American Journal of Advanced Technology and Engineering Solutions

Volume: 1; **Issue:** 4
Pages: 01-42
eISSN: 3067-5146

INTEGRATION OF MACHINE LEARNING MODELS AND ADVANCED COMPUTING FOR REDUCING LOGISTICS DELAYS IN PHARMACEUTICAL DISTRIBUTION

Md. Rabiul Karim[1]; Samia Akter[2];

[1]. *Assistant Manager, Senior Executive, MGH Group (FedEx), Dhaka, Bangladesh.*
Email: rabiulkarim2024@gmail.com

[2]. *Master of Business Studies, National University Bangladesh, Gazipur, Bangladesh.*
Email: musaraat75@gmail.com

Doi: 10.63125/ahnkqj11
Received: *09 September 2021;* ***Revised:*** *10 October 2021;* ***Accepted:*** *12 November 2021;* ***Published:*** *08 December 2021*

Abstract

This study examined how integrating machine learning models with advanced computing–enabled prescriptive controls reduced logistics delays in pharmaceutical distribution, with particular attention to cold-chain reliability. The conceptual and measurement framework was grounded in an evidence base of 58 peer-reviewed quantitative papers that informed delay constructs, cold-chain risk measures, and predictive–prescriptive integration logic. A retrospective, multi-source observational dataset was analyzed, consisting of 4,820 shipment episodes across 26 transportation corridors and 8 third-party carriers, covering both ambient and cold-chain products. Ambient flows represented 68.4% of shipments and cold-chain flows 31.6%, enabling criticality-stratified analysis. Descriptive results indicated clear schedule deviation: mean planned lead time was 42.6 hours (SD=11.9), mean actual lead time was 49.8 hours (SD=17.4), and mean delay was 7.2 hours (SD=9.6). The delay distribution was right-skewed with pronounced tail risk, showing a P90 delay of 18.7 hours and P95 delay of 26.4 hours. Node-wise dwell decomposition showed the largest time accumulation at distribution centers (6.8 hours) and customs/regulatory nodes (5.7 hours), confirming multi-echelon delay formation. Baseline performance before integration recorded OTIF of 83.9%, lead-time variance of 92.1, coefficient of variation of 0.36, emergency shipment frequency of 7.4%, and cold-chain excursion incidence of 2.9%. Correlation and multicollinearity checks supported the retained predictor structure, with congestion and corridor volatility most strongly associated with lateness. Regression findings showed lane volatility and queue pressure as significant positive predictors of delay (β=2.41 and β=3.09, $p<0.001$), while carrier reliability reduced delay (β=-1.68, $p<0.001$). The integrated predictive–prescriptive mechanism was significantly associated with lower delay magnitude (β=-2.27, $p<0.001$), lower lead-time variance (β=-5.91, $p<0.001$), improved OTIF, and reduced severe-delay odds (OR=0.58, $p<0.001$). For cold-chain shipments, severe delay substantially increased excursion risk (OR=2.36, $p<0.001$), demonstrating coupled time–temperature vulnerability. Overall, the results confirmed that ML-driven risk forecasts embedded into prescriptive decision rules corresponded with measurable improvements in timeliness and cold-chain safety across pharmaceutical distribution networks.

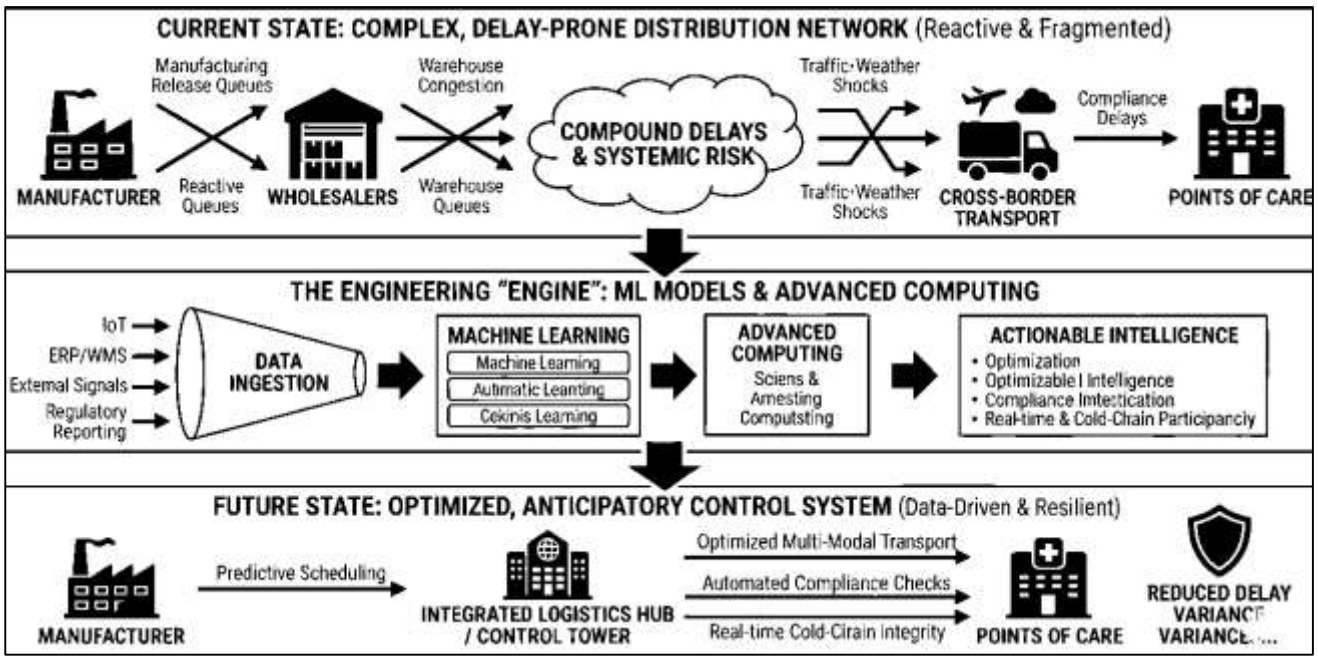
Keywords

Pharmaceutical logistics; Machine learning; Cold-chain; Delay mitigation; Prescriptive analytics

