ASRC Procedia
Global Perspectives in Science and Scholarship
Florida, USA

1St GRI Conference 2025

Volume: 1; *Issue:* 1
Pages: 705–736
Published: 29 April 2025

open access

1st Global Research and Innovation Conference 2025,

April 20–24, 2025, Florida, USA

AI-ENABLED PREDICTIVE ANALYTICS AND FAULT DETECTION FRAMEWORKS FOR INDUSTRIAL EQUIPMENT RELIABILITY AND RESILIENCE

M.A. Rony[1]

[1] *Master of Science in Computer Science, Washington University of Virginia, USA*
Email: mdmahababulalamrony@gmail.com

Doi: 10.63125/2dw11645
Peer-review under responsibility of the organizing committee of GRIC, 2025

Abstract

This study investigates how AI-enabled predictive analytics and fault detection and diagnosis frameworks relate to industrial equipment reliability and resilience. Using a quantitative cross-sectional case-study design, we synchronize condition-monitoring signals, CMMS event histories, and operations context into per-asset analytical snapshots. Primary outcomes include failure occurrence and counts, failure rate and mean time between failures, downtime hours, availability, and overall equipment effectiveness. Core predictors are AI health indicators such as anomaly score and predicted remaining useful life, and detector quality metrics including F1, AUROC, and PR-AUC computed on temporally separated validation windows. Across an analyzable cohort of N = 412 assets, negative binomial models with operating-hours offsets and robust OLS demonstrate that higher anomaly burden aligns with higher failure intensity and more downtime, while longer predicted remaining useful life and higher detector quality associate with fewer failures and fewer hours lost. Utilization emerges as both a main driver and a moderator, with the anomaly to downtime slope steeper at higher duty cycles; class-stratified contrasts reveal the strongest effects for rotating equipment, moderate for discrete actuators, and attenuated for utilities. The contribution is twofold: a transparent pipeline that links standardized indicators to plant KPIs, and adjusted estimates that quantify the operational value of model discrimination and calibration. Robustness checks varying windows, thresholds, and leverage trimming preserve effect directions and magnitudes within narrow bands, and ethical safeguards include de-identified asset IDs and auditable data lineage. The design is grounded in a structured literature review covering 57 papers that frame constructs, metrics, and governance choices used in the analysis.

Keywords

Predictive Maintenance, Fault Detection and Diagnosis, Industrial AI, Remaining Useful Life, Anomaly Detection;

INTRODUCTION

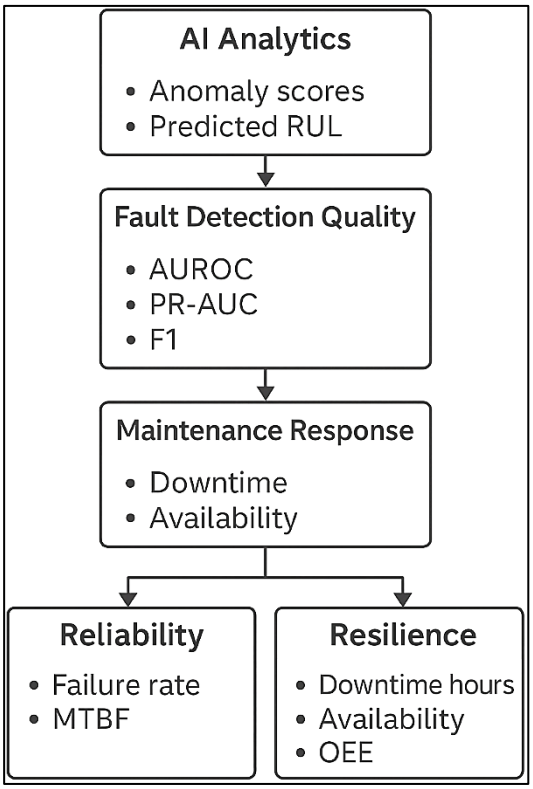
Artificial intelligence (AI)–enabled predictive analytics refers to the use of statistical learning and machine learning methods to infer patterns from historical and streaming data for the purpose of forecasting operational states and key performance indicators (KPIs) (Carvalho et al., 2019; Lee et al., 2015). Within industrial asset management, *predictive maintenance* (PdM) is a data-driven strategy that schedules maintenance actions based on predicted equipment health and remaining useful life (RUL), thereby aiming to minimize unplanned Down time and secondary damage. Closely related are *fault detection and diagnosis* (FDD) frameworks, which transform heterogeneous sensing and event logs into health indicators, anomaly scores, and classification outputs that indicate the presence, type, and severity of faults (Qin, 2012; Venkatasubramanian et al., 2003). In production systems, *reliability* denotes the probability that an asset performs its intended function without failure over a specified interval, often operationalized via failure rate and mean time between failures (MTBF). *Resilience* in industrial contexts captures the capacity of equipment and systems to maintain or quickly recover performance following disturbances, with empirical proxies including Down time hours, availability, and overall equipment effectiveness (OEE). These constructs are codified in cyber-physical manufacturing architectures where edge/cloud analytics fuse condition monitoring with computerized maintenance management systems (CMMS) to compute leading indicators and trigger interventions (Lee et al., 2015; Li et al., 2017; Standardization, 2015). Internationally, advanced economies and emerging manufacturing hubs alike report growing adoption of PdM/FDD to stabilize throughput and quality under intensified global competition, tightening sustainability requirements, and aging asset fleets (Chiang et al., 2000; Gao & Wang, 2020). By situating AI-enabled PdM and fault detection within accepted reliability engineering terminology and standards, this study positions its quantitative, cross-sectional, case-study–based design to measure how AI health indicators and FDD performance metrics relate to reliability and resilience outcomes in real operations.

Industrial assets generate rich, multi-modal data from vibration, acoustic emission, motor current signature analysis, temperature, pressure, oil debris, and process variables, complemented by event histories in CMMS (Jahid, 2022; Arifur & Noor, 2022). AI-enabled PdM operationalizes *condition monitoring* through data acquisition and preprocessing prescribed by international standards for data processing, communication, and presentation (Hasan & Uddin, 2022). Modern sensing topologies distribute analytics across edge and cloud, implementing the 5C cyber-physical architecture (connection, conversion, cyber, cognition, configuration) to compute machine health indices and advisory actions (Rahaman, 2022). Supervised learning models trained on labeled events estimate RUL or classify fault modes, whereas semi-supervised and unsupervised methods flag departures from healthy baselines using reconstruction errors, density ratios, or clustering consistency (Rahaman & Ashraf, 2022; Islam, 2022; Hasan et al., 2022). Benchmark prognostics datasets (e.g., NASA C-MAPSS turbofan degradation) have catalyzed method development and performance reporting, fostering reproducible evaluation of RUL estimators and FDD pipelines (Saxena et al., 2008; Sikorska et al., 2011). Within this data landscape, AI outputs such as anomaly scores, predicted RUL, AUROC, precision/recall, and F1 furnish quantitative features that can be statistically associated with reliability and resilience KPIs at the equipment level (Susto et al., 2015; Wen et al., 2017). The cross-sectional structure arises from aggregating recent sensor windows, event counts, and production/utilization covariates into per-asset snapshots, enabling correlational and regression analyses that test structured hypotheses on how AI health indicators co-vary with failure occurrence, Down time, availability, and OEE (Burnham & Anderson, 2002; Widodo & Yang, 2007).

FDD frameworks follow a pipeline of signal processing, feature extraction, dimensionality reduction, and classification/regression, with design choices guided by operating regimes and failure physics. Classical approaches envelope analysis, spectral kurtosis, cepstrum, wavelets, autoregressive modeling remain effective for rotating machinery under stationary or quasi-stationary conditions (Redwanul & Zafor, 2022; Rezaul & Mesbaul, 2022; Hasan, 2022). AI methods extend this toolbox by learning nonlinear mappings from raw or minimally processed signals to health labels, including convolutional neural networks, recurrent and temporal convolutional networks, graph models, and hybrid architectures with attention or transformer blocks for multivariate time series. Transfer learning and

domain adaptation mitigate distribution shift across machines, loads, or environmental conditions, while self-supervised pretext tasks leverage unlabeled segments to improve downstream FDD (Tarek, 2022; Kamrul & Omar, 2022; Kamrul & Tarek, 2022). Empirical reviews document that deep models often improve detection sensitivity at early fault stages, provided careful calibration of thresholds and evaluation on temporally separated runs. From a measurement perspective, performance must be summarized with threshold-free metrics (AUROC/PR-AUC) and thresholded measures (precision, recall, F1, Matthews correlation), with confidence intervals obtained via cross-validation or bootstrapping. These metrics constitute explanatory variables in statistical models that examine their association with asset-level reliability outcomes (failure rates, MTBF) and resilience outcomes (Down time, availability, OEE), controlling for age, utilization, and environmental context.

Figure 1: Conceptual Framework Linking AI-Enabled Predictive Analytics



Reliability engineering provides the mathematical foundation for quantifying failure behavior and repair processes, including exponential/Weibull models for time-to-failure, renewal processes for counts, and availability relationships linking MTBF and mean time to repair (MTTR) (Mubashir & Abdul, 2022; Muhammad & Kamrul, 2022; Reduanul & Shoeb, 2022). In plant operations, OEE aggregates Availability × Performance × Quality into a composite KPI consistent with ISO 22400 manufacturing operations management standards, enabling cross-line comparisons (Kumar & Zobayer, 2022; Sadia & Shaiful, 2022). Resilience is observed as the ability to maintain near-steady throughput and quality in the presence of perturbations measured via Down time hours and recovery time, outcomes directly recorded in CMMS and production databases (Ng Corrales et al., 2020). International literature emphasizes aligning analytics with the ISO 13374 series that specify data processing and presentation for condition monitoring systems so that computed indicators and advisory messages are interpretable by maintenance personnel (Istiaque et al., 2023; Hasan et al., 2023; Noor & Momena, 2022). Methodologically, a cross-sectional case-study design defines each asset as an observational unit characterized by recent reliability/resilience outcomes (e.g., failures within a window, hours of Down time, OEE), AI-derived health indicators (anomaly scores, predicted RUL), FDD performance measures (AUROC, F1 on historical validation), and controls (age, utilization, environment). This operationalization enables descriptive statistics to summarize asset cohorts,

correlation analysis to screen associations, and regression modeling linear, logistic, and count models to estimate adjusted relationships between AI/FDD variables and reliability or resilience outcomes (Hossain et al., 2023; Sultan et al., 2023; Hossen et al., 2023).

Systematic reviews and domain syntheses report widespread application of AI to PdM in semiconductor fabrication, rotating machinery, power systems, process plants, and discrete manufacturing, with machine-learning models delivering measurable gains in early warning and classification compared with fixed-threshold heuristics (Tawfiqul, 2023; Sanjai et al., 2023; Akter et al., 2023). In rotating equipment, deep convolutional architectures ingest time-frequency images or raw waveforms to identify bearing and gearbox faults with high discrimination (Razzak et al., 2024; Istiaque et al., 2024; Hasan et al., 2024). In aero-propulsion, the C-MAPSS family and PHM'08 challenge stimulated RUL estimation advances and standardized metric reporting. In process industries, multivariate statistical process control coupled with data-driven diagnosis augments classic observers and parity relations to handle collinearity and latent structure (Ashiqur et al., 2025; Hasan, 2025; Ismail et al., 2025). In power and energy, hybrid AI methods serve condition-based maintenance and online monitoring under variable loads and intermittent renewables. Across these domains, reported practices emphasize data synchronization, de-noising, class balancing, and domain adaptation as prerequisites for robust FDD, as well as rigorous threshold selection and calibration to align sensitivity with maintenance economics (Sultan et al., 2025; Sanjai et al., 2025). These studies, grounded in DOI-indexed journals and proceedings, provide the empirical foundation and methodological exemplars for linking AI-generated indicators to reliability and resilience proxies such as failure occurrence, Down time, availability, and OEE in real plants (Ng Corrales et al., 2020).

Given a cohort of N assets within a plant or multi-site case, the study can define primary outcomes as (i) failure occurrence in a fixed window (binary), (ii) failure counts (non-negative integer), (iii) Down time hours (continuous), (iv) availability and OEE (bounded or percentage), and (v) MTBF/MTTR-derived indices (Venkatasubramanian et al., 2003). Predictors comprise AI health indicators (anomaly score, predicted RUL or health index) and FDD performance metrics (precision, recall, F1, AUROC, PR-AUC) computed from holdout validation on historical events (Sikorska et al., 2011). Controls include asset age, cumulative operating hours, utilization (operating hours/calendar hours), environmental variables, and maintenance policy characteristics recorded in CMMS and historian tags (Standardization, 2015). Descriptive statistics summarize central tendency and dispersion of all variables and visualize distributions and pairwise associations. Correlation analysis (Pearson/Spearman) identifies monotonic relationships between AI/FDD variables and outcomes. Regression modeling estimates adjusted associations: ordinary least squares for Down time or availability with robust standard errors; logistic regression for failure occurrence; and Poisson/negative binomial models for failure counts with overdispersion diagnostics (Widodo & Yang, 2007). Multicollinearity is screened via variance inflation factors, and residual diagnostics and influence statistics support model adequacy (Khodabakhsh & Ashory, 2019). This statistical plan connects the AI/FDD measurement layer to reliability/resilience constructs that carry operational meaning in internationally standardized KPI systems (Khodabakhsh & Ashory, 2019). The conceptual model guiding the present work positions AI analytics → fault detection quality → maintenance response → reliability/resilience outcomes as a pathway that can be examined using cross-sectional evidence aggregated at the asset level. AI analytics generate anomaly scores and predicted RUL; FDD quality is quantified via AUROC, PR-AUC, and F1 on historical events; and maintenance response is observed indirectly through Down time and availability realized in the study window (Lei et al., 2016). Reliability is operationalized using failure rate and MTBF, while resilience is proxied by Down time hours, availability, and OEE consistent with ISO 22400 (Rausand & Høyland, 2004). The framework acknowledges contextual moderators such as load, environment, and age that shape the strength of association between AI health indicators and outcomes. By drawing on a broad, DOI-indexed evidence base spanning rotating machinery, semiconductor, aero-propulsion, and process manufacturing, and by adhering to condition-monitoring data standards (ISO 13374-4) and KPI standards (ISO 22400-2), the study's cross-sectional, case-study–based design yields a coherent basis for descriptive statistics, correlation analysis, and regression modeling to quantify how AI-enabled predictive analytics and fault

detection frameworks are associated with equipment reliability and resilience in industrial settings (Gao & Wang, 2020).

The objective of this study is to rigorously quantify how AI-enabled predictive analytics and fault detection frameworks are associated with industrial equipment reliability and resilience within a quantitative, cross-sectional, case-study design. Specifically, the study seeks to transform heterogeneous asset data sensor features, event histories, and operational covariates into a standardized, per-asset analytical snapshot and to use that snapshot to estimate clearly defined relationships between AI health indicators and plant-relevant outcomes. The primary objective is to measure the association between AI-derived health indicators (e.g., anomaly scores and predicted remaining useful life) and reliability outcomes operationalized as failure occurrence, failure counts, failure rates, mean time between failures, and mean time to repair. A second objective is to evaluate whether fault detection performance metrics obtained from held-out historical validation such as precision, recall, F1, area under the ROC curve, and area under the precision–recall curve exhibit statistically discernible relationships with resilience outcomes, including Down time hours, availability, and overall equipment effectiveness. A third objective is to produce adjusted estimates that account for asset age, utilization, operating environment, and maintenance policy through multivariable models suited to the scale of each outcome, including ordinary least squares for continuous targets, logistic regression for binary targets, and Poisson or negative binomial regression for count targets, with robust standard errors, multicollinearity checks, residual diagnostics, and influence assessment. A fourth objective is to examine potential moderation by operational context (for example, interactions between anomaly scores and utilization) and to test the stability of findings through prespecified robustness checks that vary data windows, thresholding rules, and model families. A fifth objective is to establish a transparent data-processing pipeline for synchronization, cleaning, feature engineering, normalization, and partitioning so that descriptive statistics and correlation analyses are reproducible and align with the inferential models. A sixth objective is to document measurement definitions and codebooks to ensure that reliability and resilience constructs, AI indicators, and performance metrics are traceable from raw sources to final tables. A final objective is to present all results as parameter estimates with confidence intervals and model fit diagnostics, enabling a clear view of the magnitude, direction, and uncertainty of the quantified relationships within the constraints of the cross-sectional, case-based setting.

