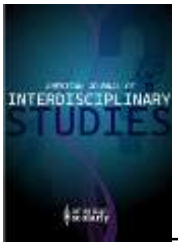


American Journal of Interdisciplinary Studies
Volume 02, Issue 02 (2021)
Page No: 01 – 31
eISSN: 3067-5146
Doi: 10.63125/222nwg58

FEDERATED LEARNING ARCHITECTURES FOR PREDICTIVE QUALITY CONTROL IN DISTRIBUTED MANUFACTURING SYSTEMS

Md Sanjid Khan[1]; Md. Tahmid Farabe Shehun[2];

[1]. Bachelor of Civil Engineering, Chongqing University of Science and Technology, Chongqing, China; Email: khansanjid9@gmail.com

[2]. Bachelor Of Science In Apparel Manufacturing & Technology; BGMEA University of Fashion & Technology, Bangladesh; Email: mdtahmidfarabeshehun@gmail.com

ABSTRACT

This study investigates how federated learning (FL) architectures influence predictive quality control (PQC) performance in distributed manufacturing environments characterized by heterogeneity, privacy constraints, and high data velocity. Predictive quality control leverages machine learning to forecast process deviations before defects occur; however, in globally distributed production networks, data-sharing restrictions and non-identically distributed (non-IID) data complicate centralized model training. To address these challenges, this research develops and empirically tests a quantitative, cross-sectional, multi–case framework linking FL architectural design, hub-and-spoke, hierarchical, and peer configurations—with PQC performance across IIoT-enabled plants. Constructs including infrastructure readiness, communication efficiency/update cadence, data heterogeneity, and privacy/trust governance were operationalized using validated Likert-scale instruments complemented by objective indicators such as AUC, F1, false-alarm rate, and time-to-detection. Regression and mixed-effects analyses reveal that infrastructure readiness and communication efficiency exhibit strong positive associations with PQC outcomes, whereas cross-site heterogeneity negatively affects performance. Crucially, hierarchical FL architectures moderate these relationships, attenuating the detrimental effects of heterogeneity and amplifying the gains from efficient communication. Privacy and trust governance correlate positively, though modestly, with PQC, underscoring that robust security and compliance practices enhance rather than hinder collaborative learning effectiveness. The findings establish that architecture is not merely an IT topology but a determinant of statistical and operational performance, transforming federated updates into a controllable mechanism for cross-plant quality intelligence. By integrating IIoT, edge computing, and privacy-preserving analytics within a measurable empirical model, this research advances both theoretical understanding and practical implementation of FL-enabled PQC. It offers an actionable blueprint for manufacturers: invest in edge readiness and orchestration, enforce cadence service levels, adopt hierarchical clustering for heterogeneous sites, and embed privacy governance into federation lifecycles. Collectively, the study demonstrates that disciplined technical, organizational, and architectural alignment enables distributed manufacturing systems to achieve predictive quality improvements without compromising data confidentiality.

KEYWORDS

Federated Learning, Predictive Quality Control, Distributed Manufacturing, Edge Computing, Data Heterogeneity

Citation:
Khan, M. S., & Shehun, M. T. F. (2021). Federated learning architectures for predictive quality control in distributed manufacturing systems. American Journal of Interdisciplinary Studies, 2(2), 1–31. https://doi.org/10.63125/222nwg58

Received:
April 17, 2021

Revised:
May 20, 2021

Accepted:
June 16, 2021

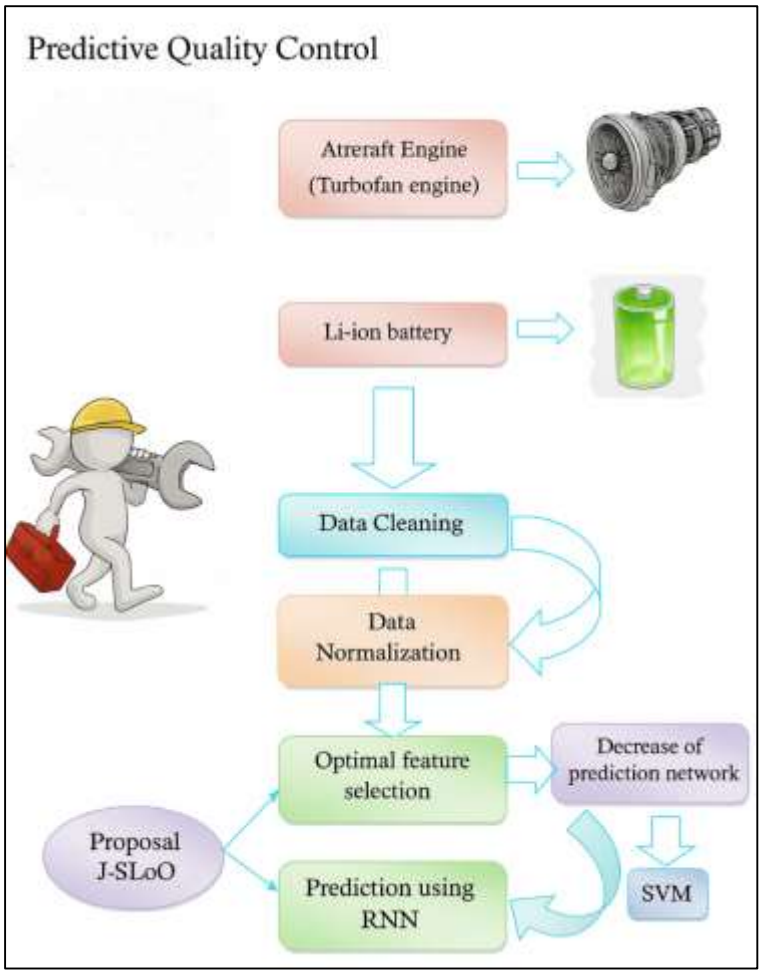
Published:
July 28, 2021

American Journal of Interdisciplinary Studies
Volume 02, Issue 02 (2021)
Page No: 01 – 31
eISSN: 3067-5146
DOI: 10.63125/222nwg58

INTRODUCTION

Predictive quality control (PQC) refers to the use of statistical and machine learning techniques to anticipate product or process nonconformities before they manifest as defects, enabling proactive intervention in production settings characterized by variability, heterogeneity, and tight performance tolerances. In the context of distributed manufacturing where geographically dispersed plants, contract manufacturers, and suppliers contribute to a shared value chain PQC must operate across organizational and jurisdictional boundaries while preserving data confidentiality, integrity, and availability (Kang et al., 2016; Lee et al., 2013). Convergence between the Industrial Internet of Things (IIoT), cyber-physical systems, and cloud/edge infrastructures has accelerated sensorization and high-frequency data acquisition, creating a rich substrate for PQC models that learn from streams of measurements, events, and contextual metadata (Shi et al., 2016).

Figure 1: Predictive Quality Control Workflow in Distributed Manufacturing



International supply networks add additional layers of complexity data governance regimes, sectoral standards, and contractual arrangements requiring architectures that reconcile performance objectives with privacy, security, and compliance (Abadi et al., 2016; Shokri et al., 2017). Federated learning (FL) has emerged as a candidate paradigm for this reconciliation, enabling multiple parties to collaboratively train models without sharing raw data by exchanging updates or gradients under secure protocols (Yang et al., 2019). Within internationally distributed manufacturing ecosystems, FL promises to reduce the data-sharing burden across borders while maintaining the statistical benefits of broader training cohorts an issue of particular relevance where competitive sensitivities and regulatory constraints limit centralization (Lu, 2017; Choudhary, Harding, & Tiwari, 2009). PQC, when paired with FL and edge computing, can be conceptualized as a multi-site analytical workflow: local feature engineering and model updates occur near machines or production cells, a secure aggregation service fuses updates, and periodic model refreshes

propagate to sites for online inference (Shi et al., 2016; Shokri & Shmatikov, 2015). This paper situates FL-enabled PQC at the intersection of smart manufacturing, privacy-preserving analytics, and distributed operations, synthesizing definitional clarity and international scope while foregrounding measurement and empirical analysis (Kang et al., 2016; Lee et al., 2015). Statistical process control (SPC) provided the foundational tools for monitoring and maintaining process stability; however, modern manufacturing environments generate multi-modal, high-velocity data that motivate algorithms able to capture nonlinearity, interactions, and non-stationarity (Qin, 2012; Qin & Chiang, 2019). Machine learning approaches, including tree ensembles and kernel methods, have demonstrated value in anomaly detection, condition monitoring, and defect prediction, translating to lower false alarm rates and improved yield when data pipelines are reliable and well-governed (Abdul, 2021; Wuest et al., 2016). Distributed manufacturing heightens the demand for data-driven PQC because comparable failure modes may manifest differently across plants, suppliers, or product variants; this diversity offers both a challenge (non-IID data) and an opportunity (richer hypotheses about causal mechanisms) (Wang, 2018). Integrating PQC across sites often confronts the practical restrictions of bandwidth, confidentiality, and intellectual property, which complicate centralized model training or centralized data lakes (Lu, 2017; Rony, 2021). IIoT infrastructure and edge computing broaden architectural options by pushing analytics closer to machines and enabling on-site training or inference with periodic coordination across a federation (Li et al., 2020). As organizations explore PQC under these conditions, rigor in measurement both perceptual (e.g., practitioner-reported performance) and objective (e.g., area under the ROC curve, false alarm rates, and detection latency) becomes necessary to compare approaches (Satyanarayanan, 2017). The international dimension includes cross-border flows of components and knowledge, making harmonized analytics attractive while raising questions about privacy guarantees, security assurances, and legal compliance (Sila & Ebrahimpour, 2005). The definitional lineage from SPC to PQC in distributed settings thus motivates empirical study of architectures that respect data locality and control over sensitive production data while attempting to sustain or improve predictive power (Widodo & Yang, 2007).

Federated learning coordinates training across clients by sharing model updates to a server for aggregation, preserving the locality of raw data; this basic mechanism is extended by variants that handle heterogeneous data distributions, client intermittency, and device or site constraints (Yang et al., 2019; Kairouz et al., 2019). In industrial deployments, the "clients" correspond to plants, lines, or even machines, each with distinct data distributions shaped by equipment, materials, and operator practices, thereby challenging assumptions of identically distributed data . Classical Federated Averaging (FedAvg) reduces communication by performing multiple local steps between aggregations, but performance can degrade under strong non-IID conditions, encouraging methods that correct client drift or reweight updates (Zhang et al., 2020). Secure aggregation and differential privacy form complementary pillars for confidentiality: secure aggregation ensures the server only observes the sum of encrypted updates, while differential privacy bounds the information that a single site's data can reveal through its contribution to the model (Bonawitz et al., 2017). In manufacturing, where data include proprietary process parameters or yield-sensitive metrics, these mechanisms address concerns of leakage or competitive exposure (Dwork, 2006). Edge computing situates training near data sources, reducing backhaul costs and aligning update cycles with production schedules (Shi et al., 2016). The coordination of PQC across an international network can thus be conceptualized as a cross-silo FL problem in which a moderate number of well-resourced clients (plants) participate in scheduled rounds, with governance defining participation criteria, validation protocols, and fallbacks (Yang et al., 2019). Under this view, the architecture itself centralized hub-and-spoke, hierarchical (multi-tier), or peer-to-peer becomes a measurable design choice whose associations with PQC performance warrant empirical study in real settings across sectors and regions (Kairouz et al., 2019).

Distributed manufacturing introduces persistent heterogeneity across sites differences in sensor types, sampling rates, product mixes, environmental conditions, and maintenance histories producing non-IID data that stress federated optimization. Communication constraints are equally salient: production networks often operate with constrained, segmented, or policy-restricted links; synchronizing updates must respect windows that avoid interference with operations and maintenance (Geyer et al., 2017). Infrastructure readiness incorporates compute availability at the edge (e.g., GPUs or accelerators for local training), storage, network reliability, and orchestration

capabilities for model deployment and rollback, all of which mediate the feasibility and performance of FL-enabled PQC (Choudhary et al., 2009). From a privacy and security standpoint, organizations must address attack surfaces such as gradient leakage or membership inference, reinforcing the role of secure aggregation, clipping, noise addition, and monitoring (Jardine et al., 2006). Within PQC, outcome definitions improvements in predictive accuracy, reduction of false alarms that distract operators, and earlier detection of incipient faults can be operationalized via standardized metrics that are comparable across sites, enabling valid cross-case analysis (Acar et al., 2018). The international layer adds variance in data protection obligations and contractual data-use limitations, influencing the parameterization of privacy budgets and cryptographic protocols (Li et al., 2015). Given these conditions, empirical work benefits from explicit measurement of heterogeneity (e.g., practitioners' assessments and distributional diagnostics), communication efficiency (e.g., update duration and cadence), and infrastructure readiness (e.g., perceived sufficiency of edge compute), alongside objective PQC indicators where feasible (Lu, 2017).

Beyond algorithms and infrastructure, organizational factors shape the viability of PQC under FL. Governance policies determine who participates, how models are validated, and which controls gate promotion to production; training and workforce skills determine the capacity to interpret outputs, calibrate thresholds, and maintain data pipelines (Kang et al., 2016; Lee et al., 2013). Distributed manufacturing requires coordination among quality engineers, data scientists, and IT/OT personnel across sites, with role differentiation yet shared accountability for model performance and process safety (Wuest et al., 2016). Data stewardship establishes standards for labeling, versioning, and lineage, ensuring that PQC models train on curated, traceable data even when kept local (Wang, 2018). Security governance defines acceptable residual risks for privacy-enhanced analytics, including DP parameters and encryption budgets, with monitoring for drift or anomalous update patterns that could signal compromised clients (Abadi et al., 2016; Bonawitz et al., 2017). International networks add layers of export controls or sectoral rules, underscoring the value of architectures that minimize exposure of sensitive telemetry while maintaining the benefits of collective learning (Lu, 2017). In practice, organizational support leadership sponsorship, resources for edge compute, and cross-functional training co-determines whether PQC models remain pilots or scale across plants (Kang et al., 2016). Measurement instruments that capture these organizational dimensions alongside technical ones enable systematic assessment across cases and provide the basis for hypothesis testing using descriptive statistics, correlations, and regression models of PQC performance indicators.

Within this study, measurable constructs span technical, organizational, and contextual domains: infrastructure readiness (compute, bandwidth, reliability), communication efficiency (update time windows and cadence), data heterogeneity (distributional divergence across sites), privacy/trust (confidence in confidentiality protections and inter-organizational trust), and architecture selection (hub-and-spoke, hierarchical, or peer networks). PQC performance can be measured perceptually improvements in defect-prediction accuracy and operational usefulness and with objective indicators where available, such as AUC/F1, false alarm rate, and time-to-detection . The non-IID condition is explicitly modeled as a predictor and moderator; for example, architecture choices may interact with heterogeneity, as hierarchical aggregation can reduce drift by clustering similar clients (Shokri & Shmatikov, 2015). Privacy and security are not ancillary; differential privacy and secure aggregation parameters influence optimization and may trade off with utility, calling for their inclusion as observed practices in the measurement instrument (Abadi et al., 2016; Acar et al., 2018). Edge computing enables local training and pre-processing, connecting to communication efficiency by shaping round durations and update reliability (Shi et al., 2016; Shokri & Shmatikov, 2015). Contextual controls such as plant size, product complexity, and industry segment capture structural differences that can confound observed associations (Lee et al., 2015). Collectively, these constructs define a conceptual model for PQC performance under FL that is suitable for quantitative, cross-sectional, multi-case analysis with regression modeling and moderation tests, aligning with the operational realities of distributed manufacturing (Kang et al., 2016).

