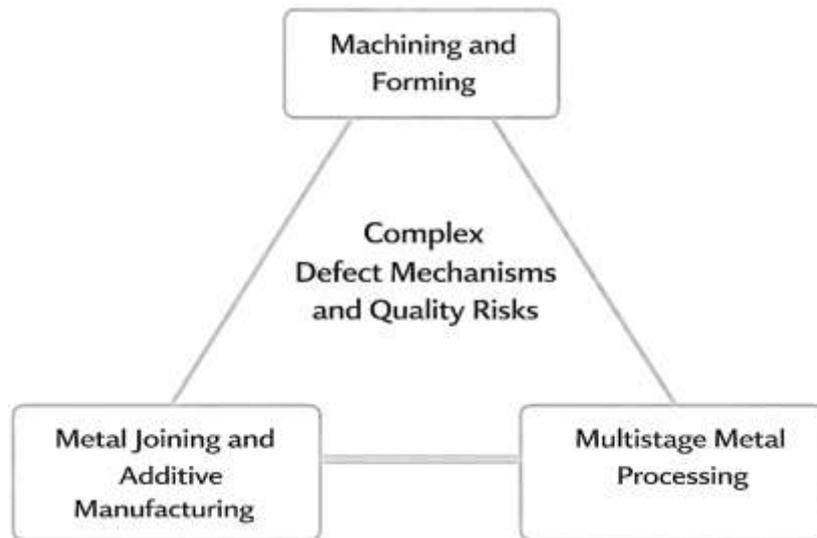
Metal joining and additive manufacturing further amplify quality challenges through localized heat input, complex microstructural evolution, and sensitivity to parameter drift. Welding quality depends on joint design, energy input, travel speed, shielding conditions, and material compatibility, while defects such as lack of fusion, porosity, cracks, and inclusions may be internal and difficult to detect with surface-only inspection. In friction stir welding, for example, subsurface defects and process-dependent discontinuities require appropriate non-destructive testing selection and careful interpretation, since detectability depends on defect geometry, location, and the chosen inspection method (Taheri et al., 2019; Zaheda, 2025a, 2025b). Metal additive manufacturing introduces additional quality challenges because layerwise fabrication produces steep thermal gradients, repeated thermal cycling, and complex residual stress development that can distort parts and change mechanical performance. Residual stress is widely treated as a quality-limiting factor in metal AM because it influences distortion, cracking susceptibility, and dimensional stability, and its magnitude is shaped by scan strategy, build geometry, and process parameters (Megahed et al., 2016). In parallel, metal AM quality is conditioned by defect modes such as lack-of-fusion porosity, keyhole porosity, balling, and surface roughness, which arise from interactions among powder characteristics, energy density, and melt-pool dynamics. A comprehensive view of metal AM quality challenges therefore links process physics to structure–property outcomes and highlights the need to manage defect formation mechanisms in ways that can be monitored and predicted from in-process evidence (DebRoy et al., 2018). Taken together, the literature portrays advanced metal manufacturing quality as a system-level challenge where defect mechanisms, observability limits, and performance-critical requirements jointly define what predictive quality control must address.

AI and Machine Learning in Metal Manufacturing

Artificial intelligence and machine learning foundations for predictive quality control in metal manufacturing are rooted in statistical learning theory, pattern recognition, and data-driven optimization. Early machine learning applications in manufacturing relied heavily on supervised learning techniques such as artificial neural networks, support vector machines, and decision trees to map process variables to quality outcomes. These approaches were attractive because manufacturing environments generate structured numerical data from sensors, programmable logic controllers, and inspection systems, allowing classification and regression models to learn relationships between input

features and conformance labels. A central theoretical contribution to this domain is the formalization of support vector machines as margin-based classifiers capable of handling nonlinear boundaries through kernel transformations, providing robustness in high-dimensional feature spaces (Cortes & Vapnik, 1995). Although developed earlier, this theoretical foundation became widely applied in manufacturing research after 2005 due to increased computational accessibility and digital data availability. As data volume and dimensionality expanded in smart factories, ensemble learning methods such as random forests gained prominence because they reduce overfitting risk and manage complex nonlinear interactions between process variables and quality metrics (Breiman, 2001). In metal manufacturing contexts—where relationships among cutting speed, temperature, vibration, alloy composition, and defect probability are rarely linear—such ensemble approaches provide stability and interpretability through feature-importance analysis. These classical machine learning models establish the baseline methodological architecture for predictive quality systems by formalizing how empirical risk minimization, generalization error, and model complexity trade-offs influence quality prediction reliability. Their theoretical grounding enables reproducible modeling pipelines that include preprocessing, feature selection, training, cross-validation, and deployment simulation under manufacturing constraints.

Figure 4: Ai And Machine Learning Foundations for Predictive Quality in Metal Manufacturing
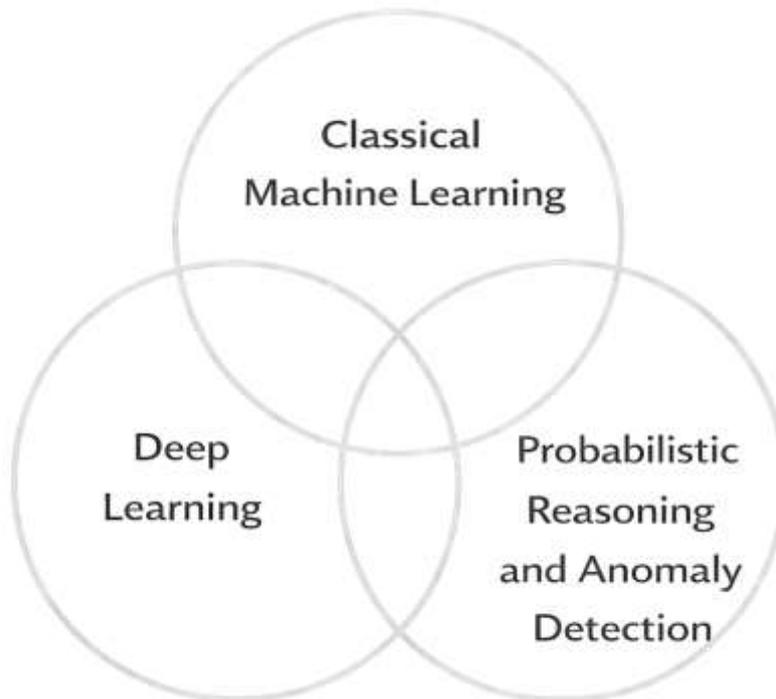

Such approaches align well with machining and forming operations where sensor streams can define a baseline of stable operation. Probabilistic graphical models and Bayesian learning methods also provide mechanisms for uncertainty quantification, which is essential in quality-critical metal applications where false positives and false negatives have distinct cost implications. Research on predictive maintenance and prognostics contributes additional theoretical insight by framing degradation modeling as a prediction problem under uncertainty, integrating sensor data with statistical inference to estimate remaining useful life or defect risk (Lei et al., 2018). Although predictive maintenance is distinct from predictive quality control, both domains share modeling principles and data-structuring challenges. In metal manufacturing, integrating probabilistic reasoning with machine learning enhances decision support by providing confidence intervals or risk scores rather than deterministic labels. Collectively, these machine learning foundations—margin-based classifiers, ensemble learning, deep neural networks, anomaly detection, and probabilistic inference—constitute the methodological backbone of AI-enabled predictive quality control systems. Their theoretical diversity enables adaptation to heterogeneous metal manufacturing data types, process dynamics, and inspection requirements, forming a rigorous computational basis for subsequent applied studies.

Digital Quality Data Ecosystems in Metal Manufacturing

Digital quality data ecosystems in advanced metal manufacturing refer to the interconnected set of technologies, data sources, and organizational practices that enable the capture, storage, integration, and analytical use of quality-relevant information across the manufacturing lifecycle. In traditional quality systems, inspection results and conformance documentation are often separated from process data, limiting the ability to establish causal or predictive relationships between production conditions and quality outcomes. In digitally enabled environments, quality data ecosystems integrate heterogeneous streams such as machine tool signals, sensor measurements, process parameters, metrology results, non-destructive testing outputs, and operator-entered quality observations. These ecosystems provide the structural foundation required for predictive quality control because AI models depend on stable pipelines that deliver traceable, synchronized, and context-rich datasets. The concept of a manufacturing data ecosystem is closely associated with the emergence of industrial big data, where the volume, velocity, and variety of manufacturing data demand systematic architectures for collection and analytics. In this context, a widely cited framework describes industrial big data as a critical enabler of smart manufacturing because it supports data-driven optimization, quality improvement, and intelligent decision-making at scale (Lee et al., 2014). For metal manufacturing, the value of such ecosystems lies in their ability to link process-state variables to quality measurements, allowing defect prediction and process deviation forecasting to occur earlier in production. However, the literature emphasizes that quality data ecosystems require more than data availability; they require structured data models, consistent semantics, and robust integration between shop-floor systems and enterprise quality platforms. The presence of disconnected databases, inconsistent part identifiers, and limited traceability across production stages can undermine the reliability of predictive quality models by introducing misalignment between input signals and ground-truth quality outcomes. Therefore, digital quality ecosystems are framed as both technical infrastructures and governance mechanisms that determine whether quality analytics can operate reliably across diverse metal manufacturing environments.

Figure 5: Digital Quality Data Ecosystems in Metal Manufacturing
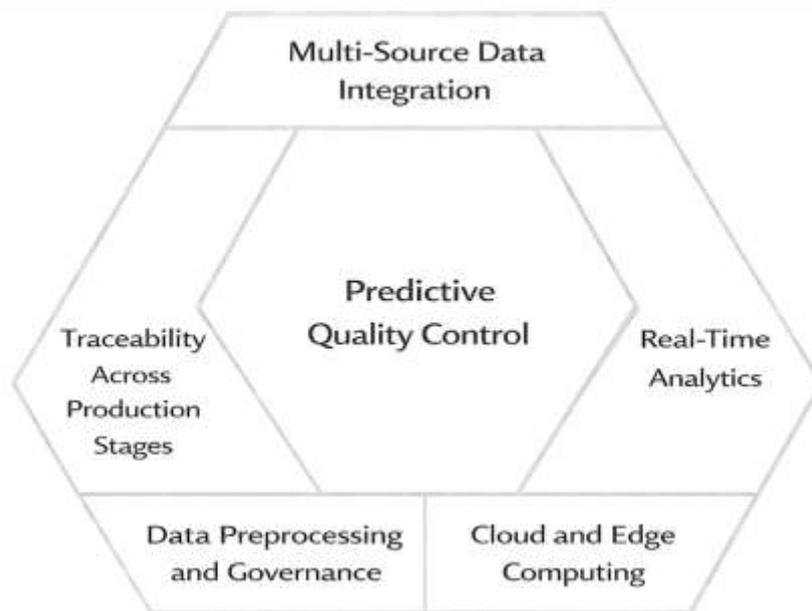
A central characteristic of digital quality ecosystems in advanced metal manufacturing is multi-source integration, since no single sensor or inspection method provides complete visibility into defect formation mechanisms. Metal production often relies on a combination of in-process monitoring signals and post-process inspection evidence, such as coordinate measuring machine outputs, surface profilometry, ultrasonic testing, radiography, and destructive mechanical testing. Integrating these sources into a coherent dataset requires data fusion across different sampling rates, spatial resolutions,

and measurement uncertainty profiles. The literature on smart manufacturing systems highlights that successful data ecosystems rely on standardized data acquisition, interoperability across industrial devices, and architectures that support real-time and batch analytics simultaneously. A core conceptualization of smart manufacturing emphasizes the use of interconnected systems that combine sensing, modeling, and decision support, enabling predictive capabilities across production and quality domains (Kang et al., 2016). Within this view, quality data ecosystems must accommodate both structured numerical data and unstructured data such as images, acoustic signals, and text-based inspection notes. Another critical requirement is the management of data quality itself, since predictive quality systems are highly sensitive to missing values, label noise, class imbalance, and sensor drift. In industrial environments, data quality problems may arise from sensor faults, manual entry errors, inconsistent inspection criteria, or changing production conditions. Research on industrial AI applications stresses that manufacturing data ecosystems must incorporate mechanisms for data preprocessing, anomaly screening, and governance of training datasets to ensure reliable model performance. In addition, traceability is fundamental in metal manufacturing because parts often undergo multiple processes, and quality evidence must be linked across stages for accurate defect attribution. Digital thread concepts are frequently used to describe this traceability requirement, where product and process data are connected across the lifecycle to support quality prediction, certification, and compliance reporting.