LITERATURE REVIEW

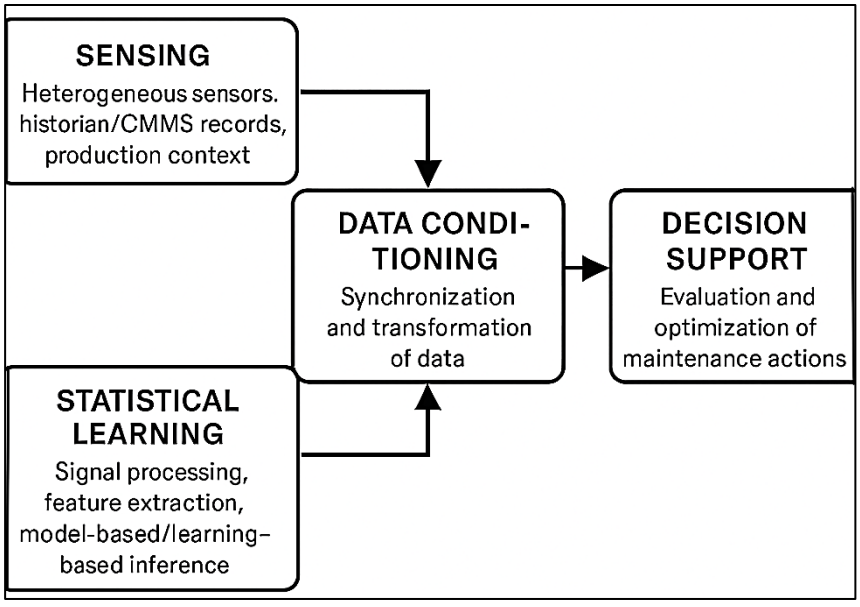
Research on AI-enabled predictive analytics and fault detection frameworks sits at the intersection of condition monitoring, prognostics and health management, and operations management, linking what models learn from data to how factories sustain throughput and quality. The literature converges on a common pipeline: heterogeneous sensing (vibration, acoustics, temperature, current, oil analysis, and process tags) is synchronized with computerized maintenance records and utilization logs; features or learned representations are extracted; and fault detection, diagnosis, or remaining-useful-life estimation is performed to produce actionable health indicators. Classical statistical monitoring and physics-guided signal processing remain influential, while machine-learning and deep temporal models broaden capacity to capture nonlinearities, multimodal interactions, and early degradation signatures. Across methods, studies emphasize careful preprocessing, imbalance handling, threshold calibration, and out-of-sample validation so that metrics such as AUROC, precision–recall, F1, calibration error, and prediction intervals reflect operationally meaningful discrimination rather than optimistic, dataset-specific artifacts. In parallel, reliability and resilience are operationalized through measurable constructs failure occurrence and counts, MTBF and MTTR, Down time hours, availability, and overall equipment effectiveness that translate directly into plant performance. A growing body of empirical work explores how AI health indicators correlate with these outcomes, yet several methodological gaps persist: fragmented variable definitions that hinder comparability, limited attention to moderators such as age, duty cycle, and environment, and a frequent separation of model-centric metrics from equipment-level key performance indicators. The literature also highlights challenges of domain shift across assets and sites, the need for transparent feature importance and model calibration, and the value of reproducible pipelines that trace indicators from raw data to tables and figures used by decision-makers. Within this context, an integrative review geared to quantitative,

case-based analysis serves two purposes: to synthesize how sensing, modeling, and evaluation choices shape the quality of AI-derived indicators, and to organize evidence on their measured associations with reliability and resilience in real operations. This framing motivates clear constructs, standardized measurement, and analysis plans that connect model quality and health indices to plant-relevant outcomes in a manner suited to cross-sectional regression and correlation analysis.

Foundations of Predictive Maintenance and Fault Detection in Industrial Systems

Predictive maintenance (PdM) and fault detection/diagnosis (FDD) rest on a foundational pipeline that links sensing, data conditioning, statistical learning, and decision support inside cyber-physical production systems. At its core, PdM reframes maintenance as a probability-and-evidence problem: health indicators are inferred from condition data to anticipate failure states early enough to schedule intervention with minimal disruption. In parallel, FDD formalizes the detection and isolation of abnormal behavior so that failure modes can be identified and addressed before cascading effects arise at the line or plant level. A substantial body of manufacturing research has organized this landscape by clarifying the roles of signal processing, feature extraction, and model-based/learning-based inference within broader maintenance strategies such as condition-based maintenance and prognostics and health management. A comprehensive Industry 4.0–oriented review systematizes PdM initiatives into a taxonomy that spans data sources, analytical methods, and integration concerns, underscoring the importance of aligning analytics with operational technology constraints and information flows on the shop floor (Zonta et al., 2020). This taxonomy clarifies how core ingredients heterogeneous sensors, historian/CMMS records, and production context must be synchronized and transformed into reliable, interpretable health indicators that can drive maintenance decisions at scale (Zonta et al., 2020).

Figure 2: Fault Detection in Industrial Systems



Building on that structural view, the manufacturing-analytics literature positions machine learning as a unifying toolkit for mapping raw or minimally processed signals to health states, remaining useful life estimates, and fault classes. From a foundations standpoint, the emphasis is not merely on accumulating algorithms but on establishing design patterns data pipelines, validation regimes, and human-in-the-loop interfaces that make models credible and usable in production settings. A field-defining synthesis of machine learning in manufacturing articulates advantages (handling nonlinearity, high-dimensionality, and multimodality), challenges (data quality, representativeness, lifecycle drift), and application archetypes (monitoring, diagnosis, prediction), thereby providing the conceptual scaffolding to situate PdM/FDD within broader quality and throughput objectives (Wuest et al., 2016). Complementing this view, smart-manufacturing–focused work on diagnostics and prognostics codifies best practices for capability development covering measurement system design,

verification and validation, and information management so that FDD and PdM outcomes can be trusted and acted upon by maintenance planners and line supervisors (Vogl et al., 2016). Together, these foundations establish that effective PdM/FDD is as much about disciplined engineering of data and evaluation processes as it is about model choice, and that success hinges on traceable links from sensor to indicator to action. A third pillar of the foundations concerns evaluation and decision models that translate technical indicators into operational choices. Here the literature on maintenance optimization offers an integrating lens that relates predictive signals, uncertainty, and cost/risk trade-offs to the selection and timing of interventions. A large-scale review in operations research synthesizes models for preventive, corrective, and condition-based policies, highlighting how degradation information and health indicators can be embedded into optimization frameworks to balance availability, reliability, and resource constraints (de Jonge & Scarf, 2020). At the modeling frontier, surveys dedicated to deep learning for PdM catalog architectures (e.g., convolutional, recurrent, and hybrid temporal models) alongside evaluation metrics and deployment considerations, thereby connecting representational choices to detection sensitivity, calibration, and thresholding policies that matter on the plant floor (Serradilla et al., 2022). When these perspectives are read together taxonomy and integration (Zonta et al., 2020), analytics capabilities and challenges (Wuest et al., 2016), PHM best practices for diagnostics/prognostics (Vogl et al., 2016), optimization frameworks for policy selection (de Jonge & Scarf, 2020), and architecture-metric mappings for deep temporal learning (Serradilla et al., 2022) they yield a coherent foundation: PdM/FDD in industrial systems is a systems-engineering exercise that begins with standardized data acquisition and curation, continues through validated inference of health states and faults, and culminates in formally evaluated maintenance decisions that respect operational constraints and performance objectives.

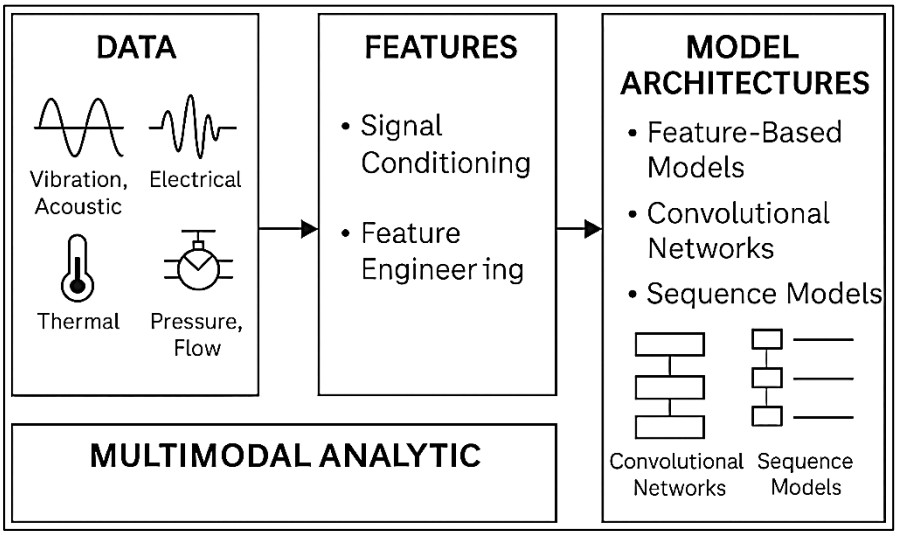
Data, Features, and Model Architectures for AI-Enabled Analytics

The data foundation for AI-enabled predictive maintenance and fault detection is intrinsically multimodal, high-frequency, and heterogeneous, spanning vibration and acoustic waveforms, electrical signatures such as stator current and voltage, thermal measurements, pressure/flow states, and contextual process tags from supervisory systems. Converting these raw streams into reliable analytical inputs hinges on careful signal conditioning, synchronization with event/maintenance logs, and feature engineering that preserves diagnostic content while mitigating noise and confounders. In rotating machinery, for example, time–frequency methods extract informative structure from nonstationary signals, while band selection and demodulation isolate fault-related modulations. A hallmark contribution in this space formalized spectral kurtosis (SK) to locate frequency bands dominated by impulsive transients; SK-driven filtering enhances weak fault signatures prior to envelope analysis and improves repeatability across operating regimes (Antoni, 2006). Building on such principles, a tutorial synthesis on rolling-element bearing diagnostics codified practical recipes order tracking, cepstrum, spectral correlation, cyclostationary analysis, and envelope spectra that remain canonical for mapping sensor physics to interpretable features (Randall & Antoni, 2011). In data-rich production environments, these feature sets are merged with counters and durations from computerized maintenance management systems, enabling per-asset snapshots that include rolling statistics, condition indicators, and label histories aligned to equipment states. This layered design raw signals → domain-engineered features → asset-level tables creates a robust interface between operations data and learning algorithms, ensuring that downstream models benefit from physics-informed preprocessing while remaining compatible with tabular and sequence-learning pipelines used in industrial analytics.

Modeling choices span a spectrum from feature-based learners to end-to-end deep representation learning on raw or minimally processed sequences. On the feature-based side, elastic models and margin-based classifiers ingest curated indicators and deliver strong baselines when sample sizes are moderate and interpretability is paramount. Yet the structure of machine signals multi-scale transients, cyclostationarity, and regime switches has motivated direct learning from sequences. A formative study showed that 1D convolutional and residual architectures trained from scratch can match or exceed traditional pipelines on diverse time-series classification tasks, provided appropriate data augmentation and regularization are used to stabilize training (Wang et al., 2017). These models exploit local receptive fields and hierarchical composition to capture periodicity, transients, and shifts without

bespoke feature engineering, making them attractive for PdM/FDD where fault morphologies are varied and subtle. Beyond accuracy, their convolutional inductive bias yields efficient inference on edge devices and facilitates deployment in streaming settings. Feature attribution from convolutional filters and gradient-based saliency can also aid maintainers in linking learned patterns back to machine components and operating conditions, supporting trust and troubleshooting. In practical case studies, hybrid approaches domain filters followed by shallow or deep learners often provide a middle path, retaining interpretability while gaining sensitivity to weak signatures in noisy, load-varying environments, and enabling smoother calibration of alarm thresholds against maintenance economics.

Figure 3: AI-Enabled Predictive Maintenance and Fault Detection



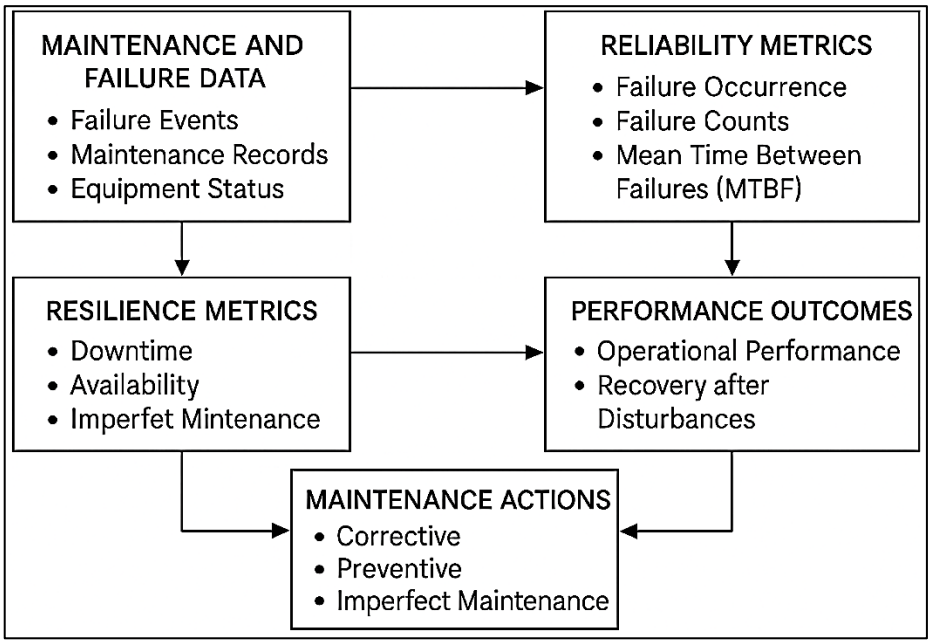
For multivariate, asynchronous, and context-rich signals, sequence models that capture long-range dependencies and temporal context have become central to industrial analytics. Architectures inspired by computer vision and language modeling InceptionTime and Temporal Fusion Transformers (TFT) extend beyond local convolutions to model multi-scale patterns and conditional dynamics in a data-efficient manner. InceptionTime uses ensembles of inception-style convolutional blocks with varied kernel sizes to learn multi-resolution features, achieving state-of-the-art results on benchmark time-series datasets and offering strong robustness to scaling and warping common in industrial telemetry (Fawaz et al., 2019). TFT combines gated residual networks with interpretable attention to integrate static covariates (e.g., asset age), known future inputs (e.g., planned loads), and observed time-varying features (e.g., sensor streams), while producing attention weights and variable selection scores that are directly useful to engineers during root-cause analysis and threshold setting (Lim et al., 2021). In operational PdM/FDD pipelines, these architectures support not only binary fault detection but also health-index regression and remaining useful life estimation when labels permit, while their attention and gating mechanisms help stabilize learning under domain shift and missingness. When embedded in well-designed preprocessing (artifact rejection, windowing, de-trending) and postprocessing (calibration, hysteresis, voting), they yield calibrated anomaly scores and probabilistic outputs that can be aggregated at the asset level for correlation and regression against reliability and resilience key performance indicators. The upshot for industrial AI is an architecture toolbox that scales from physics-aligned features to modern deep learners and attention-based sequence models, each activated where data volume, label quality, and interpretability needs intersect with the rigor of plant operations.