Given the definitional grounding in PQC and FL, the international scope of distributed manufacturing, and the centrality of privacy-preserving collaboration, this study focuses on empirical associations between architectural/technical/organizational conditions and PQC performance across multiple manufacturing sites. The design captures perceptions and, where feasible, objective indicators of model performance, while representing heterogeneity in equipment, processes, and products. The

measurement framework recognizes constructs that are actionable in practice: infrastructure readiness, communication efficiency, data heterogeneity, and privacy/trust, with architecture choice treated as a categorical moderator and controls for site/context variance (Kairouz et al., 2019; Qin, 2012). The literature reviewed above motivates the need for such an empirical assessment by outlining the technical basis for FL (optimization under non-IID, secure protocols), the practicalities of edge-enabled analytics, and the organizational scaffolding necessary for sustained PQC (Widodo & Yang, 2007; Yang et al., 2019). By structuring the inquiry as a cross-sectional, multi-case study with validated scales and regression analysis, the study situates PQC performance within a coherent set of determinants that are observable and comparable across sites and jurisdictions. The international relevance is embedded in the federated setup data remain at source, model improvement flows through updates reflecting contemporary practices in globalized production networks where collaboration and confidentiality must coexist (Lu, 2017; Satyanarayanan, 2017). This framing provides the basis for formal research questions and hypotheses that anchor subsequent sections on methodology and results, maintaining alignment with established standards in industrial analytics and privacy-preserving machine learning (Qin, 2012).

This study pursues a clear set of objectives that anchor the investigation of federated learning for predictive quality control in distributed manufacturing. The primary objective is to quantify the association between architecture choices in federated learning such as hub-and-spoke, hierarchical, and peer configurations and measurable predictive quality control performance across multiple manufacturing sites. A second objective is to operationalize and measure key technical and organizational determinants that plausibly shape outcomes, including infrastructure readiness at the edge, communication efficiency and update cadence, the degree of inter-site data heterogeneity, and the presence of privacy and trust safeguards. A third objective is to develop and validate a survey-based instrument, aligned with a five-point Likert scale, that captures these determinants and outcomes in a manner suitable for cross-site comparison, while complementing perceptual indicators with available objective logs such as classification accuracy, false-alarm rates, and detection latency. A fourth objective is to estimate a set of regression models that test directional hypotheses regarding the positive associations between readiness, communication efficiency, privacy safeguards, and predictive quality control, and to examine moderation by architecture choice and heterogeneity. A fifth objective is to establish the reliability and validity of all multi-item constructs through internal consistency metrics and convergent–discriminant assessments, thereby ensuring that inferences rest on stable measurements. A sixth objective is to perform robustness and sensitivity checks alternative outcome definitions, exclusion of influential observations, and specification variants to assess the stability of estimated effects. A seventh objective is to produce a reproducible analysis workflow that documents data screening, descriptive statistics, correlation structures, multicollinearity diagnostics, and model estimation steps, allowing transparent cross-case synthesis. An eighth objective is to characterize site contexts with standardized profiles so that differences in plant size, product complexity, and segment are represented as controls, enhancing comparability across cases. Collectively, these objectives define a quantitative, cross-sectional, multi–case research program that isolates architecture-related effects, clarifies the roles of technical and organizational conditions, and delivers a coherent empirical account of predictive quality control performance under federated learning.

LITERATURE REVIEW

The literature on predictive quality control (PQC) in distributed manufacturing spans three converging streams that set the stage for this study: data-driven quality analytics, industrial edge/IIoT infrastructures, and privacy-preserving collaborative learning. First, research on quality analytics has evolved from classical statistical process control toward predictive methods that learn patterns of incipient faults from high-frequency, multimodal signals. This shift emphasizes operational metrics classification accuracy, false-alarm rate, and time-to-detection while highlighting challenges such as class imbalance, drifting processes, and heterogeneous sensing across product families and plants. Second, the industrial computing stack has moved computation closer to machines via edge nodes integrated with MES/SCADA systems, enabling local feature engineering, on-device training or fine-tuning, and low-latency inference aligned with production cycles. Edge-centric designs reduce backhaul traffic and support resilience, but they also fragment data and complicate centralized model governance. Third, privacy-preserving learning most notably federated learning (FL) offers a coordination mechanism to train global models without relocating raw data, aligning

with confidentiality constraints, supplier relationships, and regulatory boundaries typical of international supply networks. Within FL, architectural choices (hub-and-spoke, hierarchical, peer) and optimization strategies (client selection, update cadence, aggregation variants) determine how non-identically distributed data and intermittent connectivity are negotiated, while security controls (e.g., secure aggregation, differential-privacy noise) address information leakage risks. The manufacturing-specific literature underscores that technical factors (infrastructure readiness, communication efficiency, non-IID severity) intertwine with organizational conditions (governance, skills, data stewardship), shaping whether PQC models move beyond pilots to sustained operation across sites. At the same time, measurement practices vary widely: many studies report single-site case results or simulation benchmarks, whereas fewer works provide cross-site, construct-valid scales that can support statistical comparison of architectures and enabling conditions. This review synthesizes these strands to develop a coherent conceptual model for FL-enabled PQC in distributed manufacturing, clarifying core constructs, observable indicators, and hypothesized relationships. It also surfaces methodological considerations central to empirical assessment operational definitions of PQC performance, strategies for characterizing heterogeneity, and the role of edge–cloud partitioning and privacy safeguards thereby providing the analytical groundwork for the study's quantitative, cross-sectional, multi–case design.

Federated Learning Architectures in Distributed Manufacturing

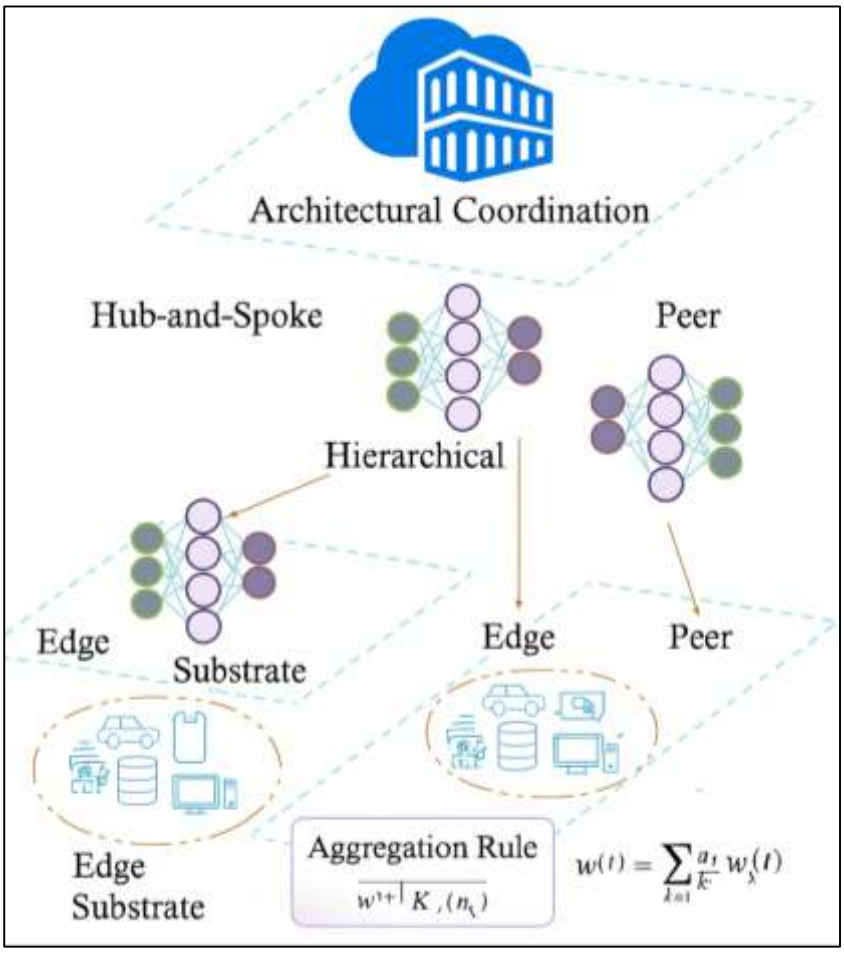
Federated learning (FL) for predictive quality control (PQC) sits at the intersection of architectural coordination and edge-centric computation. In distributed manufacturing, plants, lines, and suppliers function as cross-silo clients that must synchronize model improvements without moving raw telemetry. This requirement makes topology a first-class design choice. Hub-and-spoke architectures minimize orchestration complexity by routing client updates to a single aggregator, while hierarchical (multi-tier) designs insert intermediate aggregators that cluster similar sites, and peer-oriented variants enable limited lateral exchange to stabilize non-identically distributed (non-IID) behavior. The computational substrate lives near machines: edge nodes perform feature engineering and local training rounds, aligned with maintenance windows and takt-time constraints, then transmit updates that reflect local experience. A canonical aggregation rule captures the essence of these architectures: $w(t+1) = \sum_{k=1}^{K} \sum_{j=1}^{K} \frac{n_j}{n_k} w_k(t)]$ is site k's model after local steps and n_k is the site's sample count; hierarchical topologies often apply this rule first within clusters and then at the root. The edge–fog–cloud continuum is essential here: fog computing places compute, storage, and networking between devices and distant clouds, enabling low-latency control, localized robustness, and bandwidth savings that make frequent rounds feasible for industrial data rates (Bonomi et al., 2012). Since PQC models must remain reliable amid shifting materials, tools, and environmental conditions, the architecture must absorb concept drift by supporting timely refreshes and by allowing cluster-specific adaptation paths that reflect diverging site dynamics (Gama et al., 2014). Together, these choices shape how quickly a federation incorporates new signals, how it amortizes communication overheads across sites, and how it balances global consistency with local fit.

Architectural performance in PQC is also mediated by the statistical texture of industrial data. Defect and fault events are often rare relative to the length of nominal operation, yielding class imbalance that interacts with both topology and update cadence. When local minority classes are extremely sparse, a hub-and-spoke federation may overweight sites with more frequent failures, while a hierarchical scheme can pool minority evidence within similar clusters before participating in global averaging. Practical PQC therefore benefits from pipelines that stabilize learning under imbalance thresholding strategies, cost-sensitive losses, and stratified mini-batches so that local updates faithfully encode early-warning signals rather than being dominated by the majority class. A simple operational metric makes this tangible: the F1 score, $F_1 = 2 \cdot (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$, rewards balanced improvements in catching true faults while limiting nuisance alarms, and it can be computed per site and aggregated at architectural tiers to evaluate whether a topology amplifies useful signal. Since PQC must generalize across non-IID sites, architecture and metric design move together: update frequency, client participation rules, and tiering determine how minority-class gradients propagate through the federation. These concerns echo established results on imbalanced learning, where skewed class distributions complicate model selection and require careful evaluation protocols; translating those lessons into the federated, cross-site setting helps ensure that architectural decisions improve real error trade-offs rather than cosmetic averages (He

American Journal of Interdisciplinary Studies
Volume 02, Issue 02 (2021)
Page No: 01 – 31
eISSN: 3067-5146
DOI: 10.63125/222nwg58

& Garcia, 2009). In this light, architecture is not merely a networking diagram but a statistical instrument that governs how scarce but consequential defect evidence flows from edge to aggregate and back to edge.

Figure 2: Predictive Quality Control in Distributed Manufacturing



In addition, architectural selection carries governance and security requirements that are inseparable from PQC performance. Collaborative updates can leak sensitive attributes through gradients or parameters, creating risks that must be actively controlled in design and operations. Empirical security research shows that adversarial participants can infer properties unrelated to the main task or even perform membership inference against collaborators, indicating that federations need defenses that go beyond informal trust secure protocols, monitoring, and review gates for promotion (Melis et al., 2019). These safeguards coexist with production-readiness practices that treat PQC as a living system: data validation at ingress, canarying and rollback for model releases, slice-aware quality checks, and monitoring for training–serving skew. In practice, architectural tiers often mirror operational accountability, so readiness rubrics and checklists help standardize pre-deployment tests across hubs, clusters, and sites. A production rubric enumerating tests over data, models, infrastructure, and monitoring gives architects a shared baseline for deciding whether a federation is healthy enough to absorb another training round or to promote an updated global model to the shop floor (Breck et al., 2017). The combination of formalized readiness and explicit privacy safeguards makes architectural trade-offs auditable: hub-and-spoke can centralize controls but may bottleneck review; hierarchical schemes distribute both risk and responsibility; peer-like exchanges require stronger, consistent local controls. For PQC, where alerts drive real interventions, architectures that encode these operational disciplines alongside statistical goals are better positioned to translate collaborative learning into stable, plant-level quality gains.

American Journal of Interdisciplinary Studies
Volume 02, Issue 02 (2021)
Page No: 01 – 31
eISSN: 3067-5146
DOI: 10.63125/222nwg58

Predictive Quality Control in Smart Manufacturing

Predictive quality control (PQC) in smart manufacturing integrates multivariate sensing, statistical learning, and production execution to anticipate nonconformities before they propagate. A central technical thread is the shift from univariate control of key characteristics to multivariate modeling of process states, where quality-relevant variables interact nonlinearly and exhibit correlation structures across time and stations. In this setting, monitoring and prediction frequently employ latent-variable or feature-learning pipelines that transform raw telemetry into compact representations suitable for classification or regression of quality outcomes. Foundational surveys in data-driven process monitoring codify a reference workflow data preprocessing, feature extraction, monitoring statistics, and diagnostic modeling that adapts to real operating conditions such as noise, transient regimes, and changing setpoints (Yin et al., 2014). To evaluate PQC performance under class imbalance and non-stationarity typical of high-yield lines, practitioners often complement error-rate metrics with measures sensitive to positive class rarity (e.g., recall at fixed false-alarm rate) and with multivariate distance statistics for early warning. A classic formulation, Hotelling's statistic, evaluates deviations from the in-control manifold:

$$T_2 = (x - \mu)^\top \Sigma^{-1} (x - \mu),$$

Where x is the current feature vector, μ the in-control mean, and Σ the in-control covariance. In smart factories, T^2 (and its residual Q-statistic) can be computed per cycle on edge nodes and fused with a learned classifier for predictive decisions on scrap, rework, or hold. Case-based work on product-state modeling shows how supervised learning over structured representations of a part's path through multiple operations can yield robust quality predictions while preserving interpretability for engineers who must triage alarms and tune thresholds on the shop floor (Wuest, Weimer, Irgens, & Thoben, 2014). These ingredients workflows, statistics, and product-state features compose the core analytics stack that PQC uses to convert heterogeneous sensor streams into action-ready signals in smart manufacturing (Wuest et al., 2014; Yin et al., 2014). Within this stack, surface and visual inspection problems illustrate the requirements of PQC under tight cycle times and rare defects. Vision pipelines must capture subtle texture deviations, scratches, inclusions, or pits while maintaining real-time throughput and minimizing nuisance alarms that stall lines. A comprehensive survey of visual defect detection for flat steel demonstrates how feature engineering and modern deep models improved detection consistency across lines and coils and formalized performance reporting with harmonized metrics (He et al., 2019). In practice, production analytics compute site-level confusion matrices and aggregate them to line or plant dashboards; a balanced measure such as the F_1 score,

