

From Digital Twin to Competitive Advantage: Designing a Digital Thread-Enabled Predictive Quality Data Ecosystem for Aerospace Manufacturing

Renata ȘEPTICHITA*

National University of Science and Technology POLITEHNICA Bucharest, Romania

*Corresponding author, renata.septichita@stud.fiir.upb.ro

Augustin SEMENESCU

National University of Science and Technology POLITEHNICA Bucharest, Romania

augustin.semenescu@upb.ro

Abstract. *The competitive advantage in aerospace manufacturing is more and more based on ensuring high conformance at high throughput and audit-ready traceability through complex supply networks. Advanced technologies like laser powder bed fusion (PBF-LB/M) generate rich multi-modal data (manufacturing intent, execution logs, in-situ monitoring, and ex-situ verification), but value is often compromised by fragmented repositories, weak semantic/coordinate alignment, and limited governance of analytics. In this study, we first present a **Digital Thread-enabled Predictive Quality Data Ecosystem (PQDE)** that leverages digital twin capabilities (state estimation and quality-risk forecasting) with measurable operational outcomes. PQDE combines (i) digital-thread continuity from manufacturing intent to inspection evidence through identifier binding and multi-modal registration, (ii) reproducible DataOps and MLOps controls (schema, lineage, acceptance gates, monitoring and controlled changes), and (iii) decision integration through **risk-guided inspection triage** with evidence packaging in alignment with aerospace quality expectations. This approach is validated with the public NIST Additive Manufacturing Metrology Testbed (AMMT) pilot “Overhang Part X4,” which contains a fully registered in-situ/ex-situ dataset, where all of this data is matched with machine coordinates and XCT-derived verification. By leveraging a governance-friendly baseline model and a rare-event label created based on the 1st percentile threshold of the training split, we obtain a strong level of discrimination among validation and cross-part test slices (ROC-AUC 0.953-0.982; PR-AUC up to 0.448). More specifically, Top-10% triage captures **98.20%** of defect-labeled points for validation and **96.88%** for late-layer test slice while only reducing the inspection scope by **90%**; Top-10% triage captures **83.07%** on that more challenging early-layer test slice. Lift values of **~8.31x – 23.82x** reflect the significant defect enrichment in inspected subsets. Finally, it demonstrates KPI proxy templates (inspection reduction, defect capture, enrichment, false-alarm load, decision-latency proxy, and evidence completeness) to transfer the predictive outputs into competitive advantage without needing proprietary cost-of-quality data.*

Keywords: Digital Thread, Digital Twin, Aerospace Manufacturing, Additive Manufacturing, MLOps Governance, Competitive Advantage Inspection Triage.

Introduction

Aerospace firms operate in environments of recurring limitations which add cost to non-conformance and cost of assurance in one, through high conformity requirements, long qualification and change-approval periods, complex procurement networks and the need to prove audit-ready traceability to customers and regulators. In that context, competitive advantage is rarely rooted in a discrete process improvement – rather than in isolated process improvements, competitive advantage arises from being able to identify and address quality risk early, limit

unwarranted non-value-added control (including rework costs) and scale production at a range of sites and partners while protecting the integrity of evidence and the configuration.

The current pressures are highly significant for mature and data-heavy processes such as metal additive manufacturing, where process variability can be subtle, mechanism defects may start early, and verification is costly and time-consuming. However, the traditional quality paradigms based on end-of-line inspection and reactive corrective actions, continue to fail to meet the complex challenges in the high-mix, high-assurance manufacturing landscape. For instance, defect in laser powder bed fusion (PBF-LB/M) such as its porosity or lack of fusion signatures, can appear at the melt pool level and can follow layers to remain hidden until the later stage inspection. Focusing mainly on quality as a principle: inspection that is broad and uniform can put an organization at a balance between throughput and assurance, as high inspection coverage reduces risk but causes bottlenecks and high appraisal costs, whereas by decreasing inspection coverage the risk of downstream escape and expensive disposition actions is increased. As such, the competitive strategy must be more focused and proactive—reflecting the mindset of risk and decisions driven by evidence that are already embedded in aerospace quality (i.e. the IAQG 9100 family on systematic risk management and performance improvement in quality, schedule and cost). (International Aerospace Quality Group - IAQG, 2025)

In this context, predictive analytics and Machine Learning (ML) are frequently offered as the answer. On the other hand, in high-assurance aerospace, the predictive quality is not just "training an ML model". A model which does well in offline notebook can still suffer from operational failure if (i) the data pipeline is not stable and traceable, (ii) feature definitions drift between training and serving, (iii) missingness and outliers are uncontrolled, (iv) the model lifecycle lacks controlled change and monitoring, or competitive advantage in this domain is not built off model accuracy alone; it is built upon AI becoming a repeatable, governed capability within a digital ecosystem -bridging people and processes, data and decisions across the development, adoption, and operation of a business. (Okaro, 2019)

This paper aims to fill this void by introducing a **Digital Thread-enabled Predictive Quality Data Ecosystem (PQDE)** to convert Digital Twin predictive capacity to measured operational performance and business value. The “ecosystem” framing is critical in context of the conference theme Driving competitive advantage through AI and digital ecosystems: PQDE describes predictive quality as a socio-technical system that combines data lineage with governance controls and decision integrations as primary design values rather than an afterthought. The Digital Thread literature has underscored the critical necessity for authoritative information flows and lifecycle connection as the foundation for enterprise-level decision making on board (theoretically, of a company’s value and operational potential); in fact, evidence lineage in a quality perspective seems the most effective for an industry using for example on how to bind manufacturing intent, execution logs, monitoring elements, inspector results, disposition action, maintenance activities or disposition actions using stable identity markers (e.g., with stable data collection and ID. build / part / layer / location / operation).

Complementary to Digital Thread the Digital Twin concept enables predictive and decision-support if realized as a holistic model of the manufacturing elements - product/process/resources - that are able to estimate state and predict it. ISO 23247 shapes digital twins for manufacturing around specific requirements and exchange of information, which in the context of predictive quality entails that risk forecasting and model updates should be able to be made **decisionable, governable, and reconstructable** - with an explicit tie to data and transformations that generated these. That is, **Digital Thread offers connectivity and**

traceability, Digital Twin gives you predictive state and risk models combined and together they provide audit ready predictive quality at ecosystem level.

This paper introduces an ecosystem of digital thread: Digital Thread enabled PQDE as a platform that enables the digital twin capability to be translated (and delivered) into competitive advantage. It is the contribution is architectural but also operationalized with a public pilot and provides quantified results and business-facing KPI proxy templates. (ŞEPTICHITA Renata, SEMENESCU Augustin, 2025)

The aim of the research is to develop and assess a predictive quality ecosystem using Digital Threads, generating outputs from AI that they can translate in an objective way into manufacturing competitiveness.

Literature

1. Concept and Framework of the Future: Digital Thread & Digital Twin as Drivers of Predictive Quality

Lifecycle, evidence lineage across digital threads.

Digital Thread scholarship generally characterizes the latter construct as a well-integrated and authoritative information flow throughout the product lifecycle, enabling decision makers to have access to and turn disparate pieces of data into information they can use. However, there is no universal definition that was clearly discussed elsewhere in literature. Only several repeated findings emerge in the review of the literature: authoritative data sources and linkage across lifecycle stages; integration and transformation of heterogeneous data to inform decision making. (Qiang Zhang, Jihong Liu, Xu Chen, 2024)

For aerospace quality, the most meaningful interpretation is evidence lineage: the capacity of associating manufacturing intent, execution logs, monitoring features and inspection outcomes with disposition actions through stable identifiers (part, build, layer, operation). Such lineage is a prerequisite of traceability, root-cause analysis, supplier communication and audit reconstruction in multi stakeholder ecosystems.

