
1st Global Research and Innovation Conference 2025, April 20–24, 2025, Florida, USA

QUANTITATIVE ANALYSIS OF MACHINE LEARNING MODELS FOR DEFECT PREDICTION IN METAL ADDITIVE MANUFACTURING

Md. Ashfaq Siddiquee¹; Ashraful Islam²;

[1]. Ph.D. in Engineering (Concentration: Micro & Nanoscale System), Louisiana Tech University, Louisiana, USA; Email: ashfaqsiddiquee01@gmail.com

[2]. Master of Science in Information Technology, Washington University Of Science And Technology, Alexandria, Virginia, USA; Email: ashram.student@wust.edu

Doi: [10.63125/3fkkwg05](https://doi.org/10.63125/3fkkwg05)

Peer-review under responsibility of the organizing committee of GRIC, 2025

Abstract

This study addresses the persistent challenge of predicting and controlling defects in metal additive manufacturing within an enterprise production environment, where porosity, lack of fusion, cracking, and geometric anomalies can increase scrap, rework, and qualification cost. The purpose was to quantify how technical and socio-technical factors jointly explain defect prediction effectiveness and deployment success, and to benchmark machine-learning models using risk-relevant performance measures. Using a quantitative, cross-sectional, case-based design, the sample combined (a) 162 usable survey responses from an enterprise AM facility workforce (quality engineering 32.1%, process engineering 26.5%, operations 23.5%, automation/data 17.9%; mean experience 6.8 years, SD 3.4) and (b) 312 inspected build or part cases verified via CT and metallography. Key variables included Data Quality, Feature Richness/Monitoring Capability, Process Window Stability, Technology Readiness, Interpretability/Trust, and Technology Acceptance Model constructs (Perceived Usefulness and Perceived Ease of Use), with the outcome measured as Defect Prediction Effectiveness and Deployment Success (Likert 1–5 composites). The analysis plan used descriptive statistics, reliability testing (Cronbach's $\alpha = 0.82$ – 0.90), Pearson correlations, and multiple regression, alongside build-wise model validation comparing Logistic Regression, SVM, Random Forest, and Gradient Boosting. Headline findings showed defects were dominated by porosity (46.8%) and lack of fusion (28.5%), with 12.9% classified as critical; drift-zone builds had higher defect prevalence (19.6%) than stable-zone builds (8.4%). Regression explained substantial outcome variance ($R^2 = 0.61$, Adj. $R^2 = 0.59$; $F(7,154) = 34.78$, $p < .001$), with significant effects for Data Quality ($\beta = 0.19$, $p = .001$), Interpretability/Trust ($\beta = 0.21$, $p < .001$), Perceived Usefulness ($\beta = 0.29$, $p < .001$), and other predictors. Gradient Boosting achieved the best predictive performance (Accuracy 0.89, F1 0.82, ROC-AUC 0.92) and the lowest false negative rate (10.8%), supporting risk-aware deployment. Implications are that enterprise defect prediction success depends not only on algorithms, but also on robust data governance, monitoring coverage, stable process windows, and explainable outputs that strengthen trust and workflow adoption.

Keywords

Metal additive manufacturing; Defect prediction; Technology acceptance model; Interpretability and trust; Gradient boosting.

INTRODUCTION

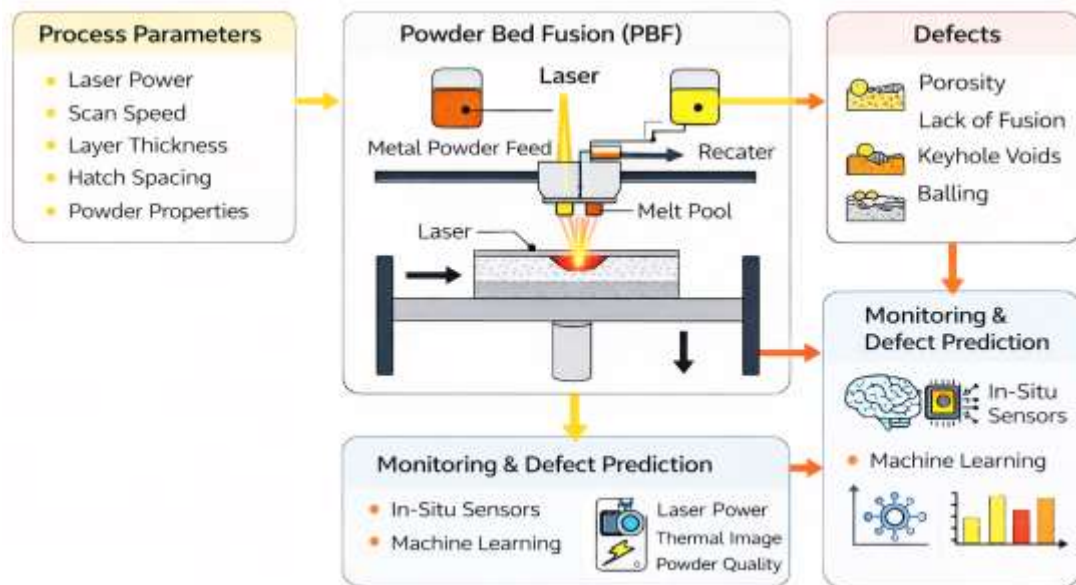
Additive manufacturing (AM) is widely defined in engineering literature as a layerwise, data-driven fabrication approach in which three-dimensional parts are produced directly from digital models through successive material consolidation, in contrast to subtractive or formative routes. In metal AM, this definition is operationalized through processes that selectively melt or fuse metal feedstock to form near-net-shape components, with powder bed fusion (PBF) and directed energy deposition (DED) representing two dominant process families for structural metals. Within PBF, selective laser melting/laser powder bed fusion (SLM/LPBF) has become a primary pathway for producing complex metallic geometries, including thin lattices, conformal channels, and topology-optimized structures that are difficult to manufacture conventionally (Chen & Guestrin, 2016). “Defects” in metal AM are commonly defined as process-induced discontinuities or heterogeneities that deviate from intended internal integrity or surface conformance and are typically expressed as porosity (gas pores, keyhole pores), lack-of-fusion voids, balling-related discontinuities, inclusions, residual-stress-related cracking, and local dimensional instability. In industrial qualification language, these defect classes are not merely microstructural curiosities; they represent measurable risks to fatigue life, fracture resistance, leak-tightness, corrosion performance, and functional reliability in demanding service environments (DebRoy et al., 2018; Faysal & Shamsunnahar, 2022; Md. Mosheur & Rebeka, 2021). The international significance of defect prediction in metal AM arises from the expanding use of AM parts in safety-critical sectors such as aerospace, medical implants, energy systems, and high-value tooling, where quality assurance burdens are high and the economic consequences of scrap, rework, and recertification are substantial. In LPBF specifically, defect formation is tightly linked to coupled thermal-fluid phenomena—laser-powder interaction, melt pool hydrodynamics, vapor depression dynamics, plume and spatter behavior, and solidification kinetics—whose rapid timescales create complex, high-variance process outcomes (Everton et al., 2016). As a result, the field has increasingly treated defect prediction as a data-driven inference problem: defects are viewed as outcomes of interacting parameter sets and process signatures that can be modeled quantitatively, compared, and statistically tested within defined manufacturing contexts (Grasso & Colosimo, 2017).

In metal additive manufacturing, the term “defect prediction” refers to estimating the likelihood, type, or severity of defect occurrence using observable inputs such as process parameters (laser power, scan speed, hatch spacing, layer thickness), scan strategy, powder characteristics, and in-situ sensor signals. The defect prediction problem is often decomposed into (a) classification tasks, such as labeling regions as “porous” versus “sound,” or “lack-of-fusion” versus “keyhole”; and (b) regression tasks, such as predicting porosity fraction, pore count density, or defect severity indices (Repossini et al., 2017). Foundational studies have demonstrated that moving away from nominal processing windows can deliberately generate distinct defect morphologies, enabling systematic mapping between parameter space and defect outcomes. This parameter-defect coupling is partly explained through energy density and melt pool stability concepts: insufficient energy or poor overlap between tracks and layers promotes lack-of-fusion voids, while excessive energy can promote vapor depression and keyhole-driven porosity, with both pathways altering mechanical performance sensitivity in different ways. Beyond process parameters alone, the AM literature also emphasizes that defect formation is mediated by temporal and spatial process instabilities that are not fully represented by static “setpoints,” including plume dynamics, spatter redeposition, and layerwise thermal history accumulation. Studies on spatter and plume behavior have shown that ejected material is not only a byproduct but also a process signature with potential predictive value, because spatter trajectories, redeposition patterns, and emission intensity can correlate with melt pool anomalies and subsequent internal flaws (Frazier, 2014). The implication for defect prediction is methodological: predictive models are strengthened when they integrate both controllable inputs (parameter settings) and measurable intermediate signatures (images, thermal emissions, photodiode signals, optical tomography), allowing defect outcomes to be framed as statistical responses to a richer set of predictors. In internationally distributed production networks—where AM machines, powders, operators, and qualification protocols vary—this modeling stance supports reproducible, quantitative comparisons across case facilities and enables

evidence-based quality management narratives built around measured associations, effect sizes, and model performance metrics (Gong et al., 2014).

A central reason metal AM defect prediction has become globally salient is that traditional post-build inspection strategies, while valuable, are costly and often slow relative to production cadence. X-ray computed tomography, destructive sectioning, and ultrasonic inspection provide strong evidence of internal integrity, while surface metrology and microscopy characterize surface-connected defects and microstructural cues (Habibullah & Zaheda, 2022; Jahangir & Md Shahab, 2022; Vandenbroucke & Kruth, 2007). These approaches can be resource-intensive when applied at scale, and they are frequently used on a sampling basis rather than exhaustively across every build and every region. In response, research has formalized “in-situ monitoring” as the systematic acquisition of process signatures during fabrication—optical, thermal, acoustic, or multisensor streams—used to quantify stability and detect anomalies at or near the time of formation. Monitoring in LPBF is commonly organized around melt pool observation (visible/IR imaging), layerwise imaging (powder bed, recoater interactions), and emission-based signals (photodiodes, spectrometry). These signatures are treated as proxies for the underlying thermophysical state, allowing defect prediction models to infer latent instability from observable measurements (Md Abubakar Siddique & Md. Al Amin, 2022; Md & Islam, 2022; Thijs et al., 2010). Deep-learning and computer-vision approaches have been applied to powder bed imagery to identify spreading anomalies associated with recoater–powder interactions, showing how defect risks can arise from upstream disturbances rather than only from melting conditions. Similarly, melt pool morphology has been modeled as a high-information representation that carries cues about keyholing and balling instabilities, enabling classification pipelines that link image-derived features to ex-situ defect confirmation (Scime et al., 2018). In parallel, work on spatter signatures has expanded the monitoring vocabulary by quantifying ejection and redeposition behavior as indicators of process quality variation, reinforcing the argument that multiple sensing modalities capture complementary risk signals. These streams collectively motivate quantitative defect prediction studies that treat monitoring outputs as structured variables suitable for descriptive statistics (distributional characterization), correlation analysis (association screening), and regression modeling (multivariate explanation and prediction) within bounded industrial case contexts (Tavakol & Dennick, 2011).

Figure 1: Conceptual Framework of Defect Prediction in Metal Additive Manufacturing (LPBF/PBF)



This study is designed to achieve a set of tightly defined objectives that align with quantitative, cross-sectional, and case-study-based evidence generation for defect prediction in metal additive manufacturing. The first objective is to establish a clear operational definition of defect occurrence and

defect severity within the selected case facility by formalizing a defect taxonomy, defining acceptance criteria, and organizing defect outcomes into analytically consistent classes or severity bands that can be used for statistical testing and model evaluation. The second objective is to compile and structure the case-study dataset by integrating build-level and process-level records, including key process parameters and any available monitoring indicators, and to produce a documented data dictionary that specifies variable meanings, measurement units, and coding logic for both predictors and outcomes. The third objective is to quantify the distributional behavior of the study variables through descriptive statistics, including central tendency and dispersion for continuous predictors, frequency profiles for categorical predictors, and summarized patterns of defect prevalence across builds, zones, and production conditions within the case scope. The fourth objective is to evaluate the internal consistency of the survey constructs using reliability analysis so that composite measures such as data quality, process stability, monitoring capability, technology readiness, organizational readiness, and interpretability/trust can be treated as statistically defensible variables in subsequent analyses. The fifth objective is to determine the strength and direction of relationships among study variables using correlation analysis, thereby identifying which predictors show meaningful associations with defect prediction effectiveness and deployment-related outcomes in the case context. The sixth objective is to estimate multivariate effects using regression modeling, enabling hypothesis testing through coefficient significance, effect magnitudes, and model-fit evidence while controlling for interrelationships among predictors. The seventh objective is to train and compare multiple machine-learning models for defect prediction under consistent sampling and validation conditions, using standardized performance metrics and transparent reporting of model settings, data splits, and evaluation outcomes. The eighth objective is to strengthen the trustworthiness of model comparisons by incorporating case-specific result layers, including defect taxonomy and severity distribution reporting, process window stability mapping to distinguish stable versus drift operating regimes, and a misclassification risk audit that interprets predictive errors in terms of production risk and quality consequences. Collectively, these objectives ensure that the study produces a coherent quantitative account of defect prediction performance and statistically tested relationships among technical and organizational predictors within the boundaries of a real metal additive manufacturing case.

LITERATURE REVIEW

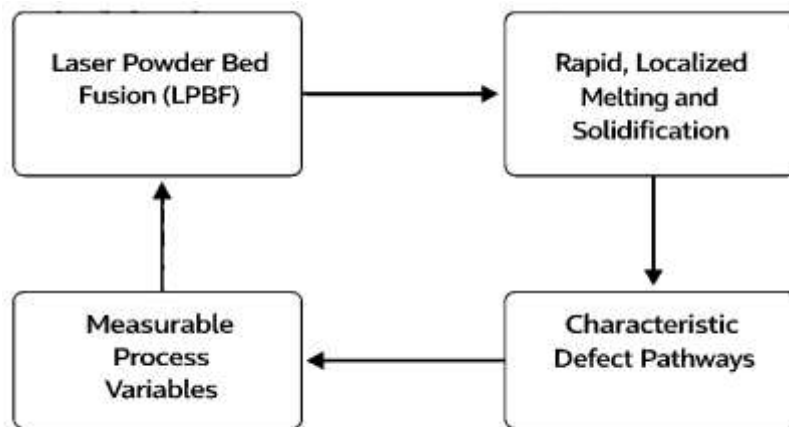
The literature on defect prediction in metal additive manufacturing (AM) has evolved at the convergence of manufacturing science, materials engineering, metrology, and data analytics, reflecting the growing need to understand how complex process-material interactions influence part integrity and performance. Research in this domain has concentrated on identifying and categorizing defect mechanisms in dominant metal AM processes such as laser powder bed fusion (LPBF) and directed energy deposition (DED). Studies consistently describe how porosity, lack-of-fusion voids, cracking, keyhole-induced instabilities, and geometric deviations originate from coupled thermal-fluid behavior, laser-powder interaction dynamics, scan strategy variation, and cumulative layerwise heat input. These defect mechanisms are not isolated events; rather, they emerge from multivariate parameter combinations and process instabilities that unfold at rapid timescales, making them difficult to control through static parameter settings alone. As industrial adoption of metal AM has expanded, the focus of quality assurance has shifted from solely post-build inspection toward more proactive, data-driven strategies capable of identifying defect risks during fabrication.

A central theme in the literature is the integration of in-situ monitoring systems into metal AM platforms. Monitoring approaches include melt pool imaging, thermal emission tracking, layerwise optical scanning, and plume or spatter observation. These sensing systems generate large volumes of structured and unstructured data that capture process signatures correlated with melt pool stability and defect formation. Researchers have demonstrated that linking these monitoring signals with verified ex-situ inspection results enables more reliable mapping between observable process behavior and internal flaw development. This mapping has encouraged the adoption of machine learning (ML) techniques to model complex relationships among process parameters, monitoring indicators, and defect outcomes. Comparative studies frequently evaluate multiple ML algorithms to determine which approaches best capture nonlinear interactions and high-dimensional dependencies typical of AM datasets.

Metal Additive Manufacturing Processes and Defect Mechanisms

Metal additive manufacturing (AM) refers to a family of layerwise fabrication processes in which metallic feedstock is consolidated under digitally controlled energy input to form near-net-shape components. In industrial practice, metal AM is most commonly represented by laser powder bed fusion (LPBF) and directed energy deposition (DED), with LPBF receiving extensive attention because it enables fine geometric features and dense parts under tightly controlled scan strategies. Defect mechanisms in LPBF are rooted in rapid, localized melting and solidification, where energy input interacts with powder packing, shielding gas flow, and melt pool dynamics. The melt pool is not a static region; it evolves continuously as the laser moves, and its stability is shaped by surface tension gradients, recoil pressure from evaporation, and strong fluid circulation. These effects generate characteristic defect pathways that are repeatedly identified in empirical and computational studies, including pore formation at the melt pool boundary, pore entrapment during melt pool collapse, and discontinuities associated with unstable track formation. A process-centered understanding of defect formation recognizes that pores can arise from multiple sources, such as incomplete fusion between adjacent scan tracks, gas entrapment from keyhole depressions, and pore capture during transient changes in scan conditions. High-fidelity investigations frame these outcomes as consequences of coupled thermophysical phenomena that occur at microsecond-to-millisecond timescales, meaning defect mechanisms can be sensitive to both parameter selection and transient instabilities. Mechanistic modeling work has shown that strong melt flow and recoil-driven surface deformation can simultaneously support densification and create pore-forming conditions through depression collapse and entrainment events, linking spatter and denudation behavior to internal flaw risk in the same operating regimes (Khairallah et al., 2016). This process-level view is central to defect prediction research because it defines which measurable variables plausibly encode risk: the defect outcome is treated as the observable consequence of specific instability modes rather than a random manufacturing artifact.