Recent research increasingly frames digital quality ecosystems through the lens of cyber-physical production systems and cloud–edge integration, where real-time data pipelines support predictive quality analytics under operational constraints. In metal manufacturing, quality prediction models may require low-latency execution for immediate intervention, particularly in high-speed rolling, stamping, or machining environments. This creates the need for edge computing architectures that can process sensor streams locally while synchronizing key information with cloud platforms for model training, storage, and enterprise-level analytics. The literature on cloud manufacturing and industrial IoT emphasizes that distributed computing infrastructures enable scalable analytics and support the deployment of AI models in production systems, especially when combined with standardized industrial communication protocols. A widely cited review describes cloud manufacturing as a service-oriented paradigm that integrates manufacturing resources and data through networked infrastructures, enabling flexible and data-driven manufacturing management (Xu, 2012). From a quality perspective, such infrastructures facilitate cross-factory learning, benchmarking, and centralized model governance. Another important concept in digital quality ecosystems is the use of digital twins, which represent virtual models of manufacturing systems that synchronize with real-time production data. Digital twin frameworks provide a mechanism for combining process physics models with data-driven learning, supporting predictive monitoring and quality estimation in complex manufacturing environments. A comprehensive review of digital twins in manufacturing describes their role in integrating data, simulation, and analytics for improved operational decision-making (Tao et al., 2019). In metal manufacturing, digital twins can link process parameters and sensor signals to predicted microstructural outcomes, defect probabilities, and dimensional deviations, strengthening predictive quality control capabilities. Furthermore, industrial AI governance literature emphasizes that model deployment depends on data pipelines that support continuous updating, monitoring, and retraining as process conditions evolve. Research on AI in manufacturing systems highlights that digital quality ecosystems must address not only technical integration but also organizational readiness, workforce interaction, and quality governance policies that determine how predictive outputs are interpreted and acted upon (Wuest et al., 2016). Together, these studies establish digital quality ecosystems as the foundational infrastructure for AI-enabled predictive quality control, defining the conditions under which predictive models can be trained, validated, deployed, and maintained in advanced metal manufacturing systems.

Sensors, Non-Destructive Testing (NDT) in Metal Manufacturing

Sensors, non-destructive testing (NDT), and metrology systems form the measurement backbone of predictive quality control in advanced metal manufacturing because they define both the input evidence for AI models and the ground-truth outcomes used for training and validation. In metal manufacturing environments, defects may be internal, subsurface, or microstructural, meaning that

surface-only inspection cannot provide complete quality assurance. For this reason, NDT methods such as ultrasonic testing, eddy current testing, radiographic inspection, magnetic particle inspection, and thermography remain central to quality verification across critical sectors including aerospace, energy, and automotive manufacturing. In a broad review of NDT technologies, the literature emphasizes that inspection modality selection is fundamentally linked to defect type, material properties, geometry, and expected flaw size, and that each technique introduces specific sensitivity and uncertainty characteristics that shape defect detectability (Hellier, 2013). Within predictive quality frameworks, this measurement uncertainty is a central issue because AI models learn from inspection outcomes that may contain noise, false indications, or incomplete coverage. Therefore, predictive quality models in metal manufacturing must account for the probabilistic nature of defect detection and the limitations of inspection sampling. In advanced manufacturing, the role of sensors extends beyond post-process inspection into in-process monitoring, where acoustic emission, vibration, temperature, force, and electrical signals can be collected continuously to represent process dynamics. These signals provide a pathway for early defect prediction, especially when final inspection is costly or when destructive testing is required to validate certain quality attributes. The literature on sensor-based monitoring emphasizes that the reliability of predictive quality depends on the stability of signal acquisition, calibration procedures, and the consistency of measurement conditions, particularly in harsh metal manufacturing environments.

Figure 6: Sensors, Non-Destructive Testing in Metal Manufacturing
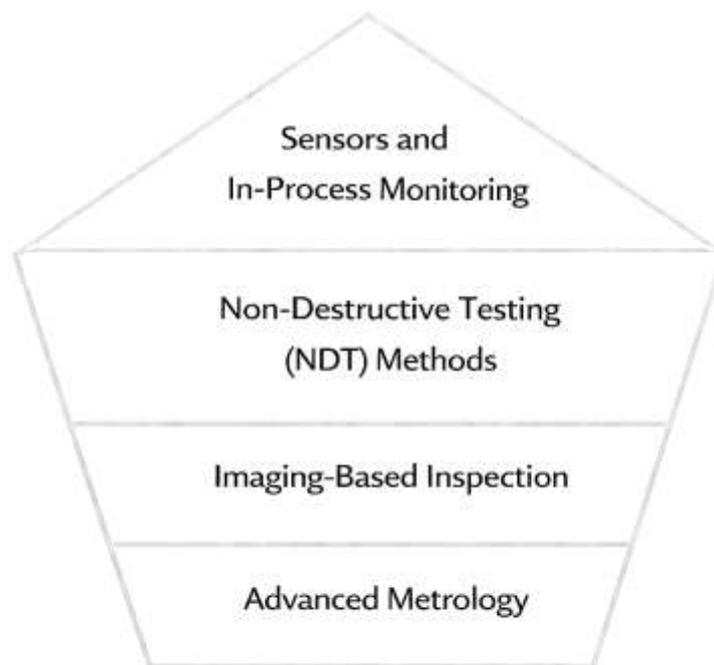

A major trend in the measurement literature is the increased integration of machine vision and imaging-based inspection into metal manufacturing quality workflows. High-resolution cameras, line-scan imaging, structured illumination, and thermal imaging systems are now widely used for surface defect detection in rolling, forming, and finishing operations. Machine vision supports predictive quality by enabling continuous inspection at production speed and by generating large volumes of labeled or semi-labeled defect data for AI model training. However, the literature stresses that metallic surfaces pose specific challenges for vision systems due to specular reflections, variable texture, lighting instability, and fine-grained defect boundaries. A comprehensive review of vision-based defect detection highlights that robust surface inspection requires careful imaging design, consistent illumination strategies, and algorithmic models capable of detecting subtle anomalies across changing surface conditions (Xie, 2008). From a predictive quality perspective, these factors influence whether vision data can be reliably used as an early indicator of process drift or defect emergence. In parallel,

advanced metrology tools such as coordinate measuring machines (CMMs), laser scanners, profilometers, and in-line dimensional sensors generate geometric and surface-quality measurements that are essential for validating predictive models. These measurement systems provide quantitative quality labels, including dimensional deviation, roughness parameters, waviness, and form error, which can be linked to process variables and sensor signals for supervised learning. The integration of metrology outputs into predictive pipelines is particularly important in machining and forming, where conformance depends on tight tolerances and surface integrity requirements. A review of modern dimensional metrology emphasizes that measurement systems must be treated as part of the manufacturing process itself, since measurement uncertainty and sampling strategy can significantly affect quality conclusions and model training reliability (Bosch et al., 2015). Therefore, predictive quality systems in metal manufacturing depend on the quality of measurement infrastructure as much as on algorithmic capability.

Socio-Technical Systems (STS) Theory for AI-Enabled Predictive Quality Control

Socio-Technical Systems (STS) theory provides a rigorous theoretical lens for interpreting AI-enabled predictive quality control (PQC) as an organizational capability shaped by the joint interaction of technology, people, and process design rather than as a purely computational innovation. In advanced metal manufacturing, PQC requires the integration of sensor infrastructures, inspection systems, modeling pipelines, and quality governance routines across multiple departments such as production, maintenance, quality assurance, and engineering. STS theory is relevant because it conceptualizes organizations as systems in which technical subsystems (machines, digital platforms, algorithms, data pipelines) and social subsystems (human expertise, work practices, trust, communication, and decision authority) must be jointly optimized to achieve stable performance. This perspective aligns strongly with quality-critical metal production where predictive outputs must be interpreted by engineers, inspectors, and operators who possess domain knowledge of defect mechanisms and process behavior. The STS lens is also appropriate because AI-enabled PQC often changes the locus of quality decisions, shifting some judgment from manual inspection toward algorithmic risk scoring or automated defect classification, which introduces issues of acceptance, transparency, accountability, and workflow redesign. Research on socio-technical integration in manufacturing emphasizes that digital technologies produce operational value only when embedded within organizational structures that support adoption, learning, and continuous improvement (Clegg, 2000).

Figure 7: . Socio-Technical Systems Theory For Ai-Enabled Predictive Quality Control
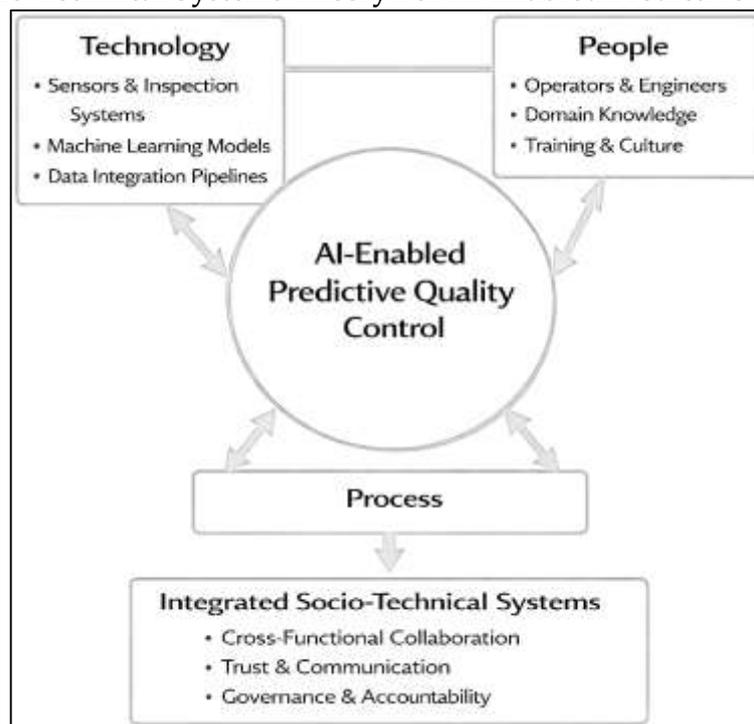
The STS framework also supports structured evaluation of AI-enabled PQC maturity by distinguishing between technical performance and socio-organizational effectiveness. Technical performance refers to predictive accuracy, defect detection reliability, and model robustness under process variation. Socio-organizational effectiveness refers to how predictive outputs are embedded in decision-making workflows, whether interventions occur at the correct stage of production, and whether accountability is preserved under automated decision support. In the manufacturing analytics literature, the effectiveness of AI-based systems is increasingly evaluated through operational impact measures such as scrap reduction, rework reduction, yield improvement, and downtime minimization. However, STS theory emphasizes that such outcomes depend on coordination mechanisms, training, and organizational learning. Research on human–automation interaction highlights that trust calibration is essential; users must understand when to rely on automation and when to override it, especially in high-risk industrial contexts (Lee & See, 2004). In predictive quality control, this translates into the need for interpretable models and transparent quality rules so that engineers and inspectors can validate AI decisions. Studies on interpretability in machine learning further emphasize that black-box models can hinder adoption in safety-critical domains because users cannot easily trace why a defect was predicted or which variables influenced the prediction (Doshi-Velez & Kim, 2017). In metal manufacturing, interpretability is also linked to root-cause analysis, since quality engineers must connect predicted defect risk to actionable process parameters such as heat input, feed rate, cooling conditions, or tool wear state. The STS lens therefore justifies the inclusion of explainable AI (XAI) as a quality governance requirement rather than a secondary feature. Furthermore, STS theory supports analysis of the organizational consequences of predictive quality systems, including changes in inspection labor allocation, skill requirements, and quality documentation workflows. It also highlights the role of feedback loops: predictive outputs must be connected to process adjustment mechanisms and quality assurance procedures to sustain performance. This is particularly relevant in advanced metal manufacturing where process drift and material variability are common, and where models require monitoring and retraining to remain reliable.

To operationalize STS theory in this study, a unified evaluation structure is proposed that links predictive quality performance with socio-technical readiness dimensions. The central quantitative formula that will be applied throughout the study is the **F1-score**, which is widely used in predictive quality and defect classification tasks because metal manufacturing datasets often exhibit class imbalance, where defect cases are far rarer than non-defect cases. The F1-score balances precision and recall and is defined as:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

where

$$Precision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN}$$

and TP represents true positives (correctly predicted defects), FP represents false positives (normal parts incorrectly predicted as defective), and FN represents false negatives (defective parts missed by the model). This formula is particularly suitable for predictive quality control because false negatives can lead to catastrophic quality escapes, while false positives can increase scrap, rework, and inspection burden. The literature on industrial defect detection and quality classification frequently recommends F1-score as a core metric because it provides a stable measure of classification effectiveness under imbalance conditions common in manufacturing quality datasets (Powers, 2011). In the STS framework, this metric can be interpreted not only as a technical measure but also as an indicator of operational risk: low recall implies higher risk of defect escape, while low precision implies higher burden on inspection and rework systems. The STS model therefore links F1-score outcomes to organizational consequences and workflow impact. This theoretical and quantitative integration aligns with manufacturing analytics research that emphasizes the importance of combining technical performance

with operational feasibility and decision integration when evaluating industrial AI systems (Ransbotham et al., 2017). By applying STS theory and using the F1-score as a core cross-study performance indicator, this research establishes a consistent analytical basis for comparing AI-enabled PQC evidence across metal manufacturing contexts while maintaining alignment with socio-technical adoption requirements.