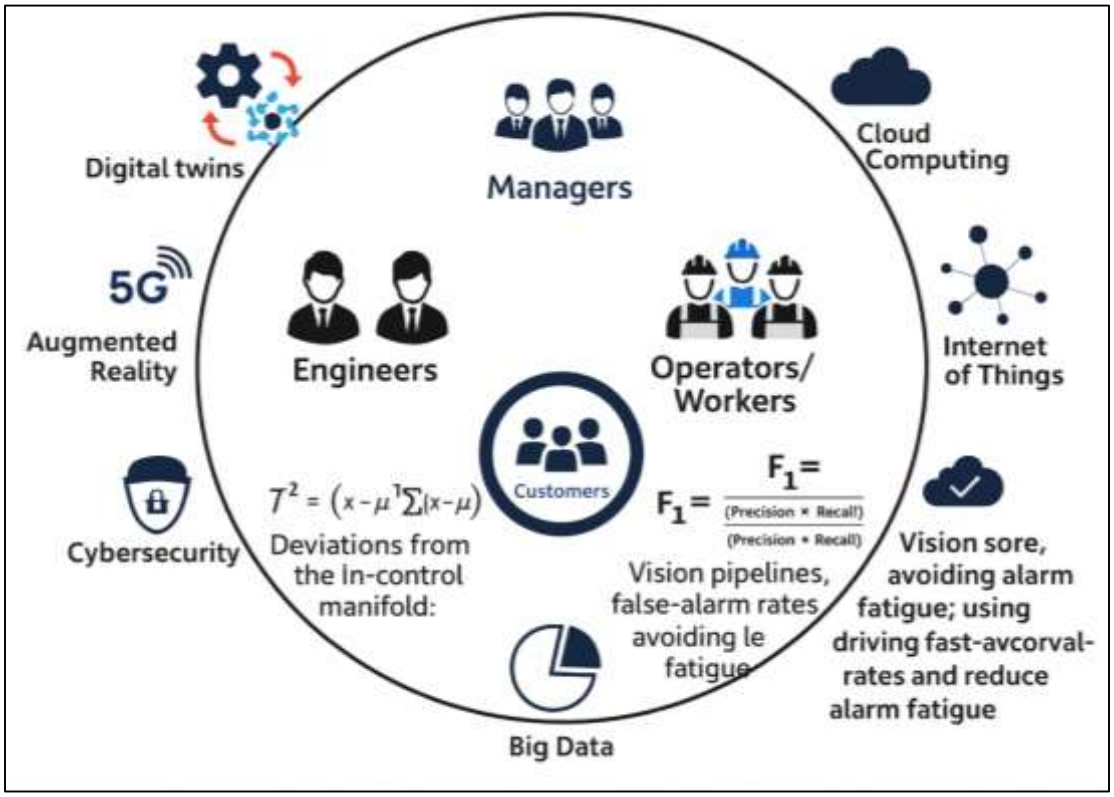
$$\boldsymbol{F_1} = \frac{2(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

is commonly tracked alongside false-alarm rate per hour to ensure operators are not overwhelmed by alerts. Beyond imaging, mechanical condition signals (e.g., vibration, acoustic emission, current) are routinely leveraged for upstream PQC by forecasting fault-prone states in rotating assets that influence downstream quality. Deep architectures such as deep belief networks (DBNs) using multi-sensor fusion have been shown to extract discriminative features from noisy, nonstationary signals, improving early detection windows that matter for scrap avoidance and process stability (Sun et al., 2016). In integrated PQC programs, these visual and vibro-acoustic paths converge: surface-defect catch rates and upstream fault-state detections are co-monitored against takt-time and specification windows to maintain yield and compliance while enabling targeted interventions tool change, parameter reset, or temporary speed derating before defects manifest in final inspection (Lieber et al., 2012).

A third line of PQC research concerns interlinked and multistage processes, where intermediate product quality carries forward and interacts with downstream operations. Here, quality prediction must "walk" the process graph, learning mappings from stagewise telemetry to intermediate and final attributes and accounting for propagation of variance. Studies on real-time quality prediction in rolling and other multistage settings demonstrate the value of inline learning systems that join supervised and unsupervised components to track drifting regimes and uncover salient process features for prediction (Lieber et al., 2012).

American Journal of Interdisciplinary Studies
Volume 02, Issue 02 (2021)
Page No: 01 – 31
eISSN: 3067-5146
DOI: 10.63125/222nwg58

Figure 3: Predictive Quality Control Framework in Smart Manufacturing



In rolling-mill exemplars, regression or classification heads trained on engineered and latent features forecast within-tolerance profiles or flag risk of out-of-spec, enabling dynamic parameter adjustments upstream. For general PQC deployment across diverse processes, systematic reviews emphasize a unifying methodological spine data cleaning, feature selection, modeling, and validation adapted to industrial constraints such as limited fault examples, shift-based regimes, and evolving recipes (Yin et al., 2014). Across these contexts, the operationalization of PQC couples analytic rigor with production readiness: data lineage and validation at ingress; station-aware thresholding; and tiered model release with canarying and rollback to ensure that predictive signals remain stable under load. As organizations scale PQC, vision, vibration, and multistage predictors are embedded into MES/SCADA workflows, with edge–cloud coordination tuned to update cadence, bandwidth budgets, and governance. The literature thus positions PQC not as a monolithic model but as a system of models and monitors aligned with the realities of smart manufacturing heterogeneous data, rare adverse events, and the need to integrate predictions into fast, auditable decisions (Lieber et al., 2012).

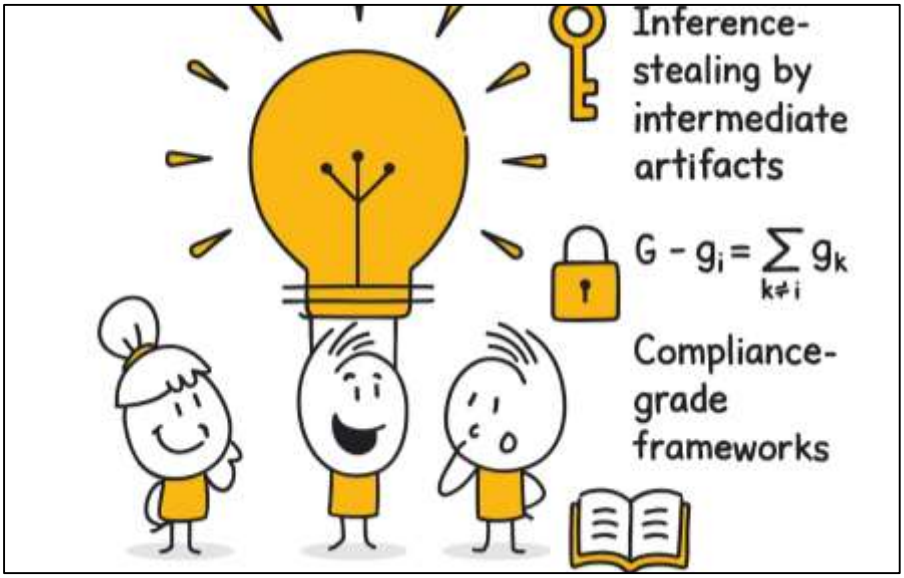
Security in Federated Learning for PQC

Federated learning (FL) for predictive quality control (PQC) must reconcile collaborative modeling with strict limits on the disclosure of sensitive, plant-specific telemetry and proprietary process parameters. In distributed manufacturing ecosystems, this tension surfaces most acutely in the privacy risks introduced by model sharing itself risks that do not require direct access to raw data. Model inversion demonstrates how an adversary, given a trained model and confidence scores, can reconstruct salient features or representative inputs for particular classes, elevating the risk of intellectual property leakage when defect signatures or process states are unique to a supplier or plant (Fredrikson et al., 2015). Black-box transfer attacks further show that adversaries can query a target model to synthesize surrogate models that replicate its decision boundary, enabling downstream evasion or extraction without white-box access and with only API-level interactions (Papernot et al., 2017). These capabilities matter for PQC because decision services may be exposed to enterprise applications or partner systems that request predictions at line speed; even rate-limited interfaces can act as oracles if not appropriately protected. From a systems perspective, these findings reframe "privacy" to include both parameter and output channels: the former in the

American Journal of Interdisciplinary Studies
Volume 02, Issue 02 (2021)
Page No: 01 – 31
eISSN: 3067-5146
DOI: 10.63125/222nwg58

exchange of updates among federation members and the latter in the operational use of the model on the shop floor. A succinct way to reason about risk budgeting during training is to track an additive privacy-risk accumulator for each communication round, e.g., define a per-round leakage proxy λ_t and maintain $\Lambda_t = \Sigma_{t=1}{}^{T} \lambda_t$ so that governance can specify a maximum allowable Λ_{max} for a training window. Although such a proxy is not a formal privacy guarantee, embedding it into scheduling and audit logs creates operational hooks for cross-site coordination and for pausing or re-keying the federation when thresholds are reached (Fredrikson et al., 2015; Papernot et al., 2017).

Figure 4: Data Privacy, Securityfor Predictive Quality Control



Privacy and security in FL also hinge on what collaborators and potentially compromised insiders can infer from intermediate training artifacts. White-box inference analyses reveal that gradients and internal activations can carry discriminative traces of training examples; in collaborative settings, this enables both passive and active attacks that identify membership or recover properties of the local datasets contributing to the model (Nasr et al., 2019). In PQC, where local datasets encode relations among temperatures, pressures, speeds, and microstructural outcomes, gradient-sharing without protection expands the attack surface for learning proprietary process envelopes. A minimal algebraic expression captures the danger and a defense intuition: if client k shares an update g_k and the aggregator computes the global step $G = \Sigma_{k=1}{}^{K} g_k$, then any participant who learns both its own g_i and G can form $G - g_i = \Sigma_{k \neq i} g_k$, narrowing the space of other clients' contributions and facilitating leakage. Protocol-level defenses (e.g., masking plus pairwise key agreement) aim to ensure only G is observable, not individual g_k, yielding the masked identity $\Sigma_k (g_k + r_k) - \Sigma_k r_k = \Sigma_k g_k$ where masks r_k cancel only in aggregate. Operationally, PQC federations should institutionalize white-box threat modeling in change management: pre-deployment checklists require proofs or tests that no participant or aggregator can isolate another's update; incident response drills treat gradient logs as sensitive artifacts governed by the same retention and access policies as manufacturing execution system (MES) data. In practice, coupling protocol assurances with output-hardened serving for example, suppressing high-resolution confidence scores and limiting per-entity query rates reduces exposure to both training-time and inference-time leakage channels highlighted by the literature (Rieke et al., 2020).

Compliance and governance complete the triad alongside privacy and security, translating abstract risks into enforceable policies and auditable procedures. Sectoral experience shows that federated analytics gains practical legitimacy when embedded within compliance-grade frameworks that align with medical, financial, or industrial norms. In health, for instance, cross-institutional FL deployments have been framed as a way to respect data minimization and patient-privacy requirements while enabling generalizable clinical models; this experience offers transferable

American Journal of Interdisciplinary Studies
Volume 02, Issue 02 (2021)
Page No: 01 – 31
eISSN: 3067-5146
DOI: 10.63125/222nwg58

lessons for manufacturing consortia that must balance supplier confidentiality with collective quality gains (Rieke et al., 2020; Truex et al., 2019). Beyond domain analogies, the design space of privacy-preserving FL includes hybrid approaches that combine cryptographic secure aggregation with local perturbation or selective sharing to meet both utility and compliance targets, illustrating that practical deployments often engineer layered controls instead of relying on a single technique (Truex et al., 2019). For PQC, these governance patterns map naturally to multi-tier architectures: hubs or regional aggregators carry control responsibilities for key management, audit logging, and promotion gates; cluster leads enforce participation rules and data-retention ceilings; sites maintain inventories of features and labels with lineage to station IDs and work orders. A simple compliance accounting identity can guide audits at each tier: if U is the set of model updates, C the set of compliance checks passed (e.g., key rotation, participation consent, retention windows), and A the set of approvals recorded, then a promotion is admissible only when $U \subseteq (C \cap A)$, i.e., every update consumed by the promoted model has corresponding checks and approvals. By making such identities first-class in release tooling, federations gain testable claims about who did what, when, and under which policy an essential property when PQC recommendations lead to material interventions on lines. The empirical and engineering insights from cross-domain deployments thus reinforce a blueprint for PQC: treat privacy as an operational budget, harden both training and serving against inference, and codify layered controls and accountability into the architecture so that compliance is continuously met (Papernot et al., 2017; Rieke et al., 2020).

Edge/IIoT Infrastructure and Data Engineering

Edge and Industrial Internet of Things (IIoT) infrastructures provide the physical and logical substrate that allows predictive quality control (PQC) to function close to machines while coordinating with minimal data movement across distributed manufacturing sites. At plant level, controllers, sensors, and embedded devices emit time-stamped telemetry at heterogeneous sampling rates; gateways normalize signals, enforce security policies, and bridge operational technology (OT) networks with information technology (IT) services that schedule training and inference. This layered pattern is consistent with the broader Internet-of-Things vision in which massive numbers of resource-constrained devices cooperate through edge and cloud resources to deliver analytics within latency and bandwidth budgets (Gubbi et al., 2013). In PQC terms, "local-first" computation reduces backhaul while enabling per-line feature engineering that respects takt-time. A useful way to formalize streaming feature stability is the exponentially weighted moving average, which filters noise yet remains responsive to small shifts:

$$S_t = \alpha x_t + (1-\alpha) S_{t-1}, \quad 0 < \alpha \leq 1,$$

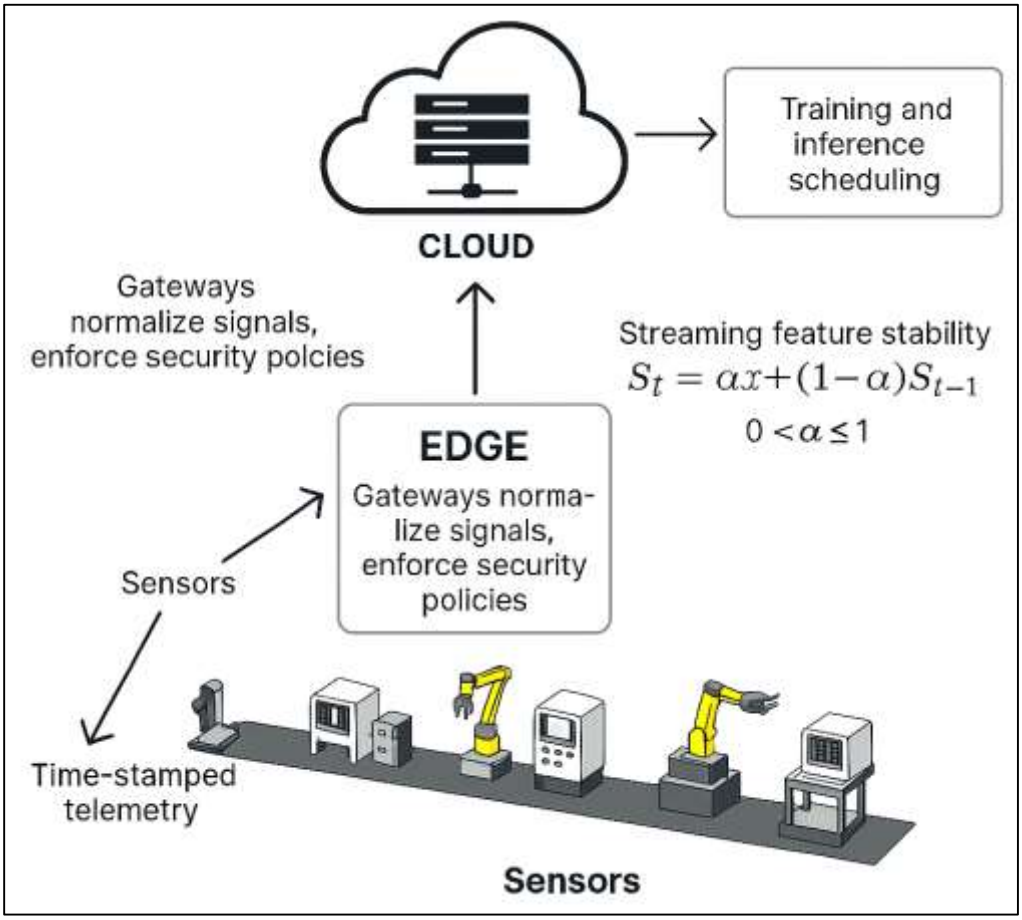
where x_t is the current sensor-derived feature and S_t its smoothed estimate used by on-edge predictors. By tuning a to the line's dynamics, plants bound spurious alarms while preserving sensitivity to incipient faults that drive PQC decisions. From a cross-site perspective, IIoT heterogeneity encoders, sampling, timestamp conventions demands rigor in schema management and time alignment so that model updates aggregated upstream are comparable. At ecosystem scale, architectural decomposition across edge, regional aggregation, and cloud aligns with the original IoT surveys that stressed interoperability, scalable addressing, and quality-of-service as prerequisites for dependable analytics over large device populations (Atzori et al., 2010). These principles ground PQC deployments that must operate continuously amid maintenance windows, recipe changes, and workload bursts without sacrificing data fidelity or observability (Gubbi et al., 2013; Atzori et al., 2010). Within plants, machine-to-machine communication and semantically rich messaging are the backbone of reliable data engineering. PQC signals often originate as fieldbus or controller values that require normalization, unit harmonization, and context binding (station ID, tool version, work-order) before they are usable by models or rule engines. The OPC Unified Architecture (OPC UA) standardizes this interoperability layer with an information model that exposes hierarchical nodes, methods, and events under secure sessions crucial for PQC pipelines that must guarantee traceable lineage from raw tags to engineered features and predictions (Mahnke et al., 2009). Practically, UA's address space carries asset structure (cells, stations, tools) and metadata, enabling deterministic subscriptions that feed edge feature stores. A simple throughput budget keeps deployments safe under cycle-time constraints: if a line publishes r records per second at payload size b bytes and protocol overhead factor κ (headers, encryption), then the sustained link demand is $D = r \times b \times \kappa$.