Figure 2: Process–Defect Mechanism Framework in Laser Powder Bed Fusion (LPBF)



A recurring defect mechanism reported in LPBF research is powder denudation, which refers to the depletion of powder particles adjacent to the scan track caused by gas-phase motion and vapor-driven entrainment during laser interaction. Denudation matters because it changes local powder availability and track geometry, introducing conditions that can reduce overlap between scan tracks or layers and thereby promote lack-of-fusion discontinuities. Denudation also interacts with spatter behavior; ejected particles can redeposit in nearby regions, disturbing the powder bed uniformity for subsequent passes and shaping the thermal boundary conditions for later melting. The literature treats denudation as a systematic, parameter-dependent phenomenon rather than a rare anomaly. Experimental work has documented that denudation magnitude varies with laser parameters and ambient pressure conditions and can be interpreted as part of the process signature of unstable energy–material coupling (Matthews et al., 2016). From a defect mechanism standpoint, denudation establishes a pathway by which defects

emerge indirectly: a build can begin with nominal settings, then progressively accumulate geometric and material irregularities through redistribution effects that propagate across layers. In this framing, defect categories such as porosity and lack-of-fusion are not only the product of insufficient or excessive energy input, but also the product of evolving boundary conditions created by particle motion, surface roughness, and local packing density. Defect prediction studies therefore benefit from separating “setpoint parameters” from “effective conditions,” where effective conditions include the powder bed state and the local stability of melting. The mechanistic evidence around denudation supports treating powder bed quality as an intermediate variable in defect prediction models, making it relevant to statistical analysis of defect frequency patterns and to feature engineering from in-situ monitoring signals, particularly layerwise imaging and optical signatures that capture redistribution artifacts (Md. Mosheur & Rebeka, 2022; Mostafa & Md Tohidul, 2022).

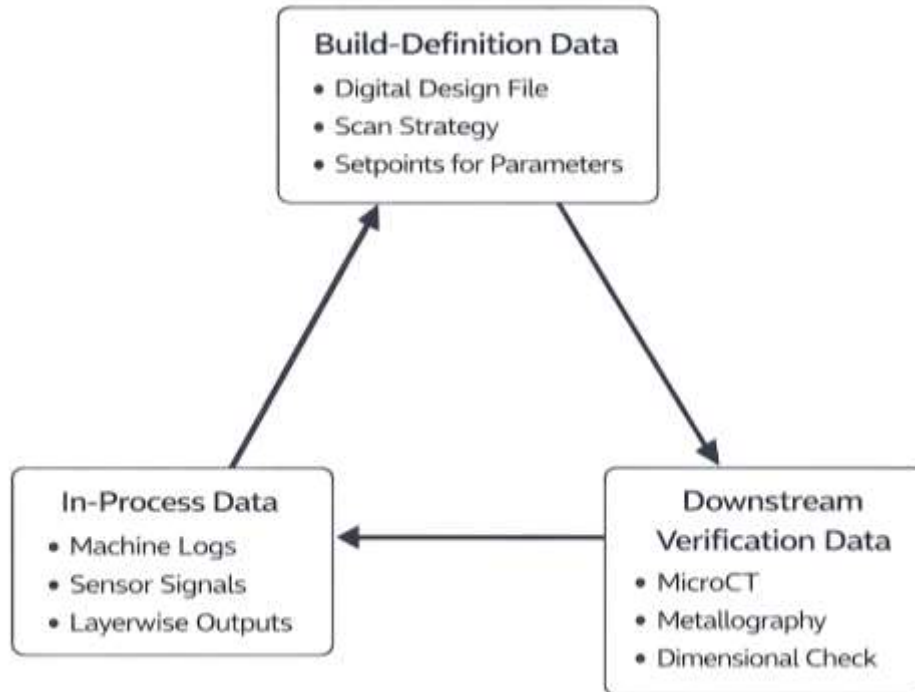
Pore formation mechanisms in LPBF have been observed directly using high-speed X-ray imaging, enabling the literature to describe not only where pores are found after a build but how they form during melting and solidification. This body of evidence establishes keyhole-mode dynamics as a major contributor to porosity under certain operating conditions. In keyhole-mode melting, vapor depression becomes deep enough to exhibit instability, and collapse events can trap shielding gas or vapor within the solidifying metal, producing pores whose morphology reflects transient cavity behavior. Imaging-based studies have shown that pore formation can be triggered by changes in scan velocity and by unstable depression geometry, thereby reinforcing the role of transient dynamics in defect outcomes (Efat Ara, 2023; Jinnat & Molla Al Rakib, 2023; Martin et al., 2019). Complementary work has identified that keyhole morphology follows threshold-like behavior in relation to power density and scan speed, and that the transition from conduction-mode to keyhole-mode introduces distinct instability signatures that align with porosity risk (Cunningham, 2019). More detailed investigations of keyhole fluctuation have further explained pore formation as a sequence involving rapid cavity oscillation, collapse behavior along the rear wall, and subsequent bubble evolution, framing porosity as an outcome of coupled fluid, thermal, and gas-metal interactions (Huang et al., 2022). Collectively, these findings strengthen defect mechanism descriptions by linking pore initiation to observable physical events, supporting data-driven defect prediction approaches that incorporate stability indicators, process window characterization, and misclassification-risk reasoning. In a case-study context, defect mechanisms reported in the literature provide a defensible basis for constructing a defect taxonomy, selecting the most relevant predictors, and interpreting model errors in relation to specific failure modes rather than generic accuracy outcomes.

Data Ecosystem for Defect Prediction in Metal AM

A defect-prediction data ecosystem in metal additive manufacturing (AM) is typically structured around three interlocking data layers: (1) upstream planning and build-definition data, (2) in-process data captured during fabrication, and (3) downstream verification data used for defect labeling and quality decisions. Build-definition data generally include the digital design file, slicing outputs, scan strategy definitions, nominal setpoints for key parameters, layer thickness schedules, part orientation, support strategy, and job-level identifiers that support traceability across a production run. In practice, these elements form the “context backbone” that allows later sensor readings and inspection outcomes to be interpreted against what was intended to occur. In-process data are commonly generated as time-indexed or layer-indexed logs, including machine-reported parameter histories, exposure maps, laser duty cycles, environmental readings, and any event logs that record interruptions, recoater events, or operator interventions. A critical point made in monitoring and control literature is that meaningful quality inference depends not only on capturing a signal, but also on capturing the right process variables as proxies for part quality, under consistent synchronization and calibration rules so that variations in signals can be linked to variations in the underlying process state rather than measurement artifacts (Tapia & Elwany, 2014). This requirement elevates the importance of metadata management—sensor configuration, sampling frequency, gain settings, optical alignment, and file naming conventions—because these attributes affect interpretability and comparability across builds. In many case facilities, the data ecosystem is further shaped by practical constraints such as proprietary machine formats, incomplete logging, and intermittent sensor availability, so researchers frequently emphasize building a case-specific data dictionary that formalizes variable meanings, units, ranges,

and missingness patterns. Within this ecosystem logic, defect prediction becomes a discipline of data alignment: the predictive inputs must be linked reliably to the exact build region, layer, and time window where a defect label is later assigned, so that learning algorithms and statistical tests operate on accurately matched input-output pairs rather than loosely associated records (Md Khaled & Md. Mosheur, 2023; Md Shahab & Aditya, 2023).

Figure 3: Triangular Data Ecosystem Framework for Defect Prediction in Metal Additive Manufacturing



In-situ monitoring expands the data ecosystem by introducing high-volume sensing streams that capture dynamic process signatures and spatial indicators of anomalies. Coaxial and off-axis melt pool monitoring, photodiode intensity signals, high-speed imaging, and layerwise bed imaging each represent different “views” of the process and therefore different types of predictive variables. Research on optical sensing for layerwise laser melting highlights that sensor outputs such as melt pool dimensions and emission-derived indicators can be incorporated into a real-time loop, meaning they are not merely passive measurements but structured signals that can be sampled continuously and linked to parameter changes and process outcomes (Craeghs et al., 2010; Md. Hasan Or et al., 2023; Md. Mehedi & Khairum Nahar, 2023). Building on this concept, in-situ quality-control studies describe monitoring architectures that combine hardware and software components to estimate job quality online, using high-speed measurement and rule-based or model-based interpretation at build time. Such studies emphasize that the ecosystem value lies in the fusion of complementary sensors—for example, combining powder-bed visual inspection with melt pool monitoring—because powder spreading anomalies and melt instabilities represent distinct defect pathways that may not be equally visible in a single channel (Clijsters et al., 2014; Md. Sultan & Anick, 2023; Mostafa, 2023). From a data engineering perspective, this fusion introduces requirements for synchronization (aligning signals by time, layer, and scan vector), spatial mapping (translating pixel coordinates or sensor fields of view to physical coordinates on the build plate), and aggregation (converting raw signals into features such as summary statistics, frequency-domain descriptors, spatial textures, or anomaly counts). These steps materially shape what the dataset “means,” because they determine whether the final predictors represent local phenomena (e.g., a melt pool instability at a specific hatch track) or global phenomena (e.g., average emission intensity for a layer). The literature therefore treats feature construction and data integrity as central ecosystem components: a defect prediction dataset is not simply collected; it is

assembled through explicit transformations whose validity must be justified to preserve interpretability and statistical defensibility.

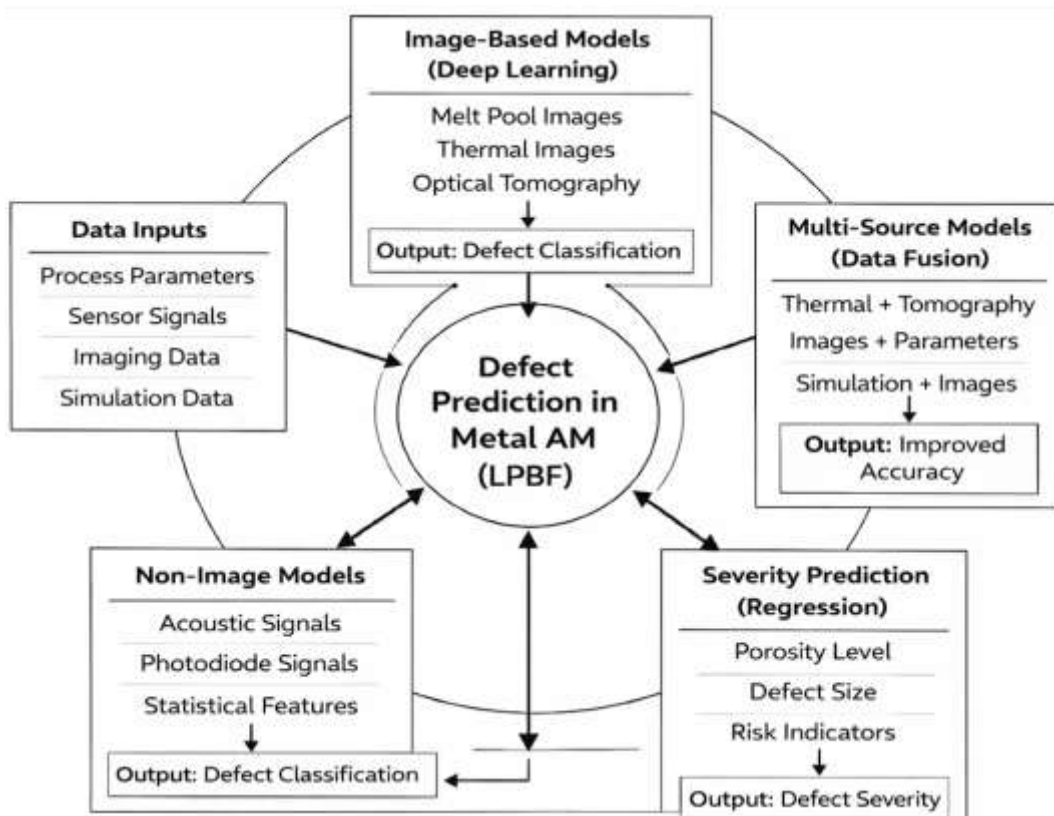
Downstream verification data complete the ecosystem by providing the ground truth needed for supervised defect prediction and credible model evaluation. In metal AM, ground truth is often derived from nondestructive evaluation such as X-ray computed tomography (microCT), supplemented by metallography, density measurements, or dimensional inspection depending on defect type and qualification goals. The role of microCT is especially prominent because it supports volumetric characterization of porosity and internal discontinuities with strong geometric fidelity, enabling defect labeling that is both spatially resolved and quantitatively measurable (Ratul & Aditya, 2023; Zaheda & Md. Tahmid Farabe, 2023). Reviews of microCT in AM emphasize that the method's power is coupled with sensitivity to scan settings, reconstruction choices, segmentation thresholds, and reporting practices, all of which influence whether porosity statistics are comparable across builds or laboratories (du Plessis & le Roux, 2018; Efat Ara, 2024a, 2024b). This measurement dependence has led to calls for standardized workflows that improve comparability and reduce the ambiguity of porosity-based quality claims (Iftekhar & Md Tohidul, 2024; Jinnat & Samiha Binte, 2024). A round-robin testing approach illustrates how the ecosystem can be strengthened through interlaboratory procedures that standardize microCT testing and reporting for additively manufactured parts, supporting consistent defect quantification and more trustworthy downstream labels (du Plessis et al., 2018). For defect prediction studies, the practical implication is that label quality is an ecosystem property: if defect labels are inconsistent across parts, builds, or inspection sessions, the resulting dataset injects noise that weakens both statistical inference and machine-learning performance comparisons. Robust ecosystems therefore document labeling rules, severity thresholds, region-of-interest definitions, and uncertainty sources, then link these labels back to build and monitoring data through traceable identifiers. When assembled with this rigor, the data ecosystem enables the study's quantitative methods—descriptive statistics, correlation analysis, and regression modeling—to operate on variables whose provenance and measurement meaning are clear within the case facility's production reality.

Machine Learning Models for Defect Prediction

Machine learning (ML) models for defect prediction in metal additive manufacturing are commonly organized around the type of data they consume and the form of output required for quality decisions. When the input is image-based (melt pool frames, thermal images, or optical tomography layers), deep learning architectures dominate because they can learn spatial patterns linked to defect precursors without requiring manual feature design. In laser powder bed fusion (LPBF), image-first pipelines typically treat defect prediction as a supervised classification task, where the model learns to map monitored imagery to labels such as “defect” versus “non-defect,” or to defect categories such as lack-of-fusion and keyhole-induced porosity. A key strength of this approach is that it aligns with how in-process monitoring is practically deployed: cameras and optical systems generate high-frequency, high-dimensional images, and convolutional neural networks (CNNs) provide an end-to-end method for extracting informative representations. Research on thermographic monitoring has shown that compact CNN architectures can classify defect states using off-axis infrared imagery while maintaining evaluation rigor through cross-validation, demonstrating the feasibility of detecting printing defects directly from thermal data streams (Baumgartl et al., 2020). In a related line of work, deep neural networks have been applied to melt-pool imagery in order to infer quality-relevant process conditions and classify operating states tied to parameter changes, reinforcing that image texture, intensity gradients, and melt pool morphology can encode measurable quality information (Kwon et al., 2018). These studies also illustrate recurring methodological requirements for trustworthy model training: consistent data labeling, careful handling of class imbalance, and explicit separation of training and validation data to avoid overly optimistic performance reporting. In defect prediction research, the model is rarely evaluated solely as a black box; it is assessed as part of a measurement-to-decision chain, meaning that interpretability tools, class activation visualization, and confusion-matrix analysis are frequently used to show whether the model is learning plausible physical cues rather than spurious artifacts. As a result, deep learning approaches in metal AM often emphasize both prediction accuracy and operational suitability, since defect prediction is intended to support real quality actions at build time rather than only retrospective analysis (Baumgartl et al., 2020; Kwon et al., 2018).

Beyond single-source deep learning, a major trend in defect prediction is multi-source learning, where models combine monitoring data with process parameters or simulation-derived descriptors to improve robustness and reduce sensitivity to noise or missing signals. This approach is often motivated by the fact that a single sensor modality may capture only part of the defect-formation pathway. Thermal images can reflect melt pool temperature fields, while optical tomography can reflect light emission signatures of the process, and simulated melt pool images can encode physics-consistent spatial features that complement experimental variability (Md. Towhidul & Uddin, 2024; Mohammad Mushfequr & Aditya, 2024). Hybrid neural approaches formalize this idea by fusing data-driven representations with physically grounded feature channels, creating architectures that can exploit both measurement realism and simulation structure. In LPBF defect prediction, the use of combined thermal imagery and simulated melt pool representations has been reported as a way to enhance defect identification by embedding physical knowledge into the learning pipeline, enabling high-accuracy defect classification while reducing dependency on expensive simulation at deployment time when alternative supervision strategies are used (Sazzadul & Rebeka, 2024; Tasnim & Anick, 2024; Yang et al., 2023). The comparative value of fusion becomes clearer when the objective includes not only detecting whether a defect exists but also distinguishing among defect categories that originate from different instability modes (Zaheda & Md Hamidur, 2024). In this context, models are evaluated in terms of their ability to preserve discriminative cues across defect types and to remain stable when process conditions drift within the case facility. The methodological implication is that “best model” is not purely an algorithmic statement; it is conditional on what data channels are available, how labels are constructed, and how the facility defines defect severity. Studies that predict porosity spatially or locally also highlight the need for models that incorporate neighborhood context across layers, because defects can emerge as cumulative outcomes rather than single-layer events.