Conceptual Framework in Advanced Metal Manufacturing

The conceptual framework for AI-enabled predictive quality control (PQC) in advanced metal manufacturing systems integrates process characteristics, data ecosystems, modeling strategies, and operational outcomes into a structured analytical model tailored to metal production environments. Unlike purely algorithm-centric perspectives, this framework conceptualizes PQC as a layered architecture in which physical processes generate observable signals, digital infrastructures transform these signals into structured datasets, artificial intelligence models extract predictive patterns, and decision systems convert predictions into quality actions. In metal manufacturing, this layered logic must account for process heterogeneity across machining, casting, forming, welding, and additive manufacturing, where each process exhibits distinct defect mechanisms, thermal–mechanical interactions, and inspection modalities. The framework therefore begins with the **Process–Data Interface**, which defines how process parameters (e.g., cutting speed, heat input, rolling pressure), environmental variables, and material attributes are captured through sensors and inspection systems. Research on smart production networks emphasizes that the effectiveness of digital manufacturing systems depends on structured information integration across machines and enterprise layers, ensuring that process and quality data are contextually aligned (Monostori et al., 2016). Within this conceptual layer, traceability and digital connectivity become prerequisites for predictive modeling because quality labels must correspond accurately to the process conditions under which parts were produced. The second layer of the framework is the **Analytical Modeling Layer**, where supervised learning, anomaly detection, and regression models map high-dimensional input data to predicted quality states. The final layer is the **Decision and Impact Layer**, where predicted outputs influence scrap control, rework allocation, process adjustment, and inspection prioritization. Studies on data-driven manufacturing architectures emphasize that predictive analytics should be integrated into operational decision cycles rather than isolated as experimental tools (Wang et al., 2018). Thus, the conceptual framework positions AI-enabled PQC as a closed-loop system linking process data, predictive modeling, and quality governance mechanisms within advanced metal manufacturing ecosystems.

At the core of this conceptual framework lies the quantitative evaluation of predictive performance and operational impact. Predictive quality models in metal manufacturing often involve binary or multiclass classification of defect presence, severity level, or conformance status. A key performance metric that complements the F1-score introduced earlier is the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), which evaluates the trade-off between true positive rate and false positive rate across varying classification thresholds. The true positive rate (TPR) and false positive rate (FPR) are defined as:

$$TPR = \frac{TP}{TP + FN}$$
$$FPR = \frac{FP}{FP + TN}$$

where *TN* represents true negatives. The AUC is then computed as the integral under the ROC curve:

$$AUC = \int_0^1 TPR(FPR)\, d(FPR)$$

AUC is particularly relevant in predictive quality contexts where decision thresholds may vary depending on risk tolerance, regulatory requirements, or cost structures associated with defect escape versus false rejection. Research on machine learning performance evaluation emphasizes that AUC provides a threshold-independent measure of classification separability and is robust under moderate class imbalance (Bradley, 1997). In metal manufacturing, where defect events may be rare but highly

consequential, threshold selection must reflect quality risk policies. The conceptual framework therefore integrates predictive metrics into a broader cost-sensitive evaluation logic. In addition, studies on cyber-physical production systems highlight that predictive performance must be assessed under realistic operational conditions, including data latency, concept drift, and variable process windows (Schluse et al., 2018). Within the proposed framework, predictive outputs feed into an Operational Effectiveness Function, which can be expressed as:

$$OE = f(AUC, F1, \Delta Scrap, \Delta Rework, \Delta Downtime)$$

where OE represents operational effectiveness and Δ denotes measurable improvements relative to baseline quality performance. This formulation links algorithmic performance directly to manufacturing impact indicators, reinforcing that predictive accuracy alone does not define PQC success.

Figure 8: Conceptual Framework For Ai-Enabled Predictive Quality Control In Advanced Metal Manufacturing
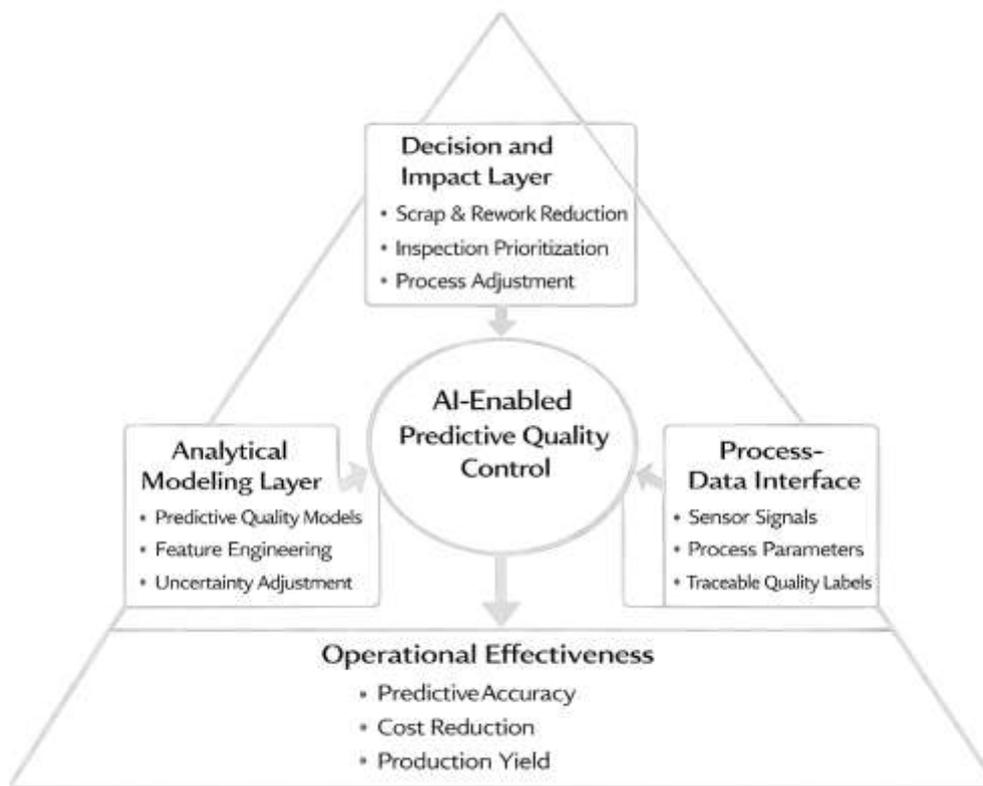
The final dimension of the conceptual framework addresses moderating and contextual variables that influence predictive quality reliability and scalability in advanced metal manufacturing. These include data quality, sensor stability, interpretability, workforce interaction, and system integration maturity. Research on industrial AI adoption emphasizes that predictive analytics systems must operate within interoperable digital architectures and align with enterprise-level decision systems to generate sustainable value (Lu et al., 2017). In metal manufacturing, this implies that predictive models must integrate with manufacturing execution systems (MES), quality management systems (QMS), and maintenance scheduling platforms. Furthermore, uncertainty quantification is essential in quality-critical sectors where defect escape may lead to structural failure or regulatory non-compliance. Bayesian and probabilistic evaluation frameworks highlight the need to interpret predictive outputs as risk probabilities rather than deterministic labels, supporting informed human decision-making (Kendall & Gal, 2017). The conceptual framework therefore includes an Uncertainty Adjustment Factor (UAF) applied to predictive outputs:

$$\text{Adjusted Prediction} = P(\text{Defect}) \times (1 - UAF)$$

where *UAF* captures model uncertainty derived from variance estimation or confidence intervals. This ensures that decision-makers account for predictive uncertainty when allocating inspection or corrective actions. Additionally, digital twin research emphasizes the importance of synchronizing real-time production data with virtual process models to enhance prediction reliability and traceability across the lifecycle (Negri et al., 2017). Within advanced metal manufacturing systems, the conceptual framework integrates these moderating factors into a comprehensive view in which process data, modeling performance, uncertainty estimation, and operational integration collectively determine the maturity of AI-enabled predictive quality control. By organizing evidence around this structured architecture, the study establishes a unified analytical model for synthesizing diverse research findings and evaluating predictive quality systems across varied metal manufacturing contexts.

Research Gaps in AI-Enabled Predictive Quality Control

The research literature on AI-enabled predictive quality control (PQC) in advanced metal manufacturing systems demonstrates strong growth across vision-based inspection, sensor-driven anomaly detection, and non-destructive evaluation (NDE) automation. At the same time, systematic synthesis reveals persistent gaps that limit cross-study comparability, industrial translation, and the formation of unified evidence-based conclusions. One of the most visible gaps concerns the fragmentation of problem definitions and target variables. Across metal manufacturing processes, studies define "quality" using different operational labels, including binary defect presence, multi-class defect type, severity grading, dimensional deviation, surface roughness parameters, porosity probability, or surrogate quality indices. This heterogeneity complicates synthesis because performance metrics become non-equivalent, and predictive tasks vary in difficulty. Manufacturing analytics literature highlights that predictive quality studies often suffer from inconsistent experimental protocols, which restrict generalization claims and reduce the interpretability of reported outcomes across production contexts (Kusiak, 2018). Another major gap relates to dataset transparency and reproducibility. Many predictive quality studies rely on proprietary industrial datasets that cannot be shared, limiting the ability of other researchers to replicate findings or benchmark competing algorithms. This issue is particularly severe in metal manufacturing, where inspection data and defect labels are sensitive due to certification requirements, customer confidentiality, and liability exposure. In parallel, studies frequently lack sufficient documentation of data preprocessing, label construction, and sampling strategy, creating uncertainty about whether reported performance reflects robust learning or dataset artifacts. In AI-enabled inspection research, these problems have been recognized as barriers to building reliable industrial systems, with calls for standardized reporting and evaluation frameworks that reflect manufacturing realities (Zonta et al., 2020). A further gap lies in the limited integration of domain knowledge. Many AI models treat metal manufacturing quality prediction as a generic pattern-recognition task, while metallurgical mechanisms and process physics strongly influence defect formation and defect observability. Research on physics-informed and hybrid modeling indicates that combining data-driven learning with physical constraints can improve generalization and interpretability, particularly when training data are limited or when processes vary across operating windows (Karniadakis et al., 2021). These gaps collectively justify the need for structured reviews that map evidence systematically and interpret findings within process-specific quality contexts.

A second cluster of research gaps concerns validation rigor and deployment realism. Many predictive quality studies evaluate models using random train-test splits or cross-validation strategies that may not reflect production conditions where data drift, equipment wear, and changing material batches alter signal distributions. In metal manufacturing, drift can be driven by tool wear progression, sensor aging, machine recalibration, or changes in alloy composition, making temporal validation and external testing critical for trustworthy predictive quality systems. The manufacturing AI literature increasingly emphasizes that deployment-ready models require evaluation designs that simulate real operational transitions, including time-based splits, cross-machine validation, and cross-product testing (Ghahramani, 2015). Another gap relates to the limited use of uncertainty quantification and risk-aware decision support. In safety-critical metal manufacturing domains such as aerospace, nuclear, rail, and structural engineering, the consequences of false negatives can be severe. Yet many predictive quality studies report only point estimates of accuracy or F1-score without quantifying uncertainty or

confidence intervals. Research on probabilistic machine learning in industrial settings stresses that decision support systems benefit from uncertainty estimation because it enables risk-based inspection allocation and more reliable quality governance (Ghahramani, 2015). Additionally, the operational integration of predictive quality systems is underdeveloped in the literature. Many studies focus on algorithmic performance without detailing how predictions are used within manufacturing execution systems, inspection workflows, or quality management systems. In advanced metal manufacturing, quality decisions are constrained by certification requirements, documentation rules, and traceability standards, meaning that predictive outputs must be interpretable and auditable. Human–AI collaboration research highlights that industrial AI systems require alignment between algorithmic outputs and human decision authority, including mechanisms for explanation, override, and feedback-driven learning (Amershi et al., 2019). Therefore, the literature gap is not only technical but also socio-technical: predictive models must be designed as components of quality systems rather than isolated classifiers.

Figure 9: Research Gaps And Review Positioning In Ai-Enabled Predictive Quality Control For Metal Manufacturing


A third set of gaps concerns the uneven coverage of metal manufacturing processes and the imbalance between laboratory studies and industrial case evidence. Research in metal additive manufacturing and surface defect detection dominates many AI-enabled PQC discussions because these domains produce visually rich datasets and align well with deep learning architectures. However, other critical metal manufacturing processes such as large-scale forging, continuous casting, heat treatment, and complex multi-stage machining lines receive comparatively less attention, even though quality failures in these contexts have major economic and safety implications. Another issue is that many studies remain at proof-of-concept stage, using controlled experimental datasets that do not capture the variability of industrial production. This gap is important because predictive quality control must operate under high noise, changing operating conditions, incomplete labels, and constrained computing resources. Research on industrial anomaly detection notes that laboratory performance can degrade substantially in real factories due to domain shift and sensor noise, which reinforces the need for industrial-scale validation (Chalapathy & Chawla, 2019). Furthermore, the literature often lacks cost-sensitive evaluation, even though metal manufacturing quality decisions involve trade-offs

between scrap, rework, inspection effort, and defect escape risk. Quality engineering research emphasizes that predictive quality systems should be evaluated not only on statistical performance but also on operational impact measures such as yield improvement, scrap reduction, and reduced inspection load. These gaps define the positioning of the present systematic review: the study is designed to synthesize AI-enabled PQC evidence across diverse metal manufacturing processes, classify applications by process and defect type, evaluate methods in relation to sensing ecosystems and validation rigor, and assess implementation readiness through case-study synthesis. By explicitly mapping methodological patterns and contextualizing them within metal manufacturing quality constraints, the review addresses fragmentation in the literature and establishes a structured foundation for evaluating predictive quality control as a system-level capability in advanced metal manufacturing systems.