Digital Twin as a managed and aligned representation.

The Digital Twin as a concept is sometimes confused with simulation. In manufacturing operation, its advantage lies in a digital representation that remains synchronized with a manufacturing element (e.g., product, process, resource), allowing for state estimation, and forecasting, as well as possible what-if analyses. ISO 23247 establishes a clear framework for a global digital twin model for manufacturing, with specified requirements and information exchanges. (Guodong Shao, 2023)

In the area of predictive quality, the Digital Twin is very useful only when it generates decisionable and governable outputs: calibrated risk estimates, explainable drivers and controlled lifecycle updates. In other words, Digital Thread offers continuity and evidence lineage, and Digital Twin offers predictive state and risk models. They enable predictive quality at scale for an ecosystem, that is audit ready. (Renata SEPTICHITA, 2025)

The research questions:

- What reference architecture that integrates Digital Thread continuity and Digital Twin predictive capability into an operational predictive quality ecosystem for aerospace manufacturing?
- Is a public metrology-grade pilot able to produce consistent rare event quality-risk prediction under cross-part and temporal robustness constraints?

- How to translate model generated outputs into KPI proxies and competitive advantage narratives without proprietary CoQ data (Shao, 2023)

2. Reference Architecture: Digital Thread-Enabled Predictive Quality Data Ecosystem

Layer A: Data Acquisition, Interoperability and Semantic Binding. Aerospace manufacturing data are multi-modal and multi-timescale: commanded parameters, machine feedback, in-situ monitoring and ex-situ inspection. Without registration and semantic binding, analytics become brittle and non-auditable. (Hedberg Jr., 2016)

The Predictive Quality Data Ecosystem needs:

- Stable identifiers: build, part, layer, XY location, operation IDs.
- Cross-system mapping: intent (PLM/MES), execution, monitoring, verification.
- Registration/alignment: mapping signals to a common coordinate/time basis, enabling pixel/voxel-level linking where applicable.

A characteristic practical aspect of the chosen pilot dataset is exactly this alignment: the final registered dataset consists solely of numerical values aligned with the machine coordinate system and includes process parameters, laser power, in-situ melt pool characteristics, layerwise optical intensity and ex-situ XCT voxel values. (Zhuo Yang, 2025)

Layer B – Governed analytics. Operational failure of predictive quality happens when training/serving feature definitions drift but also when missingness/outliers go untended. DataOps controls therefore should cover:

- Schema versioning (feature names, allowed missingness),
- Reproducible transformations (imputation/scaling),
- Sampling policies (if required for tractability),
- Documentation of label construction and thresholds.

In the aerospace world, interpretability and calibration aren't "nice-to-haves". Outputs should be defensible, decisionable and reconstructable.

Layer C – MLOps governance: validation, drift monitoring, controlled change. Manufacturing conditions change (tool wear, process updates, powder batches, environmental drift). PQDE treats every model like a controlled artifact with: (Breck, 2017)

- Training dataset identifiers and feature schema version,
- Validation protocol and acceptance gates,
- Post-deployment monitoring for data drift and performance drift,
- Retraining triggers and controlled release steps.

This logic lines up well with the risk-based thinking that has been highlighted in the aerospace QMS space; firms should recognize and address risk proactively rather than purely reactively. (Group, 2023)

Layer D – Decision integration and evidence packaging. PQDE adds to business value only when its predictions are factored into decision making. A scalable solution is risk-guided triage: allocate limited inspection capacity to the highest-risk regions and document the decision basis. (Davis, 2006)

Evidence packaging should minimally capture:

- Data slice identifiers (build, part, layers),
- Model version and feature schema version,
- Thresholds/policies,
- Key performance indicators (ROC-AUC, PR-AUC, calibration),
- Action logs and outcomes.

3. Architectures-to-Experiment Mapping

To address the potential disconnect between conceptual architecture and empirical contribution, this study implements a minimal but auditable instantiation of PQDE. Layer A (data acquisition / registration) is treated as an externalized prerequisite satisfied by NIST’s published registration pipeline for AMMT Overhang Part X4. The present work operationalizes Layers B-D through explicit governance artifacts: (i) a feature schema and preprocessing contract (Layer B), (ii) acceptance gates and drift / calibration monitoring reports (Layer C), and (iii) a risk-guided inspection triage policy bound to an audit-ready evidence package for each run (Layer D). All layers are described in Table 1.

Table 1. – PQDE layer-to-experiment traceability

PQDE Layer	What it claims (role)	Implemented in this study (minimal)	Artifact produced	Where reported
A. Data acquisition and linkage	Traceable data slices and registration	Use of NIST AMMT registered datasets (machine-coordinate alignment + IDs)	Data-slice spec (part/layer / window)	Section 4 (Public Pilot and Data) + Table 3
B. Governed analytics	Data contract + reproducible features / prep	Features manifest, schema version, preprocessing contract	Subsection Preprocessing contract (Methods) Label specification and threshold value (Section Methods) Split sizes - Table	Section 5 (Methods)
C. MLOps	Monitoring + acceptance gates	Acceptance gates + drift checks + calibration monitoring	Gate report + drift report + calibration summary	Section 6 Results
D. Decision integration	Risk-guided action and auditability	Top-k triage + evidence package for each run	Precision at k, Recall at k, Lift at k, IR=1-k (Table 4-6) Cost/impact translation using Top-10% (Table 4-6)	Section 6 (Results) + Table 4-6

Source: Authors’ own research.

4. Public Pilot and Data: AMMT Overhang Part X4 as a Replicable Validation Case

In order to prevent dependence on proprietary pilots while maintaining research robustness, this work employs a public metrology-grade pilot: the NIST Additive Manufacturing Metrology Testbed “Overhang Part X4”. The AMMT’s “Overhang Part X4” process monitoring dataset details a metal AM build that occurred at NIST, including commanded input files, process monitoring signals and metadata. (Brandon Lane, 2020) Simultaneously providing the relevant post-build X-ray CT dataset for the parts at NIST allowing ex-situ inspection evidence. (Maxwell Praniewicz, 2020)

NIST also presents a fully registered in-situ and ex-situ dataset for Overhang Part X4, documenting data processing, extraction of features, registration and treatment of uncertainties, generating numerical values that are aligned to the machine coordinate system (i.e., an explicit instantiation of digital thread alignment). (Zhuo Yang, 2025) This fully registered dataset is the main enabler of predictive quality research as it enables a traceable connection of process signatures with verification outputs without any effort of proprietary integration. (Tapia, 2014)

The pilot is explicitly intended to enable training, validation, and testing with nominally identical parts built in the same build, allowing for part-to-part comparison and predictive modelling. It also covers: commanded scan strategy, laser parameters; melt pool monitoring; layerwise imaging; and feedback monitoring. (Brandon Lane, 2020) The process monitoring dataset presents four nominally identical parts, each with 250 layers at 20 μm layer thickness, produced in a documented build at NIST AMMT (2019). A scan plan consists mostly of X/Y oriented scans, 90° rotations between layers along with part layout & scan sequence. NIST also publishes ex-situ XCT data for AMMT parts, providing verification of internal features and defect-related signals. (Maxwell Praniewicz, 2020)

An important step towards functioning digital thread consists of converting heterogeneous raw artefacts into a registered and analysis-ready representation. In its published 2025 version, NIST documents the pipeline from processing and feature extraction and multi-modal registration to alignment and uncertainty considerations to generate a numerical dataset to machine coordinates (Zhang, 2024)

This serves as an excellent proxy for aerospace digital ecosystems since it looks very similar to what organizations try to do internally: bind intent, monitoring and verification into a traceable data product and therefore support governed analytics.