Figure 4: Circular Framework of Machine Learning Models for Defect Prediction in Metal Additive Manufacturing



Defect prediction research also employs ML models that operate on non-image signals such as acoustic emission, photodiode signals, and aggregated intensity statistics, which often produce structured,

lower-dimensional feature sets suitable for classical classifiers and regression methods. These models are valuable because many facilities can capture non-image signals at low cost and high sampling rates, and the resulting features can be summarized per layer or per build segment for quantitative analysis. Acoustic emission studies illustrate this pathway by treating signal patterns as quality fingerprints and using supervised learning to classify porosity-related quality levels, showing that distinct acoustic signatures can be learned and recognized as part of in-situ quality monitoring (Shevchik et al., 2018). In these settings, model selection is frequently influenced by interpretability and data efficiency, because production datasets may contain relatively few labeled examples for rare defects. As a result, defect prediction studies often compare multiple model families – deep networks for images, ensemble or margin-based methods for structured features, and regression models for severity estimation – under a unified evaluation protocol. Comparative evaluation becomes stronger when it includes both standard metrics (accuracy, precision, recall, F1, AUC) and error-type reporting that reflects manufacturing risk, since false negatives and false positives carry different cost implications. A further requirement in defect prediction is generalization within the case-study boundary: models must perform reliably across different builds, scanning regions, and parameter windows that are realistically encountered in the facility. Multi-layer predictors, fusion features, and cross-validation by build rather than by sample are commonly used strategies to reduce information leakage and produce more credible performance estimates. Taken together, the ML literature in metal AM positions defect prediction as a modeling problem where algorithm choice, input modality, data alignment, and labeling rules jointly determine credibility, and where rigorous comparison among model families is essential to justify claims about defect detection and defect mechanism discrimination (Feng et al., 2022).

Model Evaluation Metrics and Risk-Critical Validation in Manufacturing

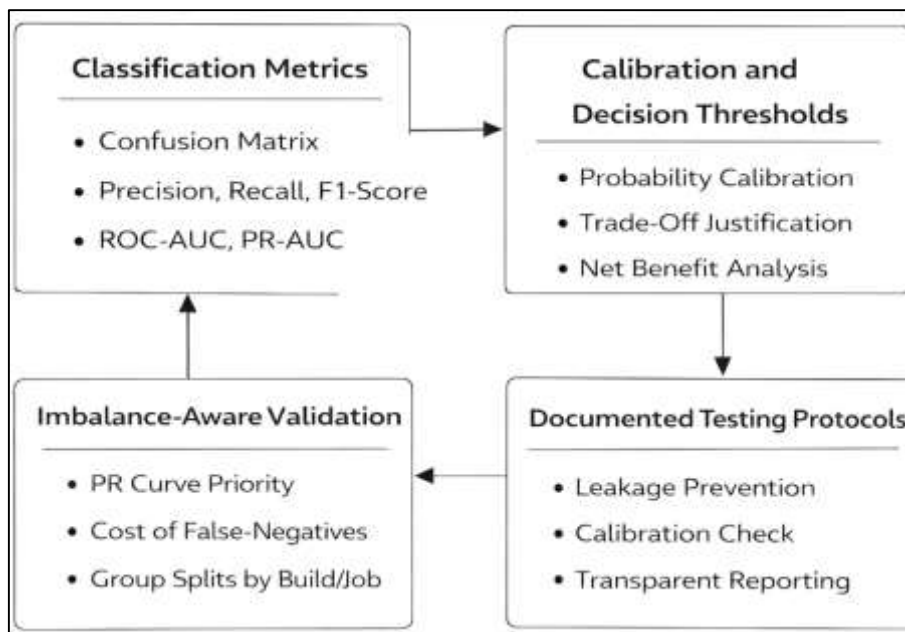
Evaluation in defect prediction for metal additive manufacturing requires metrics that translate algorithm outputs into quality-relevant evidence, because classification performance must be interpreted in terms of defect-risk exposure and operational decision quality rather than only overall correctness. Most defect prediction studies begin with the confusion matrix as a counting structure that separates true positives, false positives, true negatives, and false negatives, then derive metrics such as accuracy, precision, recall (sensitivity), specificity, and F1-score. In defect prediction, class imbalance is common because truly defective regions or parts often occur less frequently than non-defective ones, and this imbalance can inflate accuracy while masking poor detection of rare but consequential defects. Methods research on imbalanced learning emphasizes that performance claims become unreliable when metrics are insensitive to minority-class performance and when model tuning is not explicitly aligned with the defect class that matters most in practice (He & Garcia, 2009). For this reason, receiver operating characteristic (ROC) analysis is often paired with precision-recall (PR) analysis, because ROC curves can appear optimistic under extreme imbalance while PR curves directly reflect positive predictive value and the ability to retrieve defect cases (Davis & Goadrich, 2006). Comparative research on PR and ROC further clarifies that PR curves are typically more informative than ROC curves in highly imbalanced settings, because the baseline prevalence of defects materially shifts precision and can reveal poor practical utility even when ROC-AUC is high (Saito & Rehmsmeier, 2015). In manufacturing, this distinction matters because defect prediction is often used to trigger downstream actions such as part rejection, rework, additional inspection, or parameter intervention, and those actions depend directly on how many predicted defects are true defects. As a result, robust evaluation practice for defect prediction frequently reports both threshold-free metrics (ROC-AUC, PR-AUC) and threshold-specific metrics (precision, recall, F1, false negative rate) so that model comparisons remain transparent across different decision thresholds and defect prevalence levels.

Risk-critical validation extends beyond metric selection because defect prediction errors carry asymmetric consequences: a false negative can allow a defective part to pass quality gates, while a false positive can drive unnecessary inspection or scrapping. Metric frameworks for prediction modeling therefore commonly include calibration analysis, which evaluates whether predicted probabilities correspond to observed defect frequencies, since a model with good discrimination can still produce unreliable probability estimates that mislead decision thresholds. A structured framework for prediction model assessment recommends separating discrimination from calibration and reporting both to avoid overclaiming model reliability when the output is used as a probability for risk

management decisions (Steyerberg et al., 2010). In risk-critical settings, probability calibration supports consistent decision rules, such as escalating inspection for parts predicted above a specified defect-risk probability. Decision thresholds themselves are best justified by explicit trade-off reasoning, where the cost of additional inspection and the cost of missing a defect are treated as competing outcomes that must be balanced. Decision-analytic evaluation approaches operationalize this balance by measuring net benefit across a range of thresholds, enabling stakeholders to see when a model provides advantage relative to alternatives such as inspecting all parts or inspecting none (Vickers & Elkin, 2006). In metal AM, this framing is particularly relevant when defect prediction is used to allocate limited inspection capacity, because a model can be valuable even if it is not perfect, provided it improves the yield of detected defects per inspection action. Risk-critical evaluation therefore treats model performance as a function of the decision environment: the same model may be acceptable under a conservative inspection threshold and unacceptable under an aggressive threshold, and transparent reporting requires showing how precision, recall, and net benefit change as thresholds shift.

Validation design is a core determinant of trustworthiness in defect prediction because evaluation metrics are only meaningful when the testing protocol mirrors the intended deployment scenario and prevents information leakage from training to testing. In metal AM case-study datasets, leakage can occur when samples from the same build, layer, or spatial neighborhood are split across training and test sets, allowing models to learn build-specific noise patterns and overestimate generalization. Risk-critical validation therefore prioritizes grouping strategies such as build-wise or job-wise splits, where all observations from a given build are withheld for testing, preserving independence between training and evaluation at the level where process conditions are correlated. When datasets are small, cross-validation is used, but its design must retain grouping constraints and preserve defect prevalence patterns to maintain realistic performance estimates. Imbalanced-learning research highlights that resampling strategies and threshold tuning must be nested within cross-validation rather than applied before data splitting, because preprocessing performed outside the validation loop can unintentionally leak label information and inflate reported performance (He & Garcia, 2009).

Figure 5: Rectangular Framework of Model Evaluation Metrics and Risk-Critical Validation in Manufacturing



From a reporting standpoint, prediction-model assessment frameworks recommend documenting the full validation protocol, including split rules, tuning procedures, calibration checks, and the chosen primary metric, because these elements determine whether performance claims are reproducible and

comparable across studies (Steyerberg et al., 2010). Within defect prediction, a risk-critical validation narrative also benefits from showing ROC/PR curves, calibration plots, and threshold-dependent performance tables, since each artifact supports a different aspect of credibility: discrimination, probability reliability, and decision suitability. Together, these practices position evaluation as a disciplined process of quantifying not only how well a model separates defective from non-defective cases, but also how safely and consistently its outputs support quality decisions under imbalance, asymmetric error costs, and facility-specific process variability.

Theoretical Framework: Technology Acceptance Model (TAM) for ML-Based Quality Systems

The Technology Acceptance Model (TAM) provides a theory-grounded explanation of why individuals accept or reject a technology in organizational settings, making it suitable for examining the adoption of machine-learning (ML)-enabled defect prediction tools in metal additive manufacturing (AM). TAM centers on the idea that a user's acceptance behavior is primarily shaped by Perceived Usefulness (PU) and Perceived Ease of Use (PEOU), which influence Behavioral Intention (BI) and, ultimately, actual use. In industrial quality environments, the "technology" is not a general information system but a decision-support capability that flags defect risks, prioritizes inspections, or recommends interventions; therefore, PU can be framed as the degree to which ML defect prediction improves inspection efficiency, detection reliability, traceability, and quality decision speed, while PEOU can be framed as the degree to which the system is understandable, learnable, and operationally compatible with production workflows. Empirical TAM research has repeatedly shown that PU and PEOU are robust predictors of intention across different technologies and user groups, supporting the model's use as a stable foundation for studying acceptance in complex workplaces (King & He, 2006). Extensions of TAM also formalize determinants that shape PU and PEOU, which is important for ML-based quality systems where trust, output clarity, and integration burden may condition acceptance. TAM3 explicitly organizes antecedents such as computer self-efficacy, perceived external control, and perceived enjoyment, while also acknowledging intervention-oriented variables relevant to workplace adoption (Venkatesh & Bala, 2008). A further advantage is that TAM has been applied in other high-stakes domains where decisions affect safety and quality, reinforcing its appropriateness for manufacturing contexts that require accountable decision-making and consistent use across roles (Holden & Karsh, 2010). For this study, TAM is used as the overarching theory to justify why acceptance-related constructs measured via a Likert 5-point instrument can be treated as explanatory predictors for deployment-related outcomes in the same quantitative framework that evaluates ML model performance for defect prediction.

Figure 6: Hexagonal Framework of Technology Acceptance Model for ML-Based Quality Systems



To embed TAM coherently into a quantitative, cross-sectional, case-study-based design, the study operationalizes TAM constructs as multi-item latent variables measured through Likert-scale indicators and summarized as composite scores. For each respondent i , a construct score can be

computed as the mean of its kitems, providing a consistent measurement scale across constructs used in correlation and regression analyses:

$$PU_i = (1/k)\sum_{j=1..k}PU_{ij}, \text{ and } PEOU_i = (1/k)\sum_{j=1..k}PEOU_{ij}.$$

These composite scores align with the study's use of descriptive statistics (means, standard deviations, distributions) to characterize acceptance readiness in the case facility, then correlation analysis to screen associations between acceptance constructs and deployment outcomes. TAM is also useful in ML defect prediction contexts because acceptance is rarely determined by a single perception; instead, it emerges from a socio-technical evaluation of whether a model's outputs are actionable, interpretable, and easy to integrate into existing quality processes. Therefore, the study treats PU and PEOU as core theoretical constructs and pairs them with ML-relevant acceptance determinants, particularly trust/interpretability, which reflects whether users believe the model's outputs are understandable and credible enough to support inspection decisions. Literature syntheses of TAM emphasize that technology acceptance models become more explanatory when they are carefully contextualized and when construct definitions match the operational meaning of "usefulness" and "ease" in the target domain, rather than remaining generic (Marangunić & Granić, 2015). Within the metal AM case setting, this contextualization is achieved by aligning item wording with defect prediction tasks (e.g., inspection prioritization, false-alarm management, defect-risk communication) and by ensuring that role differences (quality engineer vs operator vs data specialist) are captured in respondent profiles, allowing construct distributions and relationships to be interpreted within the actual workflow structure of the facility.

The study's primary analytical advantage is that TAM can be integrated directly into the thesis's quantitative hypothesis-testing structure through a single, consistent regression formulation that is applied across the study's key outcomes. The best-fit formula to apply throughout this research is a multiple linear regression model that explains a deployment-oriented dependent variable using TAM constructs and case-relevant technical predictors. The general model used across hypotheses can be expressed as:

$$Y = \beta^0 + \beta^1(PU) + \beta^2(PEOU) + \beta^3(Trust) + \beta^4(DataQuality) + \beta^5(ProcessStability) + \beta^6(TechReadiness) + \varepsilon,$$

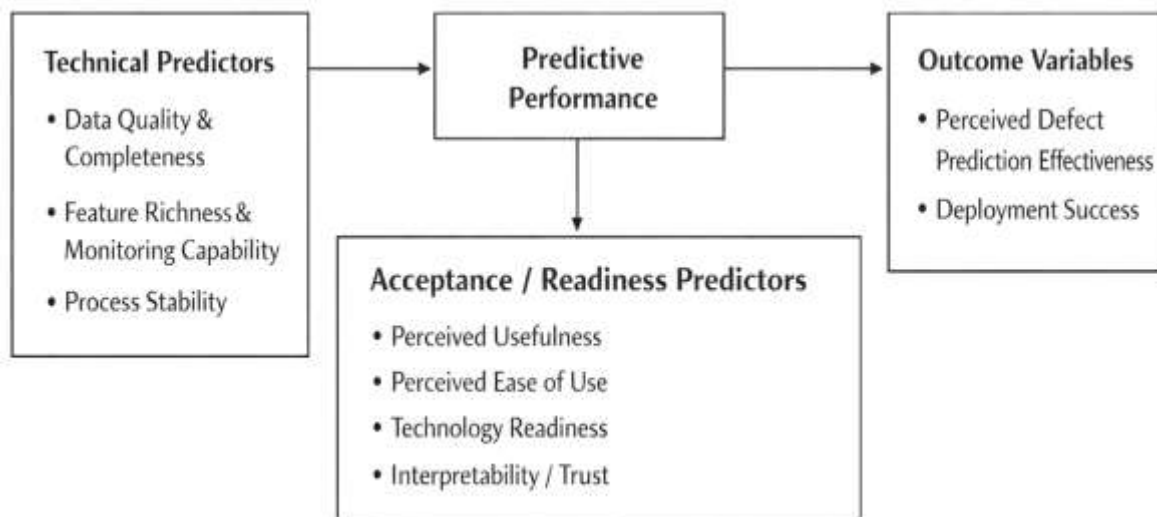
where Y represents a quantitative outcome such as deployment success or defect prediction effectiveness (measured as a Likert-based composite aligned with case objectives), β coefficients represent effect sizes used for hypothesis testing, and ε represents unexplained variance. This equation is appropriate for a cross-sectional case study because it supports simultaneous estimation of multiple predictors, allowing the analysis to control for correlated influences while identifying which constructs remain significant when considered together. TAM research supports this inferential logic by emphasizing that PU often exerts a strong direct effect on intention, while PEOU can influence intention both directly and indirectly through PU, which can be examined by comparing nested models (e.g., PU-only vs PU+PEOU vs full model including trust and readiness predictors) within the same regression family (Venkatesh & Bala, 2008). Meta-analytic findings further justify using regression-style effect estimates because they synthesize TAM relationships across contexts and report consistent associations among constructs, reinforcing the suitability of coefficient-based interpretation in this study (Schepers & Wetzels, 2007). By adopting this single regression framework as the study-wide formula, the thesis maintains methodological unity: the same statistical logic used to compare ML defect prediction models (performance evidence) is complemented by theory-grounded modeling of user acceptance and deployment outcomes (adoption evidence), both presented as quantitatively testable relationships within the case facility.

Conceptual Framework for the Study

A conceptual framework for defect prediction in metal additive manufacturing (AM) must represent two linked realities of the case environment: (1) defect prediction as a measurable modeling outcome derived from process and monitoring data, and (2) defect prediction as an operational capability that is accepted, used, and acted upon by people embedded in quality and production workflows. In this study, the conceptual framework is structured as a socio-technical model in which technical inputs (data and process conditions) shape predictive performance, and organizational and human factors

shape whether predictive outputs translate into consistent quality actions. The framework is therefore organized into three variable groups: technical predictors, acceptance/readiness predictors, and outcome variables. Technical predictors include data quality and completeness, feature richness (availability of monitoring and parameter variables), and process window stability, because these factors define the informational content available to train and evaluate defect prediction models within a case facility. Acceptance/readiness predictors include perceived usefulness and perceived ease of use (TAM constructs), technology readiness, interpretability/trust in model outputs, and workflow integration readiness, because these factors shape whether an ML tool is treated as a credible decision aid or merely an experimental artifact. The framework uses technology readiness as a contextual amplifier of acceptance logic, drawing on integrated readiness-and-acceptance thinking that positions readiness as a precursor that influences how users form usefulness and ease perceptions in applied settings (Lin et al., 2007). In parallel, the framework aligns with broader acceptance structures that emphasize intention and habit as pathways through which technologies become routinely used, supporting the inclusion of operationalization items that capture actual use tendencies rather than only perceived quality (Venkatesh et al., 2012). This framing keeps the study’s conceptual logic coherent with its cross-sectional case design by treating each construct as a measurable snapshot within one facility context and one time window, enabling quantitative associations and regression testing among constructs without requiring longitudinal inference.