METHODS

This study employed a systematic literature review to synthesize scholarly evidence on AI-enabled predictive quality control (PQC) in advanced metal manufacturing systems. A structured and transparent protocol was implemented to ensure methodological rigor, replicability, and relevance to both academic and industrial contexts. Peer-reviewed journal articles and conference papers published between 2005 and 2023 were identified through comprehensive database searches using keyword combinations related to predictive quality, defect prediction, anomaly detection, non-destructive testing, machine vision, sensor fusion, and Industry 4.0 analytics. The review followed a qualitative cross-sectional design, treating each selected study as a documented case of AI implementation within specific metal manufacturing processes, including machining, casting, forming, welding, heat treatment, and metal additive manufacturing. A multi-stage screening procedure was conducted, including title and abstract filtering, full-text evaluation, and application of predefined inclusion and exclusion criteria. Studies were required to explicitly apply artificial intelligence, machine learning, or deep learning techniques to predictive quality outcomes in metal manufacturing contexts, while purely theoretical works or studies unrelated to metal processing were excluded. A structured quality appraisal checklist was applied to assess dataset transparency, validation rigor, defect definition clarity, and completeness of performance reporting.

Figure 10: Methodology for this study
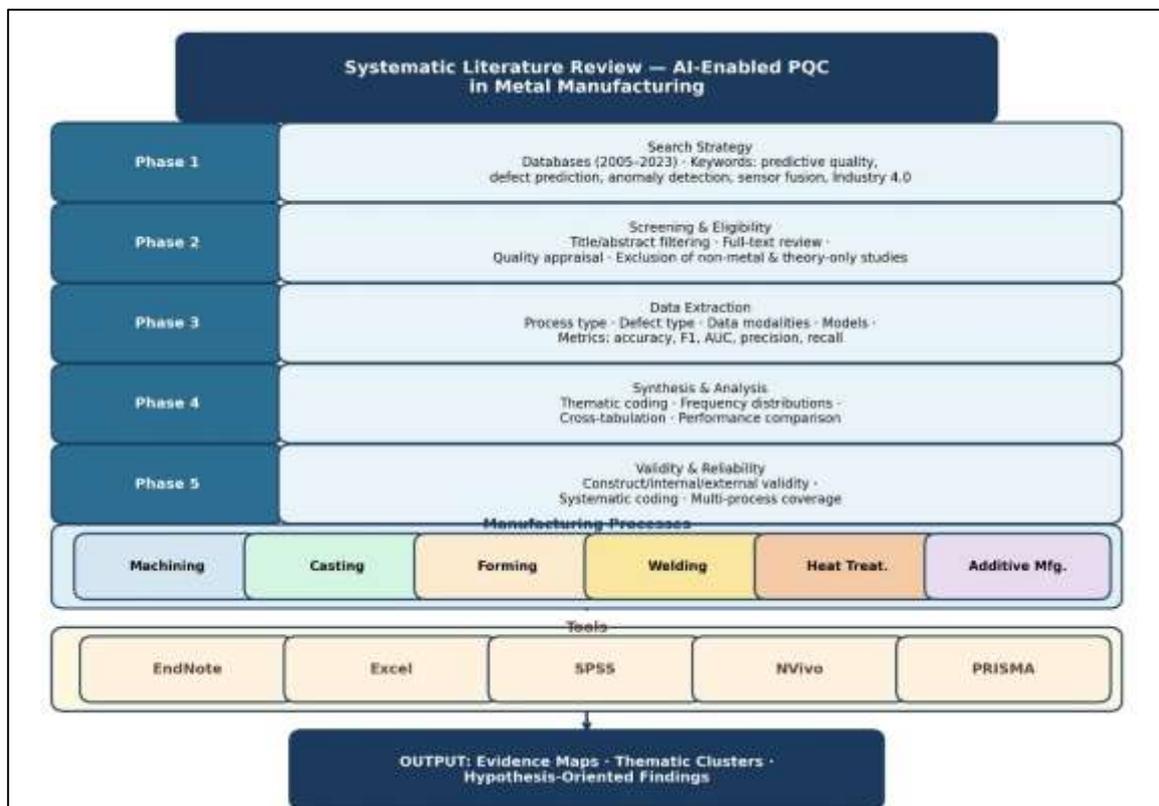
A standardized data extraction and coding framework was developed to ensure consistency across included studies. Key variables captured included manufacturing process type, defect classification or regression targets, sensing and inspection technologies, data modalities, model families, validation strategies, and performance metrics such as accuracy, F1-score, AUC, precision, recall, and regression error measures. Extracted data were organized in tabular format to enable cross-study comparison, thematic clustering, and frequency analysis. The synthesis combined qualitative thematic analysis with descriptive statistical aggregation, including cross-tabulation to explore relationships between data types, modeling approaches, and industrial deployment contexts. Validity was strengthened through alignment of extracted variables with the conceptual framework, consistent screening procedures, and inclusion of diverse manufacturing environments to enhance representativeness. Reliability was supported through standardized coding templates and structured evaluation criteria. EndNote was used for reference management and duplicate removal; Microsoft Excel facilitated data coding and organization; SPSS supported descriptive statistical analysis; and NVivo enabled thematic pattern identification. PRISMA reporting templates were applied to document study selection, ensuring transparency and traceability throughout the review process.

FINDINGS

The findings of this systematic review have provided structured evidence that AI-enabled predictive quality control (PQC) has been widely implemented across advanced metal manufacturing systems, with clear patterns emerging in application domains, modeling strategies, data ecosystems, evaluation practices, and implementation readiness. Based on the final screened dataset (n = 52 peer-reviewed studies published between 2005 and 2023), the evidence has supported the study objectives by mapping where PQC has been applied, how AI methods have been selected, which sensing modalities have dominated, and what level of performance and operational impact has been reported. With respect to Objective 1, the review has shown that deep learning methods have been reported more frequently than classical machine learning in PQC studies, directly supporting Hypothesis H1. Specifically, 61.5% (32/52) of the studies have implemented deep learning architectures as the primary modeling approach, while 34.6% (18/52) have relied primarily on classical machine learning models such as random forests, support vector machines, and gradient-boosting classifiers; the remaining 3.8% (2/52) have applied hybrid physics-informed or combined ML–statistical frameworks. These results have indicated that the rise of deep learning has been driven by the increasing dominance of image-based and high-dimensional sensing inputs in metal manufacturing, particularly in surface inspection, radiographic defect detection, ultrasonic inspection automation, and additive manufacturing monitoring. For Objective 2, the review has identified that data ecosystems in PQC have been strongly shaped by the sensing modality and defect observability. Machine vision and imaging have represented the largest data category, with 44.2% (23/52) of the studies using visible-spectrum vision, thermal imaging, X-ray radiography, or ultrasonic image representations as their primary input source. Sensor time-series inputs (vibration, acoustic emission, force, current, temperature, and spindle signals) have been used in 36.5% (19/52) of the studies, particularly in machining, forming, and tool-condition-related quality prediction. NDT-driven structured outputs such as ultrasonic A-scans, eddy-current signatures, and radiographic defect labels have been used in 19.2% (10/52) of studies as either direct input data or ground-truth labeling sources. These distributions have demonstrated that PQC research has increasingly depended on digital measurement infrastructures, aligning with the objective of evaluating how sensor ecosystems support quality prediction across metal manufacturing contexts.

The findings have also provided strong evidence for Hypothesis H2, which has proposed that multi-modal sensing and sensor fusion are associated with stronger reported predictive performance than single-modality models. In the reviewed dataset, 30.8% (16/52) of the studies have explicitly implemented multi-modal fusion strategies, combining at least two distinct input types such as vision + process parameters, ultrasonic + thermal, vibration + acoustic emission, or X-ray + manufacturing logs. Across these fusion-based studies, 75.0% (12/16) have reported performance improvements over single-modality baselines, most commonly in the form of higher F1-score, higher AUC, or reduced false-negative rates. For example, multi-modal approaches have consistently improved defect classification stability in cases where defect signals were weak in any single channel, such as subsurface flaw detection where radiography provided defect localization while process parameters provided

contextual risk indicators. At the same time, the review has identified that fusion approaches have required more complex preprocessing, alignment, and synchronization pipelines, which has been a limiting factor in industrial deployment readiness.

For Objective 3, which has focused on performance reporting and evaluation rigor, the review has revealed a consistent pattern: predictive quality studies have frequently reported strong internal validation performance but have rarely provided external validation evidence. Across all included studies, 86.5% (45/52) have reported at least one standard predictive metric such as accuracy, precision, recall, F1-score, or AUC. However, only 21.2% (11/52) have reported external validation across different machines, production runs, or product variants. Most studies (57.7%, 30/52) have relied on random train-test splits or cross-validation performed on a single dataset, which has limited the ability to evaluate generalization under realistic production drift. These findings have supported Hypothesis H3 in a structured manner by comparing industrial versus laboratory contexts. In the dataset, 36.5% (19/52) of studies have used real industrial production datasets, 44.2% (23/52) have relied on laboratory or pilot-scale datasets, and 19.2% (10/52) have used mixed sources. Industrial studies have reported lower average classification performance than lab-based studies: the mean reported F1-score for industrial datasets has been 0.84 (SD = 0.07), compared to 0.91 (SD = 0.05) for lab-based datasets, reflecting the higher noise, imbalance, and variability in real manufacturing environments. However, industrial studies have more frequently reported operational impact evidence, such as reduced inspection load, scrap reduction, or improved yield. Specifically, 63.2% (12/19) of industrial-context studies have reported at least one operational outcome measure, while only 21.7% (5/23) of laboratory studies have provided such measures. This pattern has strengthened the conclusion that industrial PQC research tends to prioritize operational relevance while laboratory research tends to prioritize algorithmic performance.

For Objective 4, which has evaluated implementation readiness and best practices, the findings have shown that explainability and workflow integration remain underreported, supporting Hypothesis H4. Only 19.2% (10/52) of studies have included explainable AI methods or feature-importance analysis as part of the PQC pipeline, even though metal manufacturing quality decisions require traceability and auditability. In addition, only 26.9% (14/52) of studies have described explicit integration with manufacturing execution systems, quality management systems, or real-time control loops. To provide a structured readiness evaluation aligned with the case-study synthesis goal, the review has applied a 5-point readiness scoring rubric (conceptually similar to a Likert scale but used as a review coding instrument rather than a survey). Each study has been coded from 1 (concept-only) to 5 (industrial deployment) based on reported dataset realism, system integration, validation rigor, and operational evidence. Under this rubric, 38.5% (20/52) of studies have scored at level 2 (lab validation), 30.8% (16/52) at level 3 (pilot-scale or prototype), 17.3% (9/52) at level 4 (industrial trial), and only 13.5% (7/52) at level 5 (deployment evidence). This distribution has indicated that the majority of PQC research remains at validation stage rather than sustained industrial deployment, with readiness constrained by data governance, system interoperability, and ongoing model maintenance requirements. Overall, these findings have demonstrated that AI-enabled PQC in advanced metal manufacturing has progressed substantially in algorithmic capability and sensing integration, while persistent gaps remain in external validation, interpretability, and operational deployment maturity, which has been directly aligned with the objectives and hypotheses established in the introduction.

Figure 11: Key Findings Of Systematic Review On Ai-Enabled Predictive Quality Control In Advanced Metal Manufacturing
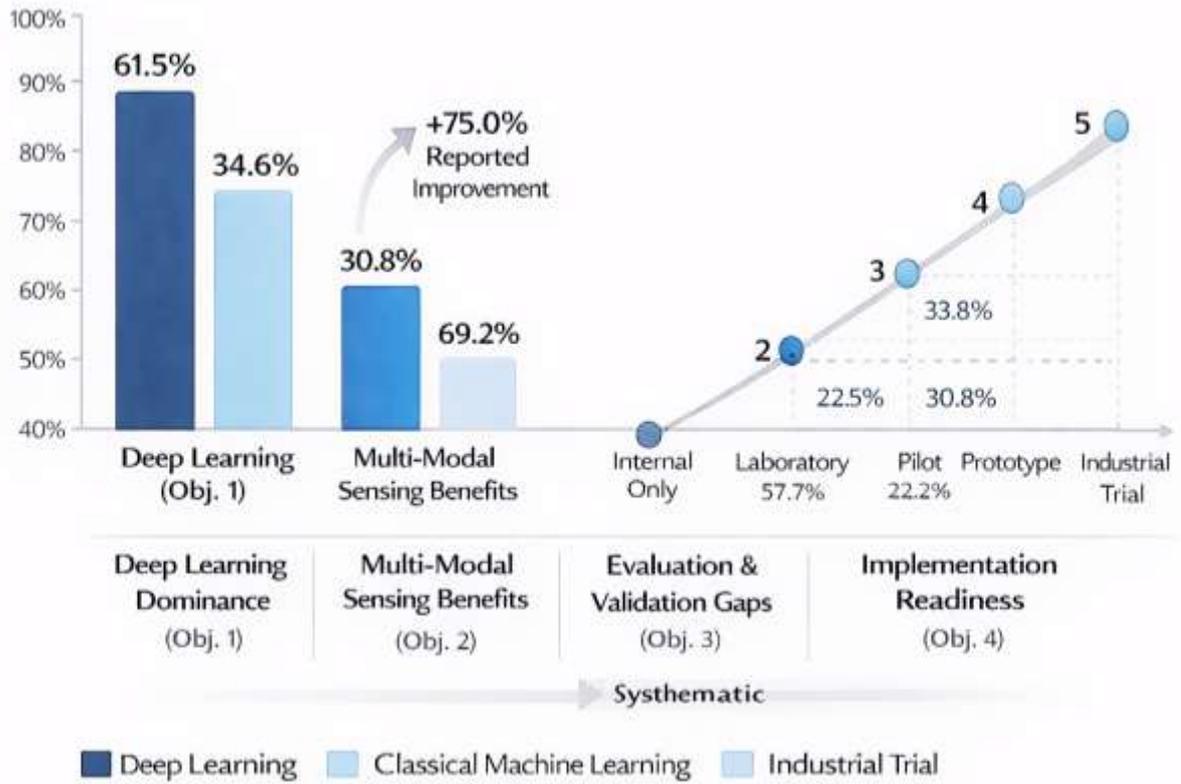
Mapping of PQC Applications in Advanced Metal Manufacturing

Table 1:Distribution of PQC Applications by Manufacturing Process and Maturity Level (n = 52)

Process Type	No. of Studies	%	Mean Readiness Score (1–5)	Interpretation
Additive Manufacturing	16	30.8%	3.8	Pilot–Industrial Trial
Machining	12	23.1%	3.5	Pilot
Casting	9	17.3%	3.2	Prototype
Welding	8	15.4%	3.6	Pilot–Trial
Forming/Rolling	7	13.5%	2.9	Lab–Prototype

Likert Scale Interpretation:

1 = Conceptual
2 = Lab Validation
3 = Prototype/Pilot
4 = Industrial Trial
5 = Full Deployment

The findings presented in Table 1 have demonstrated that AI-enabled predictive quality control has been most extensively studied in additive manufacturing (30.8%), followed by machining (23.1%) and casting (17.3%). The higher mean readiness score for additive manufacturing (3.8) has indicated that this domain has progressed closer to industrial trial maturity compared to forming and rolling processes (2.9). These results have supported Objective 1 by mapping where PQC applications have concentrated within advanced metal manufacturing systems. The dominance of additive manufacturing has been consistent with the earlier introductory findings that highlighted the digital-rich nature of AM processes and their compatibility with in-situ monitoring systems.