INTRODUCTION

Pharmaceutical distribution refers to the coordinated set of physical, informational, and regulatory processes through which medicines move from manufacturers to end users through wholesalers, third-party logistics providers, hospitals, clinics, and retail pharmacies (Jordon et al., 2019). Logistics delays within this domain are deviations from planned lead times in procurement, warehousing, transportation, customs clearance, cross-docking, and last-mile delivery that reduce the ability of a supply system to deliver the right product, in the right quantity, at the right time, and under the right conditions.

Figure 1: Pharma Logistics Optimization



In pharmaceuticals, delay is not merely a cost or efficiency issue; it is a risk variable embedded in patient safety, national health security, and public trust. The definition of machine learning models in this setting centers on algorithmic systems that learn statistical patterns from historical and real-time logistics data to predict, classify, or optimize outcomes such as delivery times, route disruptions, inventory stockouts, cold-chain excursions, or demand shocks. Machine learning differs from classical optimization by enabling nonlinear relationships, high-dimensional interactions, and adaptive recalibration as new data arrive (Asadi et al., 2015). Advanced computing for logistics includes high-performance computing, cloud-native architectures, edge analytics, Internet of Things (IoT) sensor networks, digital twins, and GPU-accelerated simulation environments that expand the speed, scale, and granularity of decision-support (Arfan et al., 2021; Jahid, 2021). Integration in this paper's context means a tightly coupled operational pipeline where ML outputs feed computing-driven optimization or automation layers, and these layers, in turn, generate new data that refine ML performance.

This integration produces an end-to-end system that moves from monitoring to prediction to action under complex constraints. The pharmaceutical distribution environment makes such integration uniquely demanding because it must simultaneously respect service-level agreements, temperature thresholds, batch traceability, regulatory reporting, controlled substance handling, and multi-echelon inventory rules. Logistics delays arise from both predictable variations, such as seasonal demand or traffic congestion, and stochastic shocks, such as port interruptions, political disruptions, or manufacturing quality holds. Therefore, the operational problem is best framed as a probabilistic, multi-objective decision system rather than a single deterministic routing or scheduling task. In quantitative terms, delay reduction can be measured through lead-time variance, on-time-in-full rates, time-temperature compliance, stockout incidence, and total landed cost, while system-level risk is captured through service criticality indices and patient outcome proxies (Knoll et al., 2016). An extended introduction thus begins by treating delays as measurable system deviations and ML-

advanced computing integration as a structured analytical response embedded in regulated healthcare supply chains.