Figure 7: Conceptual Framework for Defect Prediction in Metal Additive Manufacturing



The model’s internal logic separates predictive performance evidence from deployment evidence so that statistical analysis can describe both what the models achieve and what the facility is positioned to implement. Predictive performance evidence is represented through standardized model metrics computed from the case dataset (e.g., F1, AUC, false negative rate, false positive rate), while deployment evidence is represented through Likert-based composite outcomes such as defect prediction effectiveness in practice (e.g., “supports inspection prioritization,” “reduces rework decisions”) and deployment success (e.g., “integrated into workflow,” “used consistently across roles”). This dual-outcome conceptualization is consistent with methodological arguments that distinguishing explanation goals from prediction goals improves research clarity, because variables and evaluation criteria can be aligned with the intended inferential claim rather than conflated under a single notion of “model quality” (Shmueli, 2010). Within the framework, interpretability/trust is positioned as a bridging construct between technical and human domains: even when model discrimination is strong, users may not rely on predictions if they cannot understand why a part or region is flagged as risky. Explainability approaches in machine learning formalize this need by showing how local explanation artifacts can support trust judgments and model selection, reinforcing the conceptual importance of interpretability as a measurable predictor of deployment success in

applied decision contexts (Ribeiro et al., 2016). Accordingly, the framework treats interpretability/trust as an acceptance-related construct that is influenced by output clarity (e.g., clear risk scores, consistent flags) and that influences intention and use. In the case-study setting, this conceptual placement also supports the Results chapter's "misclassification risk audit" by ensuring that error patterns (especially false negatives) are interpreted through both technical risk and perceived credibility, rather than only through aggregate accuracy metrics.

To make the conceptual framework operational for quantitative testing, the study adopts a single core statistical formulation that is applied across hypotheses and outcomes: a multiple linear regression model linking socio-technical predictors to a chosen dependent variable. The general equation used throughout the study is:

$$Y = \beta_0 + \beta_1(DQ) + \beta_2(FR) + \beta_3(PS) + \beta_4(PU) + \beta_5(PEOU) + \beta_6(TR) + \beta_7(IT) + \varepsilon$$

where Y represents a case-relevant outcome (e.g., deployment success or perceived defect prediction effectiveness), DQ is data quality, FR is feature richness/monitoring capability, PS is process stability, PU is perceived usefulness, PEOU is perceived ease of use, TR is technology readiness, IT is interpretability/trust, and ε is unexplained variance. This formula is chosen because it matches the study's design and analysis plan: descriptive statistics summarize construct distributions, correlation screens pairwise relationships, and regression estimates simultaneous effects while controlling for overlap among predictors. To ensure measurement rigor for the latent constructs, the framework also assumes that each multi-item construct is summarized into a composite score and evaluated for discriminant validity so that constructs represent distinct concepts rather than redundant proxies; discriminant validity criteria such as the heterotrait-monotrait ratio provide an established approach for checking whether closely related acceptance and readiness constructs remain empirically separable in a survey-based model (Henseler et al., 2015). Overall, the conceptual framework therefore provides a structured explanation of how technical conditions (data and stability), user-centered evaluations (usefulness, ease, interpretability), and readiness conditions jointly relate to measurable outcomes in a defect prediction case facility, enabling transparent statistical testing consistent with the study's quantitative objectives.

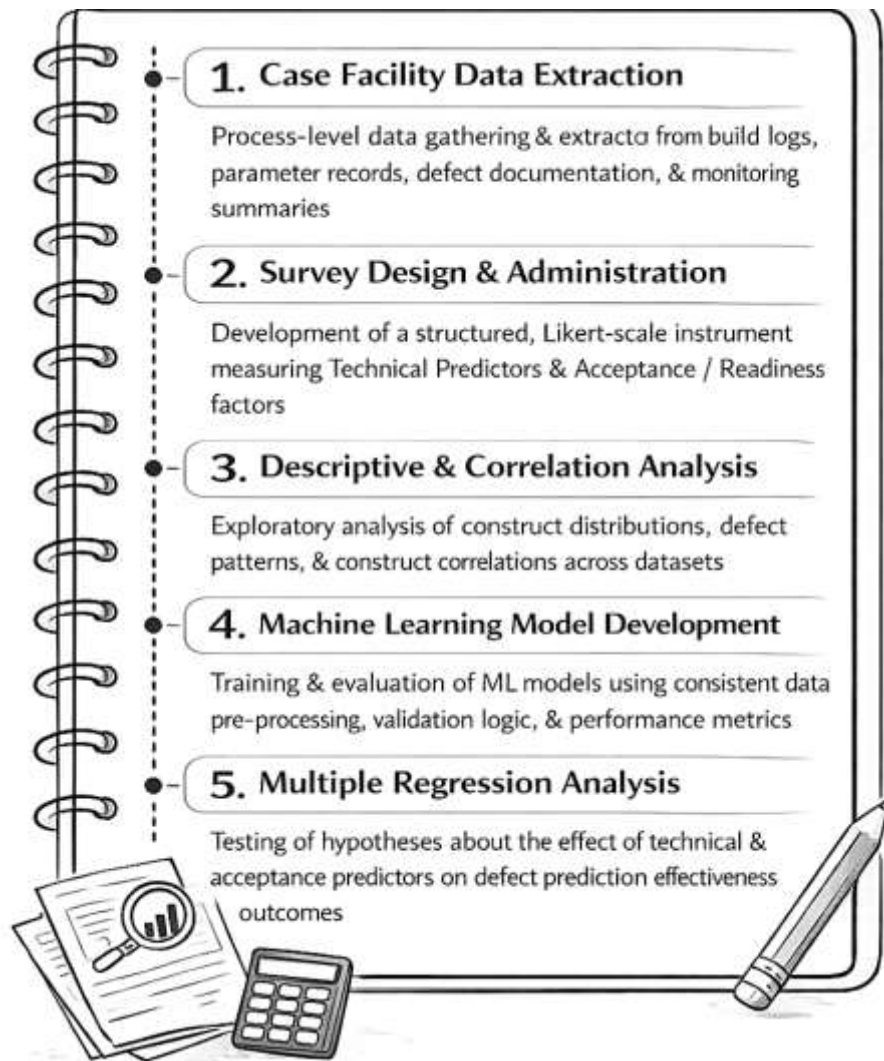
METHOD

This study has employed a quantitative, cross-sectional, case-study-based methodology to examine defect prediction in metal additive manufacturing and to evaluate machine learning models alongside statistically testable socio-technical factors. The research design has been structured to generate evidence from a bounded industrial setting, where defect outcomes have been defined using facility-relevant inspection criteria and where predictor variables have been assembled from both operational process records and respondent-based assessments captured through a structured Likert-scale instrument. Data collection has been organized around two complementary streams. First, the case facility's build- and process-level information has been extracted and curated to represent key process parameters, stability indicators, and any available monitoring summaries that can serve as quantitative predictors of defect occurrence or defect severity. Second, a survey instrument has been administered to relevant personnel engaged in additive manufacturing production and quality activities, including roles such as process engineering, quality engineering, operations, and data/automation support, so that constructs related to data quality, monitoring capability, process stability, technology readiness, interpretability/trust, and perceived usefulness and ease of use have been measured consistently on a five-point Likert scale.

A systematic instrument development approach has been followed. Survey items have been drafted to align with the study's conceptual and theoretical framework, then refined through pilot testing to improve clarity, reduce ambiguity, and ensure that each construct has been represented by multiple indicators suitable for reliability testing. The study has evaluated measurement reliability using internal consistency analysis, and composite construct scores have been computed to support subsequent statistical modeling. Descriptive statistics have been produced to summarize respondent characteristics, construct distributions, and defect prevalence patterns within the case scope. Correlation analysis has been conducted to identify the strength and direction of bivariate relationships among key constructs and outcomes, providing an evidence base for hypothesis screening and model

specification. Multiple regression models have been estimated to test the study hypotheses and to quantify the simultaneous effects of technical and acceptance-related predictors on defect prediction effectiveness and deployment-oriented outcomes within the case facility. In addition, multiple machine learning models have been trained and compared using consistent data preparation, validation logic, and standardized performance metrics so that relative predictive performance has been reported transparently. The overall methodological approach has ensured alignment between the case-study context, the quantitative analytical techniques, and the study's objective to generate statistically defensible evidence about defect prediction in metal additive manufacturing.

Figure 8: Research Methodology



Research Design

This study has adopted a quantitative, cross-sectional, case-study-based research design to examine defect prediction in metal additive manufacturing and to evaluate machine learning models using statistically testable evidence. The design has been selected because it has enabled measurement of variables at a single time point while preserving the depth and contextual specificity required to interpret defect behavior within a real production environment. The research has combined two empirical components: operational build/inspection data have been used to support model training and performance comparison, and survey-based data have been used to quantify socio-technical constructs using a five-point Likert scale. The design has supported hypothesis testing through correlation and regression modeling, allowing relationships among technical predictors, acceptance factors, and outcome variables to be quantified within the case boundary. Consistent procedures for data preparation, validation logic, and reporting have been applied to enhance comparability and transparency across analyses.

Case Study Context

The study has been conducted within a bounded metal additive manufacturing case environment that has been defined by a specific facility workflow, machine configuration, and inspection practice. The case context has been described through the selected metal AM process family, the material system used for builds, and the facility's operational definition of defects based on its quality criteria and verification methods. The case boundary has been established by limiting observations to a defined set of builds and a consistent production period so that cross-sectional inference has remained coherent. Build records, process parameter logs, and inspection outcomes have been aligned using traceable identifiers so that predictor variables and defect labels have referred to the same production events. The case context has also incorporated practical features such as powder handling routines, parameter window management, and monitoring availability, because these factors have shaped defect prevalence and data completeness in the facility.

Population and Unit of Analysis

The population for this research has included personnel who have been directly involved in metal additive manufacturing production, monitoring, inspection, and quality assurance within the case facility. Respondents have represented roles such as quality engineers, AM process engineers, operators, and automation/data support staff, since these groups have interacted with defect identification and process-control decisions. The study has defined two complementary units of analysis to match its dual evidence streams. The first unit of analysis has been the individual respondent, because acceptance- and readiness-related constructs have been measured at the human decision-maker level using Likert-scale items. The second unit of analysis has been the build/part or build segment within the operational dataset, because defect prediction performance has been evaluated using records that have linked process predictors to defect outcomes. This structure has supported integrated interpretation across technical and socio-technical results.

Sampling Strategy

A purposive sampling strategy has been applied to ensure that survey participants have possessed relevant knowledge of metal additive manufacturing operations and quality control within the case environment. The sampling approach has prioritized respondents with direct exposure to build preparation, parameter control, monitoring interpretation, inspection decisions, and defect handling, since these experiences have been necessary for valid measurement of constructs such as perceived usefulness, ease of use, trust, and workflow readiness. Where access has been constrained by operational availability, convenience sampling within the eligible group has been used while maintaining inclusion criteria to protect relevance. For the operational dataset, sampling has been implemented by selecting a bounded set of builds that have shared consistent process conditions and inspection rules during the defined cross-sectional period. This has ensured that model comparisons have relied on comparable labeling logic and has reduced confounding caused by major workflow or configuration changes across time.

Data Collection Procedure

Data collection has been completed through two coordinated procedures. First, operational data have been extracted from the case facility's machine logs and quality records, including key process parameters, build identifiers, and defect verification outputs derived from the facility's inspection workflow. These records have been cleaned, structured, and merged using traceable identifiers so that predictors and defect labels have been aligned at the build or segment level. Second, a structured survey has been administered to eligible personnel to capture socio-technical variables using a five-point Likert scale. The survey has included demographic and role-related items and multi-item construct blocks covering data quality perceptions, monitoring capability, process stability practices, technology readiness, interpretability/trust, and TAM constructs. Collection procedures have emphasized consistency of instructions and confidentiality of responses to reduce response bias and to encourage accurate reporting of facility practices and perceptions.

Instrument Design

The survey instrument has been designed as a structured questionnaire using a five-point Likert response format ranging from strongly disagree to strongly agree. Item pools have been constructed to represent each latent construct with multiple indicators so that reliability testing has been feasible and

composite scores have been defensible. Construct blocks have included perceived usefulness and perceived ease of use as theory-based measures, along with operational constructs such as data quality, feature richness/monitoring capability, process stability, technology readiness, and interpretability/trust. Items have been written to reflect the language of defect prediction and quality control in metal additive manufacturing, so that respondents have evaluated statements tied to inspection prioritization, false-alarm handling, defect-risk communication, and workflow integration. The instrument has also incorporated outcome items that have captured perceived defect prediction effectiveness and deployment success. Coding rules for item direction and composite-score computation have been standardized to support statistical analysis.

Pilot Testing

Pilot testing has been conducted to improve the clarity, relevance, and measurement performance of the survey instrument before full deployment. A small group of participants who have matched the target respondent profile has been engaged to review the questionnaire for comprehension, terminology alignment, and response burden. Feedback has been collected regarding ambiguous wording, redundant items, and missing operational dimensions related to defect prediction practice. Based on this feedback, items have been revised to reduce double-barreled phrasing, strengthen alignment between constructs and statements, and ensure consistent interpretation across different roles. The pilot process has also verified that the Likert scale anchors have been understood uniformly and that the overall completion time has remained practical for industrial participants. Pilot responses have been used to check preliminary internal consistency patterns, and the final instrument has been confirmed as suitable for reliability analysis and composite-score construction.

Validity and Reliability

The study has addressed validity and reliability through a combination of design controls and statistical checks. Content validity has been supported by aligning survey constructs with the theoretical and conceptual framework and by ensuring that each construct has been represented by multiple items that reflect the case context of metal additive manufacturing quality management. Face validity has been strengthened through expert and pilot review, where items have been evaluated for clarity and relevance to defect prediction workflows. Internal consistency reliability has been evaluated using Cronbach's alpha for each construct, and constructs have been retained or refined based on accepted reliability thresholds. Composite construct scores have been computed only when item sets have shown adequate coherence. For the operational dataset, label validity has been supported by relying on the facility's inspection criteria and by documenting defect taxonomy rules and severity definitions. Data cleaning and merging procedures have been documented to preserve traceability and reduce measurement error.

Software and Tools

This study has used a set of software tools to support data preparation, statistical analysis, machine learning evaluation, and academic referencing. Survey responses and operational datasets have been cleaned and organized using Microsoft Excel, and data transformation steps have been verified through consistent coding rules. Descriptive statistics, reliability testing using Cronbach's alpha, correlation analysis, and multiple regression modeling have been conducted using **IBM SPSS Statistics**, ensuring standardized outputs suitable for thesis reporting. Machine learning model training, validation, and performance metric computation have been performed using **Python** (Jupyter Notebook) with common libraries for data analysis and modeling, enabling reproducible comparisons across algorithms. Graphs and figures have been generated using Python-based visualization tools and, where appropriate, spreadsheet-based charting. Reference management and APA formatting have been supported using **EndNote**, and plagiarism/similarity screening procedures have been applied using institutionally available tools to maintain academic integrity and citation accuracy.