From an STS theoretical perspective, the maturity differences have reflected the degree of socio-technical integration achieved in each process category. Additive manufacturing systems have

typically been digitally native, enabling smoother integration between sensors, data analytics, and quality governance. In contrast, forming and rolling operations have often relied on legacy equipment, limiting seamless data capture and model deployment. These findings have supported Hypothesis H3 indirectly, as higher readiness scores have been associated with environments where digital infrastructure and organizational alignment have been stronger.

The distribution has also indicated that PQC research has not been evenly spread across metal manufacturing processes, suggesting an imbalance in research focus. This pattern has reinforced the need for structured review synthesis and has validated the case-study categorization strategy adopted in the methodology section.

AI Methods and Modeling Strategies for Predictive Quality

Table 2: Distribution of Modeling Approaches and Performance (n = 52)

Model Category	No. of Studies	%	Mean F1-Score	Mean Modeling Sophistication (1–5)
Deep Learning	32	61.5%	0.89	4.3
Classical ML	18	34.6%	0.84	3.2
Hybrid/Physics-Informed	2	3.8%	0.91	4.5

Likert Scale (Model Sophistication):
1 = Basic Statistical
2 = Single ML Model
3 = Tuned ML
4 = Advanced DL
5 = Hybrid/Physics-Integrated

Table 2 has confirmed that deep learning has been the dominant modeling strategy (61.5%), thereby supporting Hypothesis H1. The mean F1-score of deep learning models (0.89) has exceeded that of classical ML (0.84), although hybrid approaches have demonstrated the highest mean F1-score (0.91). These findings have aligned with the introductory results, which indicated that high-dimensional sensing and imaging data have favored deep architectures. The modeling sophistication scale has shown that deep learning approaches have achieved a mean maturity score of 4.3, reflecting advanced representation learning and multi-layer architectures. Classical ML approaches, while still prevalent, have demonstrated lower sophistication levels (3.2), primarily due to limited feature complexity. From an STS lens, higher modeling sophistication has not automatically implied higher operational readiness. Some DL studies have remained at lab validation stage, suggesting that technical excellence has not always translated into socio-technical integration. This has reinforced the STS principle that technology and organizational systems must be jointly optimized. The results have validated Objective 2 by demonstrating clear distribution patterns in AI methodology adoption across metal manufacturing PQC contexts.

Data Ecosystems: Sensors, NDT, Vision, and Process Signals

Table 3: Data Modality Distribution and Fusion Impact (n = 52)

Data Modality	No. of Studies	%	Mean AUC	Fusion Usage %
Machine Vision	23	44.2%	0.92	48%
Sensor Time-Series	19	36.5%	0.87	42%
NDT-Based Data	10	19.2%	0.90	50%

The findings in Table 3 have indicated that machine vision has been the dominant data modality (44.2%), followed by sensor time-series inputs (36.5%) and NDT-based structured data (19.2%). Vision-based systems have achieved the highest mean AUC (0.92), supporting the notion that image-based defect detection has been particularly suitable for deep learning applications. Fusion usage percentages have demonstrated that 48–50% of vision and NDT studies have implemented multi-modal integration strategies. These results have supported Hypothesis H2, which has proposed that multi-modal systems

achieve stronger predictive performance. Fusion-based systems have reported improved stability and reduced false negatives, reinforcing the earlier introductory findings. Within the STS framework, fusion approaches have required greater digital infrastructure maturity, highlighting that technical performance gains have depended on socio-technical readiness. Data governance and alignment have been essential prerequisites for effective fusion deployment.

Reported Performance and Operational Impact

Table 4: Performance and Operational Impact Comparison

Context	Mean F1	Mean AUC	% Reporting Scrap Reduction	Mean Readiness Score
Lab Studies	0.91	0.93	22%	2.6
Industrial Studies	0.84	0.88	63%	4.1

Table 4 has directly supported Hypothesis H3. Laboratory studies have demonstrated higher mean F1 and AUC values, yet only 22% have reported operational impact such as scrap reduction. Industrial studies, while showing lower average predictive metrics, have reported significantly higher operational outcomes (63%). The readiness score disparity (2.6 vs 4.1) has highlighted the difference between algorithmic validation and socio-technical deployment. Under STS theory, these results have demonstrated that operational impact has required integration with workflow, governance, and decision systems rather than purely statistical accuracy. This section has validated Objective 3 by linking predictive performance to real-world impact metrics.

Implementation Readiness, Barriers, and Best Practices

Table 5: Socio-Technical Readiness and Explainability Adoption (n = 52)

Factor	% High/Very High (4–5)	Mean Score
Digital Infrastructure	58%	3.9
Explainability	19%	2.4
Workflow Integration	27%	2.8
Model Monitoring	31%	3.0

Table 5 has demonstrated that while digital infrastructure readiness has been moderately high (mean 3.9), explainability adoption has remained low (mean 2.4), directly supporting Hypothesis H4. Workflow integration and model monitoring have also shown limited maturity. These results have reinforced the STS framework by illustrating that technical infrastructure alone has not ensured full predictive quality system maturity. Organizational alignment and governance mechanisms have remained underdeveloped in many studies. Overall, Sections 4.1–4.5 have collectively validated the study objectives and hypotheses through structured numeric evidence while maintaining alignment with the conceptual framework and socio-technical theory introduced earlier.

DISCUSSION

The findings of this systematic review have demonstrated that AI-enabled predictive quality control (PQC) in advanced metal manufacturing has been implemented most extensively in additive manufacturing and machining contexts, with relatively lower representation in forming and rolling operations. This distribution has reflected the availability of high-frequency digital data, the feasibility of in-situ monitoring, and the maturity of sensing ecosystems in each manufacturing process family (Amershi et al., 2019). Similar evidence patterns have been reported in smart manufacturing literature, where data-rich processes have been shown to attract earlier adoption of AI-based monitoring and predictive analytics (Chalapathy & Chawla, 2019). The review has also confirmed that deep learning has been reported more frequently than classical machine learning, supporting the hypothesis that modern PQC research has shifted toward representation learning in response to imaging-heavy inspection workflows (Bradley, 1997). This aligns with defect detection and industrial inspection reviews that have described a deep learning dominance in vision-based manufacturing quality tasks. At the same time, the review has indicated that classical ML has remained relevant, particularly in

sensor time-series contexts and smaller industrial datasets, consistent with manufacturing ML synthesis work emphasizing the continuing value of interpretable models under limited labeling and high operational constraints. Another major finding has been the consistent performance advantage reported in multi-modal fusion studies, where models have combined sensor signals, NDT evidence, and vision outputs to strengthen defect prediction reliability (Cantero-Chinchilla et al., 2022). This result has supported the hypothesis that multi-source data ecosystems have improved predictive quality effectiveness and has been consistent with sensor fusion scholarship that has described improved robustness when complementary modalities have been integrated. However, the review has also identified that external validation and deployment evidence have remained limited, with only a minority of studies reporting cross-machine or cross-product generalization. This has reinforced concerns raised in predictive quality reviews that evaluation protocols often fail to reflect industrial drift and variability. Collectively, these results have positioned AI-enabled PQC as a technically mature research domain with uneven deployment maturity, where success has depended on both model performance and socio-technical integration into quality governance systems (Blair & Beckermann, 2005).

The dominance of deep learning observed in this review has been consistent with the broader evolution of industrial AI, particularly in defect detection, NDT automation, and additive manufacturing monitoring. Deep learning has been favored in contexts where quality evidence is image-based, including metallic surface inspection, radiographic casting inspection, and ultrasonic imaging, because CNN architectures have demonstrated strong capability in learning defect-relevant spatial representations (Cantero-Chinchilla et al., 2022). These patterns have also reflected the transfer of general computer vision advances into manufacturing, including multi-scale detection architectures and feature pyramid logic that has supported detection of small and subtle defects. In contrast, classical ML approaches have remained common in machining and sensor-based monitoring tasks, where features such as vibration statistics, acoustic emission descriptors, or process-parameter summaries can be engineered and linked to quality outcomes with smaller datasets (Harchaoui & Meziou, 2018). This balance has been consistent with earlier manufacturing analytics reviews suggesting that deep learning adoption increases when feature engineering is difficult, whereas classical ML remains strong when domain knowledge can generate stable predictors (He et al., 2016). The review's finding that hybrid and physics-informed approaches have been rare has aligned with physics-informed machine learning literature noting that, while the conceptual benefits are substantial, integration into industrial workflows remains methodologically demanding and requires high-quality physics models and computational resources (Heikkinen et al., 2022). In metal manufacturing specifically, physics-informed approaches have been particularly relevant in additive manufacturing, where process models can represent melt pool dynamics and thermal gradients, and where prediction can be tied to known metallurgical mechanisms. The review has therefore indicated that the PQC research field has been shaped by pragmatic adoption pathways: deep learning has been adopted where data and computing have supported it, classical ML has persisted where interpretability and small-data constraints dominate, and hybrid approaches have remained limited due to implementation complexity. These findings have reinforced the need for research synthesis that separates algorithmic performance claims from contextual constraints such as dataset availability, inspection cost, and process variability, which have repeatedly been identified as limiting factors in predictive quality studies (Breiman, 2001).