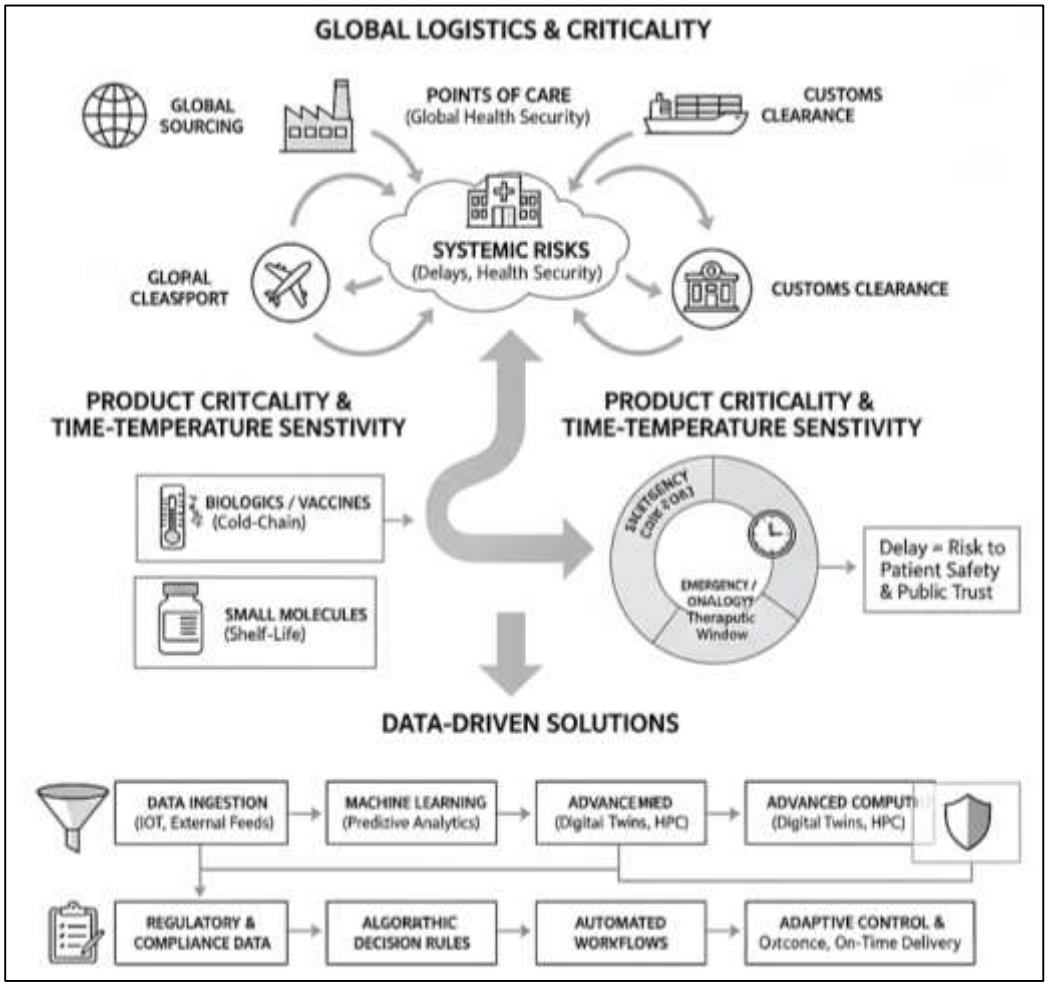
The international significance of reducing logistics delays in pharmaceutical distribution is tied to the globalized nature of medicine production and consumption (Akbar & Farzana, 2021; Reza et al., 2021). Active pharmaceutical ingredients, excipients, packaging materials, and finished formulations are sourced from geographically dispersed clusters, then consolidated into cross-border distribution networks that feed national health systems. Any delay in these networks magnifies as medicines move across multiple nodes, often shifting from ocean freight to air freight, from customs warehouses to regional depots, and from urban hubs to rural health posts (Cao et al., 2016; Saikat, 2021; Shaikh & Aditya, 2021). At the global level, delays translate into lost therapeutic windows, interrupted chronic disease management, compromised vaccination campaigns, and reduced resilience to outbreaks. Low- and middle-income countries are especially exposed because their distribution corridors can include limited cold-chain coverage, variable road reliability, and dependency on imported stock. International standards such as Good Distribution Practices, serialization rules, and pharmacovigilance reporting make delays operationally visible and legally consequential (Kanti & Shaikat, 2021; Zobayer, 2021a). A temperature breach triggered by late delivery can render vaccines unusable, and delays in oncology or emergency antibiotics can generate avoidable mortality. These effects elevate logistics from a back-office function to a public health determinant. In addition, contemporary pharmaceutical products are increasingly sensitive to handling conditions (Alcantara et al., 2017; Zobayer, 2021b).

Biologics, insulin analogs, blood products, cell and gene therapies, and mRNA vaccines require strict, time-bounded temperature control, making delay a primary risk driver. The world has also seen that health emergencies intensify logistics volatility: demand surges, border restrictions, transport capacity shortages, and rapid policy changes increase lead-time uncertainty. Supply networks that depend on manual planning or low-resolution forecasting struggle under such volatility because they cannot detect disruption early or propose adaptive routes quickly (An et al., 2018). International donors, humanitarian organizations, and global health alliances likewise rely on predictable delivery windows to manage funding cycles and equitable allocation models. Thus, delay reduction is a shared global objective spanning commercial supply, humanitarian response, and universal health coverage agendas. A quantitative research framing is appropriate because the problem involves measurable uncertainty, multi-node flows, and a need for algorithmic decision rules that can generalize across diverse geographies. Integration of ML and advanced computing offers a mechanism to encode global-scale data into localized operational decisions, enabling distribution systems to shift from reactive recovery to data-driven anticipatory control.

Logistics delays in pharmaceutical distribution originate from interacting layers of operational, infrastructural, informational, and regulatory complexity (An et al., 2018). At the upstream level, manufacturing release cycles, batch testing queues, and packaging line constraints can shift planned dispatch dates. At the midstream level, warehouse congestion, pick-pack inaccuracies, labor shortages, and equipment failures create internal processing delays. Transportation delays arise from route volatility, carrier capacity mismatch, traffic and weather shocks, security checkpoints, and port or airport dwell times. Cross-border movements add customs clearance uncertainty, documentation mismatches, inspection holds, and tariff-related disruptions. In the downstream environment, last-mile constraints include address ambiguity, fragmented provider networks, limited cold storage at receiving points, and variable patient demand, especially in decentralized health systems. These delay sources rarely act alone; they compound across the chain, producing nonlinear patterns that are difficult to model with simple averages or static rules (Fairley et al., 2019). Moreover, pharmaceutical distribution is a multi-echelon system in which decisions at one tier affect delay probabilities at another. For example, upstream safety stock policies change replenishment frequencies, which change warehouse workload distributions, which in turn alter outbound dispatch reliability. Delay also interacts with product criticality. A short delay in low-risk over-the-counter medicines may be tolerable, while a similar delay in antivenom, organ-transplant immunosuppressants, neonatal antibiotics, or temperature-sensitive vaccines can be catastrophic. The cold-chain dimension adds a time-temperature joint constraint; a delivery that is only slightly late can still be acceptable if thermal stability holds, but a late delivery during a heatwave may cross degradation thresholds. Information gaps are another

central delay driver. Forecast errors, missing telemetry, data silos between manufacturers and logistics providers, and weak visibility in transit reduce the ability to anticipate bottlenecks (Lancaster & Sobie, 2016). Human decision-making under such gaps tends to rely on heuristics, which can amplify bias and instability when systems are stressed.