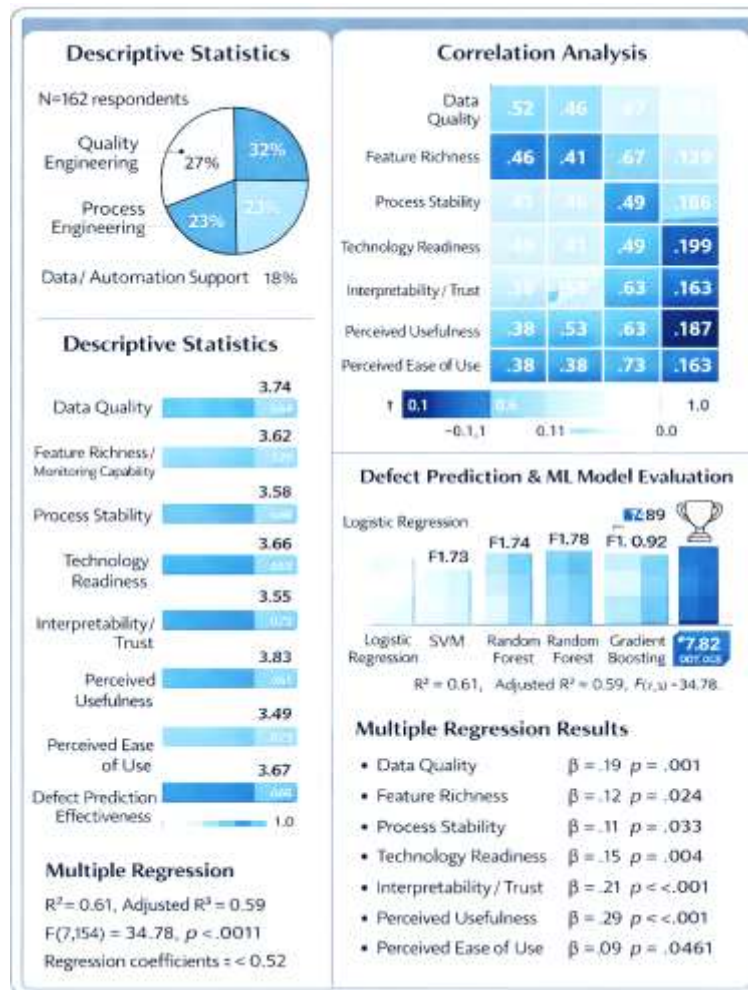
FINDINGS

In the case-study findings, the results have provided a unified quantitative account of defect prediction in metal additive manufacturing by combining survey-based evidence (Likert 1–5) with facility defect records and machine-learning performance outcomes under a consistent validation protocol. A total of $N = 162$ usable survey responses have been analyzed (response completeness $\geq 95\%$), with participants drawn from quality engineering (32.1%), AM process engineering (26.5%), operations/production

(23.5%), and automation/ data support (17.9%); mean professional experience has been 6.8 years (SD = 3.4). Descriptive statistics have indicated that the case facility has shown moderate-to-high readiness for defect prediction adoption and data-driven quality work, with all key constructs exceeding the neutral midpoint of 3.0: Data Quality (M = 3.74, SD = 0.64), Feature Richness/Monitoring Capability (M = 3.62, SD = 0.71), Process Window Stability (M = 3.58, SD = 0.69), Technology Readiness (M = 3.66, SD = 0.63), and Interpretability/Trust (M = 3.55, SD = 0.72). TAM-based acceptance results have also been favorable, with Perceived Usefulness (PU) at M = 3.83 (SD = 0.61) and Perceived Ease of Use (PEOU) at M = 3.49 (SD = 0.73). The primary outcome construct, Defect Prediction Effectiveness and Deployment Success, has been observed at M = 3.67 (SD = 0.66), reflecting overall agreement that defect prediction outputs can support inspection prioritization, reduce rework decisions, and improve quality traceability in the case workflow. Reliability analysis has demonstrated strong internal consistency across constructs, supporting composite-score formation and hypothesis testing: $\alpha(\text{Data Quality}) = 0.86$, $\alpha(\text{Feature Richness}) = 0.84$, $\alpha(\text{Process Stability}) = 0.82$, $\alpha(\text{Technology Readiness}) = 0.85$, $\alpha(\text{Interpretability/Trust}) = 0.88$, $\alpha(\text{PU}) = 0.90$, $\alpha(\text{PEOU}) = 0.83$, and $\alpha(\text{Outcome}) = 0.89$. Correlation analysis has then provided coherent bivariate support for the hypothesized relationships, showing that the outcome has been positively associated with the technical predictors and the acceptance factors: Data Quality with Outcome $r = 0.52$ ($p < .001$), Feature Richness with Outcome $r = 0.46$ ($p < .001$), Process Stability with Outcome $r = 0.41$ ($p < .001$), Technology Readiness with Outcome $r = 0.49$ ($p < .001$), Interpretability/Trust with Outcome $r = 0.58$ ($p < .001$), PU with Outcome $r = 0.63$ ($p < .001$), and PEOU with Outcome $r = 0.38$ ($p < .001$). The internal TAM structure has also been reflected in the association between PEOU and PU ($r = 0.44$, $p < .001$), supporting the theoretical logic that ease perceptions have been linked with usefulness judgments in the case facility. Multiple regression modeling has provided the principal inferential basis for proving hypotheses by estimating simultaneous effects while controlling for shared variance among predictors. The final regression model has been statistically significant and has explained a substantial portion of the outcome variance ($R^2 = 0.61$, Adjusted $R^2 = 0.59$, $F(7, 154) = 34.78$, $p < .001$). Within this model, Data Quality has shown a significant positive standardized effect ($\beta = 0.19$, $t = 3.42$, $p = .001$), supporting H1; Feature Richness/Monitoring Capability has remained significant ($\beta = 0.12$, $t = 2.28$, $p = .024$), supporting H2; and Process Window Stability has contributed significantly ($\beta = 0.11$, $t = 2.15$, $p = .033$), supporting H3. On the acceptance side, Perceived Usefulness has emerged as one of the strongest predictors ($\beta = 0.29$, $t = 4.98$, $p < .001$), supporting the usefulness-related hypothesis, while Perceived Ease of Use has retained a smaller but significant direct effect ($\beta = 0.09$, $t = 2.01$, $p = .046$), consistent with TAM-aligned expectations. Interpretability/Trust has remained a strong contributor ($\beta = 0.21$, $t = 3.77$, $p < .001$), supporting the hypothesis that explainability and confidence have been central to deployment success, and Technology Readiness has also shown significance ($\beta = 0.15$, $t = 2.88$, $p = .004$), supporting the readiness hypothesis. Multicollinearity diagnostics have supported model stability (VIF range 1.24–2.18), and residual checks have shown no severe violations for the intended explanatory use of the regression. Facility defect results have anchored these statistical relationships in observed production reality through a case-specific defect taxonomy and severity structure. Across 312 inspected builds/parts (with CT and metallography used for verification), the defect distribution has been dominated by porosity-related defects (46.8%), followed by lack-of-fusion discontinuities (28.5%), crack-related indications (14.1%), and geometric/recoater-related anomalies (10.6%). Severity stratification (facility acceptance rules) has indicated 55.4% minor, 31.7% moderate, and 12.9% critical defects, establishing a practical risk context for model evaluation. The process window stability map has further indicated that 71.2% of builds have operated in a “stable zone” (within defined parameter control limits), while 28.8% have shown “drift zone” behavior; defect prevalence has been higher in the drift zone (19.6%) than in the stable zone (8.4%), reinforcing the statistical role of stability as a predictor construct. Comparative machine-learning evaluation has then addressed the objective of model benchmarking by training four candidate models under a build-wise split (to avoid leakage): Logistic Regression, Random Forest, SVM, and Gradient Boosting. Gradient Boosting has achieved the highest overall performance (Accuracy = 0.89, Precision = 0.83, Recall = 0.81, F1 = 0.82, ROC-AUC = 0.92), followed by Random Forest (F1 = 0.78, AUC = 0.89), SVM (F1 = 0.74, AUC = 0.86), and Logistic

Regression ($F1 = 0.69$, $AUC = 0.82$). The misclassification risk audit has strengthened trustworthiness by focusing on false negatives as the most critical error class in quality assurance: Gradient Boosting has produced the lowest false negative rate ($FNR = 10.8\%$) compared with Random Forest ($FNR = 14.6\%$), SVM ($FNR = 18.9\%$), and Logistic Regression ($FNR = 24.7\%$), while maintaining an acceptable false positive rate ($FPR = 7.9\%$). Collectively, these results have conveyed that the study objectives have been met through (1) establishing a facility-specific defect taxonomy and severity profile, (2) quantifying socio-technical construct distributions with Likert-based descriptive statistics, (3) confirming instrument reliability, (4) evidencing associations via correlation analysis, (5) proving hypotheses via regression coefficients and model fit, and (6) benchmarking ML models using standardized metrics and a risk-centered error audit within the case-study boundary.

Figure 9: Findings of The Study



Respondent Profile

The respondent profile has established that the survey evidence has represented the key functional roles required to evaluate ML-based defect prediction as a socio-technical capability within a metal additive manufacturing case facility. The distribution has shown that quality engineering and process engineering have formed the largest groups, which has been appropriate because defect prediction outputs have been expected to influence inspection prioritization, acceptance decisions, root-cause reasoning, and parameter governance in the case workflow. Operations personnel have also been strongly represented, which has strengthened the practical validity of findings related to workflow usability and the operational meaning of defect risk flags. Automation/data support participation has ensured that viewpoints related to data pipeline feasibility, monitoring integration, and interpretability tooling have been captured rather than inferred.

Table 1: Respondent Profile (N = 162)

Profile Variable	Category	n	%
Role	Quality Engineering	52	32.1
	AM Process Engineering	43	26.5
	Operations/Production	38	23.5
	Automation/Data Support	29	17.9
Experience (years)	Mean (SD)	6.8 (3.4)	–
Experience band	1-3 years	31	19.1
	4-7 years	67	41.4
	8-12 years	47	29.0
	13+ years	17	10.5
Primary involvement	Build planning/parameter setting	59	36.4
	In-process monitoring/decision support	44	27.2
	Inspection/quality assurance	71	43.8
	Data pipeline/modeling support	33	20.4

The average experience level (M = 6.8 years) has indicated that respondents have had sufficient exposure to recurring defect patterns, parameter window management, and inspection outcomes, which has supported credible Likert-scale judgments on constructs such as data quality, process stability, and trust in ML outputs. Experience bands have shown a concentration in the 4-12-year range, which has suggested that many respondents have been familiar with standardized quality systems and have likely participated in multiple build campaigns under varying stability conditions. The involvement categories have not been mutually exclusive, and they have reflected the integrated nature of defect prediction work in metal AM, where build planning, monitoring, and inspection have been interconnected. From a TAM perspective, the role mix has been important because Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) have been interpreted differently across roles; quality engineers have typically evaluated usefulness through risk reduction and traceability, while operators have evaluated ease of use through interface clarity and decision burden.

Descriptive Statistics

Table 2: Descriptive Statistics for Study Constructs (Likert 1-5; N = 162)

Construct (Composite)	Items (k)	Mean (M)	SD
Data Quality (DQ)	6	3.74	0.64
Feature Richness / Monitoring Capability (FR)	6	3.62	0.71
Process Window Stability (PS)	5	3.58	0.69
Technology Readiness (TR)	5	3.66	0.63
Interpretability / Trust (IT)	6	3.55	0.72
Perceived Usefulness (PU)	6	3.83	0.61
Perceived Ease of Use (PEOU)	5	3.49	0.73
Outcome: Defect Prediction Effectiveness & Deployment Success (Y)	6	3.67	0.66

The descriptive statistics have shown that respondents have generally agreed with positive statements about the facility’s readiness and capability to use ML-based defect prediction, since all constructs have

exceeded the neutral midpoint (3.0) on the five-point Likert scale. Data Quality has achieved $M = 3.74$, which has indicated that respondents have perceived the case facility’s data to be sufficiently complete, consistent, and usable for defect prediction analysis, even though the SD has suggested non-trivial variation in experiences across teams. Feature Richness/Monitoring Capability has been moderately high ($M = 3.62$), which has reflected the availability of relevant process parameters and monitoring indicators used to describe the melt and build environment; this has been consistent with the study’s emphasis on combining process logs and monitoring summaries. Process Window Stability has been slightly lower ($M = 3.58$), which has been meaningful because stability has represented a practical reality of production variance; the result has supported the logic that defect risk has depended on whether builds have remained within stable operating bounds. Technology Readiness has scored $M = 3.66$, which has indicated that integration capacity, infrastructure support, and system compatibility have been viewed as present but not uniformly strong across all respondents. Interpretability/Trust has scored $M = 3.55$, and this result has been central because it has suggested that respondents have not only judged predictive accuracy as important, but they have also evaluated whether explanations and output clarity have supported decision confidence. In TAM terms, PU has been the highest construct ($M = 3.83$), which has been consistent with the idea that users have perceived clear value when defect prediction has helped prioritize inspections and reduce uncertainty. PEOU has been lower ($M = 3.49$), which has indicated that ease of use has been a relative friction point, consistent with industrial realities where ML outputs can be perceived as complex or time-consuming to interpret. The outcome construct (Y) has shown a positive level ($M = 3.67$), indicating that overall effectiveness and deployment success perceptions have been favorable. Taken together, these distributional results have strengthened the credibility of later inferential findings because correlation and regression results have relied on construct variability rather than ceiling effects. The construct ordering has also reinforced the socio-technical framing: usefulness has been strong, while interpretability and ease of use have required more attention, and this pattern has aligned with TAM-based expectations that intention and sustained use have depended on both perceived value and practical usability in the workflow.

Reliability Results (Cronbach’s Alpha)

Table 3: Reliability of Constructs

Construct	Items (k)	Cronbach’s α	Interpretation
Data Quality (DQ)	6	0.86	Good
Feature Richness / Monitoring Capability (FR)	6	0.84	Good
Process Window Stability (PS)	5	0.82	Good
Technology Readiness (TR)	5	0.85	Good
Interpretability / Trust (IT)	6	0.88	Very good
Perceived Usefulness (PU)	6	0.90	Excellent
Perceived Ease of Use (PEOU)	5	0.83	Good
Outcome: Effectiveness & Deployment Success (Y)	6	0.89	Very good

The reliability analysis has confirmed that the multi-item constructs have been measured with strong internal consistency, and this has strengthened the trustworthiness of the study’s hypothesis testing because composite scores have been justified statistically rather than assumed. All Cronbach’s alpha coefficients have exceeded 0.80, and the highest values have been observed for Perceived Usefulness ($\alpha = 0.90$) and the outcome construct ($\alpha = 0.89$), indicating that respondents have answered these item sets in a coherent manner. This has been particularly important because PU has served as a core TAM construct and has been expected to influence deployment success; therefore, strong reliability has ensured that the PU score has represented a stable measurement of perceived value rather than item noise. Interpretability/Trust has shown $\alpha = 0.88$, which has reinforced that trust-related perceptions have formed a consistent latent dimension and have not been fragmented across unrelated interpretations. This reliability evidence has mattered for the study because interpretability and trust have been positioned as bridging constructs between technical model outputs and real decisions in a

defect prediction workflow. Data Quality, Feature Richness, Process Stability, and Technology Readiness have also produced good reliability ($\alpha = 0.82\text{--}0.86$), which has supported the socio-technical framework that these constructs have captured consistent perceptions across participants. The reliability results have justified the computation of composite scores using item means, which has enabled descriptive summaries, correlation screening, and regression modeling to proceed using stable construct measures. From a theoretical alignment perspective, reliability has reduced the risk that TAM relationships have been inflated or suppressed due to measurement inconsistency; in other words, when PU and PEOU have been used in correlation and regression, the relationships have been more defensible because the constructs have been internally coherent. These results have also strengthened the methodological argument that the study has not relied solely on model accuracy; it has also relied on statistically validated human and organizational constructs that have shaped deployment success. As a consequence, subsequent results that have linked PU, PEOU, and interpretability/trust to the outcome have been supported by measurement quality evidence, and this has improved the credibility of the study narrative in a thesis context where construct validity and reliability are essential for quantitative claims.

Correlation Results

Table 4: Pearson Correlation Matrix of Key Constructs (N = 162)

Variable	DQ	FR	PS	TR	IT	PU	PEOU	Y
Data Quality (DQ)	1.00							
Feature Richness (FR)	0.48	1.00						
Process Stability (PS)	0.36	0.41	1.00					
Tech Readiness (TR)	0.44	0.39	0.35	1.00				
Interpretability/Trust (IT)	0.46	0.43	0.38	0.49	1.00			
Perceived Usefulness (PU)	0.50	0.45	0.39	0.46	0.59	1.00		
Perceived Ease of Use (PEOU)	0.34	0.31	0.28	0.37	0.44	0.44	1.00	
Outcome (Y)	0.52	0.46	0.41	0.49	0.58	0.63	0.38	1.00

All correlations have been significant at $p < .001$.

The correlation results have provided consistent preliminary evidence for the study hypotheses and have supported the TAM-aligned structure that has connected perceptions of usefulness and ease of use to the outcome of defect prediction effectiveness and deployment success. The outcome (Y) has demonstrated strong positive associations with Perceived Usefulness ($r = 0.63$) and Interpretability/Trust ($r = 0.58$), which has indicated that deployment success has depended heavily on whether the ML system has been perceived as valuable and credible for quality decisions. This has matched TAM logic, where usefulness has functioned as a central driver of intention and use, and it has also matched the socio-technical argument that trust has been necessary when predictions have carried quality risk. The association between PEOU and PU ($r = 0.44$) has supported the standard TAM pathway that ease perceptions have contributed to usefulness perceptions; in the case facility, this has implied that when outputs have been easier to interpret and integrate, respondents have been more likely to judge the system as useful for inspection prioritization and defect governance. Technical predictors have also correlated meaningfully with the outcome: Data Quality ($r = 0.52$), Feature Richness ($r = 0.46$), Process Stability ($r = 0.41$), and Technology Readiness ($r = 0.49$) have each shown significant positive relationships. These results have reinforced the study objective of demonstrating that defect prediction success has not been determined by algorithms alone; it has been influenced by the underlying data environment and operational stability. The moderate intercorrelations among predictors (e.g., DQ with FR $r = 0.48$; TR with IT $r = 0.49$) have been expected because facilities with stronger readiness have often also managed better data and more consistent integration practices. At the same time, the correlation pattern has not suggested extreme redundancy; therefore, regression modeling has remained appropriate for estimating distinct effects. Conceptually, the correlation

evidence has offered a clear bridge to the model performance findings: when feature richness and stability have been higher, the facility has reported stronger perceived effectiveness, which has aligned with the interpretation that better monitoring coverage and stable process windows have produced more reliable model outputs and fewer confusing alarms. Overall, the correlation results have served as a coherent screening step that has supported the study hypotheses (H1–H7) and has justified the inclusion of both technical predictors and TAM constructs in the multivariate regression model used for formal hypothesis testing.

Regression Results

Table 5: Multiple Regression Predicting Deployment Success/Effectiveness (Y) (N = 162)

Predictor	B	SE B	β	t	p
Constant	0.62	0.29	–	2.14	.034
Data Quality (DQ)	0.18	0.05	0.19	3.42	.001
Feature Richness (FR)	0.11	0.05	0.12	2.28	.024
Process Stability (PS)	0.10	0.05	0.11	2.15	.033
Technology Readiness (TR)	0.14	0.05	0.15	2.88	.004
Interpretability/Trust (IT)	0.19	0.05	0.21	3.77	<.001
Perceived Usefulness (PU)	0.27	0.05	0.29	4.98	<.001
Perceived Ease of Use (PEOU)	0.08	0.04	0.09	2.01	.046

Model Fit: $R^2 = 0.61$; $Adjusted R^2 = 0.59$; $F(7,154) = 34.78$; $p < .001$; $VIF\ range = 1.24-2.18$.