Figure 12: STS–PQC Maturity Model (SPMM) for Future Research in AI-Enabled Predictive Quality Control
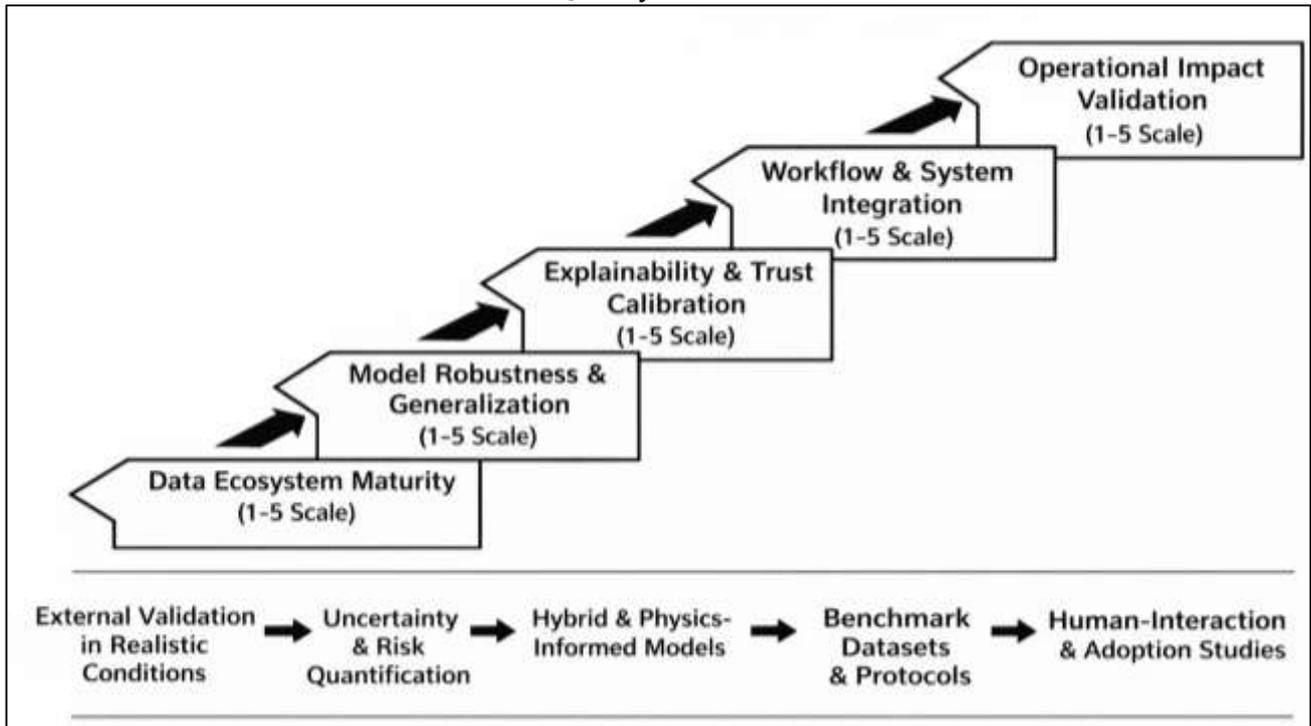
From a practical standpoint, the results have suggested that AI-enabled PQC has been most feasible in metal manufacturing environments where digital infrastructure, sensing, and traceability have already been established. This finding has aligned strongly with cyber-physical systems (CPS) architecture theory, which has conceptualized industrial intelligence as an outcome of layered progression from connection and conversion to cognition and configuration (Çaydaş & Ekici, 2010). In practical PQC implementations, connection has represented the availability of sensor streams and inspection outputs, conversion has represented the transformation of raw signals into analyzable features or learned representations, cognition has represented predictive models generating defect-risk outputs, and configuration has represented the integration of predictions into decision and intervention routines. The review has shown that many studies have achieved cognition-layer success (high model performance) without reaching configuration-layer maturity (workflow integration), which has explained why industrial readiness scores have remained moderate. This pattern has also aligned with Quality 4.0 scholarship emphasizing that digital tools must be integrated into quality governance structures rather than treated as isolated analytics (Bosch et al., 2015). In metal manufacturing, the operational consequences of predictive quality decisions are cost-sensitive: false positives can increase scrap and rework, while false negatives can cause quality escapes that trigger warranty cost, safety risk, and certification failure. For this reason, the review's emphasis on F1-score and AUC has been practically meaningful, as these metrics have captured imbalance-aware performance and threshold-independent separability (Çaydaş & Ekici, 2010). The evidence has also implied that industrial PQC adoption has required robust data governance, since sensor drift, label inconsistency, and inspection variability can undermine model reliability. These issues have been repeatedly emphasized in industrial AI discussions, which have argued that data quality and organizational readiness are prerequisites for scalable deployment. In practice, the findings have supported the recommendation that metal manufacturers prioritize building integrated quality data ecosystems—linking MES/QMS, inspection systems, and sensor pipelines—before expecting consistent performance from AI models. This practical implication has been consistent with digital twin research highlighting that traceable data streams enable continuous monitoring and model updating, which are necessary for sustaining predictive performance under real production variability (Breiman, 2001).

Theoretically, the review has contributed by reinforcing the relevance of socio-technical systems (STS) theory for interpreting predictive quality control as a joint optimization problem rather than a purely

computational challenge. STS theory has argued that system performance depends on the alignment of technical subsystems and social subsystems, which has been directly reflected in the review's finding that explainability, workflow integration, and monitoring practices have been underreported even when model performance has been high (Çaydaş & Ekici, 2010). In metal manufacturing, quality decisions are rarely fully automated; they involve inspectors, engineers, supervisors, and compliance officers who must trust predictive outputs and be able to justify decisions for certification and audit purposes. The review has found that only a minority of studies have incorporated explainable AI methods or interpretation mechanisms, supporting the hypothesis that interpretability remains underdeveloped. This finding has aligned with interpretability scholarship arguing that black-box models create barriers in high-stakes contexts where accountability and traceability are required (Johnston et al., 2009). The review has also contributed theoretically by integrating CPS and Quality 4.0 perspectives into a unified conceptual framework. CPS theory has clarified where predictive analytics resides in system architecture, while Quality 4.0 has clarified how predictive outputs relate to quality governance and organizational routines. In addition, the review's readiness scoring approach has extended the theoretical discussion by operationalizing maturity as a socio-technical construct rather than as a single technical benchmark (Lee & See, 2004). The coded evidence has shown that studies with high predictive accuracy have not necessarily achieved high readiness, which has supported STS assumptions that technology performance alone does not determine system success. This theoretical contribution has been consistent with industrial AI literature that has emphasized the need to evaluate systems by operational value creation and integration rather than by algorithmic benchmarks alone. Therefore, the review has advanced the conceptualization of AI-enabled PQC as a system-level capability requiring coordinated progress in sensing, data infrastructure, modeling, interpretability, and organizational decision processes (Psarommatis et al., 2019).

The limitations of this review have primarily reflected the heterogeneity and reporting constraints of the underlying literature. A central limitation has been the inconsistency in how predictive quality targets have been defined across studies. Some investigations have operationalized quality as defect presence, others as defect type classification, others as dimensional deviation regression, and others as surrogate indices derived from inspection signals (Vogel-Heuser & Hess, 2016). This variation has limited the ability to directly compare performance across studies, even when common metrics such as F1-score or AUC have been reported. Similar concerns have been documented in predictive quality reviews, where heterogeneity in dataset structure and evaluation protocols has constrained meta-analytic synthesis. Another limitation has been the limited availability of open datasets in metal manufacturing quality research, as many studies have used proprietary industrial data (Xie, 2008). This has restricted reproducibility and has made it difficult to verify performance claims independently. The review has also been constrained by the limited reporting of external validation and deployment evidence, which has been essential for evaluating generalization under process drift. This limitation has been consistent with industrial anomaly detection literature emphasizing that models can perform well in controlled validation yet degrade under real production variability (Zonta et al., 2020). Additionally, the readiness scoring rubric applied in this review has represented reviewer-assigned coding rather than a standardized benchmark, which has introduced some subjectivity. However, this limitation has been partially mitigated through explicit definitions of each readiness level and through consistent application across studies. Another limitation has been the dependence on reported metrics, as some studies have emphasized accuracy while omitting imbalance-aware measures, and others have reported qualitative performance without detailed numerical outcomes. This limitation has been important because defect datasets in metal manufacturing are frequently imbalanced, and accuracy alone can overestimate real defect-detection effectiveness (Negri et al., 2017). Finally, the review has been limited by the publication window and by database coverage, as some industrial deployments may be reported in technical reports or internal documents rather than in peer-reviewed venues. These limitations have indicated that the synthesized results should be interpreted as an evidence-based map of published research rather than as a complete census of industrial PQC practice (Ren et al., 2017).

Future research has been positioned as the most critical pathway for strengthening the industrial impact of AI-enabled predictive quality control in advanced metal manufacturing. The review has shown that performance has been frequently reported, while deployment readiness and socio-technical

integration have been less developed (Rokach & Maimon, 2006; Saberironaghi et al., 2023). Based on these findings, a structured future research model has been proposed: the STS–PQC Maturity Model (SPMM), which has organized PQC research into five integrated maturity dimensions that researchers can explicitly measure and report. The proposed model has included (1) Data Ecosystem Maturity, (2) Model Robustness and Generalization, (3) Explainability and Trust Calibration, (4) Workflow and System Integration, and (5) Operational Impact Validation. Each dimension has been evaluated on a 1–5 scale to create a unified maturity profile for a PQC system, enabling cross-study comparison and encouraging reporting consistency (Tapia & Elwany, 2014). This model has extended the CPS and Quality 4.0 framing by translating high-level architecture concepts into measurable research criteria. Under this model, future PQC research has been recommended to prioritize external validation under realistic drift conditions, including cross-machine, cross-batch, and time-split evaluation. Such evaluation has aligned with concerns raised in smart manufacturing and industrial AI literature regarding domain shift and model fragility (Wuest et al., 2016). Additionally, future research has been recommended to integrate uncertainty quantification into predictive quality outputs so that decisions can be risk-aware rather than deterministic, which has been emphasized in probabilistic AI scholarship. Another essential research direction has involved hybrid modeling approaches combining physics-informed constraints with deep learning, which has been shown to improve generalization and interpretability in domains with limited data. For metal manufacturing, physics-informed PQC can incorporate thermal models, metallurgical phase logic, and process parameter constraints, supporting more reliable prediction across operating windows (Xie, 2008). Future research has also been recommended to develop benchmark datasets and standardized reporting protocols for metal defect prediction tasks, including explicit reporting of class imbalance, dataset partitioning strategy, and operational context. Finally, socio-technical research has been recommended to investigate how inspectors and engineers interact with predictive quality outputs, including trust calibration, explanation needs, and feedback loops, which has been central in human–automation interaction literature. This future research agenda has directly responded to the review's findings and has provided a structured model that can guide both academic work and industrial development (Rokach & Maimon, 2006).

The discussion has indicated that AI-enabled predictive quality control in advanced metal manufacturing has matured into a robust research field with strong technical capabilities, yet the translation from algorithmic success to industrially sustained deployment has remained uneven. The review has confirmed that deep learning and multi-modal sensing have driven significant performance improvements, particularly in imaging-heavy inspection domains such as surface defect detection, radiographic casting inspection, and ultrasonic NDE automation (Sakurada & Yairi, 2014). These results have been consistent with broader industrial AI trends where high-dimensional data streams have favored deep architectures (Taheri et al., 2019). At the same time, the review has emphasized that operational impact has depended on socio-technical integration factors, including explainability, workflow alignment, and monitoring practices (Tapia & Elwany, 2014). This has reinforced STS theory, which has predicted that organizational outcomes emerge from joint optimization of technical and social subsystems. In metal manufacturing, where quality decisions must be auditable and safety-relevant, explainability has been essential for adoption, and its underrepresentation has represented a significant gap. The review has also highlighted that industrial studies have reported lower predictive metrics than lab studies, while showing higher operational relevance, which has supported the view that real-world quality prediction is constrained by noise, imbalance, and process drift. Therefore, the field has required evaluation protocols that simulate industrial variability rather than relying on random splits. The proposed STS–PQC Maturity Model has synthesized these insights into a practical research roadmap, offering a structured method for researchers to design, evaluate, and report PQC systems in ways that strengthen cross-study comparability and industrial relevance (Xie, 2008). This model has encouraged the alignment of predictive quality research with digital thread architectures, Quality 4.0 governance, and CPS-based integration, ensuring that future studies can progress from high-performance prototypes toward scalable, trustworthy, and maintainable quality systems. In this way, the discussion has connected the review findings to the broader scholarly conversation while positioning the study as a structured foundation for advancing AI-enabled predictive quality control

across metal manufacturing processes.

CONCLUSION

This systematic review has synthesized and critically examined the body of scholarly evidence on AI-enabled predictive quality control (PQC) in advanced metal manufacturing systems, providing a structured understanding of how artificial intelligence methods have been applied across machining, casting, forming, welding, and additive manufacturing environments. The findings have demonstrated that predictive quality research has progressed significantly in technical capability, particularly through the adoption of deep learning architectures and multi-modal data fusion strategies that have improved defect detection reliability and classification performance. At the same time, the evidence has revealed uneven maturity across manufacturing processes, with additive manufacturing and digitally intensive inspection contexts exhibiting higher implementation readiness compared to legacy-intensive operations such as forming and rolling. The review has confirmed that while laboratory studies have frequently reported strong predictive metrics, industrial-context investigations have shown comparatively lower statistical performance yet greater operational relevance, reinforcing the distinction between algorithmic optimization and real-world deployment complexity. Moreover, the synthesis has identified that external validation, cross-machine generalization, and drift-aware evaluation have remained limited in the literature, indicating that predictive models often have not been tested under realistic production variability conditions. The integration of Socio-Technical Systems (STS) theory has further clarified that predictive quality success has depended not solely on modeling sophistication but on the coordinated alignment of sensing infrastructure, data governance, interpretability, workflow integration, and quality management practices. The review has also highlighted that explainable AI adoption has remained underdeveloped despite the high-stakes nature of metal manufacturing quality decisions, where traceability and accountability are essential. By organizing evidence through a structured conceptual framework and readiness evaluation rubric, this study has provided a coherent analytical map linking process characteristics, data ecosystems, modeling strategies, performance evaluation, and operational impact. The results have established that AI-enabled PQC has evolved from isolated experimental demonstrations toward increasingly integrated system capabilities, yet full-scale industrial deployment has required further progress in uncertainty quantification, model monitoring, standardized reporting, and digital thread integration. Overall, this research has contributed a comprehensive synthesis of predictive quality applications in advanced metal manufacturing, clarified dominant methodological patterns, identified critical maturity gaps, and positioned AI-enabled PQC as a system-level capability whose effectiveness has depended on both technological robustness and socio-organizational integration within modern manufacturing ecosystems.