Gap and positioning. Previous work in additive manufacturing predictive quality is promising evidence that defect prediction using in-situ monitoring and ex-situ XCT characterization can lead to improved results, although most works focus on offline performance and offer limited guidance for governed deployment (traceable data contracts, controlled model change, drift monitoring, calibration reporting and audit artifacts). Concurrently, traditional MLOps frameworks outline best practices for production ML that are rarely realized in high-assurance manufacturing environments where evidence lineage and reproducibility are a must. The PQDE closes this gap by aligning Digital Thread continuity with governed analytics and decision integration, making it possible to transform predictive outputs into repeatable auditable inspection triage policies and tangible business-excellence outputs. (Everton, 2016)

Methods: prediction task, splits and metrics tied to decisions

Label construction (rare-event quality proxy). To emulate rare defect occurrence without proprietary NCR data, we construct a binary label from the XCT-derived target. Using the training split, we compute the **1st percentile (q0.01)** of the selected XCT variable and label an observation as positive ($y=1$) if its XCT value falls below this threshold ($xct < q0.01$). The same fixed threshold is then applied to validation and test splits to preserve temporal and cross-part integrity.

Predictive performance

Table 2. Predictive performance by split – Baseline logistic model

Nr. crt.	Split	ROC-AUC	PR-AUC	Brier (score)
1.	Validation (Parts 1-3, L201-250)	0.974	0.205	0.050
2.	Test (Part 4, L1-200)	0.953	0.247	0.081
3.	Test (Part 4, L201-250)	0.982	0.448	0.053

Source: Authors' own research.

As is described in Table 2 the results of predictive performance baseline logistic model is :

- Discrimination remains strong across all splits (ROC-AUC ~ 0.95-0.98).
- Rare-event detection quality substantially enhanced on late-layer test slice (PR-AUC 0.448), which is consistent with the higher base rate of defect-labelled points in that slice and stronger separability.
- Brier scores indicate probabilistic results with roughly the same calibration properties on validation and late-layer tests, and degradation on the early-layer test (distribution shift or weaker signal-to-label mapping very early in the build).

Audit-ready evidence artifact is that each experimental execution yields a package of evidence that brings together data slice identifiers, schema / version hashes, preprocessing contract, model configuration, decision policy, performance metrics (with uncertainty), and monitoring outputs (calibration and drift). This artifact provides operationalization to PQDE governance, allowing for repeatable auditable decisions under controlled change.

Distribution Shift and Drift Tracking (Governance Artifact)

Table 3. Drift summary (Train vs. evaluation splits)

Split	% PSI>0.10	% PSI>0.20	Median PSI	Max PSI
Validation	0.051282	0	0.049126	0.162048
Test Early	0.794872	0.692308	0.955674	12.451257
Test Late	0.666667	0.615385	1.389674	12.451257

Source: Authors' own research.

Shift of distribution and monitoring drift. To make PQDE Layer C (governed analytics and monitoring) practical, we measure the distribution shift in the training slice to each evaluation slice by per-feature Population Stability Index (PSI) and KS-like statistic. Table 3 presents a summary of drift magnitude, showing the level of drift severity as the fraction of features that are beyond common PSI heuristic (0.10 and 0.20), median PSI, and maximum PSI. In total, drift was significant, as evidenced by largest shifts in Test Late, such that performance differences across slices may possibly signify changes in both prevalence and the shift in feature distribution. The complete per-feature drift report is included as supplemental material and is attached to the run-level evidence package to facilitate auditable monitoring and controlled model change.

Operational triage: reduction of inspection vs defect capture

The results of a triage indicate how much inspection can be decreased without losing defect capture.

Table 4. Risk-guided triage results - Validation

Policy (Top k)	Prevalence	Precision at k	Recall at k	Lift at k	Inspection reduction
1%	1.294%	30.83%	23.82%	23.82	99%
5%	1.294%	20.68%	79.91%	15.98	95%
10%	1.294%	12.71%	98.20%	9.82	90%

Source: Authors' own research.

Table 5. Risk-guided triage results – Test Part 4 early layers

Policy (Top-k)	Prevalence	Precision at k	Recall at k	Lift at k	Inspection reduction
1%	1.875%	34.86%	18.59%	18.59	99%
5%	1.875%	21.18%	56.48%	11.30	95%
10%	1.875%	15.58%	83.07%	8.31	90%

Source: Authors' own research.

Table 6. Risk-guided triage results – Test Part 4 late layers

Policy (Top-k)	Prevalence	Precision at k	Recall at k	Lift at k	Inspection reduction
1%	2.640%	50.00%	18.94%	18.94	99%
5%	2.640%	38.25%	72.44%	14.49	95%
10%	2.640%	25.58%	96.88%	9.69	90%

Source: Authors' own research.

The results regarding risk-guided triage for validation and Part 4 are presented in Tables 4, 5, and 6 and are as follows:

- With Top-10% inspection, defect capture reaches 98.20% on validation and 96,88% on late-layer test while reducing inspection scope by 90%.
- Even with the more difficult slice on early-layer cross-parts, Top-10% inspection is still able to get ~83% of defect-labeled points, indicating strong value compared to random inspection (expected recall in 10% inspection is only 10%).
- Lift indicates that inspected points are $8.31 \times$ to $23.82 \times$ greater in enriched defect data compared to baseline prevalence (depending on split and policy) and is the operating mechanism that converts risk scores into throughput and CoQ enhancement.

Performance and triage performances vary between early and late layers on the test part, possibly indicating that predictive signal strength and/or defect onset change over the build stages. In PQDE words, this is not an error; it's precisely why drift/shift monitoring must be woven into the ecosystem, and why policy thresholds can be layer-slice aware.

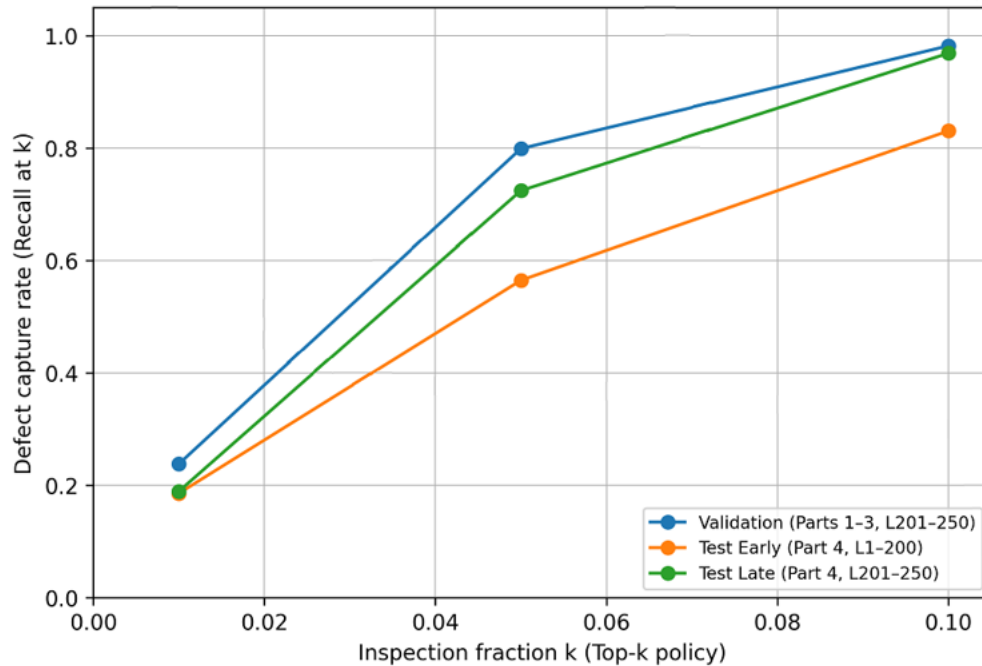


Figure 1. Inspection triage trade-off (defect capture vs inspection budget)

Source: Authors' own research.