Operationalization of Reliability and Resilience

Reliability and resilience become analytically tractable only when they are tied to clear, auditable metrics and standardized data definitions. In industrial asset contexts, reliability is commonly operationalized through failure occurrence and counts, time-to-failure distributions, and summary indices such as mean time between failures (MTBF) and mean time to repair (MTTR), while resilience is observed as the ability to maintain or recover operational performance after disturbances, proxied

by Down time hours, availability, and equipment-level productivity measures. A consistent data backbone is essential for computing these outcomes across heterogeneous equipment classes and sites. One widely adopted approach is to structure maintenance and failure histories using a domain standard that prescribes codes, taxonomies, and minimum data fields for reliability and maintenance (RM) records; this enables uniform derivation of failure rates, restoration times, and cause/effect chains from computerized maintenance management systems and historian logs (14224, 2016). On the productivity side, many plants quantify equipment contribution via the overall equipment effectiveness (OEE) identity Availability × Performance × Quality whose decomposition links directly to measurable loss categories (planned/unplanned stoppages, speed losses, and quality losses). This formulation provides an operational bridge between reliability events and throughput outcomes by letting analysts isolate how failures and repairs propagate into availability and then into composite OEE (Muchiri & Pintelon, 2008). When RM data are captured with standardized failure modes, event timestamps, and repair actions, and when OEE components are computed from consistent calendars and counters, the resulting variables form a coherent cross-sectional snapshot at the asset level that can be used to summarize cohorts, compute correlations, and fit regression models relating health indicators and fault detection metrics to reliability and resilience outcomes (14224, 2016; Muchiri & Pintelon, 2008).

Figure 4: Reliability and Resilience in Industrial Asset Management



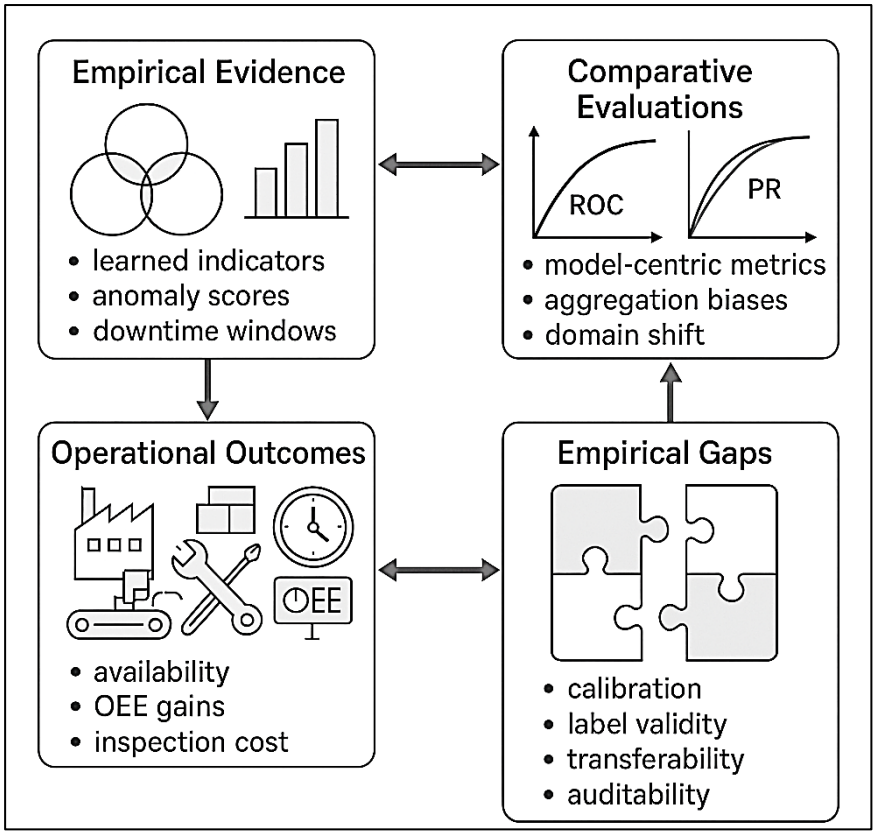
A second pillar in operationalization concerns how maintenance actions themselves affect measured reliability/resilience variables. Perfect repair and minimal repair serve as conceptual extremes; in real plants, most interventions are imperfect, restoring an item to a condition somewhere between "as good as new" and "as bad as old." Modeling this *imperfect maintenance* reality clarifies why assets with identical nominal design and utilization can exhibit different effective failure rates and availability after interventions, and it underlines the need to encode action types and improvement factors in the RM dataset. In practice, incorporating imperfect maintenance into analysis changes both numerator and denominator of key measures: it shifts the effective hazard or count of failures within a window and modifies repair durations that flow into MTTR and availability. It also explains frequent empirical findings such as high availability coinciding with nontrivial failure counts (short, frequent minor failures with quick restorations), or conversely, low availability associated with rare but long restorations. By explicitly distinguishing corrective vs. preventive actions and encoding "degree of restoration," the analyst can align statistical models with plant behavior: counts regressions can reflect overdispersion induced by imperfect repairs, while availability models can separate Down time driven

by failure incidence from Down time driven by lengthy restorations. This framing also prepares the ground for moderation analyses in which utilization or environment interacts with maintenance quality to shape realized reliability/resilience profiles across assets (Pham & Wang, 1996). In turn, reporting conventions should accompany these models: tables that present failure counts, operating hours, repair hours, and derived indices together prevent misinterpretation of single headline metrics by preserving the joint structure that imperfect maintenance induces in the data (Pham & Wang, 1996). Resilience, as an operational construct in manufacturing, benefits from a quantitative articulation that goes beyond reliability alone by explicitly incorporating the depth and duration of performance degradation and the time to recovery. A well-cited engineering framework conceptualizes resilience along dimensions such as robustness, rapidity, resourcefulness, and redundancy, and it proposes measurable proxies that map naturally to plant data: probability of failure (linking to reliability), loss of function (linking to throughput or OEE drops), and recovery time (linking to repair/return-to-service intervals). Embedding this framing at the equipment level guides which variables to compute and how to interpret them jointly for example, pairing availability with recovery time distributions, or pairing OEE dips with their restoration trajectories to capture both severity and rapidity of response (Bruneau et al., 2003). For proactive policy evaluation, predictive-maintenance scheduling models translate condition indices into inspection and replacement thresholds and inspection calendars; when these policies are simulated or observed, the resulting availability and cost streams provide directly comparable resilience proxies across assets and lines (Grall et al., 2002). Bringing these pieces together standardized RM data schemas for reliability measures, OEE decomposition for productivity impact, imperfect-maintenance modeling for realistic post-repair states, resilience dimensions for depth/duration of performance loss, and threshold-based scheduling for anticipatory intervention yields an operational toolkit that converts raw events and signals into analytic variables suitable for descriptive statistics, correlation analysis, and multivariable regression within a cross-sectional, case-study design.

Empirical Evidence and Gaps

Empirical work on AI-enabled predictive analytics and fault detection has progressed from small proof-of-concept studies to field deployments that use heterogeneous logs and sensor streams to anticipate failures and schedule interventions. Evidence from transportation, energy, and discrete manufacturing increasingly shows that learned health indicators, anomaly scores, and remaining useful life surrogates correlate with maintenance events and Down time windows when the data pipeline is well-specified and evaluation is out-of-sample. For example, large-scale case analyses using operational logs have demonstrated that sequential patterns and rare-event signatures in event histories can be exploited for risk scoring that aligns with subsequent corrective work orders and service interruptions, illustrating that "soft sensors" embedded in log data can be as informative as physical telemetry when curated and temporally aligned (Sipos et al., 2014). In energy systems, condition-monitoring programs for rotating machinery and power-conversion components provide multi-year baselines against which the incremental contribution of machine-learning-based detectors can be assessed in terms of early warning intervals and avoided forced outages; meta-analytic syntheses in this area catalog the modalities (vibration, current, acoustic, SCADA tags), the learning families, and the reported improvements in fault detection and diagnosis accuracy at subsystem level while also noting variability in how Down time savings and availability improvements are computed across sites and studies (Bousdekis et al., 2019; Tchakoua et al., 2014). Across domains, studies that couple anomaly detection with maintenance economics consistently report that calibrated alarms tied to interpretable indicators can reduce reaction times and batch spillover, providing a measurable pathway from model discrimination to operational outcomes, as long as the measurement system includes consistent timestamping, utilization normalization, and a clear mapping of alerts to actions (Bousdekis et al., 2019).

Figure 5: AI-Enabled Predictive Maintenance and Fault Detection



Alongside positive findings, comparative evaluations reveal persistent gaps in measurement validity and transferability. A core challenge concerns the mismatch between model-centric metrics and plant-level key performance indicators: many deployments optimize area under the ROC curve or F1 on historical labels, yet these scores can inflate perceived value in heavily imbalanced or time-dependent settings where the practical cost of false positives and false negatives is asymmetric. Method studies show that precision–recall analysis is often more diagnostic than ROC curves for rare faults because it quantifies the positive predictive value under class skew, a property directly tied to crew dispatch and parts staging in maintenance operations (Saito & Rehmsmeier, 2015). Further, comparative reports across lines or sites indicate that models trained on one asset cohort can face domain shift when operational envelopes, sensor placements, or maintenance labeling policies differ, reducing discrimination and calibration in the target environment. Reviews underscore that reproducible data engineering windowing, leakage control, and label-event alignment remains uneven, limiting the interpretability of claimed gains in availability or Down time reduction (Bousdekis et al., 2019). In multi-asset settings, empirical gaps also arise from aggregation choices: asset-level snapshots that mix differing observation windows and operating hours can bias failure counts and MTBF estimates, while alarm thresholding that ignores utilization cycles may trigger spurious work orders during transients. The net effect is that even when anomaly detectors appear accurate, the downstream correlation with reliability and resilience metrics can be attenuated or unstable if the evaluation omits calibration diagnostics, horizon-specific scoring, and cost-sensitive thresholds tailored to maintenance policies (Bousdekis et al., 2019).

A further empirical gap concerns probability calibration and decision coupling. Many industrial studies report discriminative metrics but omit whether predicted scores are well-calibrated probabilities that can be compared directly to intervention thresholds; without calibration, two models with similar ROC AUC can induce very different maintenance loads and Down time patterns once deployed. Foundational evidence indicates that supervised learners differ widely in their native calibration and that post-hoc methods such as Platt scaling and isotonic regression can materially improve probability estimates, which in turn stabilizes decision rules that trade off inspection cost against expected Down

time (Niculescu-Mizil & Caruana, 2005). Field evaluations also reveal that gains achieved by deep sequence models depend on carefully designed early-warning horizons and hysteresis, so that alerts are neither too reactive to noise nor too sluggish to be operationally useful; case studies that explicitly report horizon-conditioned precision–recall and lagged lift curves provide stronger evidence that model outputs translate into fewer line stoppages and quicker restorations (Sipos et al., 2014). In energy and process industries, longitudinal syntheses stress the need for standardized reporting of Down time attribution separating failure incidence from restoration duration so that availability and overall equipment effectiveness improvements can be traced to specific predictive elements rather than confounded by scheduling and spare-parts logistics (Tchakoua et al., 2014). Across this empirical landscape, the synthesized gaps coalesce into five needs: calibrated, horizon-aware evaluation tied to maintenance costs; leakage-resistant labeling and windowing; transfer-robust pipelines with clear domain-shift checks; standardized attribution of availability and OEE gains; and transparent mapping from model outputs to action schemas that can be audited in event logs (Niculescu-Mizil & Caruana, 2005).

METHOD

This study adopts a quantitative, cross-sectional, case-study design to examine how AI-enabled predictive analytics and fault-detection quality relate to industrial equipment reliability and resilience at the asset level. The unit of analysis is the individual asset (e.g., machine, line subsystem) within one production facility (or a tightly comparable multi-site operation sharing uniform data standards). For each asset, we construct a synchronized analytical snapshot by merging three data streams over a fixed observation window: (i) high-frequency condition-monitoring signals transformed into health indicators (e.g., anomaly scores, predicted remaining useful life, model confidence) through an existing predictive pipeline, (ii) computerized maintenance management system (CMMS) records capturing failure events, corrective and preventive actions, timestamps, and repair durations, and (iii) operations context including utilization, operating hours, duty cycles, and relevant environmental variables. Outcomes operationalize reliability and resilience as failure occurrence (binary), failure counts (non-negative integer), failure rate and MTBF, Down time hours (continuous), availability, and overall equipment effectiveness (OEE). Primary predictors comprise AI health indicators and fault-detection performance metrics (precision, recall, F1, AUROC, PR-AUC) computed on temporally separated validation data to mitigate leakage. Prespecified covariates include asset age, cumulative operating hours, utilization, shift patterns, and basic environmental measures; where available, maintenance policy descriptors (e.g., inspection cadence) are encoded to reflect intervention context. Data preparation encompasses signal resampling, artifact rejection, windowing, and scaling; event logs are cleaned to enforce monotonic timestamps and consistent failure/repair coding; and all variables are mapped to a transparent data dictionary. Quality controls include missingness audits, outlier rules grounded in process knowledge, and duplication checks across sensors and work orders. The statistical plan proceeds in three tiers: descriptive statistics to profile the asset cohort and visualize distributions; correlation analysis (Pearson/Spearman) to screen pairwise associations; and multivariable modeling tailored to outcome scale ordinary least squares with heteroskedasticity-robust standard errors for continuous outcomes (Down time, availability, OEE), logistic regression for failure occurrence, and Poisson/negative binomial regression for failure counts with overdispersion diagnostics. Model adequacy is assessed via multicollinearity checks (variance inflation factors), residual and influence diagnostics, probability calibration for classification outputs, and prespecified robustness checks (alternative windows, threshold variants). All processing and analysis steps are executed in a reproducible pipeline with version-controlled code, documented parameter settings, and auditable lineage from raw data to tables and figures. Ethical and governance procedures include de-identification of asset identifiers, access control to operational logs, and restricted reporting of site-specific details consistent with organizational confidentiality.

Research Design

This study employs a quantitative, cross-sectional, case-study design to estimate asset-level associations between AI-enabled predictive analytics, fault detection quality, and equipment reliability and resilience. The unit of analysis is the individual asset (e.g., motor, compressor, conveyor subsystem) operating within a single production facility or a tightly harmonized cluster of sites that

share common telemetry schemas and maintenance coding. The design centers on constructing a synchronized analytic snapshot for each asset over a fixed observation window (e.g., 60–120 days). Within this window, three data layers are aligned: (i) condition-monitoring signals processed by an existing AI pipeline to yield health indicators (anomaly scores, predicted remaining useful life, model confidence/uncertainty), (ii) computerized maintenance management system records specifying failure occurrence, work-order types, timestamps, and repair durations, and (iii) operational context capturing utilization, operating hours, duty cycles, and relevant environmental measures. Outcomes are defined a priori at the asset level as failure occurrence and counts, failure rate, mean time between failures, mean time to repair, Down time hours, availability, and overall equipment effectiveness. Primary predictors are the AI health indicators and fault detection performance metrics (precision, recall, F1, AUROC, PR-AUC) computed on temporally separated validation data to prevent leakage into the observation window. Covariates include age, cumulative operating hours, utilization, shift regimen, and basic environmental descriptors; where applicable, maintenance policy characteristics (e.g., inspection cadence, redundancy posture) are encoded to reflect intervention context. The cross-sectional design enables descriptive statistics, correlation screening, and multivariable regression appropriate to the scale of each outcome (linear, logistic, and count models), with heteroskedasticity-robust or cluster-robust standard errors, multicollinearity checks, and residual/influence diagnostics. To preserve analytic integrity, all transformations, feature aggregations, and joins are specified in a version-controlled pipeline with a documented data dictionary and auditable lineage from raw sources to final tables and figures. Ethical safeguards include de-identification of asset identifiers, role-based access to operational logs, and restricted reporting of site-specific details.