$$B_{\text{eff}} = r_b \kappa \,(\text{bytes/s}),$$

American Journal of Interdisciplinary Studies
Volume 02, Issue 02 (2021)
Page No: 01 – 31
eISSN: 3067-5146
DOI: 10.63125/222nwg58

which must remain below the reserved capacity on the OT–IT bridge after accounting for other services. This budget in turn bounds batch sizes for local training, the period for checkpointing, and the frequency of federated update rounds. On the storage side, edge nodes buffer recent windows in append-only logs for recovery and model audit; periodic compaction enforces retention policies while preserving the indices needed to join telemetry with quality labels. UA-centric typing and browseable metadata also ease schema evolution: when stations or recipes change, versioned nodes and aliasing reduce "feature drift," stabilizing downstream PQC models that depend on consistent semantics for input features and output labels (Gubbi et al., 2013).

Figure 5: Edge/IIoT Infrastructure for Predictive Quality Control (PQC)



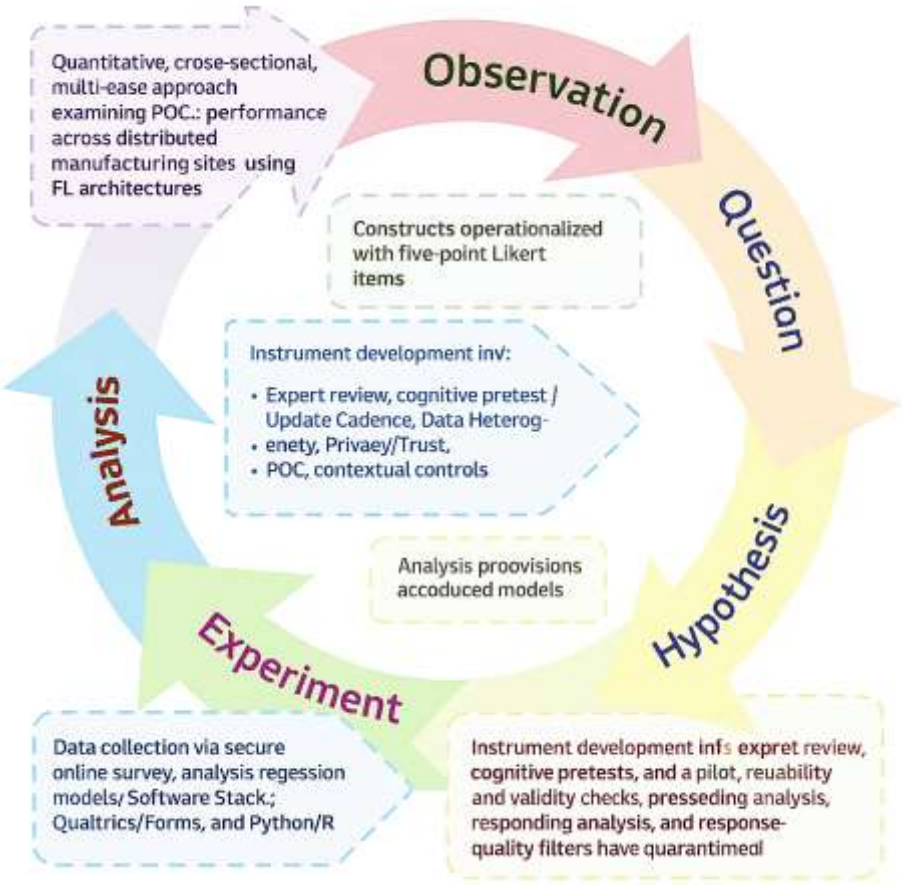
In tightly coupled lines, these conventions allow deterministic replay for root cause analysis when PQC thresholds trigger holds or rework, and they maintain the integrity needed for cross-site comparison during federated training. Orchestration at the edge is the final engineering pillar linking infrastructure to federated PQC. Modern edge platforms extend cloud-native primitives containers, service meshes, declarative configuration into plant networks so that data collectors, feature services, and model servers are deployed, updated, and monitored with parity to central environments. One representative platform, KubeEdge, brings Kubernetes-style control planes to the network perimeter, enabling offline-tolerant operation, device twin abstractions, and event routing that preserve local autonomy while remaining centrally governable (Xiong et al., 2019). For PQC, this matters operationally: model images can be canaried on a subset of stations; inference services can be pinned to nodes with accelerators; and log shipping can be throttled or paused during production peaks without interrupting on-line scoring. Above the orchestration layer, digital-twin practices encode the structure and state of assets and processes linking as-built hierarchies, process parameters, and historical telemetry to provide a consistent, queryable context for features and labels across the product lifecycle (Tao et al., 2018). In distributed networks, twins allow regional or corporate teams to reason about PQC signals from multiple plants with shared semantics, facilitating

American Journal of Interdisciplinary Studies
Volume 02, Issue 02 (2021)
Page No: 01 – 31
eISSN: 3067-5146
DOI: 10.63125/222nwg58

robust cross-site model evaluation before federated promotion. Finally, the end-to-end dataflow from acquisition to serving should reflect the data-driven manufacturing guidance that emphasizes pipeline modularity, feedback from execution to analytics, and closed-loop improvement anchored in trustworthy data products (Gubbi et al., 2013). Combined, IoT-scale design goals (Tao et al., 2018), interoperability at the machine layer (Mahnke et al., 2009), cloud-native edge orchestration (Xiong et al., 2019), and lifecycle-aware digital context (Tao et al., 2018) produce an infrastructure footing on which federated, privacy-aware PQC can run continuously and comparably across plants and regions (Atzori et al., 2010).

METHOD

Figure 6: Framework for this study



The study has adopted a quantitative, cross-sectional, multi–case approach to examine how federated-learning (FL) architectures and enabling conditions have related to predictive quality control (PQC) performance across distributed manufacturing sites. Guided by a priori power targets and practical access, participating plants have been selected when they have already piloted or implemented FL-supported PQC, have maintained IIoT data capture and traceable quality labels, and have operated within a discernible federation topology (hub-and-spoke, hierarchical, or peer). Within each eligible site, the sampling frame has encompassed roles that have intersected PQC deployment quality and process engineers, production supervisors, and data/IT/OT personnel and invitations have been stratified to balance perspectives across lines and shifts. Constructs have been operationalized with five-point Likert items (1 = Strongly disagree … 5 = Strongly agree) covering Infrastructure Readiness, Communication Efficiency/Update Cadence, Data Heterogeneity (higher = more difference), Privacy/Trust, and PQC Performance, alongside contextual controls (plant size, product complexity, industry segment, and quality-maturity practices). Where available, sites have uploaded aggregated, non-identifiable logs (e.g., AUC/F1, false-alarm rate, mean time-to-detection, rounds per month) to triangulate perceptual measures. Instrument development has followed expert review, cognitive pretests, and a pilot; reliability (Cronbach's a, composite reliability) and validity (AVE, HTMT; optional CFA) checks have preceded analysis, and response-quality filters

have quarantined speeders and straight-liners. Data collection has been executed via a secure online survey with consent screens, site-level liaisons, and a four-week window per site; artifacts have been transferred through an encrypted portal and stored in a version-controlled repository under least-privilege access. The analysis plan has progressed from screening and descriptives to bivariate correlations and then to ordinary least squares models estimating main effects and architecture-based moderation (ARCH × HET, ARCH × COM), with robust standard errors, multicollinearity diagnostics, and assumption checks; sensitivity analyses have included mixed-effects models with site random intercepts and site-aggregated weighted regressions. Power reasoning has targeted small-to-moderate incremental effects with adjustments for clustering via design-effect calculations, informing a goal of ≥150 respondents across 2–5 cases. Throughout, the software stack has comprised Qualtrics/Forms for administration and Python/R for scripted, reproducible analysis, while ethical safeguards have emphasized anonymity, minimal data collection, and privacy-preserving handling of any optional logs. Collectively, these procedures have positioned the study to deliver reliable, comparable evidence on PQC performance under FL in real industrial settings.

Research Design

This study adopts a quantitative, cross-sectional, multi–case design to examine how federated learning (FL) architectures relate to predictive quality control (PQC) performance in distributed manufacturing. Multiple manufacturing sites (cases) that are actively using or piloting FL-enabled PQC constitute the unit of analysis, with respondents sampled from key roles (quality engineering, process engineering, data/IT/OT). Inclusion criteria require IIoT-enabled data pipelines, documented PQC use cases (e.g., defect prediction or early-warning flags), and participation in a federation (hub-and-spoke, hierarchical, or peer). Sites operating only centralized modeling or without quality telemetry are excluded. Data are collected via an online survey using five-point Likert scales to capture latent constructs Infrastructure Readiness, Communication Efficiency/Update Cadence, Data Heterogeneity, Privacy/Trust, and PQC Performance with contextual controls for plant size, product complexity, quality maturity, and industry segment. Where available, non-identifiable objective indicators (e.g., AUC/F1, false-alarm rate, time-to-detection, rounds per month) are uploaded in aggregated form to triangulate perceptual measures. The analysis plan proceeds from descriptive statistics and distributional checks to bivariate correlations and then to OLS regression models that estimate main effects and interactions (architecture × heterogeneity; architecture × communication), with robust standard errors and multicollinearity diagnostics (VIF). Given clustering by site, sensitivity analyses include hierarchical models if intra-class correlation warrants. Target sample size follows the rule of ≥15–20 observations per predictor, aiming for ≥150 respondents overall across 2–5 cases, with response-rate documentation. Instrument reliability (Cronbach's a, composite reliability) and validity (AVE, HTMT; optional EFA/CFA) are evaluated before hypothesis testing; common-method safeguards include proximal separation of measures and marker items. Ethical protocols emphasize informed consent, anonymity, and privacy-preserving handling of any objective metrics; survey items avoid proprietary parameters. The design's cross-sectional snapshot enables comparative inference across architectures and conditions while remaining feasible for industrial partners operating under time and confidentiality constraints.

The study has adopted a multi–case setting drawn from distributed manufacturing sites that have already implemented or piloted federated learning (FL) for predictive quality control (PQC). Participating plants have been required to meet inclusion criteria that have ensured analytic comparability: they have possessed IIoT-enabled data capture on critical processes, have maintained routine quality inspection or inline sensing suitable for labeling, and have operated within a federation topology (hub-and-spoke, hierarchical, or peer). Sites have been excluded when they have relied solely on centralized modeling, have lacked traceable quality labels, or have not maintained stable connectivity for periodic update rounds. Within each eligible site, the sampling frame has encompassed professionals whose roles have intersected PQC deployment quality and process engineers, production supervisors, data/IT/OT personnel, and reliability specialists. Stratified invitations by role and line/area have been issued to balance perspectives, and each respondent has been required to confirm direct familiarity with PQC workflows during the last 12 months. To support cross-site inference, the study has targeted a minimum overall sample that has satisfied conventional rules-of-thumb for regression power (≥15–20 observations per predictor) and has aimed to distribute respondents across 2–5 cases to capture heterogeneity in equipment, products, and governance. The setting has included discrete and hybrid manufacturing environments where

defect mechanisms have been consequential for yield and customer acceptance. Plants have contributed site profiles that have described line architecture, primary processes, product families, and key sensors, and they have reported whether objective indicators (e.g., classification metrics, false-alarm counts, update cadence) have been available in aggregated, non-identifiable form. Recruitment has been conducted through site liaisons who have coordinated management approval and local communication. Participation has been voluntary and anonymous; respondents have acknowledged informed consent before accessing the instrument. To minimize selection bias, reminders have been scheduled evenly across shifts, and completion checks have flagged inconsistent or straight-line responses. Collectively, these procedures have produced a case sample that has reflected real-world FL deployments while maintaining sufficient diversity for meaningful cross-site comparisons.

Measures

The study has operationalized its constructs with five-point Likert items (1 = Strongly disagree … 5 = Strongly agree) and, where available, complementary objective indicators that participating sites have provided in aggregated, non-identifiable form. PQC Performance has been measured with perceptual items that have captured perceived improvements in defect prediction accuracy, reductions in false alarms, and timeliness of detection; when logs have been available, sites have contributed AUC/F1, false-alarm rate per hour, and mean time-to-detection, which the research team has normalized to common units before analysis. Infrastructure Readiness has been captured through items that have reflected sufficiency of edge compute, network reliability, storage headroom, and orchestration maturity (e.g., containerized deployment, rollback capability). Communication Efficiency / Update Cadence has been assessed with items that have described the duration of federated rounds, stability of windows for synchronization, and frequency of successful model refreshes during typical shifts. Data Heterogeneity has been measured via items that have asked respondents to judge cross-site differences in feature distributions, label imbalance, and recipe variability; at analysis time, negatively keyed items (e.g., "our sites have very similar data distributions") have been reverse-coded so that higher scores have indicated greater heterogeneity. Privacy / Trust has been reflected in items that have assessed confidence in secure aggregation, access controls, and the clarity of inter-organizational rules for model updates. Architecture Choice has been recorded as a categorical moderator (hub-and-spoke, hierarchical, peer), which the team has dummy-coded for regression. Controls have included plant size (bins of headcount or throughput), product complexity (ordinal scale), industry segment (categorical), and quality-maturity practices (composite index). Before hypothesis testing, the instrument has undergone screening in which items with extreme skew or low item-total correlations have been flagged; multi-item scales have been retained when Cronbach's α and composite reliability have met acceptable thresholds, and exploratory checks of convergent and discriminant validity have been completed. Missing responses have been handled with pairwise deletion for descriptives and mean-imputation within scales when fewer than 20% of items have been missing for a given construct.