Figure 2: Pharmaceutical Logistics Optimization Framework



Regulatory compliance also shapes delay because GDP-aligned documentation, controlled drug reconciliation, and recall-management checks require additional handling time. Therefore, delay reduction requires methods capable of learning complex interactions among demand signals, transit conditions, processing queues, and compliance steps. Machine learning models trained on multimodal operational data are aligned with this structure because they can extract hidden predictors of delay, while advanced computing enables those predictions to be operationalized in near real time (Alam et al., 2018). A rigorous quantitative introduction must therefore paint delays as emergent system properties created by interdependent, data-rich processes rather than isolated transport problems.

Machine learning has been adopted in logistics to transform delay management from retrospective reporting into predictive risk control. In pharmaceutical distribution, ML models are applied to forecast lead times, estimate probability distributions for late deliveries, classify shipment risk levels, and detect anomalies in cold-chain telemetry. Such models leverage structured variables such as historical transit times, order volumes, carrier performance, warehouse throughput, and customs dwell logs, along with unstructured or semi-structured sources such as operator notes, GPS streams, temperature sensor signals, and traffic feeds (Helo & Hao, 2019). Supervised learning approaches enable departure-to-arrival time prediction, early warning for route disruptions, and identification of delay-prone lanes. Classification models support triage systems that prioritize high-criticality shipments for expedited handling. Time-series learning is used to anticipate demand-driven congestion weeks in advance,

enabling proactive capacity adjustments. Reinforcement learning has been explored for adaptive routing and dynamic scheduling where decisions are optimized through simulation feedback. Importantly, ML contributes not only to prediction accuracy but also to uncertainty modeling. Probabilistic forecasts, quantile regression, and Bayesian learning frameworks quantify confidence intervals, offering planners a risk-weighted basis for choosing transport modes or buffer sizes. In pharmaceutical contexts, where compliance and safety depend on reliability rather than average speed, this risk quantification is especially relevant (Alanazi et al., 2017). ML also handles heterogeneous product behavior, capturing the fact that biologics and small-molecule tablets exhibit different delay sensitivities due to storage needs, shelf-life windows, and packaging constraints. Another advantage of ML is its ability to incorporate exogenous signals that classical models omit, including epidemiological trends, policy changes, or macro-economic disruptions that affect border flow and carrier capacity. Yet ML outputs must be converted into operational actions, and predictive performance alone does not reduce delays unless embedded into decision cycles. Therefore, ML forms the analytic core of an integrated system, producing delay risk scores, lead-time surfaces, and disruption probabilities that advanced computing environments can consume (Kreif et al., 2015). The quantitative paper underpinning this introduction takes ML as the statistical engine for learning delay patterns in high-dimensional logistics data, setting the stage for integration with computing-driven optimization and automation mechanisms.

Advanced computing provides the infrastructure and algorithmic acceleration required to translate ML insights into time-critical logistics decisions. Within pharmaceutical distribution, this includes cloud platforms capable of ingesting high-frequency IoT telemetry from refrigerated trucks, smart pallets, and warehouse sensors, then processing these data at scale (Hussain & Junejo, 2019). Edge computing allows temperature and location analytics to run near shipment points, enabling local alarms and corrective actions without waiting for centralized processing. High-performance computing and GPU acceleration support rapid simulation and optimization, which is valuable when planning must evaluate thousands of routing or inventory scenarios under uncertain demand and traffic states. Digital twin environments represent a particularly useful computing paradigm: they create a virtual replica of the logistics network, allowing real-time state updates and what-if experimentation. When a digital twin is linked to ML-based disruption forecasts, the system can test alternative transport corridors, warehouse reallocations, or dispatch schedules before executing them physically. Advanced computing also supports mixed-integer programming and metaheuristic solvers that incorporate ML risk scores as constraints or objective weights (Shrestha & Mahmood, 2019). For instance, predicted delay probabilities can be embedded into stochastic routing models to minimize expected late deliveries rather than average distance. Computing architectures further provide secure data pipelines, audit trails, and access controls needed for regulatory compliance, including traceability for controlled substances and accountability for cold-chain events. The pharmaceutical sector frequently requires validated systems with clear documentation of data lineage and decision rules, which modern computing stacks facilitate through automated logging and version control. Additionally, advanced computing supports interoperability between stakeholders, enabling data exchange across manufacturers, 3PLs, distributors, and health providers through APIs and standardized schemas. Without this interoperability, ML models suffer from incomplete features and delayed updates. Thus, advanced computing is not an optional enhancement but a necessary operational layer that enables prediction-to-action cycles (Chang et al., 2019). The integration theme of this paper is grounded in the view that ML provides probabilistic foresight, while advanced computing provides the speed, scalability, and system connectivity required to act on that foresight in regulated, multi-node pharmaceutical logistics.

Integration of ML models with advanced computing for delay reduction can be described through several quantitative mechanisms spanning sensing, prediction, optimization, and control. First, sensor-driven data acquisition provides the raw signals for learning, including GPS tracks for route adherence, temperature and humidity logs for cold-chain stability, and warehouse event times for process mining (Pal & Kant, 2019). Second, ML pipelines clean, fuse, and transform these signals into predictive features, then generate outputs such as expected arrival distributions and disruption probabilities. Third, advanced computing systems embed these outputs into execution models: route planners treat