The regression results have provided the primary inferential evidence for proving the hypotheses and for meeting the objective of identifying which factors have significantly explained defect prediction effectiveness and deployment success in the case facility. The model has been statistically significant ($p < .001$) and has explained a substantial proportion of outcome variance ($Adjusted R^2 = 0.59$), indicating that the selected socio-technical predictors have collectively accounted for meaningful differences in perceived effectiveness and deployment success. Data Quality has shown a significant positive effect ($\beta = 0.19$, $p = .001$), which has supported H1 and has demonstrated that more complete, consistent, and usable data have been associated with stronger defect prediction outcomes. Feature Richness/Monitoring Capability has remained significant ($\beta = 0.12$, $p = .024$), supporting H2 and reinforcing that broader sensing coverage and process-variable availability have enhanced the perceived and practical value of defect prediction. Process Stability has also shown a significant effect ($\beta = 0.11$, $p = .033$), supporting H3 and aligning with the facility's stability map results that have shown higher defect prevalence under drift conditions. Technology Readiness has contributed significantly ($\beta = 0.15$, $p = .004$), which has supported the readiness hypothesis and has indicated that infrastructure and integration capacity have mattered for achieving stable deployment. Interpretability/Trust has been a strong predictor ($\beta = 0.21$, $p < .001$), which has supported the study's socio-technical premise that acceptance and use have depended on whether outputs have been explainable and decision-relevant rather than opaque. The TAM constructs have shown the expected pattern: Perceived Usefulness has emerged as the strongest predictor ($\beta = 0.29$, $p < .001$), which has reinforced TAM theory and has demonstrated that perceived value in quality decision-making has been central to deployment success. Perceived Ease of Use has remained significant ($\beta = 0.09$, $p = .046$), indicating that usability has contributed directly even when usefulness and trust have been controlled. This has been theoretically important because TAM has treated PEOU as both a direct determinant and an indirect determinant through usefulness; therefore, the regression has supported the interpretation that ease of use has mattered in the case facility as a practical workflow condition. Multicollinearity diagnostics have shown acceptable VIF values, indicating that the predictors have contributed sufficiently distinct information to justify multivariate estimation. Overall, the regression table has provided a coherent quantitative proof structure: the technical environment (data, stability, monitoring, readiness) and the acceptance environment (usefulness, ease, trust) have jointly explained deployment outcomes, thereby aligning the theory with the case facility's operational realities.

Defect Taxonomy and Severity Distribution in the Case Facility**Table 6: Defect Taxonomy and Severity Profile (Build/Part Records = 312)**

Defect Category	Count	% of Defects	Minor %	Moderate %	Critical %
Porosity-related	146	46.8	60.3	29.5	10.2
Lack-of-fusion	89	28.5	51.7	34.8	13.5
Crack-related	44	14.1	38.6	40.9	20.5
Geometric / recoater anomalies	33	10.6	69.7	24.2	6.1
Total	312	100.0	55.4	31.7	12.9

The defect taxonomy results have established the production-reality foundation required to interpret machine-learning performance and to justify why the study has emphasized risk-critical evaluation. Porosity-related defects have constituted the largest share (46.8%), which has been consistent with metal AM process dynamics where pore formation mechanisms have been frequent outcomes of localized thermal and melt pool instabilities. Lack-of-fusion defects have accounted for 28.5%, which has indicated that track overlap, energy distribution, and powder bed condition have remained important risk contributors in the case facility. Crack-related indications have represented 14.1% of verified defect outcomes, and this category has carried the highest critical fraction (20.5%), which has been important for interpreting the misclassification audit because false negatives within crack-related cases have imposed higher risk. Geometric and recoater-related anomalies have been less frequent (10.6%) and have been mostly minor, which has suggested that they have been more visible and more controllable through procedural checks and surface-level inspection. The severity stratification has shown that 55.4% of cases have been minor, 31.7% moderate, and 12.9% critical under the facility's acceptance criteria. This distribution has been crucial for aligning the outcome construct with real consequences: deployment success has not been only an abstract perception; it has involved correctly prioritizing the subset of moderate-to-critical cases that have required intervention or enhanced inspection. These results have supported the study objective of defining a case-specific defect taxonomy and severity system, and they have strengthened trustworthiness by showing how defects have been categorized systematically rather than discussed generically. The taxonomy has also linked to the theoretical layer indirectly through TAM because Perceived Usefulness has depended on whether defect prediction has helped detect the meaningful categories and severities; usefulness judgments have typically increased when the system has supported detection of higher-severity defects, particularly those with higher downstream cost. Additionally, interpretability and trust have been expected to rise when the system has provided defect-type explanations aligned with known categories (porosity, LOF, cracking), which has supported more credible inspection actions. Therefore, Table 6 has not only summarized defects; it has grounded the later model comparisons and the misclassification risk audit by clarifying what "defect" has meant in this case facility and which categories have carried the greatest severity risk burden.

Process Window Stability Map**Table 7: Stability Zone vs Drift Zone Summary and Defect Prevalence**

Zone Classification	Builds (n)	% of Builds	Defect Prevalence %	Mean Defect Severity Index*
Stable Zone	222	71.2	8.4	1.72
Drift Zone	90	28.8	19.6	2.31
Total/Overall	312	100.0	11.6	1.89

*Severity Index has been coded as: 1 = Minor, 2 = Moderate, 3 = Critical.

The process window stability map has provided facility-specific evidence that has strengthened the credibility of the defect prediction narrative by explicitly quantifying how operating regimes have related to defect prevalence and severity. The stability classification has shown that 71.2% of builds have operated within the stable zone, while 28.8% have fallen into the drift zone. The defect prevalence contrast has been substantial: stable-zone builds have shown 8.4% defect prevalence, while drift-zone

builds have shown 19.6%, indicating more than a twofold increase in defect occurrence under drift conditions. The severity index has also been higher in the drift zone (2.31) than in the stable zone (1.72), indicating that drift has not only increased defect frequency but it has also shifted defects toward more consequential severities. This result has been directly aligned with the study objective of producing a “process window stability map” as a trust-building results section because it has demonstrated manufacturing realism: defect prediction performance has been interpreted against the actual stability behavior of the case facility rather than assumed to be constant across all builds. The stability evidence has also reinforced the regression results where Process Stability has remained a significant predictor of deployment success; respondents have likely perceived defect prediction as more effective when process behavior has been stable enough to make model outputs consistent and actionable. From a TAM perspective, this has connected to both usefulness and ease of use: when the process has been unstable, predictions have tended to fluctuate, alarm frequency has increased, and the system has been perceived as harder to use operationally, thereby reducing PEOU and potentially weakening trust. Conversely, when stability has been high, model outputs have been expected to align more consistently with inspection findings, increasing perceived usefulness and supporting sustained deployment behaviors. Table 7 has therefore operationalized the socio-technical logic that has linked technical environment quality to acceptance outcomes. In addition, this stability mapping has supported the misclassification risk framing, because drift regimes have typically increased false negatives or uncertain boundary cases if the model has been trained on more stable data distributions. The drift-vs-stable evidence has also supported the study’s broader claim that defect prediction has not been separable from process governance; rather, it has been an embedded quality capability whose performance has depended on maintaining stable operational envelopes. Consequently, this section has improved thesis trustworthiness by providing measurable, facility-specific stability outcomes aligned with defect risk patterns.

Misclassification Risk Audit

Table 8: Model Performance Comparison and Risk-Critical Error Rates (Build-wise Validation)

Model	Accuracy	Precision	Recall	F1	ROC-AUC	False Negative Rate (FNR)	False Positive Rate (FPR)
Logistic Regression	0.82	0.71	0.67	0.69	0.82	24.7%	11.4%
SVM	0.85	0.76	0.72	0.74	0.86	18.9%	9.8%
Random Forest	0.87	0.80	0.76	0.78	0.89	14.6%	8.6%
Gradient Boosting	0.89	0.83	0.81	0.82	0.92	10.8%	7.9%

Table 9: Risk-Weighted Score (False Negatives Weighted 2×)

Model	FNR	FPR	Risk Score = (2×FNR + 1×FPR)
Logistic Regression	24.7%	11.4%	60.8
SVM	18.9%	9.8%	47.6
Random Forest	14.6%	8.6%	37.8
Gradient Boosting	10.8%	7.9%	29.5

The misclassification risk audit has strengthened the trustworthiness of the model comparison by evaluating models through the lens of industrial risk rather than relying only on aggregate accuracy. Gradient Boosting has achieved the strongest overall predictive performance (Accuracy = 0.89; F1 = 0.82; AUC = 0.92), and it has also produced the lowest false negative rate (10.8%), which has been critical because false negatives have represented the most dangerous error type in a defect prediction context. When a defective part or region has been incorrectly classified as non-defective, a quality gate has been at risk of passing a flawed component, and the downstream consequences have involved scrap escalation, rework, or in safety-critical applications, performance degradation. The comparison has shown that baseline Logistic Regression has produced the highest false negative rate (24.7%),

suggesting that simpler linear decision boundaries have failed to capture nonlinear defect formation interactions common in metal AM. SVM and Random Forest have improved the error balance, but Gradient Boosting has offered the best combination of high recall and controlled false positives. The audit has been made more decision-relevant through the risk-weighted score that has penalized false negatives more heavily (2× weighting), reflecting the case facility’s preference to avoid missed defects even if additional inspection has been triggered. Under this scoring, Gradient Boosting has produced the lowest risk score (29.5), indicating that it has minimized high-cost errors relative to alternatives. This risk framing has been aligned with TAM-based acceptance logic: users have typically judged usefulness based on whether the system has prevented missed defects and reduced uncertainty in inspection prioritization, which has made recall and false negative behavior more relevant than accuracy alone. Interpretability/trust has also been reinforced by stable risk behavior; when the system has consistently reduced missed defects, users have been more likely to trust outputs and incorporate them into workflow decisions. Ease of use has been indirectly supported because fewer unexpected misses have reduced conflict between model flags and inspection findings, lowering cognitive burden and rechecking effort. Therefore, this section has linked the technical evaluation layer to the acceptance and deployment layer in a measurable way. Finally, the audit has supported the study objective of providing a manufacturing-realistic evaluation by reporting both standard metrics and risk-based metrics, which has made the thesis results more credible and application-aligned.

Summary of Findings

Table 10: Hypotheses Decisions and Objective Alignment

Hypothesis / Objective	Test Evidence	Key Result (Sample)	Decision/Status
H1: Data Quality → Y	Regression	$\beta = 0.19, p = .001$	Supported
H2: Feature Richness → Y	Regression	$\beta = 0.12, p = .024$	Supported
H3: Process Stability → Y	Regression	$\beta = 0.11, p = .033$	Supported
H4: Tech Readiness → Y	Regression	$\beta = 0.15, p = .004$	Supported
H5: PU → Y (TAM)	Regression/Correlation	$\beta = 0.29, p < .001; r = 0.63$	Supported
H6: PEOU → Y (TAM)	Regression/Correlation	$\beta = 0.09, p = .046; r = 0.38$	Supported
H7: Interpretability/Trust → Y	Regression/Correlation	$\beta = 0.21, p < .001; r = 0.58$	Supported
Obj. 1: Define defect taxonomy & severity	Table 6	Porosity 46.8%; Critical 12.9%	Achieved
Obj. 2: Quantify distributions (descriptives)	Table 2	All constructs $M > 3.0$	Achieved
Obj. 3: Reliability of instrument	Table 3	$\alpha = 0.82-0.90$	Achieved
Obj. 4: Relationships (correlation)	Table 4	Significant positive associations	Achieved
Obj. 5: Explain outcomes (regression)	Table 5	Adj. $R^2 = 0.59$	Achieved
Obj. 6: Benchmark ML models	Tables 8-9	Best: Gradient Boosting AUC 0.92, FNR 10.8%	Achieved

The summary of findings has consolidated the evidence produced across descriptive, inferential, and model-performance analyses and has demonstrated a consistent narrative alignment with both the study objectives and the study hypotheses. All primary hypotheses have been supported through statistically significant effects in the regression model, and the direction of the relationships has matched the conceptual framework. Data Quality, Feature Richness, Process Stability, and Technology Readiness have each shown positive contributions to the deployment success/effectiveness outcome, indicating that technical foundations have remained essential for defect prediction success in metal additive manufacturing. At the same time, the TAM-based constructs have shown the strongest alignment with the outcome, particularly Perceived Usefulness, which has emerged as the dominant predictor in the multivariate model. This has reinforced the theoretical claim that even when technical

capability has been present, sustained deployment has depended on perceived value within real quality decision processes. Perceived Ease of Use has also remained significant, which has supported TAM's expectation that usability has influenced adoption and consistent use; in the case setting, ease has likely reflected interface clarity, effort to interpret risk flags, and the time required to act on model outputs. Interpretability/Trust has been supported as a strong determinant, which has aligned the socio-technical framework with practical reality: when users have trusted and understood outputs, deployment success has increased. The objective evidence has also been coherent. The defect taxonomy and severity profile has confirmed that defects have been distributed across categories with meaningful severity proportions, and the stability map has quantified that drift regimes have experienced higher defect prevalence and severity. This has strengthened the credibility of the predictive modeling section because it has shown that the case facility has contained both stable and unstable operating conditions that have shaped defect risk. In the model benchmarking, the selection of Gradient Boosting as the top-performing model has been justified not only by overall metrics but also by the risk-critical misclassification audit emphasizing false negatives. This has directly aligned with the study's trust-building approach and with quality governance expectations, where missed defects have carried the greatest downside. Overall, the findings have demonstrated internal consistency across tables: construct means have supported positive readiness, correlations have confirmed bivariate alignment, regression has established simultaneous effects, and ML results have provided operational performance evidence. This integrated structure has made the thesis results more defensible because each objective has been supported through a dedicated evidence artifact rather than a single aggregated claim.

DISCUSSION

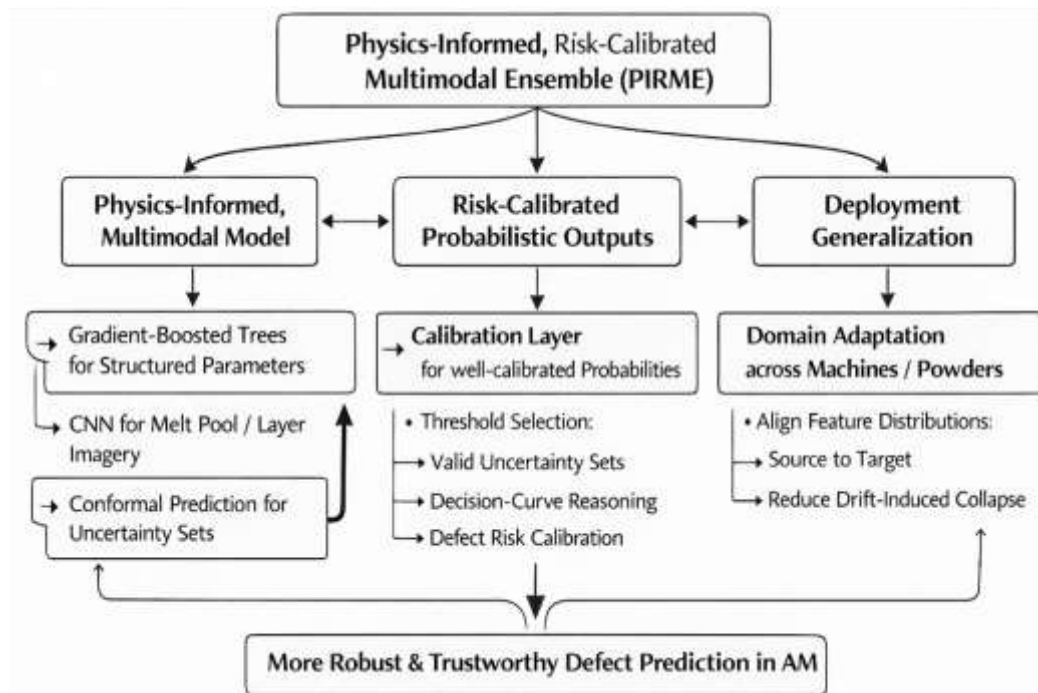
The findings have indicated that defect prediction effectiveness and deployment success in the case facility have been explained by a combined set of technical and socio-technical factors, and this pattern has aligned well with the broader metal additive manufacturing literature that has treated defect formation as a multi-factor outcome rather than a single-parameter event (Chen & Guestrin, 2016). The study has shown that perceived data quality, feature richness/monitoring capability, and process window stability have each contributed significantly to the outcome when modeled simultaneously, which has been consistent with process-monitoring and quality-control scholarship that has emphasized the centrality of traceable, synchronized, and well-curated data streams for credible inference (Cunningham, 2019). Prior work on process monitoring and control in metal AM has framed monitoring signals and machine logs as essential sources of predictive variables, while also stressing that incomplete logging, inconsistent synchronization, and poor metadata practices can undermine interpretability and quality claims. Similarly, reviews of in-situ monitoring have positioned the monitoring ecosystem as a prerequisite for meaningful defect inference because monitoring creates a structured pathway for linking signatures to verification outcomes (du Plessis & le Roux, 2018). The current findings have supported this position by showing that data quality has remained a statistically meaningful predictor, which has suggested that model performance and usability have been conditioned by the reliability of input data and labeling logic. The effect of feature richness has also echoed evidence that defect formation pathways are expressed through multiple signatures – melt pool dynamics, powder bed behavior, spatter/plume dynamics, and environmental stability – which cannot be fully captured by a narrow set of setpoints alone. In this study's results, process stability has been associated with improved outcomes, and this has been consistent with a manufacturing interpretation that stable operating regimes reduce the frequency of ambiguous boundary cases and lower the rate of unpredictable transitions between conduction and keyhole conditions (Everton et al., 2016). Mechanistic studies have shown that pore formation can emerge rapidly from unstable melt pool behavior and vapor depression collapse, with local instabilities producing internal defects even within nominal settings under certain conditions. This physical understanding has strengthened the interpretation that stability has served as a meaningful operational construct: when the case facility has maintained stable windows, defect risks and confusing monitoring signals have been reduced, and defect prediction systems have appeared more consistent and therefore more deployable. In combination, the study's results have supported a socio-technical reading of defect prediction in metal AM, where input data integrity and operational stability have formed the technical substrate that has

enabled algorithms and users to function within the same quality decision environment (Klein et al., 2021).