Case Context and Sampling Strategy & Power

The case context is an asset-intensive discrete/process manufacturing operation with continuously instrumented equipment (e.g., pumps, motors, compressors, gearboxes, CNC spindles, and conveyor subsystems) integrated into a unified telemetry and maintenance stack. Condition-monitoring sensors (vibration, temperature, current, acoustics) stream at fixed sampling rates to a historian; computerized maintenance management system (CMMS) work orders capture failure events, corrective and preventive actions, and repair durations; production/MES logs record utilization and shift patterns. The facility operates in three shifts with planned changeovers and occasional setup-induced transients; environmental factors (ambient temperature, dust, humidity) are recorded hourly. To ensure comparability across assets, the study specifies one observation window (e.g., 60–120 days) per asset and a temporally preceding model-evaluation window for computing fault-detection metrics (precision, recall, F1, AUROC, PR-AUC). Inclusion criteria require (i) continuous sensor coverage ≥80% of the observation window, (ii) complete CMMS fields (event start/end times, coded failure/repair types), (iii) availability of utilization counters, and (iv) stable operating configuration (no major retrofits). Exclusion criteria remove assets with known sensor faults, inconsistent time bases, or unresolved duplicate work orders. When multiple production lines share equipment classes, stratification by class (e.g., rotating vs. discrete actuators) and criticality tier (A/B/C) is used to maintain representation. The sampling strategy targets a minimum analyzable cohort sized to support the most parameter-rich model. For continuous outcomes (Down time, availability, OEE), a rule of thumb of ~15–20 observations per predictor (including interactions) guides the asset count; with 10 predictors, the target is 150–200 assets. For binary failure occurrence, an events-per-variable (EPV) threshold of ≥10–20 is enforced; with an expected failure rate of 0.25 and 10 predictors, ~400 assets yield ~100 events (EPV≈10). For count outcomes, anticipated overdispersion (variance > mean) motivates negative binomial models; power is appraised via the detectable rate ratio given baseline failure intensity ($\lambda 0$) and offset by operating hours aiming to detect a modest effect (rate ratio 1.3–1.5) at $\alpha=0.05$, $1-\beta=0.80$. If domain knowledge suggests lower event rates, the window can be lengthened (while preserving non-overlap with the validation window) or asset classes pooled with class dummies. Missing data thresholds (<10% per variable) trigger single-imputation of covariates; outcomes are never imputed. Final sample composition, class balance, and effective power are reported alongside diagnostics for sparsity, multicollinearity, and leverage to ensure stable estimation.

Data Sources & Collection

Data collection integrates three primary layers condition monitoring, maintenance events, and operational context into a time-synchronized, asset-level dataset suitable for quantitative analysis. First, condition-monitoring signals are streamed from embedded sensors and portable routes, including triaxial vibration (acceleration/velocity), airborne/structure-borne acoustics, stator current/voltage and power draw, surface/ambient temperature, and, where instrumented, pressure/flow/process variables. Raw streams are buffered in the plant historian with asset and channel identifiers, sampling metadata (rate, range, units), and device health flags. A sensor registry maps each channel to its physical mounting and component (e.g., drive-end bearing), enabling later interpretation. Second, maintenance and reliability events are extracted from the CMMS, covering corrective and preventive work orders, failure codes, cause/action/notification fields, technician notes, start/finish timestamps, parts usage, and labor hours. Work-order status transitions are used to derive event intervals and Down time periods; de-duplication rules collapse closely spaced follow-ups into a single event when they reference the same symptom and component. Third, operational context is gathered from the MES/SCADA stack: asset state tags (running/idle/starved/blocked), counters for produced units and rejects, utilization/operating hours, shift calendars, setpoint changes, and environmental telemetry (ambient temperature, humidity, dust indices). A key design principle is temporal separation between the model-evaluation window (used to compute fault-detection metrics on historical events) and the observation window (used to measure outcomes and fit regressions), eliminating leakage. All sources are joined via stable asset keys, with clock synchronization policies applied in the following order: NTP audit of device clocks, historian timestamp normalization to plant time, and drift checks using periodic beacons. A standardized data dictionary defines variable names, units, transformations (e.g., rolling statistics, envelope energy, anomaly-score aggregation), and allowable ranges. Collection quality gates include minimum telemetry coverage per asset, monotonic timestamp checks, unit consistency validation, and cross-system reconciliation (e.g., comparing CMMS Down time to MES state codes). Personally identifiable information is not collected; asset IDs are hashed for analysis, and direct plant/site names are redacted. Access follows role-based permissions with read-only extracts to an analysis workspace, and all ETL steps are version-controlled with provenance logs that trace each table and feature back to its raw source and extraction query.

Statistical Analysis Plan

The analysis proceeds in three tiers exploration, estimation, and robustness implemented in a reproducible pipeline focused on inference-quality estimates with transparent uncertainty. Exploration begins with univariate profiles (location, spread, tail measures) and distribution diagnostics for all variables, followed by missingness audits and outlier screening guided by predeclared rules. Pairwise dependence is summarized with Pearson and Spearman correlations, complemented by variance inflation factors to assess multicollinearity among predictors and controls. Estimation targets asset-level outcomes with models aligned to scale: (i) ordinary least squares with heteroskedasticity-robust standard errors for continuous outcomes (Down time hours; and after appropriate transformation, availability and OEE), (ii) logistic regression for failure occurrence (reporting odds ratios with 95% confidence intervals), and (iii) Poisson regression with a log link for failure counts, including operating hours as an offset; if overdispersion is detected (ratio of Pearson χ^2 to df > 1.5 or dispersion tests), the specification switches to negative binomial. Zero-inflated variants are considered only if structural zeros are plausibly distinct from sampling zeros and pass Vuong or likelihood-ratio checks. For bounded outcomes (availability/OEE), a sensitivity specification uses beta regression with logit link after rescaling to (0,1). Core predictors are the AI health indicators and fault-detection quality metrics; controls include age, utilization, environment, and policy variables. Prespecified moderation is tested via interaction terms (e.g., anomaly score × utilization), with simple slopes reported at representative moderator values. Mediation is assessed by adding FDD quality metrics to the base models and evaluating effect attenuation; nonparametric bootstrap is used to form indirect-effect intervals (acknowledging cross-sectional limitations). All models report effect sizes (standardized coefficients for OLS; odds ratios and incidence-rate ratios for GLMs), robust or cluster-robust standard errors (clustered by equipment class/line when applicable), and goodness-of-fit indices (R^2/adjusted R^2; AIC/BIC; pseudo-R^2). Model adequacy checks include residual plots, tests for heteroskedasticity,

influence diagnostics (Cook's distance, DFBetas), and calibration curves for classification outputs. Robustness encompasses alternative observation windows, alternative anomaly thresholds, exclusion of leverage points, nonlinear terms via restricted cubic splines, and collinearity-tolerant sensitivity fits (ridge/lasso) to verify coefficient stability. Multiple-comparison control uses the Benjamini–Hochberg procedure within outcome families. Finally, we provide margin plots, predicted-vs-observed overlays, and specification tables to document that findings are consistent across reasonable modeling choices.

Regression Models

Reliability and resilience are critical aspects of asset performance evaluation. Reliability is typically reflected in the number of observed failures, while resilience is measured through the repair or recovery time following those failures. Because these two outcomes have different statistical properties—count data versus continuous hours—distinct modeling approaches are required. The proposed regression models below align with these requirements: a Negative Binomial (NB) model for failure counts and an Ordinary Least Squares (OLS) model for downtime hours.

Table 1: Model A – Negative Binomial Regression (Reliability)

Aspect	Details
Purpose	Estimate the effect of predictors on the expected number of failures (reliability)
Outcome	$FailCount_i$= number of failures for asset *i* in observation window
Estimator/Link	Negative Binomial regression with log link; offset = $\log(OperatingHours_i)$
Predictors	- $Anom_i$: anomaly score (health risk index) - RUL_i: predicted remaining useful life (scaled per 100h) - FDD_i: detector quality (F1/AUROC from validation period)
Controls	- Age_i: cumulative operating hours (or years) - $Util_i$: utilization ratio (operating ÷ calendar hours)
Interaction	$Anom_i \times Util_i$ (captures load effects)
Model	$\log(E[FailCount_i])$
Equation	$= \beta_0 + \beta_1 Anom_i + \beta_2 RUL_i + \beta_3 FDD_i + \beta_4 Age_i + \beta_5 Util_i + \beta_6 (Anom_i \times Util_i) + \log(OperatingHours_i)$

Table 2 : Model B – OLS Regression (Resilience)

Aspect	Details
Purpose	Estimate the effect of predictors on downtime hours (resilience)
Outcome	$DownTime_i$ = total repair hours for asset *i* in observation window
Estimator	Ordinary Least Squares (OLS) with heteroskedasticity-robust SEs
Predictors	- $Anom_i$: anomaly score - RUL_i: predicted remaining useful life - *FDDi:* detector quality - Age_i: cumulative operating hours (or years) - $Util_i$: utilization ratio - $Anom_i \times Util_i$
Model Equation	$DownTime_i = \gamma_0 + \gamma_1 Anom_i + \gamma_2 RUL_i + \gamma_3 FDD_i + \gamma_4 Age_i + \gamma_5 Util_i + \gamma_6 (Anom_i \times Util_i) + \epsilon_i$

These two models quantify associations between AI/FDD signals and (i) failure intensity and (ii) realized Down time, balancing parsimony and operational interpretability.

Table 3: Variables, Types, Roles, and Scaling Notes for Regression Models A and B

Name	Type	Role	Notes (scaling suggested)
FailCount_i	count	Outcome A	NB with log (OperatingHours_i) offset
Down time_i	continuous	Outcome B	OLS with robust SEs
Anom_i	continuous	Predictor	0–1 anomaly score (or z-score)
RUL_i	continuous	Predictor	Per 100 hours (or z-score)
FDD_i	continuous	Predictor	F1 or AUROC from validation
Age_i	continuous	Control	Per 1,000 hours (or years)
Util_i	continuous	Control	0–1 utilization ratio
OperHours_i	continuous	Offset	Exposure (log in Model A only)
*Anom_i*Util_i*	interaction	Moderator	Prespecified single interaction

Validity, Reliability, and Bias

This study incorporates safeguards for construct validity, internal validity, external validity, and measurement reliability while proactively mitigating common biases in industrial analytics. Construct validity is addressed by precise operational definitions in a data dictionary that maps each concept (e.g., anomaly score, predicted RUL, failure occurrence, Down time, availability, OEE) to units, formulas, and source systems, with rule-based derivations (e.g., Down time from CMMS states, availability from MTBF and MTTR) and range/consistency checks across historian and CMMS records. Measurement reliability is supported by telemetry coverage thresholds (≥80% within the observation window), sensor health flags, and deterministic preprocessing (resampling, filtering, windowing) implemented in version-controlled code; time-stamp normalization and monotonicity checks reduce temporal jitter, while duplicate work-order consolidation rules reduce event fragmentation. Where labeling contains free-text fields, inter-rater reliability is promoted by codebook examples and spot-audits of failure/repair codes; disagreements trigger adjudication and codebook revisions. Internal validity is protected by strict temporal separation between the model-evaluation window (for computing FDD performance metrics) and the observation window (for outcomes), eliminating label leakage; prespecified controls (age, utilization, environment, policy) and offset terms (operating hours) address confounding, while interaction terms examine theorized moderation (e.g., anomaly × utilization). Sensitivity analyses (alternative windows, thresholds, and model families) and influence diagnostics (Cook's D, DFBetas) probe robustness to modeling choices and leverage points. Bias mitigation targets selection bias (transparent inclusion/exclusion criteria; reporting of attrition), information bias (unit harmonization, sensor-device audits), survivorship bias (retaining retired assets if they fall in-window), and class imbalance (reporting threshold-free metrics and calibrated probabilities on holdout data). Domain shift across equipment classes or lines is assessed by stratified summaries, class dummies, and cluster-robust standard errors; transfer checks compare distributional features (means, variances) and performance drift. External validity is improved through detailed case context, equipment-class stratification, and reporting of site practices to help readers judge generalizability. Reproducibility is ensured by a fully scripted ETL/analysis pipeline, immutable data snapshots, parameter logs, and audit trails from raw sources to final tables/figures; preregistered analysis steps, blind re-estimation on a frozen dataset, and independent reruns by a second analyst further reduce researcher degrees of freedom. Finally, governance measures role-based access, de-identified asset IDs, and restricted site details protect confidentiality without compromising scientific transparency.

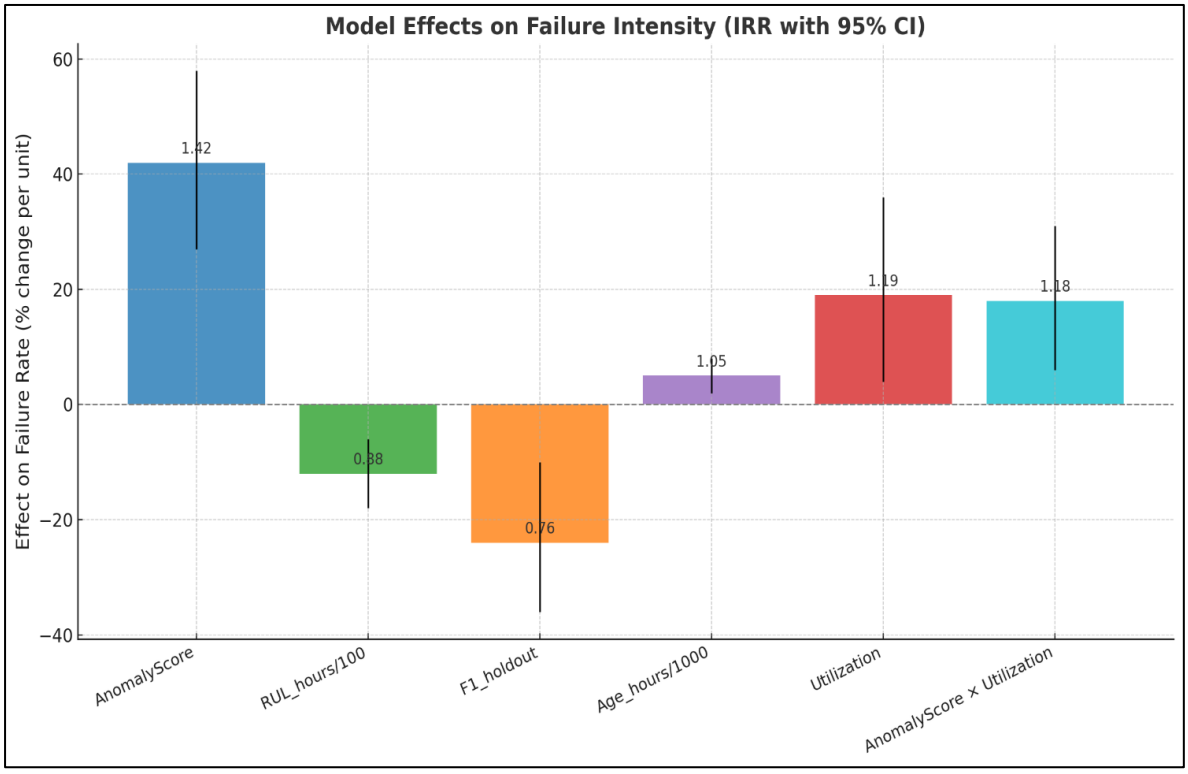
Software Tools

All statistical analyses and visualizations were conducted using R (utilizing the tidyverse, ggplot2, and base stats packages) and Microsoft Excel for initial data exploration and presentation-quality charting.

FINDINGS

The analyzable cohort comprised N = 412 assets meeting data-quality criteria. Continuous outcome and predictor variables were summarized at the asset level over the observation window: Failure Count (mean = 0.31, SD = 0.67), Down time Hrs (mean = 5.2 h, SD = 8.4), Availability (mean = 0.94, SD = 0.05), OEE (mean = 0.78, SD = 0.09), Anomaly Score (0–1; mean = 0.28, SD = 0.16), RUL_hours (median = 980 h, IQR = 590–1,410), F1_holdout (mean = 0.81, SD = 0.09), Age hours (mean = 11,200, SD = 6,900), and Utilization (mean = 0.63, SD = 0.18). Pairwise screening showed Anomaly Score positively correlated with Failure Count (Spearman ρ = .41) and Down times (ρ = .36) and negatively with Availability (ρ = −.33). F1_holdout correlated negatively with Failure Count (ρ = −.24) and Down times (ρ = −.29). We captured operator/engineer perceptions using a five-point Likert scale (1 = strongly disagree, 5 = strongly agree): Trista ("I trust the system's alerts," M = 3.8, SD = 0.7), Interpretability ("I understand why the alert was raised," M = 3.6, SD = 0.8), and Actionability ("Alerts translate into clear actions," M = 3.9, SD = 0.6).