forecast delays as penalties; scheduling algorithms re-sequence picks based on predicted carrier departure risk; and inventory controllers adjust reorder thresholds using predicted lead-time variance. Fourth, automated control loops update decisions continuously as new data arrive. A shipment delayed at a border can trigger re-routing for downstream replenishment or real-time reallocation from nearby depots. These loops create a cyber-physical supply chain where physical movements and digital decisions are synchronized. Integration also includes governance mechanisms. Model lifecycle management ensures that ML models are retrained, validated, and monitored for drift when lane conditions change (Pal & Kant, 2019). Computing platforms support A/B testing of decision rules, enabling quantitative evaluation of delay reduction effectiveness under controlled comparisons. Integration further relies on explainability and compliance alignment. Pharma stakeholders often require interpretable risk flags that can be justified to auditors and clinical partners, so ML outputs must be packaged into human-understandable dashboards with traceable logic. Advanced computing supports such packaging through visualization, rule-based overlays, and automated reporting. Another integration dimension is latency management. Predictions lose value if they arrive after a decision window has closed, therefore distributed computing and event-stream processing are central to keeping prediction and action aligned (Challita et al., 2019). The combined system can be quantified by performance indicators such as reduction in lead-time variance, improvement in on-time-in-full rates, decrease in temperature excursion incidents, and cost trade-offs between expedited shipping and safety stock. By presenting integration as a structured, multi-stage pipeline rather than a single algorithm, the introduction clarifies how quantitative methods can be mapped onto real logistics operations to reduce delays reliably and measurably.

A quantitative paper on ML-advanced computing integration for reducing logistics delays in pharmaceutical distribution rests on the premise that delays are stochastic outcomes influenced by measurable covariates and controllable decisions (Pagano & Liotine, 2019). The research problem can be formalized as a closed-loop system: observable states of the supply chain generate data, ML models map data to predictive delay risk, and computing-based optimization maps predicted risk to operational actions that reshape subsequent states. This approach treats the distribution network as a dynamic system under uncertainty. Data structures commonly involved include shipment-level records with timestamps, lane identifiers, carrier features, and product attributes; continuous IoT streams indicating temperature, vibration, and location; and contextual data such as traffic indices, port throughput metrics, and health-sector demand proxies. Quantitative modeling challenges include class imbalance because severe delays are rare relative to normal deliveries, censoring because some shipments lack complete telemetry, and heterogeneity because lane performance differs across regions and product classes (Mohassel & Zhang, 2017). A robust quantitative introduction positions ML selection, feature engineering, and validation as statistical responses to these challenges while acknowledging that the final objective is operational delay reduction rather than prediction alone. Advanced computing enters the quantitative framing through its ability to solve high-dimensional decision problems quickly enough to remain relevant in practice. When predictions change hourly and routing decisions must be updated within minutes, scalable computing shifts from technical convenience to methodological necessity. The integration viewpoint also highlights multi-objective trade-offs that can be specified mathematically, including minimizing expected delay subject to temperature compliance, cost ceilings, and service prioritization for critical medicines (Mohassel & Zhang, 2017). Pharmaceutical distribution therefore becomes an ideal domain for integrated analytics because it combines large data volumes, strict safety constraints, and high societal cost of lateness. The introduction closes its conceptual arc by establishing a clear analytical unity: ML provides statistically grounded delay foresight, advanced computing provides decision speed and system scalability, and their integration provides a measurable pathway to reduce delays in complex global medicine supply chains (Abduljabbar et al., 2019).

The objective of this quantitative study is to develop and empirically validate an integrated analytics framework that combines machine learning models with advanced computing infrastructures to reduce logistics delays in pharmaceutical distribution networks. Specifically, the study aims to (1) quantify the magnitude, distribution, and variability of end-to-end lead-time delays across multi-echelon pharmaceutical supply chains by constructing shipment-level delay metrics such as planned-

versus-actual transit time, node-wise dwell time, and lead-time variance; (2) identify and statistically model the key operational, environmental, and regulatory predictors of delays—including route congestion indicators, warehouse processing queues, carrier reliability, customs clearance durations, and cold-chain telemetry deviations—through structured feature engineering and correlation/importance testing; (3) design, train, and compare multiple machine learning approaches for delay prediction and risk scoring, including time-series forecasting, supervised regression, and classification models, to determine which techniques most accurately estimate both expected delay and delay probability under different lane and product conditions; (4) integrate the best-performing predictive models into an advanced computing layer—cloud/edge processing, high-frequency event streaming, and optimization or simulation engines—so that predictions can be operationalized into real-time routing, scheduling, and inventory control decisions; (5) evaluate, using rigorous out-of-sample validation and operational KPIs, the causal impact of the integrated ML-computing pipeline on delay reduction, measured by improvements in on-time-in-full rates, reductions in lead-time variance, decreases in severe delay incidence, and containment of temperature-excursion risk while tracking cost trade-offs; and (6) test the robustness and generalizability of the framework across product criticality classes (e.g., ambient vs. cold-chain biologics) and geographic corridors (domestic vs. cross-border lanes) to confirm that performance gains are not confined to a narrow operational niche. Collectively, these objectives position delay as a measurable stochastic outcome, prediction as a statistical learning task, and reduction as an optimization-enabled control problem, thereby linking analytic accuracy to health-system reliability and patient-critical service continuity within globally regulated pharmaceutical logistics.