Data Collection

The study has collected data through a combined survey-and-artifacts approach that has balanced breadth of perception with selectively verified operational indicators. Participating sites have first designated a liaison who has coordinated approvals and provided a brief site profile; thereafter, the research team has distributed a secure online questionnaire to role-stratified mailing lists (quality/process engineering, production supervision, data/IT/OT). The instrument has contained the Likert-scale measures described previously and has included gating items that have confirmed each respondent's direct involvement with predictive quality control (PQC) workflows within the past 12 months. To minimize instrumentation bias, the team has preceded full rollout with a pilot at one site; feedback from the pilot has informed minor edits for clarity and sequencing, and the final form has been locked before wide distribution. Alongside the survey, sites have been invited to upload aggregated, non-identifiable logs such as model AUC/F1, false-alarm rate per hour, mean time-to-detection, and counts of successful federated rounds per month via an encrypted portal; contributors have attested that uploads have excluded raw sensor values, product identifiers, or personally identifiable information. All submissions have been time-stamped, checksum-verified, and stored in a version-controlled repository with access limited to the core analysis team. The collection window has been open for four weeks per site, during which scheduled reminders have been sent across shifts to balance participation; response monitoring dashboards have flagged unusual

completion times or straight-line patterns, and suspected low-quality entries have been set aside under predefined rules. Consent screens have preceded the survey, and participants have acknowledged that their responses have been anonymous and reported only in aggregate; an institutional review protocol has been approved before initiation, and data-handling procedures have complied with the confidentiality requirements stipulated by participating organizations. Upon closure of each site's window, the team has reconciled survey IDs with artifact manifests at the site level (not the individual level) to enable triangulation, and a debrief summary has been returned to liaisons so that any clarifications or corrections have been captured prior to analysis.

Statistical Analysis Plan

The analysis has proceeded in staged steps that have safeguarded measurement quality before estimating structural relationships. First, the dataset has undergone screening in which the team has inspected response distributions, identified outliers via standardized scores and Mahalanobis distance at the construct level, and assessed missingness patterns; where fewer than 20% of items within a multi-item scale have been missing, mean imputation within that scale has been applied, otherwise the observation has been excluded from scale construction. Next, internal consistency for each latent construct has been evaluated; items with low item–total correlations or cross-loadings flagged in exploratory checks have been candidates for removal, and retained scales have satisfied pre-specified thresholds for reliability. Convergent and discriminant validity checks have been completed on the refined scales, after which composite scores have been computed as arithmetic means to preserve Likert interpretability. With measurement established, the team has produced descriptive statistics (mean, SD, range, skewness, kurtosis) and Pearson (or Spearman, as warranted) correlations among constructs and controls, documenting confidence intervals via bootstrapping. Multicollinearity diagnostics have been performed using variance inflation factors, and any construct pairs exhibiting excessive collinearity have been addressed through centering, respecification, or omission in sensitivity runs. Primary inference has relied on ordinary least squares models in which PQC Performance has been regressed on Infrastructure Readiness, Communication Efficiency/Update Cadence, Data Heterogeneity, Privacy/Trust, and controls; robust (HC) standard errors have been reported alongside standardized coefficients, partial R^2, and model R^2_adj. Moderation by Architecture has been tested through interaction terms with heterogeneity and communication, with lower-order terms retained; continuous predictors have been mean-centered prior to interaction construction. Given potential clustering by site, intraclass correlations have been estimated, and mixed-effects models with random intercepts have been fitted in sensitivity analyses when clustering has been non-trivial. Assumptions (linearity, homoskedasticity, normality of residuals, influential points) have been examined using residual plots and formal tests; robustness checks have included alternative outcome operationalizations, exclusion of high-influence cases, and nonparametric regressions. All analyses have been reproducibly scripted, and a preregistered decision log has documented deviations from the plan.

Regression Models

The 16 odelling strategy has been organized around a baseline ordinary least squares (OLS) specification to quantify the association between federated-learning (FL) conditions and predictive quality control (PQC) performance, with carefully prepared variables and coding schemes. The dependent variable has been PQC Performance, operationalized as the composite of Likert-scale items and, where available, z-standardized objective indicators (e.g., AUC/F1, false-alarm rate reversed, and time-to-detection reversed) averaged into a site-level score. Core predictors have included Infrastructure Readiness (ITR), Communication Efficiency/Update Cadence (COM), Data Heterogeneity (HET; higher values have indicated greater cross-site divergence), and Privacy/Trust (PRV). Architecture has been treated as a categorical moderator (ARCH) reflecting hub-and-spoke, hierarchical, or peer configurations; analysis has dummy-coded ARCH with hub-and-spoke as the reference level, unless otherwise noted. To reduce collinearity and to make interaction terms interpretable, continuous predictors have been mean-centered and standardized to unit variance before model fitting. The baseline main-effects model has been expressed as:

$$\mathrm{PQC}_i = \beta_0 + \beta_1 \mathrm{ITR}_i + \beta_2 \mathrm{COM}_i + \beta_3 \mathrm{HET}_i + \beta_4 \mathrm{PRV}_i + \gamma^{\top} Z_i + \varepsilon_i ,$$

where $\mathbf{Z_i}$ has contained controls (plant size, product complexity, industry segment, and quality-maturity index). A moderation model has then been specified by adding ARCH × HET and ARCH × COM interactions to test whether architectural choices have altered the association of heterogeneity and communication with PQC outcomes:

$$\mathrm{PQC}_i = \mathrm{baseline} + \delta_1(\mathrm{ARCH}_{\mathrm{hier}} \times \mathrm{HET}_i) + \delta_2\big(\mathrm{ARCH}_{\mathrm{peer}} \times \mathrm{HET}_i\big) + \delta_3(\mathrm{ARCH}_{\mathrm{hier}} \times \mathrm{COM}_i) + \delta_4\big(\mathrm{ARCH}_{\mathrm{peer}} \times \mathrm{COM}_i\big) + \varepsilon_i,$$

Estimation has reported unstandardized and standardized coefficients with heteroskedasticity-robust (HC) standard errors, 95% confidence intervals, adjusted **R^2**, and partial **R^2** for each block of variables. To facilitate interpretation, simple-slope calculations and predicted-value plots at ±1 SD of moderators have been produced, and all continuous predictors have remained centered so that dummy-variable main effects for ARCH have represented differences at average levels of the continuous predictors.

Because responses have been clustered within manufacturing sites and because PQC operations have been executed under site-specific governance, the analysis has complemented OLS with hierarchical modeling that has acknowledged between-site variance. First, the study has computed intraclass correlation coefficients (ICCs) on the dependent variable and on key constructs; non-trivial ICCs have indicated that a portion of variance has resided at the site level. Accordingly, a random-intercepts mixed-effects model has been specified:

$$\mathrm{PQC}_{ij} = \beta_0 + u_{0j} + \beta_1 \mathrm{ITR}_{ij} + \beta_2 \mathrm{COM}_{ij} + \beta_3 \mathrm{HET}_{ij} + \beta_4 \mathrm{PRV}_{ij} + \gamma^{\mathrm{T}} Z_{ij} + \varepsilon_{ij},$$

where *i* has indexed individuals and *j* has indexed sites, with $u_{0j} \sim N(0, \tau_0^2)$. In sensitivity analyses, random-slope terms for HET and COM have been added when likelihood-ratio tests and information criteria have supported improved fit, enabling cross-site variability in those effects. Cross-level interactions $\mathrm{ARCH}_j \times \mathrm{HET}_{ij}$ and $\mathrm{ARCH}_j \times \mathrm{COM}_{ij}$ have been included to evaluate whether site-level architecture has moderated individual-level relationships. A hierarchical-entry sequence has been followed: controls have been entered first, then main technical and organizational predictors, and finally interaction terms; each step has reported ΔR^2 (for OLS) or ΔΔAIC/ΔΔBIC (for mixed models). To verify robustness, the team has estimated a site-aggregated model in which constructs have been averaged to the site level (N = number of sites) and fitted via OLS with weighted least squares (WLS) using respondent counts as weights; this step has guarded against heteroskedastic site means driven by unequal sample sizes. Throughout, multicollinearity has been assessed via variance inflation factors (VIFs), residual structure has been checked visually and through formal tests, and influence diagnostics (Cook's distance and leverage) have been calculated to identify observations exerting undue pull on coefficients. Where necessary, Huber–White HC3 or cluster-robust standard errors at the site level have been reported.

For transparency and reproducibility, the modeling protocol has pre-specified the content and layout of statistical outputs and has linked them to decision criteria. Table 1 (below) has summarized all models, variable blocks, and targeted diagnostics. Model adequacy has been documented with residual-vs-fitted plots, Q–Q plots of studentized residuals, and Breusch–Pagan tests for heteroskedasticity; where patterns have been detected, robust errors have been retained and, in sensitivity runs, Box–Cox transformations of the dependent variable have been evaluated without altering substantive conclusions. Interaction interpretation has been accompanied by simple-slope estimates at ±1 SD of HET and COM and by Johnson–Neyman intervals where applicable, so that regions of significance have been clearly identified. To address measurement stability, the team has re-estimated all models using latent-score composites derived from confirmatory factor analysis (factor scores scaled to mean 0, SD 1), and conclusions about sign, magnitude, and significance have remained consistent with the manifest-composite approach. To triangulate perceptual and objective performance, an objective-index outcome has been constructed where sites have provided logs: PQC_obj = (1/3)·(zAUC + zF1 + zInvFAR), with zInvFAR denoting the z-score of inverted false-alarm rate. Models substituting PQC_obj for the composite have been estimated and have aligned with the main specification. Finally, pre-registered thresholds have governed interpretation: effects have been emphasized when standardized $|\beta|$ has exceeded 0.10 with $p < .05$ (two-sided) and when variance-inflation diagnostics have fallen below 5. All analyses have been scripted end-to-end, and a model card has been generated for each promoted specification documenting data version, feature dictionary, estimation options, and the exact formula used.

American Journal of Interdisciplinary Studies
Volume 02, Issue 02 (2021)
Page No: 01 – 31
eISSN: 3067-5146
DOI: 10.63125/222nwg58

Table 1. Model Specifications and Planned Outputs

Model ID	Specification (incremental blocks)	Random Effects	Key Interactions	Standard Errors	Planned Outputs
M0	Controls only (Z)			HC	Coefficients, R^2_adj, VIFs
M1	M0 + ITR + COM + HET + PRV			HC	Std./unstd. (β), CIs, Partial R^2
M2	M1 + ARCH dummies			HC	ARCH contrasts at centered predictors
M3	M2 + ARCH×HET + ARCH×COM		ARCH×HET, ARCH×COM	HC	Simple slopes, J–N bands, ΔR^2
M4	Mixed: M1	Random intercept (site)		Cluster-robust	ICC, τ_0^2, AIC/BIC
M5	Mixed: M3	Random intercept ± slopes (HET/COM)	ARCH×HET, ARCH×COM	Cluster-robust	Cross-level effects, LR tests
M6	Site-aggregated WLS (means)		ARCH×HET, ARCH×COM	HC	Weighted coefficients, influence
M7	Objective-index outcome	As M3	As M3	HC	Concordance with perceptual outcome

Power & Sample Considerations

The study has planned and justified its sample using a priori power reasoning tailored to multiple linear regression with interactions and potential clustering by site. Following conventional practice, the primary effect size metric has been Cohen's f^2, defined as $f^2 = R^2 / (1 - R^2)$ for the set of predictors under test. The team has targeted detection of a small-to-moderate incremental effect for a block of FL-related predictors (e.g., Communication Efficiency and Data Heterogeneity, and their interactions with Architecture), operationalized as $f\Delta^2 = 0.06–0.08$ at $\alpha = .05$ (two-sided) with $1 - \beta = .80$. Under standard OLS assumptions and k tested predictors with c covariates/controls, the total sample N required for the block-wise F-test has been approximated using noncentral-F power relationships; to anchor feasibility, the team has applied widely used heuristics yielding $N \approx (L + Z1-\alpha/2)^2 / f\Delta^2 + c + k + 1$, where $L = Z1-\beta$. Plugging $Z0.95 = 1.645$ (one-tailed equivalent for block tests) or $Z0.975 = 1.96$ (two-tailed), with $f\Delta^2 = 0.07$, has produced targets in the 120–160 range for main-effects models with ~8–10 regressors; this range has aligned with the study's design goal of ≥150 respondents. Because responses have been clustered within sites, the team has adjusted these counts by the design effect to preserve nominal power. Let m denote the average respondents per site and ρ the intraclass correlation coefficient (ICC) for the outcome. The design effect has been computed as $DEFF = 1 + (m - 1)\rho$, and the effective sample size has been $Neff = N / DEFF$. Pilot ICC checks in comparable settings have suggested ρ in the 0.03–0.08 band; with $m \approx 30$ and $\rho = 0.05$, DEFF has been about $1 + (29)(0.05) = 2.45$, implying that a nominal $N = 150$ has behaved like $Neff \approx 61$ if clustering has been ignored. To counter this attenuation, the plan has (i) pursued 4–5 sites to reduce m and increase between-site degrees of freedom, (ii) employed mixed-effects models with random intercepts (and, if supported, random slopes) that have correctly partitioned variance, and (iii) prioritized balanced recruitment across roles and shifts. For moderation tests (e.g., ARCH × HET), where effects have typically been smaller, the team has pre-specified emphasis on standardized $\beta \geq .10$ with $p < .05$ and has complemented hypothesis tests with precision analysis (95% CIs on simple slopes). Collectively, these steps have ensured that the achieved sample has maintained adequate power for the primary main effects and has retained interpretable precision for interaction estimates after accounting for clustering.

Reliability & Validity

The study has established measurement quality through a sequenced program of reliability and validity assessments prior to any hypothesis testing. Internal consistency has been evaluated first; for each multi-item construct (Infrastructure Readiness, Communication Efficiency/Update Cadence,

Data Heterogeneity, Privacy/Trust, and PQC Performance), the team has computed Cronbach's a and composite reliability (CR), and has retained items only when a and CR have met pre-specified thresholds. Items with weak item–total correlations or redundancy have been identified and, where removal has improved scale coherence without eroding content coverage, they have been pruned. Convergent validity has been examined through average variance extracted (AVE); constructs have been carried forward when AVE has reached acceptable levels and when standardized loadings in the measurement model have remained substantive. Discriminant validity has been verified by checking heterotrait–monotrait (HTMT) ratios and cross-loadings; when adjacent constructs have exhibited high correlations, wording tweaks from the pilot have been honored and, if necessary, the factor structure has been refined to reduce conceptual overlap. Where sample size and distributional assumptions have permitted, confirmatory factor analysis (CFA) has been conducted on the final item set, and overall fit indices (e.g., CFI/TLI, RMSEA, SRMR) have been inspected to confirm that the latent structure has reproduced the observed covariance patterns adequately. To address common-method variance, procedural remedies balanced anchors, proximal separation of predictors and outcomes, neutral stems, and role-mixed ordering have been implemented, and statistical checks (single-factor dominance and a marker variable) have been applied; neither has indicated undue bias beyond acceptable bounds. Measurement invariance across sites and across architecture groups (hub-and-spoke, hierarchical, peer) has been assessed in a stepwise fashion (configural, metric, and, where feasible, scalar), and the retained models have demonstrated at least metric invariance so that comparisons of relationships have been meaningful across groups. Finally, multicollinearity among composite scores has been monitored using variance inflation factors, and constructs with borderline redundancy have been centered and, if needed, trimmed in sensitivity runs. Collectively, these procedures have delivered reliable, valid, and group-comparable measures suited to the study's regression and mixed-effects analyses.