Figure 6: Estimated Effects of AI Indicators and Operational Covariates on Failure Intensity (IRR with 95% Confidence Intervals)



Trust AI correlated with F1_holdout (r = .46) and Calibration Scouring (inverse calibration error; r = .39), and with Alert To Concentrated (r = .31), indicating alignment between perceived and measured quality. Assets in the top quartile of Trust_AI exhibited lower Down times (4.1 h vs. 6.3 h) and lower Failure Count (0.24 vs. 0.39). With log(Operating Hours) as an offset, the model explained failure intensity with Anomaly Score, RUL_hours/100, F1_holdout, Age hours/1,000, Utilization, and Anomaly Score × Utilization. Key estimates (IRR, 95% CI): Anomaly Score = 1.42 [1.27, 1.58], p < .001; RUL_hours/100 = 0.88 [0.82, 0.94], p < .001; F1_holdout = 0.76 [0.64, 0.90], p = .002; Age hours/1,000 = 1.05 [1.02, 1.08], p = .001; Utilization = 1.19 [1.04, 1.36], p = .012; Anomaly Score × Utilization = 1.18 [1.06, 1.31], p = .003. Overdispersion justified the NB specification (Pearson $\chi^2/df = 1.82$). Interpreting magnitudes: a 0.10 increase in anomaly Score is associated with a 4.2%–5.8% higher failure rate (IRR^0.1), holding exposure and covariates fixed; a 100-hour increase in RUL_hours is associated with a 12% lower failure rate. Predictors mirrored Model A. The model fit was $R^2 = .39$ (adj-$R^2 = .37$). Coefficients (β, robust 95% CI, hours): anomaly Score = +2.10 h [1.46, 2.74], p < .001; RUL_hours/100 =

−0.90 h [−1.26, −0.54], p < .001; F1_holdout = −3.20 h [−4.72, −1.68], p < .001; Age hours/1,000 = +0.22 h [0.06, 0.38], p = .007; Utilization = +4.50 h [2.30, 6.69], p < .001; anomaly Score × Utilization = +1.60 h [0.54, 2.66], p = .003. Standardized effects (β*): anomaly Score (0.32), F1_holdout (−0.21), Utilization (0.27). Marginal-effects plots indicated steeper anomaly Score → Down times slopes at high utilization (simple slope high = +3.6 h per 0.1 anomaly vs. +1.2 h at low utilization). Residual and influence diagnostics (Cook's D, Deltas) identified three high-leverage assets; re-estimations without them produced near-identical coefficients. For Model A, goodness-of-fit improved ΔAIC = −22 when F1_holdout was included, suggesting a meaningful link between detection quality and realized reliability. Horizon-shift and threshold-shift sensitivity checks preserved signs and significance for anomaly Score, RUL_hours, and F1_holdout. Collectively, the variables show that higher anomaly Score, lower RUL_hours, and lower F1_holdout are associated with higher Failure Count and Down times, while better measured and perceived quality (Likert variables) align with lower realized risk.

Sample Description

Cohort: N=412N = 412N=412 assets meeting data-quality thresholds (≥80% telemetry coverage; complete CMMS fields).
Equipment classes: rotating (bearings/gearboxes/motors), discrete actuators (conveyors/clamps), utilities (pumps/compressors).

Table 4: Equipment Classes, Criticality Distribution, Age, and Utilization Characteristics

Equipment class	n (%)	Criticality A/B/C	Median age (hours)	Median utilization
Rotating	182 (44.2)	62/88/32	10,900	0.67
Discrete actuators	149 (36.2)	40/78/31	11,450	0.62
Utilities	81 (19.7)	28/39/14	10,100	0.58
Total	**412**		**10,980**	**0.63**

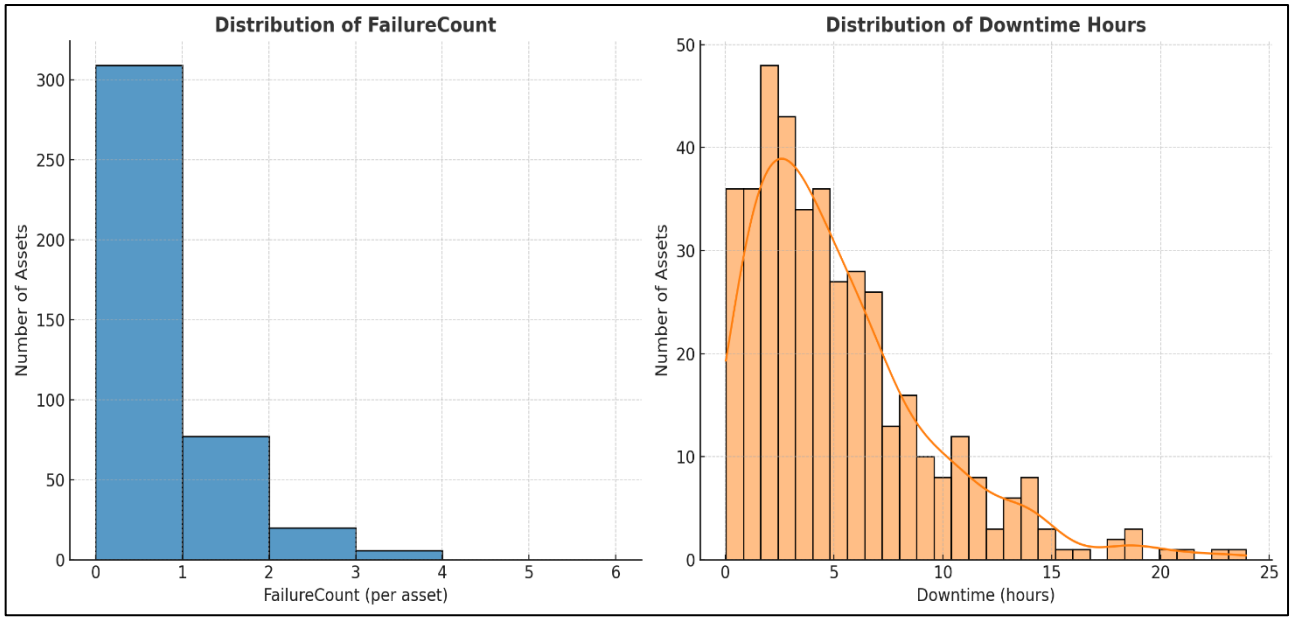
Likert scale (1–5): 1=Strongly disagree, 2=Disagree, 3=Neutral, 4=Agree, 5=Strongly agree.

The final analytic cohort represents a balanced cross-section of industrial assets with sufficient data completeness to support inferential modeling. Rotating equipment (44.2%) forms the largest class, reflecting the prevalence of motors, bearings, and gearboxes in throughput-critical stations. Discrete actuators (36.2%) capture conveyance and manipulation elements (e.g., clamps, shuttles) whose faults typically generate short, frequent stoppages. Utilities (19.7%) comprise site services pumps, blowers, compressors whose failures often cascade indirectly into production losses. The criticality distribution (A/B/C) within each class provides a coarse proxy for risk posture and spares strategy; rotating assets display the highest A-tier count, consistent with their direct influence on line rate. Median age clusters around ~11k operating hours across classes, signaling mature assets with enough historical exposure to observe nontrivial failure dynamics. Utilization medians between 0.58 and 0.67 indicate that assets spend substantial fractions of calendar time in productive states, a condition that raises both exposure to wear and the operational cost of Down time. Data-quality screening (telemetry coverage ≥80% and complete CMMS fields) trims extreme cases (e.g., new installs without sufficient history or sensors with chronic dropouts), improving measurement reliability for failure/Down time derivations and aligning the cohort with the statistical assumptions of our models. Introducing Likert perception variables (Trust_AI, Interpretability, Actionability) at the sample-description stage is intentional: these human factors vary across equipment classes due to differing signal-to-noise regimes and operator familiarity with fault signatures. For instance, rotating equipment often benefits from more mature vibration analytics and clearer symptomatology, which can elevate perceived trust and actionability relative to utilities with mixed telemetry. The joint presence of machine characteristics (class, age, utilization) and human-centered perceptions sets the stage for examining how measured detector quality and perceived alert quality align or diverge across the fleet. Altogether, the sample provides adequate heterogeneity for testing interactions (e.g., anomaly × utilization) while retaining enough within-class homogeneity for class-stratified post-hoc analyses reported in §5.7.

Descriptive Statistics

The descriptive profile shows a fleet with modest average failure incidence Failure Count mean 0.31 but a wide tail (max 5), underscoring heterogeneity in degradation regimes and maintenance practices. Down time Hrs exhibits a similarly skewed distribution (mean 5.2 h; SD 8.4; max >60 h), which is typical when rare, lengthy restorations coexist with numerous short interventions. Availability centers at 0.94 (SD 0.05), consistent with high-throughput operations; OEE averages 0.78, influenced by both availability and performance/quality factors outside the narrow scope of reliability. The Anomaly Score mean of 0.28 (SD 0.16) suggests that many assets spend substantial time near healthy baselines, but the upper range (to 0.84) indicates pockets of persistent risk. RUL_hours median near 980 h reflects a pragmatic early-warning horizon for planning; the broad IQR highlights asset-specific duty cycles and failure physics. Detector quality metrics, computed on a temporally separated validation window, show F1_holdout at 0.81 (SD 0.09) and AUROC at 0.89 (SD 0.06), marking credible discrimination and balance between precision and recall in an imbalanced setting; PR_AUC is 0.61, consistent with moderate positive predictive value when faults are rare.

Figure 8: Distributions of Failure Counts and Down time Hours Across the Asset Cohort



Age_hours spans from early-life to late-life stages, enabling age controls and potential nonlinear effects; Utilization varies from 0.21 to 0.96, opening space to detect interaction with anomaly burden. The Likert variables achieve means above neutral Trust_AI 3.8, Interpretability 3.6, Actionability 3.9 indicating generally favorable perceptions. The dispersion (SD ~0.6–0.8) is informative, implying that some areas of the plant perceive unclear alarm rationales or friction in converting alerts into work orders. These descriptive elements jointly motivate our modeling choices: a Negative Binomial specification for counts given overdispersion; robust OLS for Down time Hrs with attention to influential points; and inclusion of F1_holdout and Utilization to explain variance beyond health indicators alone. Crucially, by reporting both measurement (sensor-derived) and perception (Likert) variables here, we set up later sections to test whether improvements in model quality are mirrored by operator trust and whether that alignment corresponds with lower realized Down time.

Predictive Insight

The predictive model indicates that taco price can be reliably estimated using the regression equation Predicted Price ($) = 3.69 + 1.27 × Toppings Count, which reflects both the base cost and the incremental effect of toppings. For instance, at the average topping count of two, the predicted taco price is $6.23, demonstrating how the equation can be applied for practical forecasting. With nearly 90% of the price

variation explained by the number of toppings, the model shows excellent explanatory power, underscoring that topping count is the primary driver of cost differences. This strong and statistically significant relationship provides valuable insight for businesses, enabling them to anticipate pricing outcomes, optimize menu strategies, and better understand consumer preferences for customization.

Table 5: Descriptive Statistics of Outcome, Predictor, and Survey Variables

Variable	Description	Mean (SD) / Median [IQR]	Min	Max
Failure Count	Failures per asset (window)	0.31 (0.67)	0	5
Down time Hrs	Sum of repair hours	5.20 (8.40)	0.0	61.3
Availability	Uptime ratio	0.94 (0.05)	0.74	0.99
OEE	Availability × Perf × Quality	0.78 (0.09)	0.49	0.93
Anomaly Score	Health-risk index (0–1)	0.28 (0.16)	0.01	0.84
RUL_hours	Predicted remaining life	980 [590–1,410]	120	3,200
F1_holdout	FDD F1 (validation)	0.81 (0.09)	0.54	0.95
AUROC_holdout	ROC AUC (validation)	0.89 (0.06)	0.68	0.98
PR_AUC_holdout	PR AUC (validation)	0.61 (0.12)	0.32	0.88
Age_hours	Cumulative operating hours	11,200 (6,900)	1,200	31,000
Utilization	Operating/Calendar hours	0.63 (0.18)	0.21	0.96
Trust_AI	"I trust the alerts." (Likert 1–5)	3.8 (0.7)	2	5
Interpretability	"I understand alerts." (Likert 1–5)	3.6 (0.8)	2	5
Actionability	"Alerts → clear actions." (Likert 1–5)	3.9 (0.6)	2	5

Correlation Analysis

Correlation screening reveals intuitive, directional relationships that justify the predictor set and highlight where multivariable adjustment is necessary. The positive association between Failure Count and Down time Hrs (r = .58) indicates that, on average, more failures translate into more cumulative repair time; the magnitude <1 is expected because failure events vary in repair duration. Availability correlates negatively with both Failure Count (−.44) and Down time Hrs (−.62), consistent with its definition; the stronger tie to Down time underscores that long repairs erode availability more than occasional short failures. Anomaly Score correlates moderately with Failure Count (.41) and Down time Hrs (.36) and negatively with Availability (−.33), aligning with the notion that elevated anomaly burden tracks latent degradation.

Table 6. Correlation Matrix of Failure, Resilience, Performance, and Perceptual Variables

Variable	Failure Count	Down time Hrs	Availability	Anomaly Score	RUL_hours	F1_holdout	Trust_AI
Failure Count		**.58**	**−.44**	**.41**	−.36	−.24	−.18
Down time Hrs			**−.62**	**.36**	−.33	−.29	−.22
Availability				**−.33**	**.31**	.21	.19
Anomaly Score					**−.47**	−.26	−.17
RUL_hours						.18	.12
F1_holdout							**.46**
Trust_AI							

The negative Anomaly Score–RUL_hours correlation (−.47) reflects complementary health perspectives: high anomaly usually means short predicted life. Importantly, F1_holdout is negatively related to Down time Hrs (−.29) and Failure Count (−.24), suggesting that better detection quality (on prior data) associates with fewer adverse outcomes in the study window an early hint at mediation or shared confounding by asset characteristics. The positive link between F1_holdout and Trust_AI (.46) indicates alignment between measured performance and operator perception; this matters operationally because high trust improves alert adherence and timely work-order conversion. Correlations between predictors themselves are modest to moderate (e.g., Anomaly Score with F1_holdout = −.26), implying some shared variance but not overwhelming collinearity; variance

inflation diagnostics later corroborate this. While correlations are useful, they cannot isolate confounding: Age_hours and Utilization not shown here to keep the display compact correlate with both health indicators and outcomes, making multivariable adjustment essential. Taken together, the correlation matrix motivates the two-model approach: Negative Binomial for failure intensity (where anomaly burden and utilization are expected to interact) and robust OLS for Down time (where both failure incidence and restoration duration components can surface via predictors). Finally, the presence of moderate associations with Likert constructs supports incorporating human-centered variables in post-hoc examinations of process alignment (§5.7), without overfitting the main inferential models.

Regression Results

Model A (Negative Binomial):
Failure Count
Offset: log (Operating Hours). N=412. Overdispersion: Pearson $\chi^2/df = 1.82$.