LITERATURE REVIEW

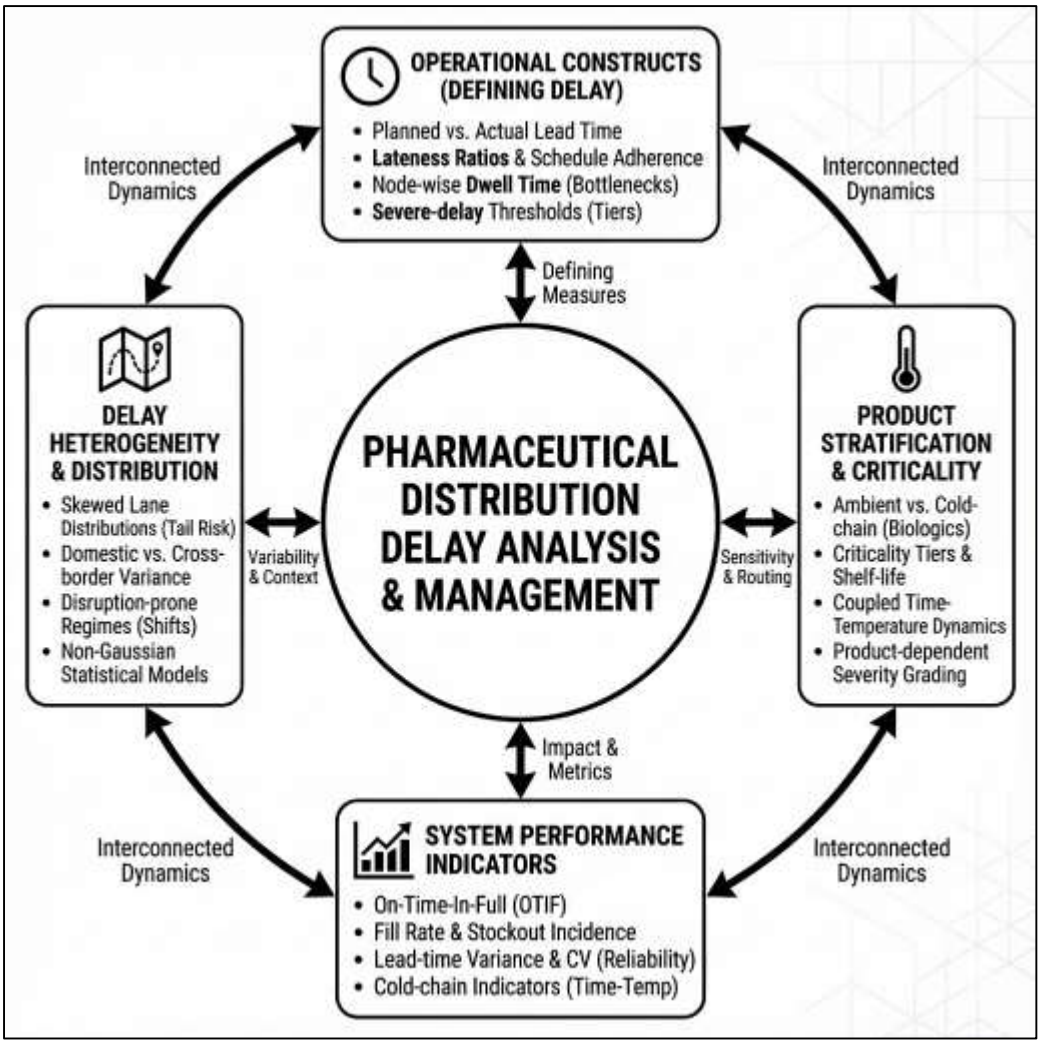
This literature review consolidates quantitative and empirical research on reducing logistics delays in pharmaceutical distribution through the integration of machine learning (ML) models and advanced computing systems. The section treats delay as an observable, statistically describable deviation within multi-echelon and regulated supply chains, where lead-time instability, cold-chain risk, and service unreliability have measurable operational and health consequences. Prior studies are examined to show how delay mitigation has progressed from deterministic optimization and classical forecasting toward data-intensive predictive–prescriptive pipelines that operate in real time. The review is organized to clarify what has been modeled (delay magnitude, probability, and variance), which covariates and data streams have been used (shipment logs, IoT telemetry, node dwell times, exogenous disruption indicators), and how performance improvements have been tested (out-of-sample accuracy, KPI deltas, significance testing). By sequencing evidence from foundational delay metrics to ML prediction, computing-enabled optimization, and fully integrated closed-loop control, the review establishes a rigorous quantitative base for the current paper's modeling choices and empirical design.

Logistics Delays in Pharmaceutical Distribution

Quantitative studies on pharmaceutical distribution consistently begin by defining delay through operational constructs that transform logistics events into measurable time deviations (Najm et al., 2019). A common construct is the comparison between planned lead time and actual lead time, where delay is observed as the gap between scheduled dispatch–arrival windows and realized delivery times. This planned–actual construct is used not only for overall shipment performance but also for mapping delay origins, since each stage contributes to total lead time. Researchers also operationalize delay through lateness ratios and schedule-adherence scores that express reliability as a proportion of shipments arriving within designated tolerances. These tolerances are typically lane-specific and product-sensitive because pharmaceutical flows involve heterogeneous transport conditions and regulatory checkpoints (Callahan & Shah, 2017). Another widely used construct is node-wise dwell time, which decomposes total delay into time spent at discrete points such as manufacturing release gates, distribution centers, cross-docking terminals, customs warehouses, and last-mile handover facilities. Node dwell measurements allow analysts to attribute delay to controllable internal processes, like picking and staging, or to exogenous bottlenecks, like inspection queues. Several works show that node-level timing is essential in multi-echelon networks because delay is cumulative and often nonlinear, with upstream lags propagating downstream into stockouts or emergency replenishments.