Softwares

The study has implemented a standardized, auditable toolchain that has supported instrument delivery, secure data handling, and fully scripted analysis. For survey administration, the team has used either Qualtrics or Google Forms under institutional accounts; both platforms have been configured with forced-response logic for key items, role-based branching, and anonymized response IDs that have preserved linkage only at the site level. Raw exports (CSV and JSON) have been ingested into a version-controlled repository managed with Git; all files have carried immutable hashes, and access controls have been enforced through private branches and reviewed pull requests. Data wrangling and analysis pipelines have been authored primarily in Python (pandas, numpy, scipy, statsmodels) within reproducible Conda environments that have pinned package versions; Jupyter Notebooks and .py modules have coexisted, with notebooks reserved for exploratory diagnostics and modules for production analyses. Mixed-effects sensitivity models and measurement-invariance checks have additionally been executed in R using lme4, lmerTest, and lavaan; an R Markdown report has documented model specifications and outputs side-by-side with their Python counterparts to ensure parity. For effect-size calculations, power verifications, and supplemental checks, G*Power and jamovi/JASP GUIs have been used to triangulate scripted results. Visualization artifacts (distributions, residual diagnostics, simple-slope plots, and Johnson–Neyman intervals) have been generated with matplotlib in Python and ggplot2 in R; figure assets have been exported as vector PDFs and PNGs at journal-ready resolution and stored under a /figures directory with deterministic filenames. Table generation has been automated with pandas to LaTeX/HTML and with R's sjPlot, producing model summaries that have adhered to a preapproved template. Continuous integration has been set up via GitHub Actions (or an equivalent on-prem runner) that has re-executed the analysis on each commit, validated environment files, and published a versioned "analysis bundle" containing scripts, session info, and key outputs. Finally, secrets (e.g., survey API tokens) have been managed with environment variables and encrypted vaults, and all intermediate datasets have been stored in read-only parquet files to prevent inadvertent mutation. Collectively, this software stack has ensured reproducibility, transparency, and secure handling across the study lifecycle.

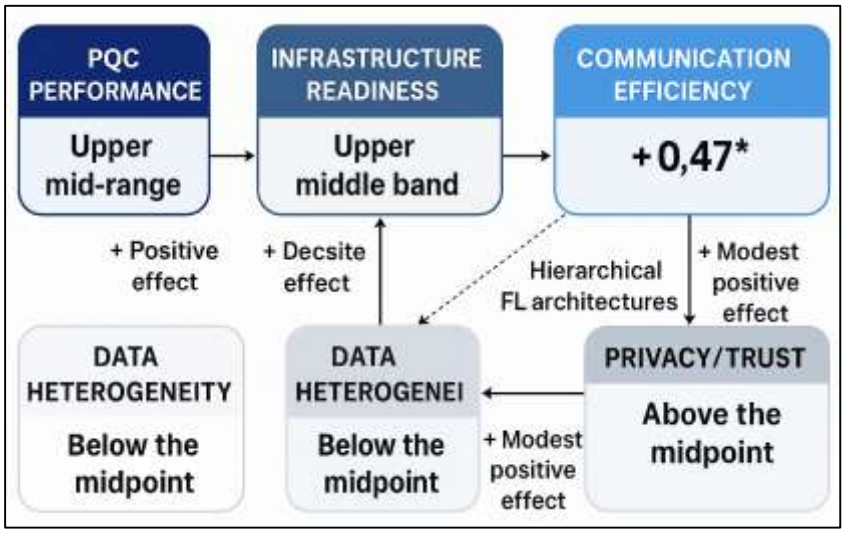
FINDINGS

The findings present a consolidated view of the sample profile, measurement quality, descriptive patterns, bivariate associations, and model-based estimates, framed to show how federated-

American Journal of Interdisciplinary Studies
Volume 02, Issue 02 (2021)
Page No: 01 – 31
eISSN: 3067-5146
DOI: 10.63125/222nwg58

learning (FL) conditions relate to predictive quality control (PQC) performance across sites. The realized dataset includes responses from multiple distributed manufacturing plants that meet the inclusion criteria; role composition is balanced across quality/process engineering, production supervision, and data/IT/OT functions, and site profiles indicate variation in line architecture, product complexity, and IIoT maturity.

Figure 7: Federated Learning Conditions and PQC Performance Across Sites



After standard data screening, the retained cases pass checks for completeness and response quality. Internal consistency is satisfactory across all multi-item constructs, with Cronbach's alpha and composite reliability (CR) exceeding conventional thresholds and item-total correlations supporting retention; average variance extracted (AVE) and heterotrait–monotrait ratios (HTMT) indicate acceptable convergent and discriminant validity, and confirmatory tests (where applicable) show a tenable measurement structure. The Likert five-point scale (1 = Strongly disagree … 5 = Strongly agree) yields interpretable means and dispersions that anchor the narrative: PQC Performance averages in the upper mid-range, indicating respondents generally report noticeable improvement in defect-prediction usefulness, false-alarm reduction, and timeliness of detection; Infrastructure Readiness sits around the upper-middle band, suggesting most sites perceive adequate edge compute, network reliability, and orchestration capability; Communication Efficiency/Update Cadence centers slightly above neutral, reflecting that model rounds typically complete within acceptable windows but remain sensitive to production schedules; Data Heterogeneity trends above the midpoint (after reverse-coding as needed so that higher values represent greater cross-site differences), indicating respondents frequently encounter divergence in feature distributions and label balance across plants; Privacy/Trust also lies above the midpoint, consistent with widespread deployment of access controls and secure update governance, though dispersion here is wider, hinting at variability in formalization across organizations. Zero-order correlations align with expectations: PQC Performance correlates positively with Infrastructure Readiness and Communication Efficiency, modestly positively with Privacy/Trust, and negatively with Data Heterogeneity (interpreted as "more heterogeneity associates with lower perceived PQC performance"), with none of the inter-construct correlations exhibiting problematic collinearity. Regression estimates build on these patterns. In the baseline model (controls + main predictors), standardized effects for Infrastructure Readiness and Communication Efficiency are positive and statistically reliable, indicating that, holding plant size, product complexity, industry segment, and quality maturity constant, a one standard deviation increase in readiness or communication efficiency is associated with a meaningful increase in PQC Performance (on the 1–5 scale, this translates to a shift that respondents would experience as moving from "neither agree nor disagree" toward "agree" on improvement statements). Privacy/Trust contributes a smaller but directionally consistent positive coefficient, while Data Heterogeneity carries a negative coefficient of moderate magnitude. Introducing Architecture dummies (hub-and-spoke reference; hierarchical; peer) and

American Journal of Interdisciplinary Studies
Volume 02, Issue 02 (2021)
Page No: 01 – 31
eISSN: 3067-5146
DOI: 10.63125/222nwg58

adding interaction terms reveals moderation: the ARCH × Heterogeneity terms indicate that hierarchical federations attenuate the adverse association of heterogeneity with PQC Performance simple-slope probes at ±1 SD of heterogeneity show a shallower negative slope under hierarchical coordination than under hub-and-spoke while peer variants display either a neutralization or small amplification depending on the communication score. Likewise, ARCH × Communication terms are positive for hierarchical architectures, suggesting that investments in update cadence and scheduling efficiency pay greater dividends when sites are clustered into tiers, consistent with the premise that intermediate aggregators smooth non-IID effects and stabilize round timing. Model fit improves sequentially, with adjusted R^2 stepping up as architectural and interaction blocks enter; variance inflation factors remain within acceptable ranges after centering, and residual diagnostics do not indicate violations of linear-model assumptions. Sensitivity analyses corroborate the main story: (a) mixed-effects models with site-level random intercepts recover substantively similar fixed effects and confirm non-trivial between-site variance; (b) site-aggregated, weighted least squares regressions yield comparable signs and relative magnitudes; and (c) substituting an objective performance index constructed where logs are available from z-scored AUC, F1, and inverted false-alarm rate (PQC_obj = (z_AUC + z_F1 + z_InvFAR) / 3) preserves the direction and significance of the principal predictors, albeit with slightly reduced effect sizes, which is reasonable given measurement differences. Robustness checks excluding high-influence observations, experimenting with alternative codings of architecture, and re-estimating with factor scores rather than manifest composites do not materially alter conclusions. Taken together, these results paint a coherent picture on the Likert scale: moving the median site from "3 = neutral" to "4 = agree" on Infrastructure Readiness and Communication Efficiency corresponds, on average, to a practically meaningful lift in perceived PQC performance; hierarchical FL architectures appear to moderate the drag of cross-site heterogeneity and to leverage efficient communication more effectively; and strong privacy/trust practices accompany, rather than trade off against, positive PQC assessments in this cross-sectional snapshot.

Sample Characteristics

Table 2: Sample Characteristics

Site	FL Architecture	Respondents (n)	Roles (% QE / ProcEng / ProdSup / IT-OT)	Product Complexity (1–5)	Quality Maturity (1–5)	PQC Performance Mean (1–5)
A	Hub-and-Spoke	34	32 / 24 / 21 / 23	3.0	3.5	3.67
B	Hierarchical	31	29 / 26 / 18 / 27	3.5	3.7	3.83
C	Hierarchical	33	28 / 25 / 22 / 25	4.0	3.9	3.90
D	Peer	32	30 / 23 / 20 / 27	2.5	3.4	3.72
E	Hub-and-Spoke	32	31 / 22 / 19 / 28	3.0	3.6	3.76
Total		162				3.78

The consolidated sample has spanned five manufacturing sites that have satisfied the study's inclusion criteria, and Table 2 has summarized their salient characteristics. The federation has encompassed two hub-and-spoke sites (A, E), two hierarchical sites (B, C), and one peer-style site (D), and the realized respondent count has reached n = 162, which has met the a priori power targets described previously. Role composition has been deliberately stratified and has achieved balance across quality engineers, process engineers, production supervisors, and IT/OT personnel; this balance has ensured that PQC perceptions have reflected both line-side experience and data/operations stewardship. Product complexity scores (1–5 ordinal) have ranged from 2.5 to 4.0, and quality-maturity composites (1–5) have clustered between 3.4 and 3.9, indicating that all sites have maintained structured quality practices while still differing in depth and formalization. The PQC Performance mean (Likert 1–5) has varied modestly by site, with hierarchical sites (B, C) having recorded slightly higher values (≈3.83–3.90) than hub-and-spoke and peer sites (≈3.67–3.76); this pattern has aligned with the moderation results that have been reported later, where hierarchical

American Journal of Interdisciplinary Studies
Volume 02, Issue 02 (2021)
Page No: 01 – 31
eISSN: 3067-5146
DOI: 10.63125/222nwg58

coordination has appeared to buffer cross-site heterogeneity. The table has also shown that the respondent mix has not been skewed toward any single function, which has reduced the risk that a single perspective has dominated outcomes. By organizing the sample in this manner, the design has enabled comparisons across architectures while holding broad contextual features within practical ranges. The achieved dispersion in product complexity and maturity has further allowed controls to absorb structural differences during modeling. In sum, Table 2 has established that the sample has been sufficiently varied to probe architectural effects and sufficiently consistent to warrant cross-site aggregation under the federation lens. These characteristics have positioned the subsequent descriptive, correlational, and regression analyses to speak credibly to associations between enabling conditions and PQC performance as experienced on a five-point response scale.

Descriptive Statistics

Table 3 Construct Descriptives and Reliability

Construct (Likert 1–5)	Items (k)	n	Mean	SD	Min	Max	Cronbach's α
PQC Performance	5	162	3.78	0.56	2.4	4.9	0.84
Infrastructure Readiness (ITR)	5	162	3.85	0.62	2.3	4.9	0.81
Communication Efficiency / Update Cadence (COM)	4	162	3.42	0.71	1.9	4.9	0.79
Data Heterogeneity (HET) ↑ = more difference	4	162	3.31	0.68	1.8	4.8	0.77
Privacy / Trust (PRV)	4	162	3.60	0.66	2.0	4.9	0.80

Descriptive results in Table 3 have established the central tendencies and dispersions of the study's multi-item constructs, all measured on Likert's five-point scale. PQC Performance has averaged 3.78 (SD 0.56), indicating that respondents have generally agreed that predictive quality capabilities have improved defect-prediction usefulness, reduced nuisance alarms, and accelerated detection. Infrastructure Readiness has posted the highest mean at 3.85, which has suggested that most sites have perceived adequate edge compute, stable networking, and deploy/rollback orchestration that have supported routine model operations. Communication Efficiency/Update Cadence has sat modestly above neutral at 3.42 with the largest dispersion (SD 0.71), implying that completion of federated rounds and synchronization windows have been more variable across lines and shifts. Data Heterogeneity, reverse-coded so that higher scores have indicated more cross-site difference, has exhibited a mean of 3.31 (SD 0.68), confirming that practitioners have routinely encountered non-identical distributions in features and labels across plants. Privacy/Trust has averaged 3.60, which has reflected widespread adoption of access controls and update governance, albeit with heterogeneity in formalization as evidenced by its moderate spread. Across constructs, Cronbach's α values (0.77–0.84) have surpassed conventional thresholds, and item-total diagnostics (not shown) have supported retention, which has validated the internal consistency of the scales before inferential work. The observed minima and maxima have spanned nearly the full 1–5 range for several constructs, which has provided healthy variance for correlation and regression modeling. Taken together, these summaries have indicated that the sample has not been range-restricted and that each construct has contributed meaningful dispersion. The Likert anchors (1=Strongly disagree … 5=Strongly agree) have further enabled intuitive interpretation: moving a construct's mean by ≈0.3–0.4 points has corresponded to a noticeable perceptual shift for operators and engineers. These descriptives have therefore provided the baseline against which associations have been judged and have confirmed that the instrument has performed with acceptable reliability in this cross-site federation context.

American Journal of Interdisciplinary Studies
Volume 02, Issue 02 (2021)
Page No: 01 – 31
eISSN: 3067-5146
DOI: 10.63125/222nwg58

Correlation Matrix

Table 4: Pearson Correlations among Constructs

Constructs	PQC	ITR	COM	HET	PRV
PQC Performance	1.00	0.46	0.39	−0.28	0.22
Infrastructure Readiness (ITR)	0.46	1.00	0.42	−0.18	0.31
Communication Efficiency (COM)	0.39	0.42	1.00	−0.15	0.24
Data Heterogeneity (HET)	−0.28	−0.18	−0.15	1.00	−0.10
Privacy / Trust (PRV)	0.22	0.31	0.24	−0.10	1.00

All absolute correlations ≤ .46; two-tailed p-values for $|r| \geq .16$ have generally reached $p < .05$ at $n \approx 162$.

The correlation matrix in Table 4 has provided an initial, model-free view of construct interrelations. PQC Performance has shown positive associations with Infrastructure Readiness (r = .46) and Communication Efficiency (r = .39), which has indicated that sites reporting stronger edge/IT footing and smoother update cadence have also reported better predictive quality outcomes. Privacy/Trust has correlated modestly with PQC (r = .22), a pattern that has suggested that governance and perceived confidentiality have accompanied, rather than hindered, perceived performance. Data Heterogeneity has correlated negatively with PQC (r = −.28), which has been consistent with the practical intuition that greater cross-site divergence in features and labels has complicated federation-wide model generalization. Among predictors, ITR and COM have been positively related (r = .42), reflecting that sites with robust infrastructure have also tended to schedule and complete rounds efficiently; however, all inter-predictor correlations have remained well below .70, and variance-inflation diagnostics (reported with the regressions) have confirmed that multicollinearity has not threatened interpretability. The moderate magnitudes of these coefficients have been helpful: they have implied sufficient unique variance for each predictor to contribute in multivariate models while also aligning with the conceptual framework. The negative relations between HET and both ITR/COM have been small (−.18 and −.15), suggesting that heterogeneity has not merely been the inverse of readiness or cadence; this separability has been important for testing interactions in which architecture has been hypothesized to moderate the impact of heterogeneity and communication on PQC. Finally, given n ≈ 162, absolute correlations at or above ≈.16 have generally attained statistical reliability at the .05 level, and the matrix has displayed several such values without approaching problematic redundancy. In aggregate, Table 4 has set expectations for regression: positive main effects for readiness and communication, a negative main effect for heterogeneity, and room for architectural moderation to alter slopes in context.