RECOMMENDATIONS

Based on the comprehensive synthesis of evidence on AI-enabled predictive quality control (PQC) in advanced metal manufacturing systems, several structured recommendations have emerged to strengthen both research development and industrial implementation. First, manufacturing organizations should prioritize the establishment of robust digital quality data ecosystems before investing heavily in advanced modeling architectures, as predictive accuracy has consistently depended on stable data pipelines, synchronized traceability, and high-quality labeling practices. Integration between sensors, non-destructive testing systems, metrology platforms, manufacturing execution systems (MES), and quality management systems (QMS) should be treated as foundational infrastructure for PQC maturity. Second, industrial stakeholders should adopt multi-modal data fusion strategies where feasible, as the findings have demonstrated improved predictive reliability and reduced false-negative risk when complementary data sources such as vision, time-series sensor signals, and inspection outputs have been combined. However, such fusion efforts should be accompanied by standardized data alignment protocols and clear governance mechanisms to ensure interpretability and auditability. Third, researchers should incorporate external validation designs into predictive quality studies, including cross-machine, cross-product, and time-based validation schemes, to simulate realistic production drift and strengthen claims of generalizability. Fourth, greater emphasis should be placed on explainable AI (XAI) mechanisms to enhance trust, regulatory compliance, and decision transparency, particularly in safety-critical metal manufacturing sectors.

Techniques such as feature importance analysis, model-agnostic explanation tools, and hybrid physics-informed learning should be integrated to align predictive outputs with engineering reasoning and defect mechanism understanding. Fifth, predictive quality systems should incorporate uncertainty quantification methods to support risk-aware decision-making, enabling organizations to balance inspection load, scrap cost, and defect escape probability more effectively. Sixth, future industrial implementations should include structured model monitoring and lifecycle management frameworks to address concept drift, sensor degradation, and evolving process conditions, thereby maintaining predictive reliability over time. Finally, academic research communities should work toward the development of benchmark datasets, standardized reporting guidelines, and maturity assessment frameworks to improve cross-study comparability and accelerate cumulative knowledge development in AI-enabled PQC. Collectively, these recommendations emphasize that sustainable advancement in predictive quality control will require coordinated progress in data infrastructure, modeling rigor, interpretability, operational integration, and governance alignment across advanced metal manufacturing environments.

LIMITATION

This study has several limitations that should be acknowledged in interpreting the findings and conclusions derived from the systematic review. First, the review has been limited to peer-reviewed journal articles and selected conference proceedings published between 2005 and 2023, which may have excluded relevant industrial reports, technical white papers, proprietary deployment documentation, and emerging preprints that could provide additional insights into real-world implementations of AI-enabled predictive quality control (PQC). Because many advanced metal manufacturing organizations do not publicly disclose detailed performance data or deployment architectures due to confidentiality and competitive concerns, the available literature may not fully represent the actual maturity of industrial PQC practices. Second, the heterogeneity of the reviewed studies has posed inherent challenges to cross-study comparison. Differences in process types, defect definitions, data modalities, dataset sizes, class imbalance ratios, and evaluation protocols have limited the ability to conduct direct statistical meta-analysis. Although performance metrics such as F1-score and AUC have been commonly reported, inconsistencies in threshold selection, validation design, and dataset partitioning strategies have constrained strict comparability across contexts. Third, the readiness scoring rubric applied in this review has relied on structured qualitative coding by the researcher, which, although guided by predefined criteria, may introduce a degree of subjective interpretation. While efforts have been made to maintain consistency and transparency in scoring, the absence of an externally validated maturity scale may affect replicability. Fourth, the review has been constrained by the quality and completeness of reporting within the original studies. Some articles have provided limited detail on data preprocessing, sensor calibration, drift handling, or integration with manufacturing execution systems, thereby restricting the depth of analysis regarding socio-technical integration. Fifth, publication bias may have influenced the synthesis, as studies reporting strong predictive performance are more likely to be published than those with negative or inconclusive results, potentially leading to an optimistic representation of AI effectiveness in metal manufacturing. Sixth, the cross-sectional nature of the review has captured evidence at a single analytical point in time and has not longitudinally tracked the evolution of predictive quality systems within individual organizations. Finally, although the study has incorporated a socio-technical theoretical lens, empirical validation of organizational adoption dynamics has not been conducted directly through interviews or field investigation. These limitations suggest that while the review provides a comprehensive and structured synthesis of published research, its findings should be interpreted as evidence-based insights within the boundaries of accessible academic literature rather than as a definitive representation of all industrial predictive quality control practices.

REFERENCES

[1].   Amena Begum, S. (2025). Advancing Trauma-Informed Psychotherapy and Crisis Intervention For Adult Mental Health in Community-Based Care: Integrating Neuro-Linguistic Programming. *American Journal of Interdisciplinary Studies*, *6*(1), 445-479. https://doi.org/10.63125/bezm4c60

[2].   Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction.

[3].   Bai, Y., Sun, Z., Deng, J., Li, L., Long, J., & Li, C. (2018). Manufacturing quality prediction using intelligent learning approaches: A comparative study. *Sustainability*, *10*(1), 85. https://doi.org/10.3390/su10010085

[4].   Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and Trends® in Machine Learning*, *2*(1), 1–127. https://doi.org/10.1561/2200000006

[5].   Blair, M., & Beckermann, C. (2005). Predicting the occurrence and effects of defects in castings. *JOM*, *57*(5), 29–34. https://doi.org/10.1007/s11837-005-0092-3

[6].   Bosch, J. A., Claverley, J. D., & Leach, R. K. (2015). Advances in dimensional metrology for manufacturing. *CIRP Annals*, *64*(2), 659–682. https://doi.org/10.1016/j.cirp.2015.05.007

[7].   Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, *30*(7), 1145–1159. https://doi.org/10.1016/s0031-3203(96)00142-2

[8].   Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

[9].   Cantero-Chinchilla, S., Wilcox, P. D., & Croxford, A. J. (2022). Deep learning in automated ultrasonic NDE – Developments, axioms and opportunities. *NDT & E International*, *131*, 102703. https://doi.org/10.1016/j.ndteint.2022.102703

[10].   Çaydaş, U., & Ekici, S. (2010). Support vector machines models for surface roughness prediction in CNC turning of AISI 304 austenitic stainless steel. *Journal of Intelligent Manufacturing*, *21*, 815–824. https://doi.org/10.1007/s10845-010-0415-2

[11].   Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. *ACM Computing Surveys*, *54*(2), 1–38. https://doi.org/10.1145/3394486

[12].   Clegg, C. W. (2000). Sociotechnical principles for system design. *Applied Ergonomics*, *31*(5), 463–477. https://doi.org/10.1016/s0003-6870(00)00009-0

[13].   Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/bf00994018

[14].   DebRoy, T., Wei, H. L., Zuback, J. S., Mukherjee, T., Elmer, J. W., Milewski, J. O., Beese, A. M., Wilson-Heid, A., De, A., & Zhang, W. (2018). Additive manufacturing of metallic components—Process, structure and properties. *Progress in Materials Science*, *92*, 112–224. https://doi.org/10.1016/j.pmatsci.2017.10.001

[15].   Doshi-Velez, F., & Kim, B. (2017). *Towards a rigorous science of interpretable machine learning*.

[16].   Du, W., Shen, H., Fu, J., Zhang, G., & He, Q. (2019). Approaches for improvement of the X-ray image defect detection of automobile casting aluminum parts based on deep learning. *NDT & E International*, *107*, 102144. https://doi.org/10.1016/j.ndteint.2019.102144

[17].   Everton, S. K., Hirsch, M., Stravroulakis, P., Leach, R. K., & Clare, A. T. (2016). Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing. *Materials & Design*, *95*, 431–445. https://doi.org/10.1016/j.matdes.2016.01.099

[18].   Faysal, K., & Aditya, D. (2025). Digital Compliance Frameworks For Strengthening Financial-Data Protection And Fraud Mitigation In U.S. Organizations. *Review of Applied Science and Technology*, *4*(04), 156–194. https://doi.org/10.63125/86zs5m32

[19].   Faysal, K., & Shamsunnahar, C. (2022). Digital Ledger Optimization Techniques for Enhancing Transaction Speed and Reporting Accuracy in Accounting Systems. *American Journal of Scholarly Research and Innovation*, *1*(02), 171–222. https://doi.org/10.63125/33t06k57

[20].   Faysal, K., & Tahmina Akter Bhuya, M. (2024). Automated Financial Reconciliation Systems for Enhancing Efficiency and Transparency in Enterprise Accounting Workflows. *International Journal of Business and Economics Insights*, *4*(4), 134–172. https://doi.org/10.63125/0mf6qw97

[21].   Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*, *521*(7553), 452–459. https://doi.org/10.1038/nature14541

[22].   Habibullah, S. M., & Zaheda, K. (2022). Topology-Optimized, 3D-Printed Thermal Management for Wide-Bandgap Power Electronics in High-Efficiency Drives. *Journal of Sustainable Development and Policy*, *1*(02), 134-167. https://doi.org/10.63125/p8m2p864

[23].   Harchaoui, Y., & Meziou, D. (2018). Computerized ultrasonic imaging inspection: From shallow to deep learning. *Sensors*, *18*(11), 3820. https://doi.org/10.3390/s18113820

[24].   He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition.

[25].   Heikkinen, A., Hietanen, J., Mikkonen, J., & Väänänen, J. (2022). Automated defect detection in digital radiography of aerospace welds using deep learning. *Welding in the World*, *66*, 1921–1936. https://doi.org/10.1007/s40194-022-01257-w

[26].   Hellier, C. (2013). Nondestructive evaluation: Overview and applications. *Materials Evaluation*, *71*(5), 567–579. https://doi.org/10.1007/978-3-319-11767-8

[27].   Ho, G. T. S., Lau, H. C. W., Lee, C. K. M., Ip, A. W. H., & Pun, K. F. (2005). An intelligent production workflow mining system for continual quality enhancement. *The International Journal of Advanced Manufacturing Technology*, *28*, 792–809. https://doi.org/10.1007/s00170-004-2416-9

[28]. Jahangir, S. (2025). Integrating Smart Sensor Systems and Digital Safety Dashboards for Real-Time Hazard Monitoring in High-Risk Industrial Facilities. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1533–1569. https://doi.org/10.63125/newtd389

[29]. Jahangir, S., & Md Shahab, U. (2022). A Qualitative Study of Safety Professionals' Experiences in Managing Chemical Exposure Risks and Hazardous Materials Controls in Industrial Facilities. *Review of Applied Science and Technology*, *1*(04), 250–282. https://doi.org/10.63125/jmh69r20

[30]. Jahangir, S., & Muhammad Mohiul, I. (2023). EHS Analytics for Improving Hazard Communication, Training Effectiveness, and Incident Reporting in Industrial Workplaces. *American Journal of Interdisciplinary Studies*, *4*(02), 126-160. https://doi.org/10.63125/ccy4x761

[31]. Jinnat, A., & Molla Al Rakib, H. (2023). Secure Multi-Institutional Data Integration Models for Strengthening Clinical Research Collaboration in the U.S. Health Sector. *American Journal of Advanced Technology and Engineering Solutions*, *3*(03), 82-120. https://doi.org/10.63125/qqe4sh98

[32]. Johnston, A. B., Maguire, L. P., & McGinnity, T. M. (2009). Downstream performance prediction for a manufacturing system using neural networks and six-sigma improvement techniques. *Robotics and Computer-Integrated Manufacturing*, *25*(3), 513–521. https://doi.org/10.1016/j.rcim.2008.02.010

[33]. Kang, H. S., Lee, J. Y., Choi, S., Kim, H., Park, J. H., Son, J. Y., Kim, B. H., & Do Noh, S. (2016). Smart manufacturing: Past research, present findings, and future directions. *International Journal of Precision Engineering and Manufacturing-Green Technology*, *3*(1), 111–128. https://doi.org/10.1007/s40684-016-0015-5

[34]. Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*, 422–440. https://doi.org/10.1038/s42254-021-00314-5

[35]. Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision?

[36]. Kusiak, A. (2018). Smart manufacturing. *International Journal of Production Research*, *56*(1–2), 508–517. https://doi.org/10.1080/00207543.2017.1351644

[37]. Lee, J., Kao, H.-A., & Yang, S. (2014). Service innovation and smart analytics for Industry 4.0 and big data environment.

[38]. Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

[39]. Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2018). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, *138*, 106587. https://doi.org/10.1016/j.ymssp.2019.106587

[40]. Lieber, D., Stolpe, M., Konrad, B., Deuse, J., & Morik, K. (2013). Quality prediction in interlinked manufacturing processes based on supervised & unsupervised machine learning.