The figure 1 visualizes Recall at k for Top-k inspection policies ($k = 1\%$, 5% , 10%) across validation (Parts 1-3, L201-250) and test slices (Part 4 early and late layers), derived from the triage results in Tables 4-6.

All experimental results reported in Table 4-6 were produced using the publicly available NIST AMMT Overhang Part X4 datasets and the corresponding XCT/registered datasets, scripts and parameter settings are available upon request.

Reproducibility: Data Preprocessing, Splits and Computational Environment

Data preprocessing. The registered AMMT dataset was contract. First, identifier fields and non-feature metadata (e.g. file path, part/layer identifiers – described in Table 7) were excluded from the feature matrix. Second, the XCT- derived target column used only for labeling was removed from the feature set to prevent leakage. Third, numeric features were retained and missing values were handled using median imputation computed on the training split only. For linear models, features were standardized (z-score normalization) using training-split statistics and then applied unchanged to validation and test splits. This preprocessing pipeline was implemented as a versioned, executable configuration to ensure training / serving consistency and auditability.

Table 7. Dataset splits and class prevalence

Split	Parts	Layers	N _{total}	N _{pos}	Prevalence
Train	1-3	1-200	269010	23340	0.008676
Validation	1-3	201-250	871164	8266	0.009488
Test Early	4	1-200	896700	14023	0.015638
Test Late	4	201-250	290388	6672	0.022976

Source: Authors' own research.

Computational environment and runtime. All experiments were executed in Python 3.14 using numpy 2.4.0, pandas 2.3.3, scikit-learns 1.8.0 and matplotlib 3.10.8. The computations were performed on a workstation with AMD Ryzen 9 7945HX with Radeon Graphics (2.50 GHz), 32 GB RAM, running Windows 11. End-to-end runtime for one complete training and evaluation pass (including metric computation, triage tables and calibration plots) was approximately 45 minutes.

The full set of artifacts required for reproduction (feature manifest, schema hash, drift report, calibration plots, triage tables and evidence package example) is generated by the provided runner script.

Feature Set and Preprocessing Contract (Reproducibility)

Feature set – The model input is defined by an explicit feature manifest generated from the NIST AMMT registered files. The manifest enumerates all numeric features used for training, their data types and missingness rates. Identifier fields (file path, part/layer keys) and the XCT-DERIVED target column used for proxy labeling are excluded from the feature matrix to prevent leakage. The full feature inventory is provided in Appendix A (Feature Manifest)

Preprocessing contract. All models use the same preprocessing policy to ensure compatibility. Numeric features are median imputed using statistics computed on the training split only and applies unchanged to validation and test splits. Linear models additionally apply z-score standardization using training statistics only. This contract is treated as part of the governed analytics layer to preserve training / serving consistency.

Model Specification and Baselines (Reproducibility)

Governance-friendly baseline (logistic regression) – Logistic regression was chosen as the baseline based on transparent coefficients and stable optimization as well as its support for audit and calibration reporting. The model was trained using solver=lbfgs, C=1.0, l1_ratio=0.0 (L2 behavior), class_weight=balance, max_iter=2000, and random_state=42. Training followed the fixed split protocol described in Section Methods—Subsection “Reproducibility: Data Preprocessing, Splits and Computational Environment”; no samples from validation/test were used for preprocessing statistics or threshold construction.

This decision also conforms with the governance goals of the papers; at the coefficient level interpretability enhances auditability and calibration and drift monitoring are also able to be leveraged alongside a robust parametric model.

The formation of the proxy label and its validity. The binary outcome utilized in this study is a proxy label calculated by XCT intensity and does not demonstrate validated manufacturing nonconformance in the real world (e.g., NCR/CAR). More specifically, positives are defined as values below q , which is a training-derived quantile threshold. The proxy allows reproducible experiments for public datasets, though in order to become a governance mechanism, we must rely on validated defect labelling, rather than certification as a defect classifier.

Uncertainty quantification. To provide some insights into the precision of performance estimates under extreme class imbalance, we report bootstrap 95% confidence intervals for ROC-AUC, PR-AUC, Brier score and ECE (Table 13) for each assessment split. Bootstrapping resamples data with replacement of the observations within each split, while keeping the trained model fixed. Bootstrapping is carried out independently within each split to retain the distributional differences between slices.

Results

From QPI to KPI to Competitive Advantage.

A problem still very much with publishable industrial research has been the unavailability of CoQ values that can be shared. We tackle it with KPI proxies organizations can instantiate by their own cost.

Thus were defined parameters:

- k – inspection fraction (e.g., 0.10 for Top - 10%),
- IR – $1 - k$ (inspection reduction),
- DC at k – Recall (defect capture rate) at k ,
- FP at k – (false alarms within the inspected set, defined as $N_{inspected} \times (1 - Precision \text{ at } k)$, where $N_{inspected} = k \times N_{total}$
- $Lift$ at k – Precision at k divided by Prevalence,

A scenario – based expected avoided CoQ proxy is:

$$EACoQ(k) = (DC \text{ (Defect capture) at } k) \cdot N_{def} \cdot C_{FN} - (FP \text{ at } k) \cdot C_{FP}$$

where C_{FN} is the cost of a missed defect (rework, scrap, escape consequence) and C_{FP} is the cost of unnecessary inspection or containment, N_{def} is the number of defect-labeled points in the evaluated slice, and FP (False alarms) at k is defined above. This proxy supports sensitivity analysis and ties directly to triage outputs.

Table 8 – KPI proxy template

KPI Proxy	What it tells you (in plain words)	How to compute from pilot outputs	Competitive advantage mechanism (why it matters)
Inspection Reduction (IR)	How much inspection work you avoid compared with “inspect everything”	$IR = 1 - k$ (k =inspected fraction). Example: Top-10% $\rightarrow k=0.10 \rightarrow IR=0.90$	Less appraisal effort and bottlenecks, frees capacity for high-value tasks, supports higher throughput
Defect Capture (DC)	Out of all defect-labeled points, how many you still catch inside the inspected subset.	DC = Recall at k. Example (Test-late, $k=0.10$): $DC=0.9688$.	Fewer escapes and fewer late surprises; lower rework/scrap risk; more stable delivery and quality performance.
Enrichment (Lift)	How “focused” your inspection is—are defects concentrated in what you inspect, versus random selection?	$Lift \text{ at } k = \frac{Precision \text{ at } k}{Prevalence}$	Faster containment and root-cause work; reduces wasted inspection on low-risk areas; makes inspection smarter, not bigger.
False Alarm (Load)	How much unnecessary follow-up you create by inspecting items that turn out OK.	FA rate at $k = 1 - Precision \text{ at } k$. (Optional count: FA count at $k = (1 - Precision \text{ at } k) \times (k \times N_{total})$)	Controls operational burden (extra checks, holds, paperwork); helps tune k /thresholds so triage stays practical.
Decision Latency Proxy	How much faster decisions could be (release/hold/contain) when you don’t inspect everything	Scenario-based: Time saved $\approx IR \times \text{baseline inspection time}$ (or queue time proxy)	Shorter cycle time and faster release; quicker feedback loops; better responsiveness under capacity constraints.
Audit Evidence Completeness	Can you later explain and reproduce the decision (data slice, model version, thresholds, logs)?	% complete = (decisions with full evidence package / total decisions) $\times 100$	Faster audits, fewer “trust gaps” with customers/regulators; scalable compliance across sites and ecosystem partners

Source: Authors’ own research.