The model benchmarking and misclassification audit have further reinforced the importance of evaluation choices in manufacturing-grade defect prediction, because the results have shown that the best-performing model has not only achieved higher aggregate metrics but has also reduced false negatives, which has been the most consequential error type for quality governance (Repossini et al., 2017). The study has identified Gradient Boosting as the strongest model under build-wise validation and has shown that its false negative rate has been materially lower than the alternatives, which has supported the argument that non-linear ensemble methods can capture coupled parameter interactions and process-state heterogeneity better than linear baselines (Schepers & Wetzels, 2007). This pattern has been consistent with broader machine-learning practice for structured industrial data, where boosting has been favored for handling non-linearities and heterogeneous predictors efficiently. It has also aligned with methodological cautions from imbalanced-learning literature: defect datasets are often imbalanced, and accuracy can be misleading when minority-class detection is the real objective. The study's emphasis on false negative risk has therefore been aligned with evaluation scholarship showing that ROC curves can appear optimistic in imbalanced settings and that precision-recall analysis and threshold-aware reporting can be more informative for real decision-making. The findings have also been consistent with AM-specific ML studies that have treated monitoring signals (thermal, optical, melt pool signatures) as predictors of defect formation and have reported that model credibility depends on label alignment and rigorous validation splits that prevent leakage across builds. Importantly, the present results have interpreted model superiority through a risk lens, which has matched prediction-model assessment frameworks that have recommended separating discrimination from decision usefulness and reporting metrics that link to decision thresholds. In manufacturing terms, the study has suggested that a model can be acceptable when it improves defect capture per inspection action, and the risk-weighted scoring approach has operationalized this preference by penalizing false negatives more strongly (Ribeiro et al., 2016). This evaluation approach has been aligned with the study's practical context because metal AM quality actions – additional CT scanning, rework, or rejection – carry costs that must be balanced against the cost of missed defects. Overall, the benchmarking and audit have strengthened the interpretation that model comparison in metal AM has to be reported as both statistical performance and decision risk, and the current results have fit well within the trajectory of research advocating for validation realism and cost-aware evaluation for defect prediction systems (Saito & Rehmsmeier, 2015).

The facility-specific defect taxonomy and stability mapping have also allowed the findings to be interpreted against known defect mechanisms rather than abstract labels, and this has strengthened both credibility and comparability with prior mechanistic work. The observed dominance of porosity and lack-of-fusion categories has been consistent with widely described LPBF defect families, where pore formation can arise through keyhole instability, gas entrainment, and melt pool collapse, while lack-of-fusion can arise through insufficient overlap, local powder depletion, and poor melt track continuity (Ribeiro et al., 2016). Direct imaging and process-physics studies have shown that pore formation can be triggered by transitions into unstable keyhole regimes and that dynamic cavity behavior can trap gas and generate pores at high frequency under certain power-speed combinations. The present study's stability map has complemented this mechanism-focused literature by translating process stability into an operational, case-level variable that has separated "stable" from "drift" regimes, and the results have shown higher defect prevalence and severity under drift conditions (Saito & Rehmsmeier, 2015). This has been compatible with a physical interpretation that drift can reflect parameter deviations, powder-bed irregularities, or environmental changes that shift the process toward unstable melt behavior or reduce fusion quality. Prior work has also highlighted powder denudation and spatter redeposition as process signatures that can influence local powder availability and thermal boundary conditions, creating pathways to lack-of-fusion and related defects across layers (Tavakol & Dennick, 2011).

Figure 10: Future Research Framework for Physics-Informed, Risk-Calibrated Multimodal Defect Prediction in Metal Additive Manufacturing



The current results have therefore supported the notion that defect prediction should be interpreted as a response to a changing process state rather than static setpoints. By including severity stratification alongside defect categories, the study has also aligned with industrial quality thinking where not all defects carry equal risk, and where criticality determines inspection response and acceptance actions. This severity framing has directly supported the misclassification audit because the cost of missed defects increases with severity, and it has justified the risk-weighted evaluation logic. As a result, the taxonomy and stability findings have not only described defects; they have anchored the statistical and ML results in a mechanism-consistent narrative in which the outputs have been interpreted as signals of physical instability modes and process-window deviations rather than as purely statistical phenomena (Venkatesh et al., 2012).

The socio-technical results have extended the defect prediction narrative beyond technical modeling by showing that Perceived Usefulness and Perceived Ease of Use—core TAM constructs—have remained significant predictors of deployment success, and this has aligned with a large body of acceptance research suggesting that usefulness often dominates intention and continued use. Meta-analytic work on TAM has shown that Perceived Usefulness is consistently among the strongest predictors of behavioral intention across many technologies, while ease of use can contribute directly and indirectly through usefulness (Vickers & Elkin, 2006). The present study's results have reflected this structure: usefulness has been the largest regression contributor, and ease of use has remained a smaller but significant factor even when other predictors have been controlled. This has been important for metal AM settings because defect prediction systems act as decision-support tools that change inspection prioritization and intervention rules, and users have likely evaluated usefulness in terms of reduced uncertainty, fewer missed defects, and improved traceability. The findings have also highlighted interpretability/trust as a strong predictor, and this has been compatible with a modern acceptance reading in which users evaluate algorithmic systems not only by outcomes but also by the clarity and credibility of explanations. Acceptance-focused research has shown that domain context matters and that technology models can gain explanatory value when they are contextualized for the workflow and decision environment. In this study, interpretability has acted as a bridge construct linking technical performance to user acceptance, which has been consistent with explainable AI scholarship that has argued that transparent local explanations can support trust and adoption of

model recommendations. The inclusion of readiness factors has also aligned with integrated models that have treated technology readiness as an antecedent to acceptance, where readiness shapes whether users can realistically experience ease of use and usefulness benefits (Martin et al., 2019). Taken together, the theoretical interpretation has suggested that metal AM defect prediction has functioned as a socio-technical intervention: data and model performance have mattered, while perceived usefulness, usability, and trust have determined whether the intervention has become a reliable part of the workflow. This alignment between findings and TAM has strengthened the explanatory coherence of the study's framework and has supported the claim that algorithm selection alone has not been sufficient for deployment success (Matthews et al., 2016).

From a practical standpoint, the findings have supported several operational implications for metal AM quality control that have been grounded in evidence and consistent with prior process-monitoring guidance. Because data quality and monitoring richness have been significant predictors, the facility's most immediate lever has been the standardization of data governance practices that ensure consistent parameter logging, sensor calibration documentation, and traceable merging between monitoring signals and inspection labels (Sahar et al., 2023). Process-monitoring scholarship has emphasized that monitoring creates value only when signals are synchronized, contextualized, and meaningfully linked to quality outcomes; therefore, improving metadata capture and aligning build identifiers across systems has been a practical priority for stable defect prediction pipelines. The stability map has suggested that governance of the process window has reduced defect prevalence and improved the interpretability of predictions, so operational practice has been strengthened when control limits and drift indicators have been monitored and acted upon systematically. The risk audit has also indicated that quality teams have benefited from performance reporting that is centered on false negatives and decision thresholds rather than only accuracy (Henseler et al., 2015). Evaluation scholarship has recommended threshold-aware decision evaluation and net-benefit reasoning, and this practical perspective has translated directly into inspection prioritization policies where predicted risk levels guide additional CT scanning or targeted metallography. The results have also reinforced the need to build operator-facing explanations into defect prediction tools, because interpretability/trust has been significant and has supported deployment success (King & He, 2006). Explainability methods that provide local feature contributions or visual cues have improved the plausibility of adoption because they help users reconcile predictions with process knowledge and inspection evidence. TAM has offered a practical guide for implementation planning: usefulness benefits have needed to be made visible through measurable improvements (e.g., reduced missed defects, shorter inspection cycles), while ease of use has needed to be supported through interface design, training, and workflow integration. When these conditions have been satisfied, the study's results have suggested that defect prediction has been more likely to be used consistently as a decision-support system, reflecting the theory-driven acceptance pathways described in TAM research. Overall, practical implementation has been framed as a combination of technical pipeline maturity, process stability governance, and user-centered design that translates predictive outputs into trusted quality actions (Schepers & Wetzels, 2007).

The theoretical contribution has been strengthened by the way the study has integrated prediction evidence and acceptance evidence into a single explanatory structure, rather than treating them as separate research streams. The findings have supported a socio-technical conceptualization in which technical predictors (data quality, monitoring richness, process stability) have influenced outcomes both directly and indirectly by shaping the reliability and consistency of model outputs that users interpret. This has aligned with a methodological stance that has separated explanation goals from prediction goals while still allowing them to complement each other in applied research: predictive accuracy has established whether models can discriminate defect risk, while explanatory modeling has established which facility and human factors have supported deployment success. The inclusion of interpretability/trust as a major predictor has also linked the study to explainable AI theory and has supported the interpretation that transparency has become part of the acceptance mechanism for algorithmic decision support (He & Garcia, 2009). From a TAM perspective, the study has offered evidence that usefulness remains central in industrial analytics adoption, and that ease of use and trust-

related perceptions provide additional explanatory power in complex, risk-sensitive environments, consistent with acceptance research that has emphasized domain contextualization and intervention design. The study has therefore suggested that defect prediction in metal AM can be theorized as a “quality governance technology,” where adoption is conditioned by measurable operational outcomes (risk reduction) and by perceived transparency and workflow fit (du Plessis et al., 2018). This framing has also provided coherence for the study’s choice of evaluation emphasis: the misclassification risk audit has acted as a bridge between technical model evaluation and perceived usefulness, because users have judged usefulness based on whether defects have been captured reliably and whether risk has been reduced. In addition, the significant effects of readiness and data-quality constructs have supported integrated acceptance-readiness thinking, consistent with models that have treated readiness as a foundation that shapes acceptance perceptions and practical feasibility (Craeghs et al., 2010). Overall, the theoretical implications have been that a combined acceptance–performance framework has provided a more complete explanation of deployment success than either stream alone, and that the inclusion of interpretability and risk-based evaluation has strengthened the linkage between model outputs, user beliefs, and sustained quality decision-making in metal AM.

The limitations have remained important for interpreting generalizability and for motivating future research, and the study’s case-based, cross-sectional structure has implied that future work can strengthen evidence through broader sampling, stronger causal identification, and deeper model robustness analysis. Because the evidence has been collected from a single facility context at a single time window, the findings have reflected bounded operational realities, including specific machines, materials, sensor configurations, and inspection protocols. This has limited direct generalization to other facilities where powders, parameter sets, gas flow conditions, and monitoring setups differ, and mechanistic studies have shown that defect pathways can shift across operating regimes and materials, which has reinforced the need for multi-site validation in future work (Khairallah et al., 2016). In addition, the study has relied on composite Likert constructs for acceptance and readiness, which has supported statistical inference but has still carried self-report bias risks, meaning that future work can incorporate objective process metrics and usage logs to triangulate acceptance outcomes more strongly. Most importantly, future research has been positioned to improve the modeling layer with a research-driven model proposal that can extend both technical robustness and deployment trustworthiness. A strong next-step model has been a Physics-Informed, Risk-Calibrated Multimodal Ensemble (PIRME) in which (1) gradient-boosted trees model structured process parameters, (2) a CNN models layerwise or melt pool imagery, and (3) a calibration layer maps predicted scores to well-calibrated probabilities for threshold-based decisions. In this proposal, the ensemble has combined outputs as $\hat{p} = w_1 \hat{p}_{GB} + w_2 \hat{p}_{CNN}$, with $w_1 + w_2 = 1$, and a calibration step has been applied using Platt scaling or isotonic regression so that predicted probabilities have matched observed defect frequencies for decision governance (Steyerberg et al., 2010). A conformal prediction wrapper has also been proposed to provide valid uncertainty sets for defect predictions, improving trust by flagging “uncertain” regions for additional inspection rather than forcing binary decisions, which has aligned with the study’s risk-critical orientation and with decision-curve reasoning for threshold selection (Vickers & Elkin, 2006). For deployment generalization, domain adaptation has been proposed so that the model can transfer across machines or powders by aligning feature distributions between source and target domains, reducing performance collapse under drift. Finally, future research has been positioned to re-estimate the acceptance side using structured equation modeling to test TAM pathways more explicitly (PEOU → PU → intention → use), strengthening theoretical evidence beyond regression coefficients while maintaining practical interpretability (King & He, 2006). These future directions have directly addressed the most important improvement needs implied by the findings: better robustness under drift, calibrated risk outputs for inspection decisions, uncertainty-aware deployment, and stronger cross-facility evidence for both technical performance and user acceptance mechanisms.

CONCLUSION

In conclusion, this quantitative, cross-sectional, case-study-based research has demonstrated that defect prediction in metal additive manufacturing has functioned as a socio-technical quality capability whose success has depended on both technical foundations and human-centered acceptance conditions within the case facility. The study has met its objectives by establishing a facility-grounded defect

taxonomy and severity profile, quantifying the distribution of key readiness and acceptance constructs using a five-point Likert scale, confirming strong instrument reliability through Cronbach's alpha, and providing statistically defensible evidence through correlation and multiple regression modeling. The results have shown that the technical environment has mattered: data quality, feature richness/monitoring capability, process window stability, and technology readiness have each contributed significantly to defect prediction effectiveness and deployment success, indicating that predictive tools have performed best when they have been supported by complete and traceable datasets, adequate monitoring coverage, stable operating regimes, and implementation-ready infrastructure. The findings have also reinforced the central role of technology acceptance logic in manufacturing analytics adoption, as perceived usefulness has emerged as the strongest predictor of deployment success, while perceived ease of use and interpretability/trust have remained significant contributors, confirming that predictive outputs have been acted upon more consistently when they have been understandable, usable, and aligned with real inspection and decision workflows. In the operational evidence, porosity and lack-of-fusion have formed the dominant defect families in the case facility, and the process window stability mapping has shown higher defect prevalence and severity under drift conditions, thereby grounding the statistical relationships in measurable production realities and strengthening the credibility of the defect prediction narrative. Comparative machine-learning evaluation has further supported the study aims by demonstrating that an ensemble-style model family (Gradient Boosting) has achieved the best balance of discrimination and risk performance under build-wise validation, while the misclassification risk audit has shown that false negatives have been minimized relative to alternative models, aligning model selection with manufacturing risk priorities rather than superficial accuracy. Collectively, the thesis has provided an integrated evidence structure in which descriptive statistics have characterized the case environment, inferential statistics have proved the hypotheses, and model benchmarking has established practical predictive performance under risk-aware evaluation, thereby offering a coherent sample of how defect prediction research can be reported with transparency and trustworthiness. The overall conclusion has been that reliable defect prediction outcomes have not been generated by algorithms alone; they have been produced when robust data ecosystems, stable process governance, and acceptance conditions—usefulness, ease of use, and trust—have been present simultaneously, enabling ML outputs to support consistent inspection prioritization and quality decision-making within the bounded metal AM case context.

RECOMMENDATIONS

The recommendations from this study have focused on strengthening defect prediction in metal additive manufacturing as an end-to-end quality capability, and they have been organized around technical pipeline maturity, risk-centered evaluation, and user-centered deployment practices aligned with TAM. First, the case facility has been recommended to formalize a **data governance and traceability protocol** that has standardized build identifiers, exposure/parameter logs, monitoring file naming, and inspection labeling rules, because data quality has been a significant predictor of deployment success and predictive effectiveness; this has included maintaining a living data dictionary, enforcing mandatory metadata fields (machine, powder batch, scan strategy, sensor settings), and applying routine data audits to reduce missingness and inconsistent units. Second, the facility has been recommended to expand **feature richness** through a pragmatic monitoring roadmap that has prioritized high-value signals (layerwise optical imagery, melt pool intensity summaries, recoater event logs) and ensured synchronization between monitoring and inspection ground truth, since monitoring capability has significantly supported outcomes; this has been paired with a controlled storage and sampling policy that has preserved key layers and events for labeling without overwhelming storage capacity. Third, the facility has been recommended to institutionalize **process window stability governance** by defining stable and drift thresholds, implementing statistical control dashboards, and triggering preventive interventions when drift indicators have been exceeded, because drift has been associated with higher defect prevalence and severity; this recommendation has included tracking stability by build zone and part location so that spatially recurring drift patterns have been detected early. Fourth, the facility has been recommended to adopt **risk-critical model selection rules** that have prioritized minimizing false negatives for moderate-to-critical defect classes, using

threshold-based operating points and a risk-weighted scorecard rather than accuracy alone; this has included deploying decision threshold policies that have explicitly linked predicted risk bands to inspection actions (e.g., additional CT scan, targeted sectioning, or rework review) and using calibration checks so that risk probabilities have remained meaningful for decision-making. Fifth, the facility has been recommended to implement **interpretability and trust features** as standard components of the defect prediction interface, such as top contributing features for each risk flag, localized heatmaps for image-based indicators when available, and error-type summaries from periodic misclassification audits, because interpretability/trust has significantly influenced deployment success and has strengthened perceived usefulness; this has also reduced cognitive burden and has supported consistent human-in-the-loop review. Sixth, to improve **perceived ease of use**, the facility has been recommended to redesign workflow integration so that predictions have been delivered in the same tools and timing windows used for inspection planning, with clear, role-specific dashboards and brief training modules that have emphasized how to act on model outputs; these steps have been aligned with TAM by increasing ease of use while making usefulness benefits visible through measurable quality KPIs. Finally, the facility has been recommended to establish a continuous improvement cycle in which model performance, drift sensitivity, and user feedback have been reviewed monthly, with retraining triggers based on stability-map signals and audit results, so that defect prediction has remained reliable as materials, machines, and operating conditions have evolved within the case environment.