A further construct is the specification of severe-delay thresholds, usually defined by time windows that reflect product criticality. Life-saving medicines, vaccines, and temperature-sensitive biologics are evaluated using tighter permissible delay ranges than ambient tablets, creating a tiered severity structure (Adhau et al., 2019). This practice converts raw time deviations into clinically meaningful categories for statistical modeling. Across these constructs, the literature portrays delay not as a single uniform measure but as a family of linked time-based indicators that capture reliability, vulnerability, and criticality within regulated pharmaceutical supply chains.

Figure 3: Pharmaceutical Distribution Delay Analysis Framework



On-time-in-full (OTIF) appears as a dominant metric because it combines timeliness with quantity correctness, aligning delay reduction with patient-facing availability (Zhou et al., 2017). Studies treat OTIF both as a dependent variable to be improved and as an explanatory lens for understanding where failure occurs. Service level indicators, including fill rate and order cycle reliability, are also tied to delay because late deliveries reduce the probability that facilities can meet demand without rationing or substitution. Stockout incidence is modeled as a downstream consequence of lead-time instability, and is frequently quantified as the fraction of SKUs or demand episodes experiencing zero inventory during replenishment windows. Emergency shipment rate, often measured through expedited-freight frequency or unplanned replenishment events, is another performance indicator that rises with delays and creates a measurable cost penalty. To capture reliability more rigorously, researchers compute lead-time variance and the coefficient of variation, framing delay reduction as a problem of minimizing dispersion rather than only shifting averages. This emphasis reflects the pharmaceutical requirement for predictable supply more than fastest possible supply. Cold-chain indicators extend this

performance set into the thermal domain (Parmata et al., 2016). Time-temperature compliance conditional on lateness is treated as a joint safety metric, acknowledging that even a small delay can elevate excursion risk for fragile products. Quantitative models link shipment lateness to probabilities of temperature breach, spoilage, or stability loss. These indicators establish a multidimensional quantitative view where delay acts as both a direct performance failure and a driver of secondary risks, creating a structured basis for delay-reduction modeling in pharmaceuticals.

Empirical studies show that pharmaceutical logistics delays exhibit heavy heterogeneity across routes, modes, and operational contexts, which motivates distribution-oriented measurement rather than reliance on single summary averages. Quantitative analyses frequently report lane-level delay distributions that are skewed, with a long tail of rare but severe late events. This skewness is treated as a critical statistical feature because it implies that mean lead time understates service risk, while tail behavior governs emergency replenishment and clinical impact (Knop, 2019). Some works estimate kurtosis or other tail-sensitivity indicators to capture the likelihood of extreme delays. Heterogeneity is also observed in the structural difference between domestic and cross-border lanes. Domestic movements tend to show tighter lead-time dispersion but remain sensitive to urban congestion and depot workload clustering, while cross-border lanes demonstrate larger variances due to customs dwell uncertainty, inspection regimes, and carrier handoffs. Quantitative comparisons emphasize that cross-border delays are not merely longer; they are more volatile and discontinuous, making probabilistic modeling necessary (A. Narayana et al., 2014). Another recurrent empirical pattern is the contrast between stable lanes and disruption-prone lanes, where the latter show regime shifts that alter lead-time distributions over time. This lane stratification supports risk-scoring approaches and targeted interventions. Overall, the literature indicates that delay behavior is context specific and statistically non-Gaussian, encouraging models that represent full distributions, tails, and lane-dependent parameters. Describing delay in distributional terms therefore becomes an essential quantitative foundation for any integrated analytics approach aimed at reliability improvement (Moons et al., 2019).