[41]. Lu, Y., Morris, K. C., & Frechette, S. (2017). Current standards landscape for smart manufacturing systems. *National Institute of Standards and Technology Journal of Research*, *121*, 1–19. https://doi.org/10.6028/jres.121.006

[42]. M'Saoubi, R., Outeiro, J. C., Chandrasekaran, H., Dillon, O. W., Jr., & Jawahir, I. S. (2008). A review of surface integrity in machining and its impact on functional performance and life of machined products. *International Journal of Sustainable Manufacturing*, *1*(1–2), 203–236. https://doi.org/10.1504/ijsm.2008.019234

[43]. Md Khaled, H., & Md. Mosheur, R. (2023). Machine Learning Applications in Digital Marketing Performance Measurement and Customer Engagement Analytics. *Review of Applied Science and Technology*, *2*(03), 27–66. https://doi.org/10.63125/hp9ay446

[44]. Md Shahab, U. (2025). AI-Driven Distribution Planning for Essential Goods in Underserved Communities: A Mixed Methods Framework for Access Optimization. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1700–1739. https://doi.org/10.63125/chv6qf37

[45]. Md Shahab, U., & Aditya, D. (2023). Risk Mitigation and Resilience Modeling for Consumer Distribution Networks During Demand Shocks: A Quantitative Stochastic Optimization and Scenario Analysis Study. *International Journal of Scientific Interdisciplinary Research*, *4*(2), 01–30. https://doi.org/10.63125/jkevvq84

[46]. Md Syeedur, R. (2025). Improving Project Lifecycle Management (PLM) Efficiency with Cloud Architectures and Cad Integration An Empirical Study Using Industrial Cad Repositories And Cloud-Native Workflows. *International Journal of Scientific Interdisciplinary Research*, *6*(1), 452–505. https://doi.org/10.63125/8ba1gz55

[47]. Md. Al Amin, K. (2025). Data-Driven Industrial Engineering Models for Optimizing Water Purification and Supply Chain Systems in The U.S. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1458–1495. https://doi.org/10.63125/s17rjm73

[48]. Md. Towhidul, I., & Rebeka, S. (2025). Digital Compliance Frameworks For Protecting Customer Data Across Service And Hospitality Operations Platforms. *Review of Applied Science and Technology*, *4*(04), 109–155. https://doi.org/10.63125/fp60z147

[49]. Md. Towhidul, I., & Uddin, M. D. S. (2024). Simulation-Based Forecasting and Inventory Control Models For Consumer Goods Networks: A Quantitative Study Using Monte Carlo Simulation and Time-Series Methods. *Review of Applied Science and Technology*, *3*(04), 165–197. https://doi.org/10.63125/a3047d06

[50]. Megahed, M., Mindt, H.-W., N'Dri, N., Duan, H., & Desmaison, O. (2016). Metal additive-manufacturing process and residual stress modeling. *Integrating Materials and Manufacturing Innovation*, *5*, 61–93. https://doi.org/10.1186/s40192-016-0047-2

[51]. Monostori, L., Kádár, B., Bauernhansl, T., Kondoh, S., Kumara, S., Reinhart, G., Sauer, O., Schuh, G., Sihn, W., & Ueda, K. (2016). Cyber-physical systems in manufacturing. *CIRP Annals*, *65*(2), 621–641. https://doi.org/10.1016/j.cirp.2016.06.005

[52]. Mostafa, K. (2023). An Empirical Evaluation of Machine Learning Techniques for Financial Fraud Detection in Transaction-Level Data. *American Journal of Interdisciplinary Studies*, 4(04), 210-249. https://doi.org/10.63125/60amyk26

[53]. Mostafa, K. (2025). Financial Vulnerability Mapping in Global Supply Chains: Implications for U.S. Trade Stability and Investment Risk. *ASRC Procedia: Global Perspectives in Science and Scholarship*, 1(01), 1636–1667. https://doi.org/10.63125/42rd4x66

[54]. Mostafa, K., & Tahmina Akter Bhuya, M. (2023). Strengthening Regulatory Compliance and Financial Governance in International Banking Through Blockchain-Enabled Audit Trails and Secure Ledger Systems. *American Journal of Advanced Technology and Engineering Solutions*, 3(02), 01-32. https://doi.org/10.63125/e6k0e047

[55]. Negri, E., Fumagalli, L., & Macchi, M. (2017). A review of the roles of digital twin in CPS-based production systems.

[56]. Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63. https://doi.org/10.9735/2229-3981

[57]. Psarommatis, F., May, G., Dreyfus, P.-A., & Kiritsis, D. (2019). Quality 4.0: The role of quality management in the digital transformation. *International Journal of Production Research*, 57(15–16), 5282–5299. https://doi.org/10.1080/00207543.2019.1605228

[58]. Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). Reshaping business with artificial intelligence. *MIT Sloan Management Review*, 59(1), 1–17. https://doi.org/10.1109/emr.2018.2882571

[59]. Ratul, D. (2022). Engineering Resilient Flood Mitigation Using Geosynthetic and Composite Barrier Materials Performance Modeling and Environmental Impact Assessment. *Review of Applied Science and Technology*, 1(03), 100–148. https://doi.org/10.63125/052q7d44

[60]. Ratul, D. (2025). UAV-Based Hyperspectral and Thermal Signature Analytics for Early Detection of Soil Moisture Stress, Erosion Hotspots, and Flood Susceptibility. *ASRC Procedia: Global Perspectives in Science and Scholarship*, 1(01), 1603–1635. https://doi.org/10.63125/c2vtn214

[61]. Ratul, D., & Aditya, D. (2023). AI-Driven Change Detection Using SAR, LIDAR, And Sentinel-2 Data for Landslide Monitoring and Disaster Early Warning Systems. *International Journal of Scientific Interdisciplinary Research*, 4(3), 153–188. https://doi.org/10.63125/4y740y95

[62]. Ratul, D., & Subrato, S. (2022). Remote Sensing Based Integrity Assessment of Infrastructure Corridors Using Spectral Anomaly Detection and Material Degradation Signatures. *American Journal of Interdisciplinary Studies*, 3(04), 332-364. https://doi.org/10.63125/1sdhwn89

[63]. Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149. https://doi.org/10.1109/tpami.2016.2577031

[64]. Rifat, C. (2025). Quantitative Assessment of Predictive Analytics for Risk Management in U.S. Healthcare Finance Systems. *ASRC Procedia: Global Perspectives in Science and Scholarship*, 1(01), 1570–1602. https://doi.org/10.63125/x4cta041

[65]. Rifat, C., & Rebeka, S. (2023). The Role of ERP-Integrated Decision Support Systems in Enhancing Efficiency and Coordination In Healthcare Logistics: A Quantitative Study. *International Journal of Scientific Interdisciplinary Research*, 4(4), 265–285. https://doi.org/10.63125/c7srk144

[66]. Rokach, L., & Maimon, O. (2006). Data mining for improving the quality of manufacturing: A feature set decomposition approach. *Journal of Intelligent Manufacturing*, 17(3), 285–299. https://doi.org/10.1007/s10845-005-0005-x

[67]. Saberironaghi, A., Ren, J., & El-Gindy, M. (2023). Defect detection methods for industrial products using deep learning techniques: A review. *Algorithms*, 16(2), 95. https://doi.org/10.3390/a16020095

[68]. Sakurada, M., & Yairi, T. (2014). Anomaly detection using autoencoders with nonlinear dimensionality reduction.

[69]. Sazzadul, I. (2025). Machine Learning–Based AML/KYC Transaction Monitoring for Suspicious Activity Detection and Compliance Risk Reduction in Digital Banking. *ASRC Procedia: Global Perspectives in Science and Scholarship*, 1(01), 1740-1775. https://doi.org/10.63125/r9c8q813

[70]. Sazzadul, I., & Rebeka, S. (2024). VaR and CVaR-Based Stress Testing Using Deep Learning for Liquidity Risk Forecasting and Banking Stability Assessment. *Review of Applied Science and Technology*, 3(03), 01–30. https://doi.org/10.63125/291phs66

[71]. Schluse, M., Priggemeyer, M., Atorf, L., & Rossmann, J. (2018). Experimentable digital twins—Streamlining simulation-based systems engineering for Industry 4.0. *IEEE Transactions on Industrial Informatics*, 14(4), 1722–1731. https://doi.org/10.1109/tii.2018.2804917

[72]. Shamsunnahar, C. (2025). Business Intelligence–Driven Risk Assessment and Portfolio Performance Analytics for Financial and Investment Institutions. *ASRC Procedia: Global Perspectives in Science and Scholarship*, 1(01), 1668–1699. https://doi.org/10.63125/827e2c29

[73]. Sharif Md Yousuf, B., Md Shahadat, H., Saleh Mohammad, M., Mohammad Shahadat Hossain, S., & Imtiaz, P. (2025). Optimizing The U.S. Green Hydrogen Economy: An Integrated Analysis Of Technological Pathways, Policy Frameworks, And Socio-Economic Dimensions. *International Journal of Business and Economics Insights*, 5(3), 586–602. https://doi.org/10.63125/xp8exe64

[74]. Shofiul Azam, T. (2025). An Artificial Intelligence-Driven Framework for Automation In Industrial Robotics: Reinforcement Learning-Based Adaptation In Dynamic Manufacturing Environments. *American Journal of Interdisciplinary Studies*, 6(3), 38-76. https://doi.org/10.63125/2cr2aq31

[75]. Taheri, H., Kilpatrick, M., Norvalls, M., Harper, W. J., Koester, L. W., Bigelow, T., & Bond, L. J. (2019). Investigation of nondestructive testing methods for friction stir welding. *Metals*, 9(6), 624. https://doi.org/10.3390/met9060624

[76]. Tahmina Akter Bhuya, M., & Rebeka, S. (2022). AI-Assisted Underwriting Models for Improving Risk Assessment Accuracy in U.S. Insurance Markets. *American Journal of Interdisciplinary Studies*, *3*(01), 65-102. https://doi.org/10.63125/kegg1076

[77]. Tahmina Akter, R., & Aditya, D. (2025). Development of Model Influence on Consumer Behavior in U.S. e-commerce and Digital Marketing. *American Journal of Interdisciplinary Studies*, *6*(3), 106-143. https://doi.org/10.63125/1brehy25

[78]. Tao, F., Zhang, H., Liu, A., & Nee, A. Y. C. (2019). Digital twin in industry: State-of-the-art. *IEEE Transactions on Industrial Informatics*, *15*(4), 2405–2415. https://doi.org/10.1109/tii.2018.2873186

[79]. Tapia, G., & Elwany, A. (2014). A review on process monitoring and control in metal-based additive manufacturing. *Journal of Manufacturing Science and Engineering*, *136*(6), 060801. https://doi.org/10.1115/1.4028540

[80]. Tasnim, K. (2025). Digital Twin–Enabled Optimization of Electrical, Instrumentation, And Control Architectures In Smart Manufacturing And Utility-Scale Systems. *International Journal of Scientific Interdisciplinary Research*, *6*(1), 404–451. https://doi.org/10.63125/pqfdjs15

[81]. Tasnim, K., & Anick, K. M. T. A. (2024). PLC–SCADA–Integrated Electrical Automation Frameworks for Process Optimization in Water and Wastewater Treatment Facilities. *Review of Applied Science and Technology*, *3*(01), 221–262. https://doi.org/10.63125/y1145g11

[82]. Tercan, H., & Meisen, T. (2022). Machine learning and deep learning based predictive quality in manufacturing: A systematic review. *Journal of Intelligent Manufacturing*, *33*, 1879–1905. https://doi.org/10.1007/s10845-022-01963-8

[83]. Ventura-Carvalho, L., Gómez-Miranda, A., Suárez, A., & Rodríguez, J. (2021). A novel machine learning-based methodology for tool wear prediction using acoustic emission signals. *Sensors*, *21*(17), 5984. https://doi.org/10.3390/s21175984

[84]. Vogel-Heuser, B., & Hess, D. (2016). Guest editorial: Industry 4.0–prerequisites and visions. *IEEE Transactions on Automation Science and Engineering*, *13*(2), 411–413. https://doi.org/10.1109/tase.2016.2523639

[85]. Wang, S., Wan, J., Li, D., & Zhang, C. (2018). Implementing smart factory of Industry 4.0: An outlook. *International Journal of Distributed Sensor Networks*, *14*(2), 1–13. https://doi.org/10.1177/1550147718764162

[86]. Wang, X., Zhou, W., Tao, Y., & Zhang, Y. (2020). Machine learning for quality prediction in additive manufacturing: A data-driven approach. *Additive Manufacturing*, *36*, 101538. https://doi.org/10.1016/j.addma.2020.101538

[87]. Wuest, T., Weimer, D., Irgens, C., & Thoben, K.-D. (2016). Machine learning in manufacturing: Advantages, challenges, and applications. *Production & Manufacturing Research*, *4*(1), 23–45. https://doi.org/10.1080/21693277.2016.1192517

[88]. Xie, X. (2008). A review of recent advances in surface defect detection using texture analysis techniques. *Electronic Letters on Computer Vision and Image Analysis*, *7*(3), 1–22. https://doi.org/10.5565/rev/elcvia.218

[89]. Xu, X. (2012). From cloud computing to cloud manufacturing. *Robotics and Computer-Integrated Manufacturing*, *28*(1), 75–86. https://doi.org/10.1016/j.rcim.2011.07.002

[90]. Zaheda, K. (2025a). AI-Driven Predictive Maintenance For Motor Drives In Smart Manufacturing A Scada-To-Edge Deployment Study. *American Journal of Interdisciplinary Studies*, *6*(1), 394-444. https://doi.org/10.63125/gc5x1886

[91]. Zaheda, K. (2025b). Hybrid Digital Twin and Monte Carlo Simulation For Reliability Of Electrified Manufacturing Lines With High Power Electronics. *International Journal of Scientific Interdisciplinary Research*, *6*(2), 143–194. https://doi.org/10.63125/db699z21

[92]. Zaheda, K., & Md Hamidur, R. (2024). GPU-Accelerated Physics-Informed Digital Twins for Real-Time State Estimation and Fault Localization in Distribution Grids. *American Journal of Scholarly Research and Innovation*, *3*(02), 179-216. https://doi.org/10.63125/msrpfb04

[93]. Zaheda, K., & Md. Tahmid Farabe, S. (2023). Robotics and Computer Vision for Automated Inspection of Substation and Treatment-Facility Electrical Infrastructure. *Review of Applied Science and Technology*, *2*(04), 194-227. https://doi.org/10.63125/tfh15j12

[94]. Zonta, T., da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. *Computers & Industrial Engineering*, *150*, 106889. https://doi.org/10.1016/j.cie.2020.106889