As you can see in Table 8 PQDE translates predictive scores into competitive advantage through risk-informed triage that illustrates a tunable trade-off: 90% inspection reduction can capture **96.88 – 98.20%** of defect-labeled points (Top-10%) while more aggressive policies (Top-1% / Top-5%) increase inspection reduction to 99% / 95% with proportionally lower capture. This shifts quality cost from broad inspection to targeted containment, reducing CoQ and improving lead time. Because the ecosystem logs model versioning, data lineage, thresholds and actions, it also produces audit-ready evidence that supports scalable operations across digital ecosystems and supply chains.

Baseline Model Comparison (Same Splits)

Table 9. Logistic regression coefficient analysis (interoperability)

Feature	Coefficient	Coefficient	Sign (Positive/Negative
C38	-21.024	21.024	↓ risk
C39	8.722	8.722	↑ risk
c0	-0.798	0.798	↓ risk
c20	0.64	0.64	↑ risk
c17	0.523	0.523	↑ risk
c19	-0.507	0.507	↓ risk
c33	0.498	0.498	↑ risk
c1	0.471	0.471	↑ risk
c36	0.451	0.451	↑ risk
c23	-0.438	0.438	↓ risk
c26	-0.42	0.42	↑ risk
c24	0.366	0.366	↑ risk
c3	0.364	0.364	↑ risk
c15	-0.358	0.358	↓ risk
c35	-0.347	0.347	↑ risk
c11	-0.344	0.344	↓ risk
c7	0.339	0.339	↑ risk
c21	-0.304	0.304	↓ risk
c10	0.295	0.295	↑ risk
c18	-0.269	0.269	↓ risk
c14	0.268	0.268	↑ risk
c27	0.265	0.265	↑ risk
c8	-0.256	0.256	↓ risk
c30	-0.254	0.254	↑ risk
c12	-0.212	0.212	↓ risk
c4	0.196	0.196	↑ risk
c6	0.166	0.166	↑ risk
c25	0.153	0.153	↑ risk
c13	0.112	0.112	↑ risk
c2	0.106	0.106	↑ risk

Source: Authors' own research.

Coefficient-based interpretability. To aid governance and auditability, we report the logistic regression coefficient (Table 9). Since features are standardized within the linear preprocessing contract, coefficient represents the change in log-odds due to a one-standard deviation increment

in a feature. Positive coefficients indicate increased predicted defect risk, while negative coefficients indicate reduced risk. This analysis complements decision-centric triage metrics with clear, native model explanations.

Table 10 – Baseline model comparison under identical splits

Model	Split	N total	N_pos	ROC-AUC	PR-AUC	Brier	ECE
Logistic regression (baseline)	Validation	871164	8266	1	0.971	0.005	0.006
Random forest	Validation	871164	8266	0.999	0.984	0.001	0.001
HistGradientBoosting	Validation	871164	8266	1	0.969	0.001	0.001
Dummy (stratified)	Validation	871164	8266	0.5	0.009	0.018	0.018
Logistic regression (baseline)	Test (early layer)	896700	14023	1	1	0.001	0.002
Random forest	Test (early layer)	896700	14023	1	0.995	0.002	0.004
HistGradientBoosting	Test (early layer)	896700	14023	1	0.99	0.001	0.001
Dummy (stratified)	Test (early layer)	896700	14023	0.5	0.016	0.024	0.024
Logistic regression (baseline)	Test (late layer)	290388	6672	1	0.999	0.002	0.004
Random forest	Test (late layer)	290388	6672	1	0.992	0.002	0.004
HistGradientBoosting	Test (late layer)	290388	6672	1	0.986	0.001	0.001
Dummy (stratified)	Test (late layer)	290388	6672	0.501	0.023	0.001	0.031

Source: Authors' own research.

Baseline comparison. To determine if findings are dependent upon the model, we tried logistic regression, random forest, histogram-based gradient boosting and stratified dummy baseline using the same data splits and preprocessing. Since the dummy baseline approximates the chance performance under extreme imbalance, improvements over dummy also suggest that the learned models are showing non-trivial signal beyond prevalence.

Calibration and Reliability of Risk Scores (Governance Metric)