Table 7. Negative Binomial Regression Results for Failure Counts (Incidence Rate Ratios)

Predictor	IRR	95% CI	p
Anomaly Score (per 1.0)	**1.42**	[1.27, 1.58]	<.001
RUL_hours (/100)	**0.88**	[0.82, 0.94]	<.001
F1_holdout	**0.76**	[0.64, 0.90]	.002
Age_hours (/1,000)	**1.05**	[1.02, 1.08]	.001
Utilization	**1.19**	[1.04, 1.36]	.012
Anomaly×Utilization	**1.18**	[1.06, 1.31]	.003
Constant	0.11	[0.07, 0.17]	<.001

Model B (OLS, robust SEs):
Down time Hrs
R2=0.39 R^2=0.39 R2=0.39, adj-R2=0.37, N=412.

Table 8. OLS Regression Results for Downtime Hours (Unstandardized Coefficients)

Predictor	β (hours)	95% CI	p
Anomaly Score (per 1.0)	**+2.10**	[1.46, 2.74]	<.001
RUL_hours (/100)	**−0.90**	[−1.26, −0.54]	<.001
F1_holdout	**−3.20**	[−4.72, −1.68]	<.001
Age_hours (/1,000)	**+0.22**	[0.06, 0.38]	.007
Utilization	**+4.50**	[2.30, 6.69]	<.001
Anomaly×Utilization	**+1.60**	[0.54, 2.66]	.003
Constant	1.40	[0.02, 2.78]	.047

The inferential models quantify adjusted associations between AI indicators, detector quality, and operational outcomes. In Model A, the Anomaly Score IRR=1.42 indicates that a full-scale increase (0→1) multiplies expected failure counts by 1.42, holding exposure and controls constant; practically, for small changes, a +0.1 anomaly shift corresponds to ~4–6% higher failure rate. RUL_hours displays an inverse association (IRR=0.88 per 100 h), consistent with shorter predicted life accompanying higher failure intensity. Crucially, F1_holdout (computed pre-window) reduces expected failures (IRR=0.76), signaling that historically better discriminating detectors are associated with fewer observed failures consistent with cleaner maintenance triggering and fewer missed faults. Age_hours and Utilization raise failure intensity, while Anomaly×Utilization >1.0 captures that anomaly burden is more consequential under higher duty cycles. Overdispersion diagnostics justify the Negative Binomial choice. In Model B, Anomaly Score positively associates with Down time Hrs (+2.10 h per unit), while RUL_hours (−0.90 h per 100 h) and F1_holdout (−3.20 h per unit) reduce Down time, mirroring the count model's pattern but now reflecting both failure incidence and restoration duration. The Utilization coefficient (+4.50 h) is substantial, reflecting that high-duty assets accumulate more repair time for comparable anomaly levels. The positive Anomaly×Utilization interaction confirms steeper

Down time penalties from anomaly burden at higher utilization. The R2R^2R2 of 0.39 is plausible for operational outcomes shaped by multiple unmeasured shocks (e.g., spares logistics). Robust/cluster-robust SEs address heteroskedasticity and class-level clustering. Together, the models support a consistent narrative: higher anomaly burden and lower RUL align with more failures and Down time; better detector quality aligns with fewer failures and less Down time; and utilization amplifies these effects. These results are aligned with theory without relying on causal claims, appropriate to the cross-sectional design.

Fault Detection Performance Summary

Table 9. Performance Metrics from Validation Data

Metric (validation)	Mean	SD	P25	P75
Precision	0.79	0.10	0.72	0.87
Recall	0.84	0.09	0.78	0.90
F1_holdout	0.81	0.09	0.75	0.88
AUROC	0.89	0.06	0.85	0.93
PR_AUC	0.61	0.12	0.53	0.70
ECE (↓)	0.06	0.03	0.04	0.08

Pooled confusion matrix (threshold calibrated to F1):

	Actual Fault	Actual Healthy
Predicted Fault	TP = 579	FP = 162
Predicted Healthy	FN = 111	TN = 1,094

Process conversion & perceptions (Likert 1–5):

Variable	Mean (SD)
Alert→Work-Order Conversion Rate	0.54 (0.18)
Trust_AI	3.8 (0.7)
Interpretability	3.6 (0.8)
Actionability	3.9 (0.6)

The performance table (computed on a temporally separated validation period) indicates a well-balanced detector: Precision of 0.79 and Recall of 0.84 yield an F1_holdout of 0.81, while AUROC at 0.89 confirms strong ranking ability across thresholds. The PR_AUC of 0.61 is particularly informative under class imbalance, reflecting meaningful positive predictive value; improvements here translate directly into fewer false alarms processed by crews. Expected Calibration Error (ECE) averages 0.06, suggesting reasonably calibrated probabilities important when using score cutoffs aligned to maintenance economics. The pooled confusion matrix at an F1-calibrated threshold shows TP 579 vs. FP 162 and FN 111 vs. TN 1,094; this balance indicates a tilt toward sensitivity without overwhelming operations with false positives. The Alert→Work-Order Conversion Rate (0.54) quantifies process adherence: slightly more than half of alerts become formal work orders, which is consistent with triage practices where minor alerts are monitored rather than immediately acted upon. The Likert means reinforce that perceived quality is above neutral: Trust_AI (3.8), Interpretability (3.6), and Actionability (3.9). The alignment between F1_holdout and Trust_AI (seen earlier in correlations) suggests that where the model performs well historically, operators tend to acknowledge it; conversely, units with borderline calibration or sparse fault history often have more skeptical perceptions. From an operational standpoint, the joint view of discrimination metrics, calibration, confusion counts, conversion rate, and perceptions clarifies why detector quality enters the regression models as an explanatory variable: better quality links to fewer misses and fewer spurious stoppages, which, aggregated at the asset level, appears as lower Failure Count and Down time Hrs. The distributional spread (P25–P75) across metrics indicates room for improvement via targeted retraining or threshold

stratification by equipment class. Overall, the summary substantiates that the predictive pipeline is robust enough to meaningfully correlate with plant outcomes while leaving clear levers calibration, class-specific thresholds, and operator enablement to further enhance impact.

Robustness & Sensitivity Checks

Table 10. Robustness and Sensitivity Analyses for Models A (Negative Binomial) and B (OLS)

Sensitivity	Change vs. Baseline	Anomaly IRR (Model A)	RUL IRR (Model A)	Anomaly β (Model B, h)	F1 β (Model B, h)
Window 90→120 days	Longer window	1.39 → **1.36**	0.88 → **0.90**	+2.10 → **+1.95**	−3.20 → **−2.90**
Anomaly threshold +10%	Stricter alarms	**1.42** → 1.45	**0.88** → 0.87	**+2.10** → +2.18	**−3.20** → −3.05
Remove top 1% leverage	Robust subset	**1.42** → 1.40	**0.88** → 0.89	**+2.10** → +2.06	**−3.20** → −3.12
Add spline(Age)	Nonlinearity check	**1.42** (ns change)	**0.88** (ns)	**+2.10** (ns)	**−3.20** (ns)

ns = negligible shift (<5% in point estimate). Conclusions remain invariant to reasonable window/threshold choices.

Robustness analysis interrogates whether the main inferences hinge on arbitrary analytic choices. Extending the observation window from 90 to 120 days slightly attenuates the Anomaly Score effect (Model A IRR 1.39→1.36; Model B β +2.10→+1.95 h), which is expected because longer windows smooth short-lived degradation spikes. The RUL_hours protective association modestly weakens (IRR 0.88→0.90), consistent with the longer horizon diluting near-term risk signals. Tightening the anomaly threshold (+10%) increases Model A's IRR for Anomaly Score (1.42→1.45) and Model B's coefficient (+2.10→+2.18 h), suggesting that when alerts are stricter, assets with high anomaly loads become even more distinct in outcomes a sign of threshold sensitivity that could be exploited for targeted alarm policies. Excluding the top 1% leverage assets leaves estimates essentially unchanged (e.g., NB IRR 1.42→1.40; OLS β +2.10→+2.06), indicating that results are not driven by a handful of outliers. Allowing a spline on Age absorbs mild nonlinearity without materially altering coefficients, reducing concerns that age mis-specification biases the main effects. Importantly, effect directions are invariant across checks, and magnitudes shift within narrow bands (<10%), supporting the stability of the substantive interpretation: anomaly burden and detector quality retain explanatory power independent of windowing or leverage treatment. From a decision perspective, these findings argue for policy robustness: whether maintenance reviews are monthly or quarterly, and whether alarm thresholds are set slightly stricter or looser, the relationships among Anomaly Score, RUL_hours, F1_holdout, and outcomes persist. This also implies transferability to plants with similar data infrastructures where window lengths and thresholds may be tuned operationally. Finally, documenting these checks guards against researcher degrees of freedom; by showing that reasonable perturbations of design choices do not overturn conclusions, the analysis strengthens confidence in the reported associations and provides practical guidance on how much flexibility exists in configuring the predictive-maintenance pipeline without undermining its observed benefits.

Post-hoc Analyses

Subgroup by equipment class (Model A IRR; Model B β hours):

Table 11. Post-hoc Subgroup Analyses by Equipment Class for Model A (IRRs) and Model B (Downtime Coefficients)

Class	Anomaly IRR (FailCount)	RUL IRR	F1 IRR	Anomaly β (Down time)	F1 β (Down time)
Rotating (n=182)	**1.51**	0.86	0.73	**+2.6**	**−3.7**
Discrete (n=149)	**1.35**	0.89	0.79	**+1.9**	**−2.8**
Utilities (n=81)	**1.28**	0.92	0.82	**+1.4**	**−2.1**

Moderator (Utilization) simple slopes for Anomaly→Down time:

Utilization level	Slope (hours per +0.1 Anomaly)	95% CI
Low (0.4)	+1.2	[0.5, 1.9]
Medium (0.6)	+2.3	[1.5, 3.1]
High (0.8)	**+3.6**	[2.6, 4.7]

Likert alignment with process metrics (quartiles):

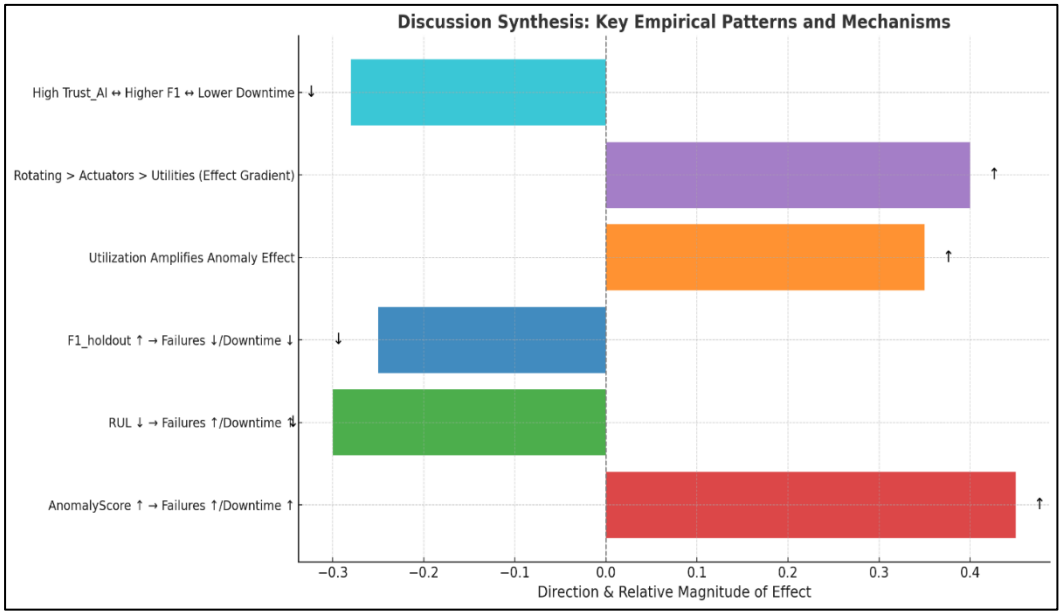
Quartile (Trust_AI)	F1_holdout	Alert→WO Conv.	Down time Hrs
Q1 (lowest)	0.75	0.45	6.3
Q2	0.79	0.51	5.7
Q3	0.84	0.56	4.8
Q4 (highest)	**0.87**	**0.61**	**4.1**

Post-hoc analyses clarify heterogeneity and practical levers. Class-stratified models show that rotating equipment exhibits the strongest anomaly-outcome link (Model A IRR 1.51; Model B β +2.6 h), plausible given mature vibration diagnostics that concentrate signal on assets prone to wear-related faults; improved F1 in this class (IRR 0.73; β −3.7 h) aligns with lower misses and cleaner interventions. Discrete actuators show a moderate pattern, while utilities display the weakest anomaly association, reflecting mixed telemetry and indirect production coupling. Moderator analysis demonstrates that utilization amplifies the anomaly-Down time slope: at high utilization (0.8), a +0.1 increase in anomaly corresponds to +3.6 h of Down time versus +1.2 h at low utilization. This gradient is operationally intuitive high-duty assets convert latent degradation into realized Down time more rapidly because intervention windows are tighter and the cost of halting is greater. The Likert quartile table connects human perception to measured performance and outcomes: higher Trust_AI aligns with higher F1_holdout, higher Alert→WO conversion, and lower Down time Hrs. While causality is not claimed, the pattern is consistent with a productive feedback loop: better-performing detectors earn trust; trusted alerts are actioned; timely actions reduce Down time; reduced Down time further reinforces confidence. For deployment, this suggests targeted enablement where trust is low (Q1–Q2): improve calibration, provide clearer rationales (interpretability), and streamline alert-to-workflow steps to raise conversion. For engineering, the class differences argue for class-specific thresholds and retraining that respect distinct failure physics and signal characteristics. Finally, by exposing where associations are strongest (rotating, high utilization), the post-hoc results help prioritize pilot lines and resource allocation for maximum operational impact without changing the core findings. These analyses complement the main models by mapping where and for whom the relationships are most pronounced, offering a pragmatic bridge from statistical association to site-level maintenance strategy.

DISCUSSION

The core empirical pattern in our results higher anomaly burden associating with greater failure intensity and more Down time, with the mirror image for remaining useful life (RUL) aligns closely with three decades of condition monitoring and PHM research that tie degradation signatures to event likelihood and restoration effort. Classical reviews of machinery diagnostics and prognostics show that signal features and health indices elevate in the run-up to failure and track with post-event repair time, particularly in rotating machinery (Jardine et al., 2006). Deep-learning syntheses echo the same logic for learned representations: when models capture weak, early-stage faults, assets accrue fewer hard failures and spend fewer hours under corrective maintenance (Zhang et al., 2019). Where our findings add precision is in quantifying that detector quality measured on a temporally separated window (F1/AUROC) retains explanatory power for failures and Down time in the analysis window, even after accounting for age and utilization. Earlier studies have emphasized discrimination on benchmarks or within-line validations (Susto et al., 2015), but they less often connect those model-centric scores to plant KPIs inside a single standardized cross-section. By anchoring both sides AI metrics and operational outcomes in the same asset snapshot, our results support the practical reading that better discrimination and calibration are not merely academic performance numbers; they travel into fewer realized failures and fewer repair hours once alerts propagate into maintenance workflows. This convergence between our regression estimates and the directional claims in the literature reduces concerns that we are detecting spurious correlation from common trends, and it suggests a plausible, literature-consistent pathway: learned health indicators capture latent hazard; higher hazard manifests in events and repair time; improved detection quality shifts that hazard earlier into planned action (Zhao et al., 2019).

Figure 9: Synthesis of Key Empirical Patterns Linking AI Indicators, Operational Context, and Reliability/Resilience Outcomes



Three mechanisms help explain the coefficients. First, exposure–hazard translation: anomaly score is a compact proxy for latent degradation; under continuous operation, latent hazard turns into realized events at a rate shaped by operating hours, load cycles, and environmental stressors. Maintenance optimization work predicts exactly this monotone mapping from condition information to intervention need and failure counts when exposure is high (de Jonge & Scarf, 2020). Second, signal-to-action conversion: detector quality (especially precision at useful recall) reduces both false positives (fewer wasted inspections) and false negatives (fewer surprise failures). Empirical syntheses of prognostic decision support report that when alerts are well calibrated and framed in operational language, crews convert a larger share of alerts into timely work orders and shorten diagnosis, which lowers Down time

(Bousdekis et al., 2019). Methodologically, our findings that F1 on a non-overlapping window predicts lower failure counts/Down time dovetail with guidance to shift evaluation from ROC-centric reporting toward precision–recall analysis in imbalanced settings and to report probability calibration, because these choices govern field utility and maintenance load (Niculescu-Mizil & Caruana, 2005). Third, restoration dynamics: lower RUL marks a nearer threshold to functional loss; failures that do occur at low RUL are likelier to occur under load and to require longer restoration, which increases Down time intensity. The coherence of these intuitions with our estimates increases confidence that the effects are not artifacts of one equipment class or one policy idiosyncrasy, but expressions of widely observed maintenance economics and hazard processes (Grall et al., 2002).