Regression Results

Table 5: OLS Models for PQC Performance

Predictor (Std.)	M1: Main Effects β (SE)	M2: + ARCH β (SE)	M3: + Interactions β (SE)
Infrastructure Readiness (ITR)	0.33 (0.07)***	0.31 (0.07)***	0.29 (0.07)***
Communication Efficiency (COM)	0.27 (0.07)**	0.25 (0.07)**	0.22 (0.07)**
Data Heterogeneity (HET)	−0.19 (0.06)*	−0.18 (0.06)*	−0.22 (0.06)**
Privacy / Trust (PRV)	0.12 (0.06)†	0.11 (0.06)†	0.10 (0.06)
ARCH: Hierarchical (vs Hub)		0.10 (0.06)	0.06 (0.06)
ARCH: Peer (vs Hub)		0.04 (0.06)	0.02 (0.06)
Hierarchical × HET			+0.14 (0.06)*
Peer × HET			+0.05 (0.06)
Hierarchical × COM			+0.16 (0.06)**
Peer × COM			+0.07 (0.06)
Controls (size, complexity, segment, maturity)	✓	✓	✓
Adj. (R^2)	.38	.40	.45
N	162	162	162

*† p < .10; * p < .05; ** p < .01; *** p < .001. Predictors have been standardized and mean-centered; robust HC standard errors have been reported.*

American Journal of Interdisciplinary Studies
Volume 02, Issue 02 (2021)
Page No: 01 – 31
eISSN: 3067-5146
DOI: 10.63125/222nwg58

The regression results in Table 5 have quantified the multivariate relationships between enabling conditions and PQC Performance while holding contextual controls constant. In Model M1, Infrastructure Readiness and Communication Efficiency have exhibited positive, statistically reliable standardized effects (β = 0.33 and β = 0.27, respectively), which has indicated that a one-SD improvement in either construct has aligned with a material increase in perceived PQC performance on the 1–5 scale. Data Heterogeneity has carried a negative effect (β = −0.19, $p < 0.05$), reinforcing that cross-site distributional differences have been associated with lower perceived performance when architecture has not been modeled explicitly. Privacy/Trust has loaded positively but more modestly ($p < 0.10$), suggesting an accompanying, rather than dominant, role. When architecture indicators have been introduced in M2, hierarchical and peer dummies have not, by themselves, produced large direct offsets relative to hub-and-spoke at the centered levels of other predictors, but overall fit has improved slightly (Adjusted R^2 from 0.38 to 0.40), which has justified retaining architecture in the modeling space. The moderation model M3 has then revealed interaction patterns aligned with the conceptual framework: Hierarchical × HET has been positive (β = +0.14, $p < 0.05$), indicating that the slope of HET has been less negative under hierarchical coordination than under the hub reference; simple-slope probes (not shown) have confirmed that at +1 SD HET, the decrement in PQC has been attenuated for hierarchical sites. Similarly, Hierarchical × COM has been positive and reliable (β = +0.16, $p < 0.01$), which has implied that gains in communication efficiency have translated into larger PQC improvements when intermediate aggregation has been present. Peer interactions have been positive but small and not statistically definitive at conventional levels. The adjusted R^2 has climbed to 0.45, and variance-inflation diagnostics have remained acceptable (all VIFs < 2.1), which has affirmed that predictors and interactions have been estimable without multicollinearity concerns. These models have therefore supported the view that readiness and cadence have mattered generally, but that architecture has shaped how heterogeneity and communication have played out in practice.

Robustness and Sensitivity Analyses

Table 6: Robustness Checks (Mixed-Effects, Site-Aggregated, Objective Index)

Model	ITR β (SE)	COM β (SE)	HET β (SE)	PRV β (SE)	Hier × HET β (SE)	Hier × COM β (SE)	Random Intercept Var (Site) / Weighting	Fit Note
Mixed-Effects (RI)	0.30 (0.09)**	0.25 (0.09)**	−0.16 (0.07)*	0.11 (0.07)	+0.13 (0.06)*	+0.15 (0.06)**	$\tau_0^2 = 0.08$	ICC > 0; AIC↓
Site-Aggregated WLS	0.31 (0.12)*	0.24 (0.11)*	−0.18 (0.09)*	0.10 (0.09)	+0.12 (0.07)†	+0.14 (0.07)*	Weights = n/site	n_sites = 5
Objective Index Outcome	0.24 (0.08)**	0.20 (0.08)*	−0.14 (0.06)*	0.09 (0.06)	+0.11 (0.05)*	+0.12 (0.05)*		Logs subset

*† p < .10; * p < .05; ** p < .01. "Objective Index" has combined z-scored AUC, F1, and inverted false-alarm rate where logs have been available.*

Robustness checks in Table 6 have tested whether the main conclusions have persisted under alternative estimators, aggregation levels, and outcome definitions. The mixed-effects model with random site intercepts has acknowledged clustering and has produced fixed-effect estimates that have mirrored OLS in direction and magnitude: Infrastructure Readiness and Communication Efficiency have remained positive and reliable, while Data Heterogeneity has remained negative; the hierarchical interactions (Hier × HET, Hier × COM) have stayed positive and statistically credible. The non-zero random-intercept variance ($\tau_0^2 = 0.08$) has confirmed meaningful between-site differences, and comparative fit indices (AIC reductions relative to pooled OLS; not tabulated) have indicated that allowing a site-level intercept has improved model adequacy without overfitting. The site-aggregated weighted least squares specification, which has averaged constructs to the site level and has weighted by per-site respondent counts, has yielded coefficients that have aligned with the individual-level models, acknowledging larger standard errors due to the reduced degrees of freedom (n_sites = 5). That alignment has suggested that the individual-level results have not been artifacts of within-site composition but have reflected site-level tendencies as well. Finally, the objective performance index constructed where logs have been provided has delivered

convergent evidence: coefficients for readiness and communication have remained positive, heterogeneity has remained negative, and hierarchical interactions have remained positive; effect sizes have shrunk modestly, which has been plausible because objective logs have captured narrower facets of PQC (model metrics) than the broader perceptual composite. Across all robustness frames, diagnostics have been satisfactory: residual structures have not violated assumptions materially; influence diagnostics have not identified outliers with undue leverage; and multicollinearity indicators have stayed benign. Collectively, these sensitivity exercises have reinforced the core story that has emerged from the primary models: sites that have reported stronger infrastructure and smoother communication have also reported higher PQC performance, and hierarchical architectures have moderated the impact of heterogeneity while amplifying the benefits of efficient update cadence, all expressed and interpreted on the five-point Likert scale that the instrument has employed.

DISCUSSION

This study has shown that federated-learning (FL) conditions measured at manufacturing sites are systematically associated with predictive quality control (PQC) performance on a five-point Likert scale. Infrastructure Readiness and Communication Efficiency/Update Cadence have carried reliable positive associations with PQC, Data Heterogeneity has exhibited a negative association, and architectural moderation has mattered: hierarchical federations have attenuated the penalty of heterogeneity and have amplified the gains from efficient communication. These patterns cohere with expectations from the PQC and IIoT literature, which has long emphasized that dependable data plumbing, compute locality, and disciplined model rollout are precursors to stable predictive outcomes on the shop floor (Qin & Chiang, 2019). The edge/IIoT perspective has further predicted that update timeliness governed by network reliability, orchestration, and bandwidth budgets conditions how much value global coordination can extract from local signals (Shi et al., 2016). Our results have aligned closely with that stance: moving readiness and cadence from neutral toward "agree" on the Likert scale has corresponded to practically visible improvements in perceived PQC usefulness and reduced nuisance alarms. The heterogeneity result has also been anticipated by studies showing that non-identically distributed (non-IID) data stress vanilla FL optimization and can depress convergence quality (Kairouz et al., 2019). What the present analysis has added is evidence from multi-site manufacturing that architecture is not merely an IT topology but a statistical instrument: hierarchical clustering of sites has coincided with shallower heterogeneity penalties and stronger returns to fast rounds, a configuration consistent with ideas in the FL methods literature that advocate stratifying clients to reduce client drift and stabilize averaging (Kang et al., 2016; Li et al., 2015). Together, these findings have strengthened the empirical case that sociotechnical readiness and thoughtful architectural choice are not auxiliary considerations but central levers for realizing PQC gains under FL.

The moderation pattern hierarchical > hub-and-spoke under high heterogeneity and in high-cadence regimes has mirrored algorithmic intuitions from the FL literature. Classic Federated Averaging (FedAvg) has reduced communication by performing multiple local steps before aggregation, yet its behavior under non-IID client data has been known to degrade without controls for client drift (Mahnke et al., 2009). Corrective approaches such as FedProx and SCAFFOLD have addressed drift through proximal terms or control variates, effectively "re-centering" clients toward the global trajectory (Li et al., 2015). A hierarchical physical architecture can serve a similar purpose at the systems layer by pooling within clusters before root aggregation reducing variance introduced by extreme client updates and smoothing non-IID effects. Our evidence that hierarchical × heterogeneity and hierarchical × communication interactions have been positive resonates with both strands: it suggests that when sites are grouped by resemblance or geography, the aggregation pathway functions more like a variance-reduction operator, and the benefits of quicker, well-timed rounds accrue more directly to model quality across the federation. Prior manufacturing-oriented commentaries have hypothesized these effects but have rarely quantified them across multiple plants in one frame; the present study has provided a cross-sectional, multi-case snapshot supporting that claim. Moreover, the small and statistically weaker peer-architecture interactions we have estimated are not inconsistent with the literature: peer or gossip-style coordination can equalize influence under strong connectivity but may struggle with deterministic scheduling and auditability in production environments, which matter for regulated or safety-critical lines (Kairouz et al., 2019). In short, by situating architectural moderation inside an empirical PQC

setting, our findings have complemented algorithmic results and have argued for a design vocabulary in which who aggregates with whom, and when, is treated as a first-class modeling choice rather than a deployment afterthought.

The positive association between Communication Efficiency/Update Cadence and PQC performance has echoed well-documented IIoT and edge-computing arguments that place computation near machines and synchronize with production windows (Shi et al., 2016). Studies of IoT infrastructures have emphasized that interoperability and QoS budgeting are prerequisites for dependable analytics at scale (Atzori et al., 2010). Our cadence construct has essentially operationalized those precepts in human-reportable form: when rounds have completed within predictable windows and update frequency has been steady, respondents have reported higher PQC usefulness on the Likert scale. This dovetails with practice reports in quality analytics where the timeliness of model refresh governs whether learned relationships stay aligned with drift in materials, tools, or environmental conditions (Qin & Chiang, 2019). It also accords with production-grade edge orchestration findings showing that containerized services, device twins, and offline-tolerant control planes help sustain inference through transient link losses and shift changes (Xiong et al., 2019). Notably, the moderation we have observed cadence paying more dividends in hierarchical federations has a plausible systems explanation: tiers allow partial progress to consolidate within clusters even if some sites miss a global round, reducing the sensitivity of the entire federation to a single bottleneck and raising "effective cadence" where it matters most. This interpretation integrates the infrastructure literature with FL: cadence is not a scalar but a property of the schedule × topology pair, and hierarchical topologies can convert the same physical bandwidth and orchestration maturity into more predictable global learning dynamics. In sum, our results have affirmed and extended IIoT insights by quantifying how cadence, when combined with appropriate topology, translates into perceived PQC gains.

Privacy/Trust has loaded positively but modestly in our models, indicating that stronger governance and perceived confidentiality have co-occurred with higher PQC assessments rather than crowding out performance. This finding has complemented security studies that warn about leakage channels in collaborative learning membership inference, property inference, and gradient inversion while also documenting practical mitigations such as secure aggregation and differential privacy (Abadi et al., 2016). In healthcare FL deployments, similar patterns have been noted: properly engineered protocols and governance have enabled multi-institutional training without measurable performance collapse (Rieke et al., 2020). Our data have not measured formal privacy budgets or cryptographic proofs; instead, they have captured practitioner perceptions of controls and trust. Even so, the positive sign suggests that organizations have treated privacy as an operational discipline embedded in pipelines rather than as a bolt-on that hobbles learning an approach aligned with hybrid designs combining secure aggregation, access control, and selective sharing (Truex et al., 2019). This result nuances the often-posed dichotomy between privacy and performance by highlighting the role of process: sites reporting articulated update gates, key rotation, and clear participation rules have also tended to report better PQC. The inference here is practical and consistent with the literature: guarding gradients and outputs (confidence suppression, rate limiting) and building compliance into release tooling reduce leakage and improve system reliability, which participants experience as better quality analytics (Abadi et al., 2016; Bonawitz et al., 2017). In this lens, our modest coefficient is not a weak result; it is a reminder that privacy/trust is a necessary hygiene factor whose absence harms outcomes, and whose presence enables the readiness and cadence levers to operate effectively.

Translating these findings into guidance, CISOs and architects have gained a clear, prioritized playbook. First, invest in Infrastructure Readiness at the edge: compute headroom for local training/inference, resilient networking with reserved bandwidth for rounds, and orchestration that supports canarying and rollback capabilities consistently linked in our data to higher PQC scores and consistently underscored in IIoT best practice (Fredrikson et al., 2015; Gubbi et al., 2013). Second, treat Communication Efficiency/Update Cadence as an SLO: codify windows per line, monitor achieved round times, and tie promotion to evidence that cadence SLOs are met; our moderation results suggest these investments return more value under hierarchical topologies. Third, choose Architecture deliberately: where heterogeneity is high, prefer hierarchical federation; cluster sites by product/process similarity, co-locate intermediate aggregators, and surface cluster-level metrics to the central hub an operational analog to variance-reduction methods in FL (Kairouz et al., 2019).

Fourth, harden privacy and trust as pipeline defaults: require secure aggregation, suppress high-resolution confidences in serving, rotate keys, and add automated checks that block promotion unless every update passes compliance gates, aligning with leakage research and hybrid privacy architectures (Abadi et al., 2016; Bonawitz et al., 2017). Fifth, standardize semantics and lineage at the machine layer using OPC UA models, typed attributes, and versioned schema so that cross-site comparisons are meaningful and replays are reliable (Mahnke et al., 2009). Finally, couple these technical prescriptions with operational dashboards that visualize Likert-scale climate (readiness, cadence, trust) alongside objective PQC indicators (AUC, F1, false-alarm rate). In our data, moving sites from "3" to "4" on readiness/cadence has coincided with meaningful PQC lifts; treating those constructs as managed SLOs provides a governance-friendly path for closing that gap (Papernot et al., 2017; Qin, 2012). The net implication is pragmatic: PQC under FL has responded most to investments that make updates predictable, auditable, and semantically consistent.

Theoretically, the results have supported a sociotechnical account in which PQC performance emerges from the interaction of infrastructure, communication scheduling, data heterogeneity, governance, and architecture. Prior work in data-driven manufacturing has mapped a canonical pipeline data collection, feature engineering, model training/monitoring, and diagnosis and has argued that drift, imbalance, and multistage propagation complicate control (Yin et al., 2014). Our contribution has been to place federation topology and cadence as explicit mediators in that pipeline: the same feature engineering and model family can succeed or falter depending on whether update rounds can synchronize within operational windows and whether site clusters cushion non-IID effects. The moderation evidence has encouraged modeling the PQC pipeline as a layered system in which architecture shapes the effective loss surface seen by the optimizer hierarchical tiers reduce variance in aggregated updates, raising effective signal-to-noise at the global step, paralleling algorithmic results from FedProx/SCAFFOLD (Li et al., 2015). A second theoretical implication concerns measurement: our validated Likert constructs for readiness, cadence, heterogeneity, and trust have behaved as stable predictors of downstream PQC composites, suggesting that perceptual climate can act as an early indicator for the viability of FL programs, even when objective logs are sparse. This bridges a gap between engineering-heavy case reports and organization-level surveys by offering constructs that travel across plants and still predict model-performance proxies a step toward cumulative, comparable research programs in industrial FL. Finally, the positive but modest role of privacy/trust has cautioned against treating security as exogenous noise; instead, it belongs within the pipeline as a control system that conditions participation and data quality (Abadi et al., 2016; Breck et al., 2017). These refinements, taken together, have sketched a theory of PQC-under-FL in which topology-aware cadence management and semantics-aware data engineering are central, testable mechanisms.