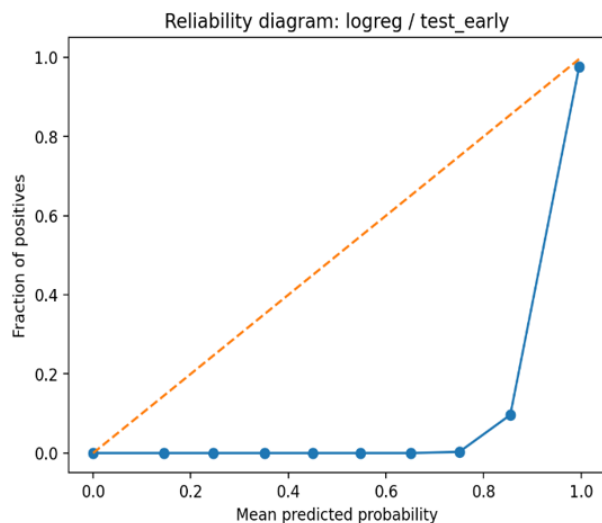


Figure 2. Calibration for logistic for test_early

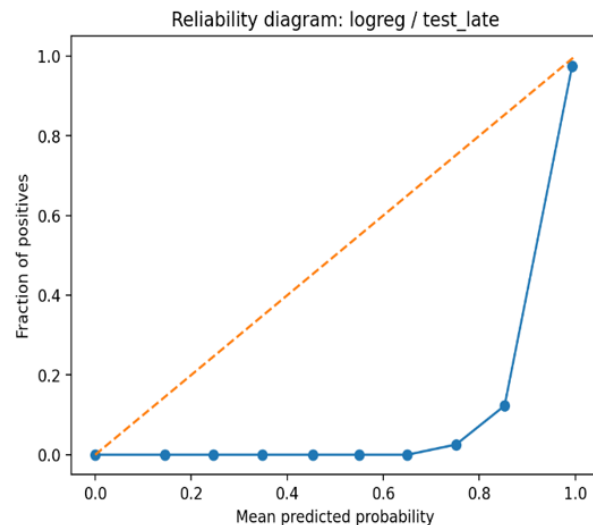


Figure 3. Calibration for logistic for test_late

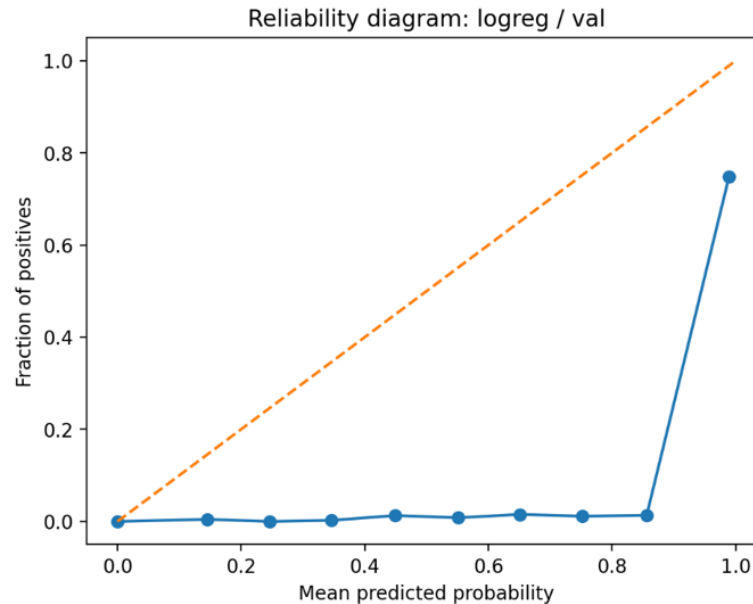


Figure 4. Calibration for logistic validation Source: Authors' own research.

Figure 2-4. Reliability diagrams (calibration) for the logistic baseline. Each plot reports the observed positive fraction versus the mean predicted probability across probability for (C1) validation, (C2) test-early and (C3) test-late slices. The diagonal line indicates perfect calibration, deviations indicate over or under -confidence.

Table 11. Calibration summary (ECE)

Split	ECE (logreg)	ECE (calibration)	Brier (logreg)
Validation	0.006	0.005	0.006
Test Early	0.002	0.001	0.002
Test Late	0.004	0.002	0.004

Source: Authors' own research.

Calibration metrics. Table 11. Expected Calibration Error (ECE) and Brier score for the logistic baseline across evaluation slices.

Because PQDE aims to assist with governed decision-making, we evaluate not only discrimination but also the reliability of predicted probabilities. Reliability diagrams for the logistic baseline are reported in Figure 2-4, and ECE and Brier score are shown across evaluation slices in Table 11. While Top-k triage predominantly depends on ranking quality, calibration is essential when probabilities are used as risk thresholds (e.g., hold / release decisions), when comparing risk across builds, and when packaging audit-ready evidence for regulators and customers. Thus, calibration outputs are incorporated as governance artifacts in the PQDE evidence package.

For ranking-based Top-k triage, miscalibration may have limited impact on Recall at k as long as the ranking is preserved, though this may mislead stakeholders when risk is interpreted in absolute terms. Therefore, PQDE treats calibration as a monitored governance metric and triggers recalibration once ECE degrades.

As a governance action, under PQDE Layer C, increases in ECE beyond a predefined tolerance (e.g., $\Delta > 0.02$) trigger recalibration (Platt scaling or isotonic regression) and re-issuance of the evidence package, while preserving traceability of model and schema versions.

Decision-Centric Comparison (Top-k Triage Across Models)

Table 12. Top-k triage performance by model (decision-centric metrics)

Model	Split	k	IR(=1-k)	Recall at k	Precision at k	Lift at k
Logistic regression (baseline)	Validation	0.01	0.99	0.938	0.89	93.778
Logistic regression (baseline)	Validation	0.05	0.95	1	0.19	20
Logistic regression (baseline)	Validation	0.1	0.9	1	0.095	10
Random forest	Validation	0.01	0.99	0.957	0.908	95.653
Random forest	Validation	0.05	0.95	0.999	0.19	19.985
Random forest	Validation	0.1	0.9	0.999	0.095	9.994
HistGradientBoosting	Validation	0.01	0.99	0.921	0.874	92.108
HistGradientBoosting	Validation	0.05	0.95	1	0.19	20
HistGradientBoosting	Validation	0.1	0.9	1	0.095	10
Dummy (stratified)	Validation	0.01	0.99	0.009	0.009	0.944
Dummy (stratified)	Validation	0.05	0.95	0.036	0.007	0.726
Dummy (stratified)	Validation	0.1	0.9	0.068	0.006	0.685
Logistic regression (baseline)	Test (early layers)	0.01	0.99	0.639	1	63.945
Logistic regression (baseline)	Test (early layers)	0.05	0.95	1	0.313	20
Logistic regression (baseline)	Test (early layers)	0.1	0.9	1	0.156	10
Random forest	Test (early layers)	0.01	0.99	0.637	0.996	63.695
Random forest	Test (early layers)	0.05	0.95	0.999	0.313	19.986
Random forest	Test (early layers)	0.1	0.9	1	0.156	9.995
HistGradientBoosting	Test (early layers)	0.01	0.99	0.636	0.995	63.638
HistGradientBoosting	Test (early layers)	0.05	0.95	1	0.313	20
HistGradientBoosting	Test (early layers)	0.1	0.9	1	0.156	10
Dummy (stratified)	Test (early layers)	0.01	0.99	0.009	0.014	0.877
Dummy (stratified)	Test (early layers)	0.05	0.95	0.009	0.003	0.175
Dummy (stratified)	Test (early layers)	0.1	0.9	0.009	0.001	0.088
Logistic regression (baseline)	Test (late layers)	0.01	0.99	0.435	1	43.523
Logistic regression (baseline)	Test (late layers)	0.05	0.95	1	0.46	19.999
Logistic regression (baseline)	Test (late layers)	0.1	0.9	1	0.23	10
Random forest	Test (late layers)	0.01	0.99	0.434	0.996	43.359
Random forest	Test (late layers)	0.05	0.95	1	0.459	19.993
Random forest	Test (late layers)	0.1	0.9	1	0.23	9.998

Model	Split	k	IR(=1-k)	Recall at k	Precision at k	Lift at k
HistGradientBoosting	Test (late layers)	0.01	0.99	0.431	0.99	43.074
HistGradientBoosting	Test (late layers)	0.05	0.95	1	0.46	19.999
HistGradientBoosting	Test (late layers)	0.1	0.9	1	0.23	10
Dummy (stratified)	Test (late layers)	0.01	0.99	0.01	0.024	1.034
Dummy (stratified)	Test (late layers)	0.05	0.95	0.031	0.014	0.623
Dummy (stratified)	Test (late layers)	0.1	0.9	0.064	0.015	0.643

Source: Authors' own research.

Decision-centric evaluation. Since PQDE operationalizes model outputs as an inspection triage policy, we report Top-k decision metrics. Across models, Top-k triage concentrates positives in the inspected subset, while the magnitude of lift and defect capture depends on ranking quality. Reporting these decision-centric KPIs directly supports the paper's claims about inspection reduction and defect capture (see Table 12).

Naïve baseline. We include a stratified random classifier that preserves the observed prevalence as a minimal baseline for comparison.

Under Top-k selection, a random ranking yields expected Recall at k and Lift at k ≈ 1 .