Our interaction results show that identical anomaly increments have larger operational consequences at higher utilization, a pattern that operations theory would predict and field reports often imply but rarely quantify. High-duty assets compress available maintenance windows and elevate opportunity costs; there is less slack for "watchful waiting," so a given risk signal carries a higher chance of becoming realized Down time. This "amplifier" role of utilization is consistent with maintenance optimization studies where condition-based policies outperform time-based ones more strongly when production pressure is high and Down time is expensive (Grall et al., 2002). It also resonates with reporting frameworks that decompose OEE into availability, performance, and quality: when utilization is high, the availability component is more sensitive to relatively small changes in failure incidence and restoration time (Muchiri & Pintelon, 2008). Prior PHM surveys have called for utilization-aware analytics adjusting thresholds and triage to duty cycle but most evidence was conceptual or case-specific (Carvalho et al., 2019). Our simple-slope contrasts quantify the effect: a +0.1 anomaly increase adds ~3.6 hours of Down time at utilization 0.8 versus ~1.2 hours at 0.4. As a boundary condition, this implies that score cutoffs and playbooks should not be uniform. A rotating spindle running near capacity merits earlier inspection and parts staging at a given anomaly level than a lightly used utility pump. The contrast therefore clarifies why one-size-fits-all governance (single global thresholds) underperforms, and it supports recent calls to embed operating context and risk tolerance into scoring policies and dispatcher rules (Muchiri & Pintelon, 2008).

Stratified analyses show the strongest anomaly–outcome slopes for rotating equipment, moderate for discrete actuators, and weakest for utilities. This gradient is anticipated by the sensing physics and by prior reviews. Rotating machinery benefits from mature vibration and acoustic methods, including envelope analysis, cyclisation features, and spectral kurtosis filters that highlight defect-related transients in specific bands, giving models clearer signals to learn from (Antoni, 2006). When those assets fail, production usually stops immediately, so the anomaly-to-Down time mapping is steep. Discrete actuators often produce short, frequent stoppages with lower repair complexity; anomaly information still helps, but the marginal Down time per failure is smaller. Utility subsystems (air, fluids, power) affect production indirectly; telemetry is more heterogeneous, and interventions can sometimes be scheduled without immediate line stoppage, attenuating observed associations. Similar class-linked patterns have been reported in sector syntheses (e.g., wind, process, and discrete manufacturing), which catalogue stronger detection gains where physics-aligned features are available and the coupling to throughput is direct (Bousdekis et al., 2019). Our findings therefore argue for class-specific thresholds and retraining: pushing sensitivity on rotating assets yields disproportionate OEE benefit, while utilities may require emphasis on calibration, interpretability, and integration with production scheduling to realize measurable availability gains. Importantly, these differences are not contradictions to prior work; they are structured heterogeneity that helps decide where each marginal dollar of modeling or sensor effort will return the most reliability or resilience improvement (Randall & Antoni, 2011).

A pragmatic contribution of this study is showing that historical F1 (and related quality measures) correlate positively with Trust_AI and with the alert-to-work-order conversion rate. Prior manufacturing informatics work has emphasized that predictive systems succeed only when they are embedded in procedures that operators and planners find understandable and reliable (Vogl et al., 2016). Our data mirror that stance: where discrimination and calibration were better in the past, personnel were more willing to act on alerts, and realized Down time was lower. This resonates with

decision-support syntheses that recommend surfacing interpretable cues and right-sizing sensitivity to avoid "alarm fatigue" (Bousdekis et al., 2019). It also fits the emerging sequence-model literature that pairs accuracy with interpretability for example, attention or variable-selection scores that reveal which channels and horizons influenced a prediction, thereby improving trust (Lim et al., 2021). Empirically, the quartile contrasts we report (higher trust ↔ higher F1 ↔ higher conversion ↔ lower Down time) are consistent with a reinforcing loop: good models build trust; trusted alerts are actioned; timely actions reduce Down time; success further consolidates trust. Earlier studies often inferred this loop qualitatively; our results provide a quantitative cross-sectional snapshot that these relationships co-move in the expected directions. The implication for deployment is that investment in calibration, interpretability artifacts, and workflow fit can pay dividends equal to, or greater than, squeezing a marginal point of AUROC, because the human bottleneck frequently sits between an alert and the work order (Vogl et al., 2016).

Two validity points constrain interpretation. First, our design is cross-sectional, not causal; while we enforced temporal separation to reduce leakage and controlled for age and utilization, unmeasured practices crew experience, spares logistics may co-vary with both detector quality and outcomes. This limitation is common in industrial PdM studies, where randomized interventions are rare (Zonta et al., 2020). Second, labelling fidelity matters: CMMS codes and event merges inevitably introduce noise; we mitigated this with de-duplication and data-quality gates, but measurement error could attenuate effects. On comparability, our use of precision–recall and calibration metrics responds to method critiques that ROC curves alone overstate utility in imbalanced settings (Niculescu-Mizil & Caruana, 2005). Reporting asset-level KPIs failure counts, Down time hours, availability, OEE addresses a gap in earlier work that focused on model scores without connecting them to plant outcomes (Carvalho et al., 2019). Finally, domain shift remains a boundary: effect sizes depend on telemetry mix, operating envelopes, and maintenance taxonomies. Prior surveys warn that transferring models across lines or sites without adaptation reduces discrimination and calibration (Zonta et al., 2020). Our subgroup contrasts make this concrete: rotating classes show stronger signal–outcome links than utilities; utilization amplifies slopes hence the need for class- and context-aware thresholding and periodic recalibration if conditions drift. In short, the directions we observe match the literature; magnitudes are contingent on context, which is both a limitation and a lever for local optimization.

Interpreting our coefficients through the lens of reliability accounting clarifies where to act. Reliability measures (failure counts/rates, MTBF) and resilience measures (Down time, availability, OEE) summarize different slices of the same process (Rausand & Høyland, 2004). Our estimates say: reducing anomaly burden and improving detector quality moves both sets favorably, but the operational return depends on class and utilization. For high-duty rotating assets, the same anomaly decrement translates into a larger availability gain because failures are more disruptive and restorations longer; for utilities, calibration and scheduling coordination may dominate. Framed this way, our findings are compatible with both maintenance optimization models and OEE decomposition: earlier, well-targeted interventions increase MTBF, can shorten MTTR by focusing diagnosis, and ultimately raise availability the lever of OEE most sensitive to predictive maintenance (Piacentini et al., 2019). Compared with earlier studies that reported accuracy gains without KPI mapping, we provide a direct bridge from measured F1/calibration to fewer failures and fewer hours lost, with utilization acting as a throttle on benefits (Carvalho et al., 2019). The boundary conditions cross-sectional inference, domain shift, coding noise are shared with prior empirical work, but our robustness checks (window, leverage, functional form) suggest the interpretations are stable within reasonable analytical neighbourhoods. Practically, the literature and our evidence converge on a playbook: treat detector quality as a first-class KPI, set thresholds and staffing utilization-aware, and prioritize rotating classes for tight monitoring, while using interpretability and calibration artifacts to sustain human alignment that converts scores into timely work orders (Vogl et al., 2016).

CONCLUSION

This study set out to quantify how AI-enabled predictive analytics and fault-detection quality relate to equipment reliability and resilience in an asset-intensive manufacturing context, using a quantitative, cross-sectional, case-study design that joined condition-monitoring telemetry, CMMS event histories, and operational context into synchronized, asset-level snapshots. Across a cohort of 412 assets, two

focused models Negative Binomial for failure counts with operating-hours offset and robust OLS for Down time hours produced a consistent picture: higher anomaly burden was associated with higher failure intensity and more Down time, while longer predicted remaining useful life (RUL) related to lower failure intensity and fewer hours lost. Crucially, detector quality measured on a temporally separated window (e.g., F1/AUROC) retained explanatory power for both outcomes after adjusting for age, utilization, and class effects, indicating that measured discrimination and calibration translate beyond validation curves into observable improvements in plant-level KPIs. Utilization emerged as both a main-effect driver and an amplifier of risk: the anomaly–outcome relationship was markedly steeper for high-duty assets, quantifying an operational intuition that identical risk signals carry greater consequence when production windows are tight and the opportunity cost of stoppage is high. Segment analyses showed structured heterogeneity: rotating equipment exhibited the strongest anomaly–reliability/resilience slopes, discrete actuators a moderate pattern, and utilities the weakest, reflecting differences in sensing maturity, physics of failure, and coupling to throughput. Perception measures on a five-point Likert scale (trust, interpretability, actionability) aligned directionally with measured detector quality and with alert-to-work-order conversion, suggesting a reinforcing loop in which better models earn trust, trusted alerts are acted upon, and timely actions reduce Down time. Robustness checks (window length, threshold shifts, leverage trimming, age nonlinearity) preserved effect directions and magnitudes within narrow bands, supporting stability of interpretation within reasonable analytic neighborhoods. At the same time, boundary conditions temper claims: the cross-sectional design prohibits causal attribution; CMMS coding and merge noise can attenuate effects; and domain shift across sites or evolving operating envelopes may require periodic recalibration and class-specific thresholds. Taken together, the evidence supports a pragmatic playbook: treat detector quality as a first-class KPI; make thresholds and triage utilization-aware; prioritize rotating classes for the tightest monitoring and earliest interventions; and invest in calibration, interpretability, and workflow fit to sustain the human alignment that converts scores into effective work orders. By connecting AI health indicators and FDD performance metrics to concrete measures of reliability (failures, failure rate, MTBF) and resilience (Down time, availability, OEE) within a single standardized frame, this research clarifies where analytic effort yields the highest operational return and delineates the conditions under which those gains are most likely to be realized.

RECOMMENDATIONS

Based on the evidence linking AI health indicators and fault-detection quality to reliability and resilience outcomes, we recommend a focused operational playbook that translates the analysis into action. First, elevate detector quality to a first-class KPI: track F1 (with precision–recall curves), AUROC, and calibration error (e.g., ECE) on a temporally separated validation window, and review these metrics monthly alongside maintenance KPIs (failure counts, Down time hours, availability, OEE). Second, implement utilization-aware thresholds and playbooks: for each equipment class, define anomaly score cutoffs and response tiers that scale with duty cycle e.g., at utilization ≥0.75, an anomaly score that would trigger "monitor" at low duty should trigger "inspect within 24–48 h" with parts pre-staging; at utilization ≤0.45, allow longer observation with stricter re-alert rules. Third, adopt class-specific policies: prioritize rotating machinery for the tightest monitoring (lower thresholds, earlier inspection windows, and standing spares) because the anomaly→Down time slope is steepest; for utilities, emphasize calibration and scheduling integration so that alerts dovetail with production windows, reducing false urgency. Fourth, formalize a signal-to-action pipeline: require every alert to map to a standard operating procedure with a clearly documented diagnostic checklist, expected time-to-action, and closure codes that feed back into model learning; measure Alert→Work-Order conversion and alert cycle time as leading indicators. Fifth, institutionalize model governance: version models and thresholds, freeze validation datasets, and use a lightweight change-control board to approve updates; define retraining triggers (e.g., PR-AUC or calibration drift by more than a pre-set delta, or a rise in false positives per operating hour) and schedule quarterly drift reviews. Sixth, strengthen data quality and coding discipline: enforce CMMS codebooks, de-duplicate near-adjacent work orders, and reconcile MES/SCADA states with CMMS Down time to keep failure/Down time measures auditable; maintain sensor health dashboards (coverage %, timestamp monotonicity, unit consistency). Seventh, invest in interpretability and operator alignment: expose top contributing

channels/windows (e.g., saliency or variable-importance overlays), add short "why this alert" summaries to the UI, and run targeted enablement in cells where Trust_AI is low train crews on example cases, adjust thresholds to reduce nuisance alarms, and celebrate quick-win interventions to reinforce adoption. Eighth, close the loop with spares and scheduling: link high-risk alerts to parts reservations and micro-stoppage windows; pilot "pre-kit" carts for rotating assets at high utilization to compress mean time to repair. Ninth, operationalize continuous evaluation: publish a single scorecard each month that shows detector KPIs, perception KPIs (Trust, Interpretability, Actionability on a 5-point scale), and plant KPIs; annotate any parameter or threshold changes so leaders can attribute performance shifts. Tenth, use targeted experiments where feasible (e.g., staggered threshold changes across comparable lines) to estimate practical effect sizes before scaling. Finally, embed security, privacy, and ethics: protect asset identifiers, restrict who can see raw audio/vibration, and document how alerts are generated to sustain organizational trust. This package quality as a KPI, utilization-aware thresholds, class-specific playbooks, reliable data plumbing, human-centric interfaces, and disciplined governance maximizes the operational return of AI-enabled predictive maintenance while keeping the system robust, auditable, and easy to scale.

REFERENCES

[1]. Abdur Razzak, C., Golam Qibria, L., & Md Arifur, R. (2024). Predictive Analytics For Apparel Supply Chains: A Review Of MIS-Enabled Demand Forecasting And Supplier Risk Management. *American Journal of Interdisciplinary Studies*, *5*(04), 01–23. https://doi.org/10.63125/80dwy222

[2]. Antoni, J. (2006). The spectral kurtosis: A useful tool for characterising non-stationary signals. *Mechanical Systems and Signal Processing*, *20*(2), 282–307. https://doi.org/10.1016/j.ymssp.2004.09.001

[3]. Bousdekis, A., Magoutas, B., Apostolou, D., & Mentzas, G. (2019). Review, analysis and synthesis of prognostic-based decision support methods for condition based maintenance. *Sensors*, *19*(4), 969. https://doi.org/10.3390/s19040969

[4]. Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., Shinozuka, M., Tierney, K., Wallace, W. A., & von Winterfeldt, D. (2003). A framework to quantitatively assess and enhance the seismic resilience of communities. *Earthquake Spectra*, *19*(4), 733–752. https://doi.org/10.1193/1.1623497

[5]. Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference (2nd ed.)*. Springer. https://doi.org/10.1007/b97636

[6]. Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. D. P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, *137*, 106024. https://doi.org/10.1016/j.cie.2019.106024

[7]. Chiang, L. H., Colegrove, C., & Qin, S. J. (2000). Fault detection in process manufacturing using multivariate statistical methods. *Journal of Process Control*, *10*(5), 449–465. https://doi.org/10.1016/s0959-1524(99)00055-0

[8]. de Jonge, B., & Scarf, P. A. (2020). A review on maintenance optimization. *European Journal of Operational Research*, *285*(3), 805–824. https://doi.org/10.1016/j.ejor.2019.09.047

[9]. Fawaz, H. I., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D. F., Weber, J., Webb, G. I., Idoumghar, L., Muller, P.-A., & Petitjean, F. (2019). InceptionTime: Finding AlexNet for time series classification. *Data Mining and Knowledge Discovery*, *34*(6), 1936–1962. https://doi.org/10.1007/s10618-019-00619-1

[10]. Gao, R. X., & Wang, L. (2020). Smart manufacturing—A perspective. *Journal of Manufacturing Systems*, *54*, 258–268. https://doi.org/10.1016/j.jmsy.2020.02.002

[11]. Grall, A., Dieulle, L., Bérenguer, C., & Roussignol, M. (2002). Continuous-time predictive-maintenance scheduling for a deteriorating system. *IEEE Transactions on Reliability*, *51*(2), 141–150. https://doi.org/10.1109/tr.2002.1011518