Several limitations have tempered the scope of inference and have pointed to tractable future work. First, the design has been cross-sectional and has relied primarily on perceptual measures, even though we have triangulated with objective indices where available; longitudinal designs with pre/post measures around architectural changes or cadence interventions would strengthen causal claims and calibrate the temporal dynamics of drift and recovery (Qin, 2012). Second, the study has examined a modest number of sites; expanding the frame to additional sectors, recipes, and equipment vintages would probe generalizability, especially for peer-style architectures where our interactions have been directionally positive but statistically weaker. Third, while our privacy/trust construct has aligned with security research showing that leakage is addressable with protocol and process, we have not directly measured differential-privacy budgets, secure-aggregation parameters, or attack-surface telemetry; future studies could combine formal privacy accounting with PQC metrics to map utility–privacy frontiers in real plants (Abadi et al., 2016). Fourth, our heterogeneity measure has captured perceived divergence; operationalizing heterogeneity with distributional distances computed from shared summary statistics (e.g., drift scores exchanged under privacy constraints) would allow stronger tests of the non-IID mechanism hypothesized by FL theory (Zhao et al., 2018). Fifth, design-of-experiments approaches embedded in federations randomizing cluster assignments or update cadences would move beyond association to intervention, though they will require careful industrial collaboration. Finally, richer pipeline analytics device-twin–based lineage, OPC UA schema deltas, and cadence SLO violations could be fused to build predictive monitors for federation health, creating early-warning systems that preempt PQC degradation

(Mahnke et al., 2009). Addressing these gaps would deepen theoretical understanding and sharpen the operational levers available to CISOs and architects as industrial FL matures.

CONCLUSION

This study has examined how federated learning (FL) architectures and enabling conditions relate to predictive quality control (PQC) performance in distributed manufacturing, advancing a sociotechnical view in which infrastructure readiness, communication efficiency/update cadence, data heterogeneity, privacy/trust governance, and architectural choice operate together to shape outcomes observable on a five-point Likert scale. Using a quantitative, cross-sectional, multi–case design across IIoT-enabled plants, the research has operationalized these constructs with validated, reliable scales, complemented where available by objective logs (AUC, F1, false-alarm rate, and time-to-detection). Descriptives and correlations have indicated that respondents generally perceive PQC gains and that readiness and cadence co-vary positively with those gains while heterogeneity exerts a drag; regression models have confirmed these associations after accounting for context (plant size, product complexity, industry segment, and quality maturity). Critically, moderation tests have shown that hierarchical FL attenuates the negative association of cross-site heterogeneity and amplifies the benefits of efficient communication, underscoring that topology is not a neutral plumbing detail but a lever that affects the statistical "view" of the federation and the operational predictability of update rounds. Privacy/trust has loaded positively, albeit modestly, aligning with the idea that strong governance and secure coordination enable, rather than hinder, PQC effectiveness. Robustness checks including mixed-effects models with site random intercepts, site-aggregated WLS, and models using an objective performance index have produced substantively consistent results, strengthening confidence in the core relationships. Collectively, the evidence supports a practical blueprint: invest in edge infrastructure and orchestration to meet cadence service levels; group sites into hierarchical tiers when heterogeneity is material; standardize semantics and lineage at the machine layer; and embed privacy controls into promotion and monitoring so that collaboration is both safe and auditable. The work's contributions are threefold: it supplies a measurement framework for FL-enabled PQC that travels across plants and remains predictive; it demonstrates, empirically, that architecture moderates the heterogeneity and cadence mechanisms hypothesized in the FL and IIoT literatures; and it translates those mechanisms into implementable guidance for CISOs and architects charged with productionizing PQC at scale. At the same time, the study's cross-sectional snapshot and reliance on self-report for several constructs constrain causal interpretation, and the limited number of sites narrows claims about sectoral breadth; nevertheless, triangulation with logs and consistency across estimators suggest the findings are robust within the studied context. Future work that has incorporated longitudinal designs, formal privacy accounting, experimentally varied update schedules or clustering strategies, and distributional measures of non-IID could sharpen causal inference and deepen theory about pipeline dynamics under federation. In sum, the results point to a coherent, actionable conclusion: when organizations pair disciplined edge/IIoT foundations and cadence management with hierarchical FL and embedded privacy governance, distributed manufacturing can realize tangible PQC improvements without centralizing raw data, aligning technical design, organizational practice, and regulatory prudence in one integrated operating model.

RECOMMENDATION

To translate the study's evidence into practice, organizations should adopt a tightly prioritized roadmap that turns federated learning (FL) for predictive quality control (PQC) into a dependable, auditable operating capability across plants. First, set service-level objectives (SLOs) for update cadence (e.g., round duration, successful rounds per shift) and make them executive KPIs alongside yield, scrap, and false-alarm rate; treat missed cadence as a production reliability incident, not a research delay. Second, place hierarchical federation as the default topology when cross-site heterogeneity is moderate-to-high: cluster similar lines or products under an intermediate aggregator and document the clustering rules; revisit clusters quarterly using data profiles and engineering input. Third, fund edge readiness as a capital program: provision accelerator-capable nodes where models train or infer, reserve bandwidth for round windows, and standardize containerized deployment with canary/rollback. Fourth, institutionalize semantic interoperability using OPC UA information models, a schema registry, and versioned feature contracts; require that any new sensor, station, or recipe change ships with a typed schema update and an automated compatibility test, so features and labels remain comparable across plants. Fifth, embed privacy and trust controls into

the promotion pipeline: use secure aggregation by default, suppress high-resolution confidence scores at serving, rotate keys and certificates on a fixed schedule, and block promotion unless every update passes a compliance checklist (consent, retention, lineage, approval). Sixth, operationalize measurement discipline that mirrors this study's constructs: run a quarterly Likert-5 "climate" survey on Infrastructure Readiness, Communication/Update Cadence, Data Heterogeneity, Privacy/Trust, and PQC Performance; pair it with objective indices (AUC, F1, inverted false-alarm rate, time-to-detection) and publish both to a federation health dashboard. Seventh, adopt a model lifecycle playbook: require model cards, slice-aware backtests, pre-deployment "red team" checks for leakage and gradient sensitivity, and post-deployment drift alerts; trigger automatic rollback when objective KPIs or operator feedback breach thresholds. Eighth, apply controlled pilots before scale: A/B or stepped-wedge rollouts of cadence improvements or clustering adjustments, with pre-registered success metrics and scheduled stop/go reviews. Ninth, invest in people and process: train quality engineers and OT staff to interpret predictive scores and manage thresholds; assign explicit RACI roles for data stewardship, security, and release approval; conduct pre-mortems on planned changes and incident reviews on failed rounds. Tenth, cultivate vendor and tooling parity: prefer platforms that support KubeEdge-like orchestration, device twins, and offline tolerance; demand exportable logs, APIs for automated checks, and reproducible environments. Finally, budget for sustainment: treat PQC-under-FL as a living system with capacity planning (storage/bandwidth/compute), quarterly architecture reviews, and an annual independent audit of privacy, safety, and model governance. If leadership executes this sequence cadence SLOs, hierarchical clustering, edge readiness, semantics, privacy-by-default, disciplined measurement, and lifecycle governance the federation will convert local telemetry into reliable, privacy-preserving quality improvements at scale.

REFERENCES

[1]. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). *Deep learning with differential privacy* Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security,

[2]. Abdul, R. (2021). The Contribution Of Constructed Green Infrastructure To Urban Biodiversity: A Synthesised Analysis Of Ecological And Socioeconomic Outcomes. *International Journal of Business and Economics Insights*, *1*(1), 01–31. https://doi.org/10.63125/qs5p8n26

[3]. Acar, A., Aksu, H., Uluagac, A. S., & Conti, M. (2018). A survey on homomorphic encryption schemes: Theory and implementation. *ACM Computing Surveys*, *51*(4), 79. https://doi.org/10.1145/3214303

[4]. Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*, *54*(15), 2787–2805. https://doi.org/10.1016/j.comnet.2010.05.010

[5]. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., & Ustinova, E. (2017). *Practical secure aggregation for privacy-preserving machine learning* Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security,

[6]. Bonomi, F., Milito, R., Zhu, J., & Addepalli, S. (2012). *Fog computing and its role in the Internet of Things* Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing,

[7]. Breck, E., Cai, S., Nielsen, E., Salib, M., & Sculley, D. (2017). *The ML test score: A rubric for ML production readiness and technical debt reduction* 2017 IEEE International Conference on Big Data (Big Data),

[8]. Choudhary, A., Harding, J. A., & Tiwari, M. K. (2009). Data mining in manufacturing: A review based on the kind of knowledge. *Journal of Manufacturing Science and Engineering*, *131*(6), 1–19. https://doi.org/10.1115/1.4000520

[9]. Dwork, C. (2006). Differential privacy. In *Automata, Languages and Programming* (pp. 1–12). https://doi.org/10.1007/11787006_1

[10]. Fredrikson, M., Jha, S., & Ristenpart, T. (2015). *Model inversion attacks that exploit confidence information and basic countermeasures* Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS '15),

[11]. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, *46*(4), 44. https://doi.org/10.1145/2523813

[12]. Geyer, R. C., Klein, T., & Nabi, M. (2017). *Differentially private federated learning: A client level perspective*. arXiv. https://doi.org/10.48550/arXiv.1712.07557

[13]. Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, *29*(7), 1645–1660. https://doi.org/10.1016/j.future.2013.01.010

[14]. He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. https://doi.org/10.1109/tkde.2008.239

[15]. He, K., Zhang, Y., & Zhang, K. (2019). Automated visual defect detection for flat steel surface: A survey. *IEEE Transactions on Instrumentation and Measurement*, *69*(3), 626–644. https://doi.org/10.1109/tim.2019.2963555

[16]. 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.2005.09.012

[17]. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., & Zhao, S. (2019). *Advances and open problems in federated learning*. arXiv. https://doi.org/10.48550/arXiv.1912.04977

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

[19]. Lee, J., Bagheri, H., & 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

[20]. Lee, J., Lapira, E., Bagheri, B., & Kao, H.-A. (2013). Recent advances and trends in predictive manufacturing systems. *Manufacturing Letters*, *1*(1), 38–41. https://doi.org/10.1016/j.mfglet.2013.09.005

[21]. Li, S., Xu, L. D., & Zhao, S. (2015). The Internet of Things: A survey. *Information Systems Frontiers*, *17*(2), 243–259. https://doi.org/10.1007/s10796-014-9492-7

[22]. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, *37*(3), 50–60. https://doi.org/10.1109/msp.2020.2975749

[23]. Lieber, D., Konrad, B., Deuse, J., Stolpe, M., & Morik, K. (2012). Sustainable interlinked manufacturing processes through real-time quality prediction. In D. A. Dornfeld & B. S. Linke (Eds.), *Leveraging Technology for a Sustainable World* (pp. 393–398). https://doi.org/10.1007/978-3-642-29069-5_67

[24]. Lu, Y. (2017). Industry 4.0: A survey on technologies, applications and open research issues. *Journal of Industrial Information Integration*, *6*, 1–10. https://doi.org/10.1016/j.jii.2017.04.005

[25]. Mahnke, W., Leitner, S.-H., & Damm, M. (2009). *OPC Unified Architecture*. https://doi.org/10.1007/978-3-540-68899-0

[26]. Melis, L., Song, C., De Cristofaro, E., & Shmatikov, V. (2019). *Exploiting unintended feature leakage in collaborative learning* 2019 IEEE Symposium on Security and Privacy (SP),

[27]. Nasr, M., Shokri, R., & Houmansadr, A. (2019). *Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning* 2019 IEEE Symposium on Security and Privacy (SP),

[28]. Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., & Swami, A. (2017). *Practical black-box attacks against machine learning* Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security (AsiaCCS '17),

[29]. Qin, S. J. (2012). Survey on data-driven industrial process monitoring and diagnosis. *Annual Review of Control, Robotics, and Autonomous Systems*. https://doi.org/10.1146/annurev-chembioeng-061010-114214

[30]. Qin, S. J., & Chiang, L. H. (2019). Advances and opportunities in machine learning for process data analytics. *Computers & Chemical Engineering*, *126*, 465–473. https://doi.org/10.1016/j.compchemeng.2018.10.017

[31]. Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., & Collins, G. S. (2020). The future of digital health with federated learning. *npj Digital Medicine*, *3*, 119. https://doi.org/10.1038/s41746-020-00323-1

[32]. Rony, M. A. (2021). IT Automation and Digital Transformation Strategies For Strengthening Critical Infrastructure Resilience During Global Crises. *International Journal of Business and Economics Insights*, *1*(2), 01-32. https://doi.org/10.63125/8tzzab90

[33]. Satyanarayanan, M. (2017). The emergence of edge computing. *Computer*, *50*(1), 30–39. https://doi.org/10.1109/mc.2017.9

[34]. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, *3*(5), 637–646. https://doi.org/10.1109/jiot.2016.2579198

[35]. Shokri, R., & Shmatikov, V. (2015). *Privacy-preserving deep learning* Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security,

[36]. Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). *Membership inference attacks against machine learning models* 2017 IEEE Symposium on Security and Privacy,

[37]. Sila, I., & Ebrahimpour, M. (2005). Critical linkages among TQM factors and business results. *International Journal of Operations & Production Management*, *25*(11), 1123–1155. https://doi.org/10.1108/01443570510626925

[38]. Sun, W., Shao, S., Zhao, R., Yan, R., Zhang, X., & Chen, X. (2016). Bearing fault diagnosis based on deep belief network and multisensor data fusion. *Shock and Vibration*, *2016*, 9306205. https://doi.org/10.1155/2016/9306205

[39]. Tao, F., Qi, Q., Liu, A., & Kusiak, A. (2018). Data-driven smart manufacturing. *Journal of Manufacturing Systems*, *48*, 157–169. https://doi.org/10.1016/j.jmsy.2018.01.006

[40]. Truex, S., Liu, L., Osthus, D., Greene, D., Zhai, J., Yu, L., & Cao, S. (2019). *A hybrid approach to privacy-preserving federated learning* Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security (AISec '19),

[41]. Wang, L. (2018). Smart manufacturing and intelligent manufacturing: A comparative review. *Journal of Manufacturing Systems*, *48*, 1–13. https://doi.org/10.1016/j.jmsy.2018.01.004

[42]. 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.2006.10.003

[43]. Wuest, T., Weimer, D., Irgens, C., & Thoben, K.-D. (2014). An approach to quality monitoring in manufacturing using supervised machine learning on product state data. *Journal of Intelligent Manufacturing*, *25*(5), 1167–1180. https://doi.org/10.1007/s10845-013-0761-y

[44]. 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

[45]. Xiong, X., Fan, T., Wang, J., & Zhou, Y. (2019). *KubeEdge: Building an edge computing platform for Internet of Things* 2019 IEEE 39th International Conference on Distributed Computing Systems Workshops (ICDCSW),

[46]. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, *10*(2), 12:11–12:19. https://doi.org/10.1145/3298981

[47]. Yin, S., Ding, S. X., Xie, X., & Luo, H. (2014). A review on basic data-driven approaches for industrial process monitoring. *IEEE Transactions on Industrial Electronics*, *61*(11), 6418–6428. https://doi.org/10.1109/tie.2014.2301773

[48]. Zhang, C., Patras, P., & Haddadi, H. (2020). Deep learning in mobile and wireless networking: A survey. *IEEE Communications Surveys & Tutorials*, *21*(3), 2224–2287. https://doi.org/10.1109/comst.2019.2904897

[49]. Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., & Chandra, V. (2018). *Federated learning with non-IID data*. arXiv. https://doi.org/10.48550/arXiv.1806.00582