Sensitivity to label threshold

Table 13. Sensitivity to proxy-label threshold q

Quantile threshold (q)	Split	Prevalence	ROC-AUC	PR-AUC	Recall at 10%	Lift at 10%	Brier	ECE
0.005	Validation	0.002	1	0.973	1	10	0.001	0.002
0.005	Test (early layers)	0.007	1	0.996	1	10	0.001	0.002
0.005	Test (late layers)	0.004	1	0.992	1	10	0.001	0.002
0.01	Validation	0.009	1	0.971	1	10	0.005	0.006
0.01	Test (early layers)	0.016	1	1	1	10	0.001	0.002
0.01	Test (late layers)	0.023	1	0.999	1	10	0.002	0.004
0.02	Validation	0.027	1	0.994	1	10	0.004	0.005
0.02	Test (early layers)	0.032	1	0.999	1	10	0.002	0.003
0.02	Test (late layers)	0.047	1	1	1	10	0.001	0.002
0.05	Validation	0.054	1	0.998	1	10	0.004	0.005
0.05	Test (early layers)	0.062	1	1	1	10	0.001	0.001
0.05	Test (late layers)	0.072	1	1	1	10	0.001	0.001

Source: Authors' own research.

Interpreting PR-AUC across splits. However, the PR-AUC is sensitive to prevalence and thus differences in validation and test slices can stem from variation of population base and shift of distribution, not merely from model quality. To frame PR-AUC differences, and to enable guided monitoring in PQDE we report prevalence per split (Table 7) and drift summary (Table 3). In such an extreme imbalance, ROC-AUC may seem optimistic and may not represent operational utility, therefore decision-centric Top-k triage metrics and PR-AUC have to be highlighted.

Discussion

Implications for Business Excellence from the Perspective of Managerial and Ecosystem Theory.

Grounding in measured outputs. The managerial implications are driven by the operational behavior of risk-guided triage. Across all evaluation slices, the Top-k inspection reduces the scope of inspection IR-1-k (e.g., 90% at $k=0.10$) while maintaining a high proportion of defect-proxy capture ($DC=Recall$ at k) and producing enrichment ($Lift$ at $k > 1$). There are three levers leveraged by such decision-centric properties: inspection capacity release, accelerated containment of high-risk regions and scalable auditability through evidence packaging and monitoring outputs.

From a business excellence perspective, the differentiator isn't the predictive model without organizational support to test, implement, optimize and make the most of AI repeatedly and safely across products, sites and partners. This capability is not only possible, but must be achieved across multiple functions (manufacturing, quality, IT and engineering) in order to provide data that is consistent, features that are reproducible and outputs that are integrated into operating decision pathways (i.e., inspection, containment and release). This also entails controlled model change (for example, established validation criteria, release gates, monitoring triggers and retraining triggers) and evidence packaging, allowing decisions to be reconstructed for internal review, customers or regulators. Digital Thread literature recognizes that integrated data linkage and authoritative information sharing across lifecycle stages is fundamental to transforming analytics to enterprise value, providing the continuity and traceability necessary for decision-making at scale.

PQDE is meant to meet standards-based governance requirements in high-assurance manufacturing. For one, the governance layer around the Digital Twin is aligned with the ISO 23247 framing of digital twins as systems with an established exchange of knowledge and a defined lifecycle; PQDE is operationalizing it through formalized data interfaces, traceable transformations, and controlled decision integration. Second, aerospace QMS practice – exemplified by the IAQG 9100 series – focuses on improved quality, schedule and cost performance across the supply chain – PQDE aligns with this intent by the translation of predictive outputs into tangible operational outcomes (inspection reductions and defect capture) while maintaining evidence lineage.

Within this standard-aligned governance context, PQDE encourages competitive advantage through four mutually reinforcing mechanisms. Cost leadership can be achieved by triage-based inspection efficiency (a large reduction in inspection workload without compromising on defect detection), where appraisal costs decrease and bottlenecks are relieved. Reliability and delivery performance improve because higher-risk regions are prioritized for verification and containment, reducing the probability of downstream escapes and accelerating disposition. Scalable compliance builds with audit-ready evidence (data slice identifiers, model versioning, thresholds and action logs) – reducing the friction by extending predictive quality to suppliers and multi-site operations. Lastly, learning speed increases: this is due to governed monitoring and retraining cycles, enabling the organization to leverage drift without affecting production, and converting continuous improvement into a structured, repeatable process in the digital-ecosystem, rather than an ad hoc analytics exercise.

Differences in PR-AUC between evaluation slices can be interpreted in context of prevalence and drift summaries. In PQDE governance, drift and calibration function like monitoring signals with the potential to trigger recalibration or retraining under controlled change; they are not ad hoc updates.

Worked Business Case (Illustrative) (Top 10% Triage).

Scenario 1 (inspection capacity and cost). N_{pos} would be the defect-proxy positives, and c_I would be the cost per unit inspected. The baseline inspection cost is N_{c_I} . If Top-10% triage is applied ($k=0.10$), inspected units will result in $0.10N_{(c_I)}$, resulting in an inspection cost reduction of $0.90*N_{c_I}$ (IR=90%). In the case where inspection is a bottleneck, the same decrease translates to around 10x less inspection workload, allowing for faster release or redirection of inspection capacity.

Scenario 2 (proxy for missed defect risk). Let N_{pos} be the number of defect-proxy positives and c_M the cost of missed defect. Under Top-k triage, the expected missed positives are $1-DC$ at k)* N_{pos} . The anticipated missed-defect cost is $(1-DC \text{ at } k)* N_{pos} * c_M$, thus offering a transparent mechanism for tuning k based on risk appetite: higher k means expected misses are reduced at the expense of higher inspection load.

Scenario 3 (inspection lead time). For r units/day inspection throughput, baseline lead time is N/r . At Top-10% triage, lead time is $0.10 * N/r$, indicating an almost 10x reduction in queue time for inspection under region-level divisible inspection.

The results are described in Table 14.

Table 14. Worked business case parameters and outputs (Top-10% triage)

Item	Meaning	Value (example)	Source
Split	Evaluation slice used for the worked example	Validation	Worked_business_case_example.csv
N_{total}	Inspection units	871,164	Worked_business_case_example.csv
N_{pos}	Defect-proxy positives	8,266	Worked_business_case_example.csv
k	Triage fraction	0.10	Fixed
$N_{inspected}$	Units routed to inspection ($k*N_{total}$)	87,117	Computed
IR	$IR=(1-k)$	0.90	Table 11
DC at k	Recall at k	1	Table 11 (logreg, chosen split)
Lift at k	Enrichment	10	Table 11
c_I	Cost per inspection unit	User-set	Assumption
c_M	Cost per missed defect	User-set	Assumption
Inspection savings	$0.90*N*c_I$	784,047 (Hedberg Jr., 2016) (Zhang, 2024)	Computed
Expected missed cost	$(1-DC)*N*c_M$	0	Computed

Source: Authors' own research.

Conclusions

In this paper, we show aerospace manufacturers how through building a Digital Thread-enabled Predictive Quality Data Ecosystem (PQDE), they can turn the ideas from Digital Twin and Digital Thread into a measurable path to competitive advantage. Instead of conceptualizing predictive quality as a discrete machine learning model, PQDE is positioned as an ecosystem capability that bridges:

- multi-modal data continuity and traceability from manufacturing intent to inspection evidence by binding IDs and registration;

- governed analytics over DataOps/MLOps controls (schema control, lineage, acceptance gates, monitoring and controlled change);

operational decision integration through risk-guided inspection triage with audit-ready evidence packaging.