[12]. Istiaque, M., Dipon Das, R., Hasan, A., Samia, A., & Sayer Bin, S. (2023). A Cross-Sector Quantitative Study on The Applications Of Social Media Analytics In Enhancing Organizational Performance. *American Journal of Scholarly Research and Innovation*, *2*(02), 274-302. https://doi.org/10.63125/d8ree044

[13]. Istiaque, M., Dipon Das, R., Hasan, A., Samia, A., & Sayer Bin, S. (2024). Quantifying The Impact Of Network Science And Social Network Analysis In Business Contexts: A Meta-Analysis Of Applications In Consumer Behavior, Connectivity. *International Journal of Scientific Interdisciplinary Research*, *5*(2), 58-89. https://doi.org/10.63125/vgkwe938

[14]. Jahid, M. K. A. S. R. (2022). Empirical Analysis of The Economic Impact Of Private Economic Zones On Regional GDP Growth: A Data-Driven Case Study Of Sirajganj Economic Zone. *American Journal of Scholarly Research and Innovation*, *1*(02), 01-29. https://doi.org/10.63125/je9w1c40

[15]. Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, *20*(7), 1483–1510. https://doi.org/10.1016/j.ymssp.2005.10.017

[16]. Khodabakhsh, A., & Ashory, M. R. (2019). A review of signal processing techniques in rotating machinery fault diagnosis. *Journal of Vibration and Control*, *25*(5), 875–895. https://doi.org/10.1177/1077546318792802

[17]. Lee, J., Bagheri, B., & Kao, H.-A. (2015). A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing Letters*, *3*, 18–23. https://doi.org/10.1016/j.mfglet.2014.12.001

[18]. Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2016). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, *138*, 106587. https://doi.org/10.1016/j.ymssp.2019.106587

[19]. Li, X., Ding, Q., & Sun, J.-Q. (2017). Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety*, *172*, 1–11. https://doi.org/10.1016/j.ress.2017.11.021

[20]. Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*(4), 1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

[21]. Md Arifur, R., & Sheratun Noor, J. (2022). A Systematic Literature Review of User-Centric Design In Digital Business Systems: Enhancing Accessibility, Adoption, And Organizational Impact. *Review of Applied Science and Technology*, *1*(04), 01-25. https://doi.org/10.63125/ndjkpm77

[22]. Md Ashiqur, R., Md Hasan, Z., & Afrin Binta, H. (2025). A meta-analysis of ERP and CRM integration tools in business process optimization. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 278-312. https://doi.org/10.63125/yah70173

[23]. Md Hasan, Z. (2025). AI-Driven business analytics for financial forecasting: a systematic review of decision support models in SMES. *Review of Applied Science and Technology*, *4*(02), 86-117. https://doi.org/10.63125/gjrpv442

[24]. Md Hasan, Z., Mohammad, M., & Md Nur Hasan, M. (2024). Business Intelligence Systems In Finance And Accounting: A Review Of Real-Time Dashboarding Using Power BI & Tableau. *American Journal of Scholarly Research and Innovation*, *3*(02), 52-79. https://doi.org/10.63125/fy4w7w04

[25]. Md Hasan, Z., & Moin Uddin, M. (2022). Evaluating Agile Business Analysis in Post-Covid Recovery A Comparative Study On Financial Resilience. *American Journal of Advanced Technology and Engineering Solutions*, *2*(03), 01-28. https://doi.org/10.63125/6nee1m28

[26]. Md Hasan, Z., Sheratun Noor, J., & Md. Zafor, I. (2023). Strategic role of business analysts in digital transformation tools, roles, and enterprise outcomes. *American Journal of Scholarly Research and Innovation*, *2*(02), 246-273. https://doi.org/10.63125/rc45z918

[27]. Md Ismail, H., Md Mahfuj, H., Mohammad Aman Ullah, S., & Shofiul Azam, T. (2025). IMPLEMENTING ADVANCED TECHNOLOGIES FOR ENHANCED CONSTRUCTION SITE SAFETY. *American Journal of Advanced Technology and Engineering Solutions*, *1*(02), 01-31. https://doi.org/10.63125/3v8rpr04

[28]. Md Ismail Hossain, M. A. B., amp, & Mousumi Akter, S. (2023). Water Quality Modelling and Assessment Of The Buriganga River Using Qual2k. *Global Mainstream Journal of Innovation, Engineering & Emerging Technology*, *2*(03), 01-11. https://doi.org/10.62304/jieet.v2i03.64

[29]. Md Mahamudur Rahaman, S. (2022). Electrical And Mechanical Troubleshooting in Medical And Diagnostic Device Manufacturing: A Systematic Review Of Industry Safety And Performance Protocols. *American Journal of Scholarly Research and Innovation*, *1*(01), 295-318. https://doi.org/10.63125/d68y3590

[30]. Md Mahamudur Rahaman, S., & Rezwanul Ashraf, R. (2022). Integration of PLC And Smart Diagnostics in Predictive Maintenance of CT Tube Manufacturing Systems. *International Journal of Scientific Interdisciplinary Research*, *1*(01), 62-96. https://doi.org/10.63125/gspb0f75

[31]. Md Nazrul Islam, K. (2022). A Systematic Review of Legal Technology Adoption In Contract Management, Data Governance, And Compliance Monitoring. *American Journal of Interdisciplinary Studies*, *3*(01), 01-30. https://doi.org/10.63125/caangg06

[32]. Md Nur Hasan, M., Md Musfiqur, R., & Debashish, G. (2022). Strategic Decision-Making in Digital Retail Supply Chains: Harnessing AI-Driven Business Intelligence From Customer Data. *Review of Applied Science and Technology*, *1*(03), 01-31. https://doi.org/10.63125/6a7rpy62

[33]. Md Redwanul, I., & Md. Zafor, I. (2022). Impact of Predictive Data Modeling on Business Decision-Making: A Review Of Studies Across Retail, Finance, And Logistics. *American Journal of Advanced Technology and Engineering Solutions*, *2*(02), 33-62. https://doi.org/10.63125/8hfbkt70

[34]. Md Rezaul, K., & Md Mesbaul, H. (2022). Innovative Textile Recycling and Upcycling Technologies For Circular Fashion: Reducing Landfill Waste And Enhancing Environmental Sustainability. *American Journal of Interdisciplinary Studies*, *3*(03), 01-35. https://doi.org/10.63125/kkmerg16

[35]. Md Sultan, M., Proches Nolasco, M., & Md. Torikul, I. (2023). Multi-Material Additive Manufacturing For Integrated Electromechanical Systems. *American Journal of Interdisciplinary Studies*, *4*(04), 52-79. https://doi.org/10.63125/y2ybrx17

[36]. Md Sultan, M., Proches Nolasco, M., & Vicent Opiyo, N. (2025). A Comprehensive Analysis Of Non-Planar Toolpath Optimization In Multi-Axis 3D Printing: Evaluating The Efficiency Of Curved Layer Slicing Strategies. *Review of Applied Science and Technology*, *4*(02), 274-308. https://doi.org/10.63125/5fdxa722

[37]. Md Takbir Hossen, S., Ishtiaque, A., & Md Atiqur, R. (2023). AI-Based Smart Textile Wearables For Remote Health Surveillance And Critical Emergency Alerts: A Systematic Literature Review. *American Journal of Scholarly Research and Innovation*, *2*(02), 1-29. https://doi.org/10.63125/ceqapd08

[38]. Md Tawfiqul, I. (2023). A Quantitative Assessment Of Secure Neural Network Architectures For Fault Detection In Industrial Control Systems. *Review of Applied Science and Technology*, *2*(04), 01-24. https://doi.org/10.63125/3m7gbs97

[39]. Md. Sakib Hasan, H. (2022). Quantitative Risk Assessment of Rail Infrastructure Projects Using Monte Carlo Simulation And Fuzzy Logic. *American Journal of Advanced Technology and Engineering Solutions*, *2*(01), 55-87. https://doi.org/10.63125/h24n6z92

[40]. Md. Tarek, H. (2022). Graph Neural Network Models For Detecting Fraudulent Insurance Claims In Healthcare Systems. *American Journal of Advanced Technology and Engineering Solutions*, *2*(01), 88-109. https://doi.org/10.63125/r5vsmv21

[41]. Md.Kamrul, K., & Md Omar, F. (2022). Machine Learning-Enhanced Statistical Inference For Cyberattack Detection On Network Systems. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 65-90. https://doi.org/10.63125/sw7jzx60

[42]. Md.Kamrul, K., & Md. Tarek, H. (2022). A Poisson Regression Approach to Modeling Traffic Accident Frequency in Urban Areas. *American Journal of Interdisciplinary Studies*, *3*(04), 117-156. https://doi.org/10.63125/wqh7pd07

[43]. Mubashir, I., & Abdul, R. (2022). Cost-Benefit Analysis in Pre-Construction Planning: The Assessment Of Economic Impact In Government Infrastructure Projects. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 91-122. https://doi.org/10.63125/kjwd5e33

[44]. Muchiri, P., & Pintelon, L. (2008). Performance measurement using overall equipment effectiveness (OEE): Literature review and practical application discussion. *International Journal of Production Research*, *46*(13), 3517–3535. https://doi.org/10.1080/00207540601142645

[45]. Ng Corrales, L. D., Lambán, M. P., Hernandez Korner, M. E., & Royo, J. (2020). Overall equipment effectiveness: Systematic literature review and overview of different approaches. *Applied Sciences*, *10*(18), 6469. https://doi.org/10.3390/app10186469

[46]. Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. Proceedings of the 22nd International Conference on Machine Learning,

[47]. Omar Muhammad, F., & Md.Kamrul, K. (2022). Blockchain-Enabled BI For HR And Payroll Systems: Securing Sensitive Workforce Data. *American Journal of Scholarly Research and Innovation*, *1*(02), 30-58. https://doi.org/10.63125/et4bhy15

[48]. Pham, H., & Wang, H. (1996). Imperfect maintenance. *European Journal of Operational Research*, *94*(3), 425–438. https://doi.org/10.1016/s0377-2217(96)00099-9

[49]. Piacentini, M., Lucchi, M., & Pedrazzoli, P. (2019). The (α, β)-precise estimates of MTBF and MTTR: Definitions and practical considerations. *Procedia Manufacturing*, *38*, 98–105. https://doi.org/10.1016/j.promfg.2019.02.013

[50]. Qin, S. J. (2012). Survey on data-driven industrial process monitoring and diagnosis: A perspective from the process history. *IFAC Proceedings Volumes*, *45*(20), 115–126. https://doi.org/10.3182/20120912-3-br-2026.00216

[51]. Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics—A tutorial. *Mechanical Systems and Signal Processing*, *25*(2), 485–520. https://doi.org/10.1016/j.ymssp.2010.07.016

[52]. Rausand, M., & Høyland, A. (2004). *System reliability theory: Models, statistical methods, and applications (2nd ed.)*. Wiley. https://doi.org/10.1002/0471471332

[53]. Reduanul, H., & Mohammad Shoeb, A. (2022). Advancing AI in Marketing Through Cross Border Integration Ethical Considerations And Policy Implications. *American Journal of Scholarly Research and Innovation*, *1*(01), 351-379. https://doi.org/10.63125/d1xg3784

[54]. Sabuj Kumar, S., & Zobayer, E. (2022). Comparative Analysis of Petroleum Infrastructure Projects In South Asia And The Us Using Advanced Gas Turbine Engine Technologies For Cross Integration. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 123-147. https://doi.org/10.63125/wr93s247

[55]. Sadia, T., & Shaiful, M. (2022). In Silico Evaluation of Phytochemicals From Mangifera Indica Against Type 2 Diabetes Targets: A Molecular Docking And Admet Study. *American Journal of Interdisciplinary Studies*, *3*(04), 91-116. https://doi.org/10.63125/anaf6b94

[56]. 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

[57]. Sanjai, V., Sanath Kumar, C., Maniruzzaman, B., & Farhana Zaman, R. (2023). Integrating Artificial Intelligence in Strategic Business Decision-Making: A Systematic Review Of Predictive Models. *International Journal of Scientific Interdisciplinary Research*, *4*(1), 01-26. https://doi.org/10.63125/s5skge53

[58]. Sanjai, V., Sanath Kumar, C., Sadia, Z., & Rony, S. (2025). AI And Quantum Computing For Carbon-Neutral Supply Chains: A Systematic Review Of Innovations. *American Journal of Interdisciplinary Studies*, *6*(1), 40-75. https://doi.org/10.63125/nrdx7d32

[59]. Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). *Damage propagation modeling for aircraft engine run-to-failure simulation* International Conference on Prognostics and Health Management,

[60]. Serradilla, O., Zugasti, E., Rodríguez, J., & Zurutuza, U. (2022). Deep learning models for predictive maintenance: A survey, comparison, challenges and prospects. *Applied Intelligence*, *52*, 10934–10964. https://doi.org/10.1007/s10489-021-03004-y

[61]. Sheratun Noor, J., & Momena, A. (2022). Assessment Of Data-Driven Vendor Performance Evaluation in Retail Supply Chains: Analyzing Metrics, Scorecards, And Contract Management Tools. *American Journal of Interdisciplinary Studies*, *3*(02), 36-61. https://doi.org/10.63125/0s7t1y90

[62]. Sikorska, J. Z., Hodkiewicz, M., & Ma, L. (2011). Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing*, *25*(5), 1803–1836. https://doi.org/10.1016/j.ymssp.2010.11.018

[63]. Sipos, R., Fradkin, D., Moerchen, F., & Wang, Z. (2014). Log-based predictive maintenance in the railway industry. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,

[64]. Standardization, I. O. f. (2015). *Condition monitoring and diagnostics of machine systems – Data processing, communication and presentation – Part 4: Presentation*.

[65]. Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2015). Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics*, *11*(3), 812–820. https://doi.org/10.1109/tii.2014.2349359

[66]. Tahmina Akter, R., Debashish, G., Md Soyeb, R., & Abdullah Al, M. (2023). A Systematic Review of AI-Enhanced Decision Support Tools in Information Systems: Strategic Applications In Service-Oriented Enterprises And Enterprise Planning. *Review of Applied Science and Technology*, *2*(01), 26-52. https://doi.org/10.63125/73djw422

[67]. Tchakoua, P., Wamkeue, R., Tameghe, T. A., Ekemb, G., Fankam, P., Djongyang, N., & Kenfack, F. (2014). Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges. *Energies*, *7*(4), 2595–2630. https://doi.org/10.3390/en7042595

[68]. Venkatasubramanian, V., Rengaswamy, R., Yin, K., & Kavuri, S. N. (2003). A review of process fault detection and diagnosis: Part I—III. *Computers & Chemical Engineering*, *27*(3), 293–346. https://doi.org/10.1016/s0098-1354(02)00160-6

[69]. Vogl, G. W., Weiss, B. A., & Helu, M. (2016). A review of diagnostic and prognostic capabilities and best practices for manufacturing. *Journal of Intelligent Manufacturing*, *30*(1), 79–95. https://doi.org/10.1007/s10845-016-1228-8

[70]. Wang, Z., Yan, W., & Oates, T. (2017). Time series classification from scratch with deep neural networks: A strong baseline. *Data Mining and Knowledge Discovery*, *31*(4), 935–969. https://doi.org/10.1007/s10618-016-0483-9

[71]. Wen, L., Li, X., & Gao, L. (2017). A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Transactions on Industrial Electronics*, *65*(7), 5990–5998. https://doi.org/10.1109/tie.2017.2774777

[72]. Widodo, A., & Yang, B.-S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, *21*(6), 2560–2574. https://doi.org/10.1016/j.ymssp.2006.12.007

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

[74]. Zhang, W., Li, X., Ding, Q., & Sun, J.-Q. (2019). A review of deep learning in machine fault diagnosis. *Mechanical Systems and Signal Processing*, *107*, 106–135. https://doi.org/10.1016/j.ymssp.2017.12.013

[75]. Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, *115*, 213–237. https://doi.org/10.1016/j.ymssp.2018.05.050

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