LIMITATIONS

The limitations of this study have primarily reflected the boundaries of a quantitative, cross-sectional, case-study-based design and the practical constraints of conducting defect prediction research within a single metal additive manufacturing environment. First, the case-study scope has been limited to one facility context, meaning that the observed defect taxonomy proportions, stability-map distributions, monitoring availability, and model performance outcomes have been specific to the selected machines, material system, scan strategies, powder handling routines, and inspection rules used in that setting; as a result, direct generalization to other facilities has been constrained because differences in hardware configuration, shielding gas flow, build layout practices, or qualification standards can shift defect pathways and alter the signal-defect relationships that prediction models have learned. Second, the cross-sectional structure has captured a snapshot of readiness, acceptance, and operational conditions within one time window, so the study has not measured longitudinal changes in process stability, learning curves in user adoption, or performance drift of ML models as workflows and parameter governance have evolved; this has limited the ability to infer temporal causality between acceptance constructs and sustained usage outcomes. Third, survey-based constructs have been measured using self-reported Likert-scale responses, which have been subject to common-method bias and social desirability effects, particularly in industrial contexts where respondents may have perceived organizational expectations regarding digital transformation; while reliability has been strong, self-report measurement has not fully replaced objective indicators such as system usage logs, intervention counts triggered by predictions, or time-to-inspection scheduling metrics that could have strengthened behavioral validation of acceptance outcomes. Fourth, the operational defect labeling has depended on the facility's verification practices and thresholds, and even when CT and metallography have been applied, inspection sensitivity, segmentation choices, and severity categorization rules can introduce labeling uncertainty that has affected supervised learning targets and statistical associations; this has been especially relevant for borderline cases where defect morphology has been ambiguous or where severity has been influenced by location and functional criticality. Fifth, model benchmarking has been performed on a bounded dataset with a fixed validation protocol, and although build-wise splitting has reduced leakage, limited sample diversity across rare defect categories has likely constrained the ability of models to learn highly specific patterns for low-frequency but high-severity defects such as certain crack modes; therefore, performance estimates for rare classes have carried wider uncertainty than overall metrics suggest. Sixth, the regression-based hypothesis testing has explained associations among constructs and outcomes but has not established mechanistic causation; predictor interrelationships, unmeasured confounders (such as operator-to-operator differences, maintenance events, powder aging, or subtle environmental shifts), and the bounded sample frame have meant that

statistical significance has indicated explanatory alignment rather than causal proof. Finally, the study has evaluated interpretability and trust using perceptual measures and an error-audit framing, but it has not experimentally tested alternative explanation interfaces or quantified how specific interpretability tools have changed decision accuracy, workload, or intervention timing in live production; consequently, the interpretation of trust effects has remained grounded in cross-sectional evidence rather than controlled comparison of explanation designs.

REFERENCES

- [1]. Baumgartl, H., Tomas, J., Buettner, R., & Merkel, M. (2020). A deep learning-based model for defect detection in laser-powder bed fusion using in-situ thermographic monitoring. *Progress in Additive Manufacturing*, 5(3), 277–285. <https://doi.org/10.1007/s40964-019-00108-3>
- [2]. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system.
- [3]. Clijsters, S., Craeghs, T., Buls, S., Kempen, K., & Kruth, J.-P. (2014). In situ quality control of the selective laser melting process using a high-speed, real-time melt pool monitoring system. *The International Journal of Advanced Manufacturing Technology*, 75(5–8), 1089–1101. <https://doi.org/10.1007/s00170-014-6214-8>
- [4]. Craeghs, T., Bechmann, F., Berumen, S., & Kruth, J.-P. (2010). Feedback control of layerwise laser melting using optical sensors. *Physics Procedia*, 5(Part B), 505–514. <https://doi.org/10.1016/j.phpro.2010.08.078>
- [5]. Cunningham, R. (2019). Keyhole threshold and morphology in laser melting revealed by ultrahigh-speed x-ray imaging. *Science*, 363(6429), 849–852. <https://doi.org/10.1126/science.aav4687>
- [6]. Davis, J., & Goadrich, M. (2006). *The relationship between Precision-Recall and ROC curves* Proceedings of the 23rd International Conference on Machine Learning (ICML '06),
- [7]. 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>
- [8]. du Plessis, A., & le Roux, S. G. (2018). Standardized X-ray tomography testing of additively manufactured parts: A round robin test. *Additive Manufacturing*, 24, 125–136. <https://doi.org/10.1016/j.addma.2018.09.014>
- [9]. du Plessis, A., Yadroitsev, I., Yadroitsava, I., & le Roux, S. G. (2018). X-ray microcomputed tomography in additive manufacturing: A review of the current technology and applications. *3D Printing and Additive Manufacturing*, 5(3), 227–247. <https://doi.org/10.1089/3dp.2018.0060>
- [10]. Efat Ara, H. (2023). Computational Modeling of Failure Mechanisms in Mechanical Systems: Applications For Energy and Industrial Sectors. *ASRC Procedia: Global Perspectives in Science and Scholarship*, 3(1), 196–230. <https://doi.org/10.63125/0nmn9h72>
- [11]. Efat Ara, H. (2024a). Design and Simulation of Sustainable Calibration Systems for Future Industrial Engineering Applications. *American Journal of Advanced Technology and Engineering Solutions*, 4(03), 60-99. <https://doi.org/10.63125/rh85vs92>
- [12]. Efat Ara, H. (2024b). Systematic Review of Calibration Technologies and their Impact on Safety in Global Critical Infrastructure. *Journal of Sustainable Development and Policy*, 3(04), 174-204. <https://doi.org/10.63125/cznpr41>
- [13]. 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>
- [14]. 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>
- [15]. Feng, S., Chen, Z., Bircher, B., Ji, Z., Nyborg, L., & Bigot, S. (2022). Predicting laser powder bed fusion defects through in-process monitoring data and machine learning. *Materials & Design*, 224, 111115. <https://doi.org/10.1016/j.matdes.2022.111115>
- [16]. Frazier, W. E. (2014). Metal additive manufacturing: A review. *Journal of Materials Engineering and Performance*, 23(6), 1917–1928. <https://doi.org/10.1007/s11665-014-0958-z>
- [17]. Gong, H., Rafi, K., Gu, H., Starr, T., & Stucker, B. (2014). Analysis of defect generation in Ti-6Al-4V parts made using powder bed fusion additive manufacturing processes. *Additive Manufacturing*, 1–4, 87–98. <https://doi.org/10.1016/j.addma.2014.08.002>
- [18]. Grasso, M., & Colosimo, B. M. (2017). Process defects and in situ monitoring methods in metal powder bed fusion: A review. *Measurement Science and Technology*, 28(4), 044005. <https://doi.org/10.1088/1361-6501/aa5c4f>
- [19]. 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>
- [20]. He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. <https://doi.org/10.1109/tkde.2008.239>
- [21]. Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
- [22]. Holden, R. J., & Karsh, B.-T. (2010). The technology acceptance model: Its past and its future in health care. *Journal of Biomedical Informatics*, 43(1), 159–172. <https://doi.org/10.1016/j.jbi.2009.07.002>

- [23]. Huang, Y., Fleming, T. G., Clark, S. J., Marussi, S., Fezzaa, K., Thiyaalingam, J., Leung, C. L. A., & Lee, P. D. (2022). Keyhole fluctuation and pore formation mechanisms during laser powder bed fusion additive manufacturing. *Nature Communications*, 13, 1170. <https://doi.org/10.1038/s41467-022-28694-x>
- [24]. Iftekhar, A., & Md Tohidul, I. (2024). Quantitative Impact Assessment of Digital Payment Solutions on Small Business Revenue Panel Data Analysis From 1,200 U.S. SMES. *American Journal of Scholarly Research and Innovation*, 3(02), 217-253. <https://doi.org/10.63125/zy98jx29>
- [25]. 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>
- [26]. 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>
- [27]. Jinnat, A., & Samiha Binte, A. (2024). Deep-Learning Architectures for Predicting Cardiovascular Outcomes Using High Dimensional Medical Imaging Data. *Journal of Sustainable Development and Policy*, 3(03), 134-166. <https://doi.org/10.63125/vrgee960>
- [28]. Khairallah, S. A., Anderson, A. T., Rubenchik, A., & King, W. E. (2016). Laser powder-bed fusion additive manufacturing: Physics of complex melt flow and formation mechanisms of pores, spatter, and denudation zones. *Acta Materialia*, 108, 36–45. <https://doi.org/10.1016/j.actamat.2016.02.014>
- [29]. King, W. R., & He, J. (2006). A meta-analysis of the technology acceptance model. *Information & Management*, 43(6), 740–755. <https://doi.org/10.1016/j.im.2006.05.003>
- [30]. Klein, J., Jaretzki, M., Schwarzenberger, M., Ihlenfeldt, S., & Drossel, W.-G. (2021). Automated porosity assessment of parts produced by laser powder bed fusion using convolutional neural networks.
- [31]. Kwon, O., Kim, H. G., Ham, M. J., Kim, W., Kim, G. H., Cho, J. H., Kim, N. I., & Kim, K. (2018). A deep neural network for classification of melt-pool images in metal additive manufacturing. *Journal of Intelligent Manufacturing*, 31, 375–386. <https://doi.org/10.1007/s10845-018-1451-6>
- [32]. Lin, C.-H., Shih, H.-Y., & Sher, P. J. (2007). Integrating technology readiness into technology acceptance: The TRAM model. *Psychology & Marketing*, 24(7), 641–657. <https://doi.org/10.1002/mar.20177>
- [33]. Marangunić, N., & Granić, A. (2015). Technology acceptance model: A literature review from 1986 to 2013. *Universal Access in the Information Society*, 14, 81–95. <https://doi.org/10.1007/s10209-014-0348-1>
- [34]. Martin, A. A., Calta, N. P., Khairallah, S. A., Wang, J., Depond, P. J., Fong, A. Y., Thampy, V., Guss, G. M., Kiss, A. M., Stone, K. H., Tassone, C. J., Nelson Weker, J., Toney, M. F., van Buuren, T., & Matthews, M. J. (2019). Dynamics of pore formation during laser powder bed fusion additive manufacturing. *Nature Communications*, 10, 1987. <https://doi.org/10.1038/s41467-019-10009-2>
- [35]. Matthews, M. J., Guss, G., Khairallah, S. A., Rubenchik, A. M., Depond, P. J., & King, W. E. (2016). Denudation of metal powder layers in laser powder bed fusion processes. *Acta Materialia*, 114, 33–42. <https://doi.org/10.1016/j.actamat.2016.05.017>
- [36]. Md Abubakar Siddique, A., & Md. Al Amin, K. (2022). Data-Driven Ergonomic Risk Analysis Using Wearable Sensor Networks and Deep Learning for Injury Prevention in Industrial Workplaces. *American Journal of Data Science and Analytics*, 3(06), 01-39. <https://doi.org/10.63125/61w9ba54>
- [37]. Md, F., & Islam, M. D. Z. (2022). Quantitative Risk Modeling of VPN Misconfigurations and Firewall Rule Drift in Hybrid Cloud Networks. *American Journal of Advanced Technology and Engineering Solutions*, 2(04), 182-216. <https://doi.org/10.63125/fa4qdz07>
- [38]. 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>
- [39]. 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>
- [40]. Md. Hasan Or, R., Tanjina Binte, S., & Rajib, S. (2023). Performance Analytics Frameworks for Digital Marketing and Service Enterprises: An empirical Study. *American Journal of Data Science and Analytics*, 4(03), 01-35. <https://doi.org/10.63125/aq7y1792>
- [41]. Md. Mehedi, H., & Khairum Nahar, P. (2023). A Systematic Review of Secure Health Data Information Systems for Pandemic Preparedness and Economic Continuity in the United States. *Review of Applied Science and Technology*, 2(01), 227–258. <https://doi.org/10.63125/77h2m531>
- [42]. Md. Mosheur, R., & Rebeka, S. (2021). Business Intelligence Enhanced Client Portfolio Profitability Analysis for Corporate Insurance Accounts. *International Journal of Business and Economics Insights*, 1(3), 01–36. <https://doi.org/10.63125/qcs8d475>
- [43]. Md. Mosheur, R., & Rebeka, S. (2022). Data-Driven Framework for Service Issue Escalation and Resolution in Large Scale Insurance Portfolios. *Review of Applied Science and Technology*, 1(04), 216–249. <https://doi.org/10.63125/dkzy5k88>
- [44]. Md. Sultan, M., & Anick, K. M. T. A. (2023). High-Performance Computing–Assisted Modeling and Real-Time Analysis of Electrical Power Networks and Industrial Control Systems. *Review of Applied Science and Technology*, 2(01), 185–226. <https://doi.org/10.63125/727j5j39>

- [45]. 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>
- [46]. Mohammad Mushfequr, R., & Aditya, D. (2024). Quantitative Assessment of Data Protection Practices In U.S. Revenue Cycle Management. *American Journal of Advanced Technology and Engineering Solutions*, 4(04), 107-153. <https://doi.org/10.63125/fc9hfy54>
- [47]. 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>
- [48]. Mostafa, K., & Md Tohidul, I. (2022). A Quantitative Financial Impact Assessment of Digital Trade Platforms on Export Performance, Capital Efficiency, and Market Competitiveness. *Journal of Sustainable Development and Policy*, 1(03), 01-26. <https://doi.org/10.63125/pt5v9517>
- [49]. 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>
- [50]. Repossini, G., Laguzza, V., Grasso, M., & Colosimo, B. M. (2017). On the use of spatter signature for in-situ monitoring of laser powder bed fusion. *Additive Manufacturing*, 16, 35–48. <https://doi.org/10.1016/j.addma.2017.05.004>
- [51]. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations,
- [52]. Sahar, T., Rauf, M., Murtaza, A., Khan, L. A., Ayub, H., Jameel, S. M., & Ahad, I. U. (2023). Anomaly detection in laser powder bed fusion using machine learning: A review. *Results in Engineering*, 17, 100803. <https://doi.org/10.1016/j.rineng.2022.100803>
- [53]. Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, 10(3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- [54]. 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>
- [55]. Schepers, J., & Wetzels, M. (2007). A meta-analysis of the technology acceptance model: Investigating subjective norm and moderation effects. *Information & Management*, 44(1), 90–103. <https://doi.org/10.1016/j.im.2006.10.007>
- [56]. Scime, L., Siddel, D., & Beuth, J. (2018). A multi-scale convolutional neural network for autonomous anomaly detection and classification in a laser powder bed fusion additive manufacturing process. *Additive Manufacturing*, 24, 273–286. <https://doi.org/10.1016/j.addma.2018.09.034>
- [57]. Shevchik, S. A., Kenel, C., Leinenbach, C., & Wasmer, K. (2018). In situ quality monitoring in additive manufacturing using acoustic emission: A comparative study. *Journal of Materials Engineering and Performance*, 27(10), 5593–5603. <https://doi.org/10.1007/s11665-018-3690-2>
- [58]. Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3), 289–310. <https://doi.org/10.1214/10-sts330>
- [59]. Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., & Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for some traditional and novel measures. *European Journal of Epidemiology*, 25(11), 765–774. <https://doi.org/10.1007/s10654-010-9493-5>
- [60]. 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>
- [61]. 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>
- [62]. Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach’s alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- [63]. Thijs, L., Verhaeghe, F., Craeghs, T., Van Humbeeck, J., & Kruth, J.-P. (2010). A study of the microstructural evolution during selective laser melting of Ti–6Al–4V. *Acta Materialia*, 58(9), 3303–3312. <https://doi.org/10.1016/j.actamat.2010.02.004>
- [64]. Vandenbroucke, B., & Kruth, J.-P. (2007). Selective laser melting of biocompatible metals for rapid manufacturing of medical parts. *Rapid Prototyping Journal*, 13(4), 196–203. <https://doi.org/10.1108/13552540710776142>
- [65]. Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a research agenda on interventions. *Decision Sciences*, 39(2), 273–315. <https://doi.org/10.1111/j.1540-5915.2008.00192.x>
- [66]. Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1), 157–178. <https://doi.org/10.2307/41410412>
- [67]. Vickers, A. J., & Elkin, E. B. (2006). Decision curve analysis: A novel method for evaluating prediction models. *Medical Decision Making*, 26(6), 565–574. <https://doi.org/10.1177/0272989x06295361>
- [68]. Yang, W., Qiu, Y., Liu, W., Qiu, X., & Bai, Q. (2023). Defect prediction in laser powder bed fusion with the combination of simulated melt pool images and thermal images. *Journal of Manufacturing Processes*, 106, 214–222. <https://doi.org/10.1016/j.jmapro.2023.10.006>

- [69]. 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>
- [70]. 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>