To test the proposed approach, it was validated in the pilot published by publicly-announced NIST AMMT “Overhang Part X4” using XCT-based verification with a rare-event label defined by the 1st percentile threshold calculated on the training split and consistently applied to validation and test slices. In both cross-part and temporal slices, the baseline, governance-friendly models delivered strong discrimination (ROC-AUC 0.953-0.982; PR-AUC up to 0.448) and relatively stable probabilistic results (Brier score 0.050-0.081). More significantly, these predictive indicators were operationalized to a decision-relevant response by Top-k triage instructions. With **Top-10% inspection**, defect capture reached **98.20%** on validation and **96.88%** on the late-layer test slice while reducing inspection scope by **90%**; Top-10% triage got 83.07% on the harder early-layer test slice as well. Lift values from **8.31x to 23.82x** indicate large enrichment of defect-labeled points in the inspected subset which also implies that finite inspection resources can be concentrated where yield is greater and high enough.

These findings shed light on a competitive advantage mechanism: PQDE promotes **throughput and cost-of-quality enhancement via decreasing low-yield inspection**, while maintaining high defect capture to limit downstream escapes and disposition delays. Furthermore, PQDE reinforces scalable assurance through the provision of decision evidence packaging (data slice identifiers, model version, threshold and logs) and cuts down friction during rolling out predictive quality in different sites and to a greater extent to suppliers.

To facilitate industrial research that is publishable and confidentiality-secure, the paper provided **KPI proxy templates** (inspection reduction, defect capture, enrichment, false-alarm load, decision-latency proxy and evidence completeness) that translate triaging outputs into business excellence indicators, without proprietary cost accounting.

These considerations can be limited to using a single public pilot and a proxy label, rather than production NCR/CAR data. Future work should validate PQDE across multiple builds and machines, and include explicit drift and uncertainty controls and cost-sensitive triage policies that are embedded in actual QMS/MES processes. This paper proposed and evaluated a Digital Thread-enabled Predictive Quality Data Ecosystem

Notations and/or Abbreviations

AMMT – Additive Manufacturing Metrology Testbed, **CoQ** – Cost of Quality, **ERP** – Enterprise Resource Planning, **MES** – Manufacturing Execution System, **PR-AUC** – Precision Recall AUC, **PQDE** – Predictive Quality Data Ecosystem, **ROC-AUC** – Receiver Operating Characteristic AUC, **QPI** – Quality Performance Indicators

References

- Brandon Lane, H. Y. (2020, September 3). Process Monitoring Dataset from the Additive Manufacturing Metrology Testbed (AMMT): Overhang Part X4. (NIST, Ed.) *Journal of Research (NIST JRES)*, 125. <https://doi.org/10.6028/jres.125.027>
- Breck, E. C. (2017). The ML test score: A rubric for ML production readiness and technical debt reduction. *IEEE Big Data 2017*, pp. 1123-1132.
- Davis, J. &. (2006). The relationship between precision-recall and ROC curves. *ICML 2006*, pp. 233-240. <https://doi.org/10.1145/1143844.1143874>.

- Everton, S. K. (2016). Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing. *Materials & Design*, pp. 431-445. <https://doi.org/10.1016/j.matdes.2016.01.099>
- Group, I. A. (2023, September). Crucial Insights into AS9100D Risk Requirements for Aerospace and Defense. *Standard Issues*. Retrieved from IAQG: <https://iaqg.org/crucial-insights-into-as9100d-risk-requirements-for-aerospace-and-defense/>
- Guodong Shao, S. P. (2023, June 16). An Analysis of the New ISO 23247 Series of Standards on Digital Twin Framework for Manufacturing. *2023 MSEC Manufacturing Science & Engineering Conference*. New Brunswick, NJ, US. Retrieved 2025, from <https://www.nist.gov/publications/analysis-new-iso-23247-series-standards-digital-twin-framework-manufacturing>
- Hedberg Jr., T. D. (2016, FEBRUARIE). Testing the digital thread in support of model-based manufacturing and inspection. *Journal of Computing and Information Science in Engineering*. <https://doi.org/10.1115/1.4032697>
- International Aerospace Quality Group - IAQG. (2025). *9100 Quality Management Systems – Requirements for Aviation, Space and Defense Organizations*. Retrieved from <https://iaqg.org/standard/9100-qms-requirements-for-aviation-space-and-defense-organizations>
- Maxwell Praniewicz, B. L. (2020, October 1). X-ray computed tomography data of Additive Manufacturing Metrology Testbed (AMMT) parts: Overhang Part X4. *Journal of Research, NIST*, 125. <https://doi.org/10.6028/jres.125.031>
- Okaro, I. A. (2019). Automatic fault detection for laser powder-bed fusion using semi-supervised machine learning. *Additive Manufacturing*, pp. 42-53. <https://doi.org/10.1016/j.addma.2019.01.006>
- Qiang Zhang, Jihong Liu, Xu Chen. (2024, February 23). *A Literature Review of the Digital Thread: Definition, Key Technologies, and Applications*. <https://doi.org/10.3390/systems12030070>
- Renata SEPTICHITA, A. S. (2025). Analysis of State-of-the-Art Technologies in the Modern Aerospace Industry. *Annals of the Academy of Romanian Scientists Series on Engineering Sciences*, 17(1/2025). <https://doi.org/10.56082/annalsarscieng.2025.1.100>
- Shao, G. F. (2023, 06 16). An analysis of the new ISO 23247 series of standards on digital twin framework for manufacturing. *ASME MSEC 2023*. Retrieved from https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=935765
- SEPTICHITA Renata, SEMENESCU Augustin. (2025). *Quality Management Standards in the Contemporary Aerospace Sector: Emerging Trends in Quality Definition*. Universitatea Națională de Știință și Tehnologie POLITEHNICA BUCUREȘTI, București.
- Tapia, G. &. (2014). A review on process monitoring and control in metal-based additive manufacturing. *Journal of Manufacturing Science and Engineering*. <https://doi.org/10.1115/1.4028540>
- Zhang, Q. L. (2024, 12). A literature review of the digital thread: Definition, key technologies, and applications. *Systems*. <https://doi.org/10.3390/systems12030070>
- Zhuo Yang, Y. L. (2025, May 8). A Fully Registered In-Situ and Ex-Situ Dataset for Metal Powder Bed Fusion Additive Manufacturing: Data Processing, Feature Extraction, Registration, and Uncertainties. *Advanced Manufacturing Series (NIST AMS) - 100-69*. <https://doi.org/10.6028/NIST.AMS.100-69>

Appendix A

Feature manifest		
Feature	Dtype	Missing rate
c38	float64	0.10129998786947556
c39	float64	0.10129998786947556
c21	float64	0.004987835779655763
c20	float64	0.004987835779655763
c19	float64	0.004987835779655763
c31	float64	0.004987835779655763
c32	float64	0.004987835779655763
c34	float64	0.004987835779655763
c33	float64	0.004987835779655763
c36	float64	0.004987835779655763
c35	float64	0.004987835779655763
c24	float64	0.004987835779655763
c23	float64	0.004987835779655763
c29	float64	0.004987835779655763
c22	float64	0.004987835779655763
c26	float64	0.004987835779655763
c25	float64	0.004987835779655763
c28	float64	0.004987835779655763
c27	float64	0.004987835779655763
c30	float64	0.004987835779655763
c2	float64	0.0
c0	int64	0.0
c1	float64	0.0
c6	float64	0.0
c5	float64	0.0
c4	int64	0.0
c3	float64	0.0
c15	float64	0.0
c16	float64	0.0
c17	float64	0.0
c18	float64	0.0
c10	float64	0.0
c9	float64	0.0
c8	float64	0.0
c7	float64	0.0
c11	float64	0.0
c12	float64	0.0
c13	float64	0.0
c14	float64	0.0