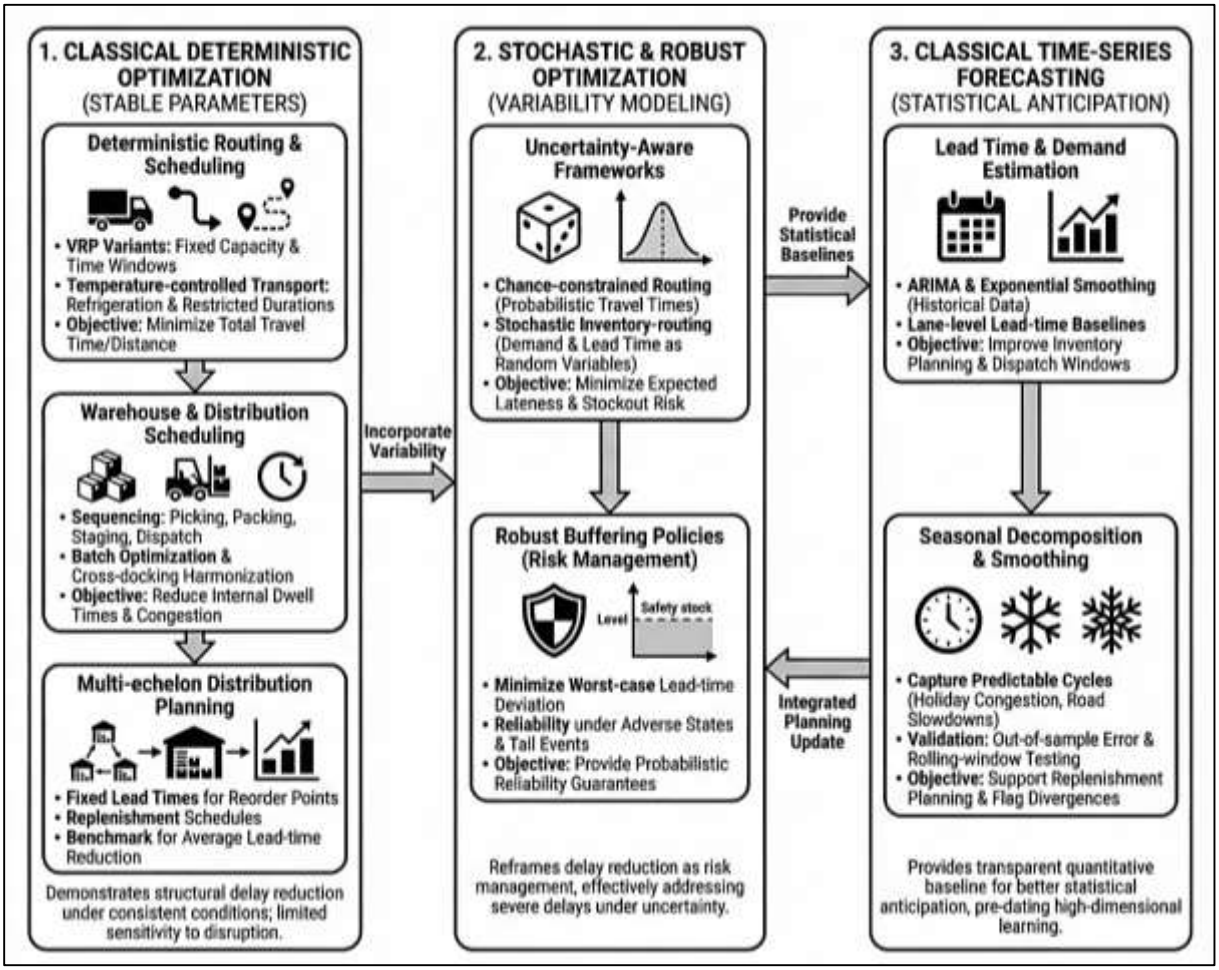
Pharmaceutical delay measurement is consistently stratified by product class because time sensitivity and handling constraints differ sharply across medicine categories. Empirical work distinguishes ambient products such as solid oral dosages from cold-chain products such as vaccines, insulin, monoclonal antibodies, and other biologics (Chen et al., 2019). Ambient products are usually evaluated with wider acceptable lead-time windows, and their delay impacts are modeled primarily through availability and inventory cost variables. Cold-chain products, by contrast, are governed by narrow stability windows, and even modest lateness can elevate spoilage risk or render doses unusable. Quantitative studies operationalize this difference through product-criticality tiers, where delay severity thresholds are adjusted to shelf-life, therapeutic urgency, and temperature fragility. Biologics and vaccines often receive the highest criticality weights, while low-risk generics are assigned lower weights. This stratification enables statistical models to reflect that not all late deliveries are equally harmful (Mehralian et al., 2015). Several works also show that cold-chain shipments present coupled time-temperature dynamics, meaning that delay measurement must be interpreted jointly with thermal exposure trajectories. Delays are thus linked to conditional probabilities of excursion, and performance dashboards for cold-chain lanes incorporate both timeliness and temperature compliance scores. Product stratification further reveals differences in empirical delay distributions: cold-chain lanes tend to exhibit stricter routing, specialized carriers, and more intensive monitoring, yet they remain more vulnerable to catastrophic failure if delays occur (Choi et al., 2019). The literature therefore treats product class as a fundamental axis of quantitative analysis, shaping delay definitions, severity grading, downstream performance linkage, and statistical modeling choices across pharmaceutical distribution networks.

Models for Delay Reduction (Pre-ML Baseline)

Classical quantitative work on delay reduction in pharmaceutical distribution initially relied on deterministic routing and scheduling models that assume stable parameters for demand, travel time, and processing capacity (Papalexi et al., 2016). Within transportation, vehicle routing problem (VRP) variants were adapted to pharmaceutical constraints by including capacity limits, delivery time windows, and product-specific handling requirements. For temperature-controlled transport, deterministic VRP models incorporate refrigeration capacity and restricted route durations to ensure

that cold-chain products reach destinations within allowable thermal exposure limits. These models often prioritize minimizing total travel time or distance while enforcing strict feasibility conditions, such as mandatory delivery windows for hospitals and clinics. In warehouse and distribution center operations, deterministic scheduling models focus on sequencing picking, packing, staging, and dispatch batches to reduce internal dwell times that contribute to late deliveries (Aldrighetti et al., 2019). Batch optimization studies demonstrate that reorganizing outbound flows, harmonizing pick waves with carrier departure schedules, and reducing congestion at cross-docking terminals can lower average processing delays. Another deterministic baseline frequently used is multi-echelon distribution planning, where fixed lead times are used to calculate reorder points and replenishment schedules. These approaches provide a clear and computationally tractable starting point for delay mitigation, and they remain common benchmarks for later machine learning comparisons. However, the deterministic literature measures success mainly through average lead-time reductions and shorter planned cycle times, while reporting feasibility trade-offs when constraints become too tight (Peng et al., 2014). The evidence suggests deterministic models reduce routine delays under stable operational conditions, yet their dependence on fixed parameters limits sensitivity to disruption-driven variability. Even so, these models established the operational vocabulary—route feasibility, time windows, batch waves, and multi-echelon coordination—that later predictive systems built upon. Their main quantitative contribution is the demonstration that delay can be structurally reduced through optimization of schedules and routes when uncertainty levels are low and system parameters remain consistent across planning horizons (Dasaklis et al., 2017).

Figure 4: Classical Quantitative Engineering Framework



As pharmaceutical networks expanded globally and uncertainty became more visible, quantitative research turned to stochastic and robust optimization methods to reduce delays by explicitly modeling variability. Chance-constrained routing frameworks treat travel times as probabilistic rather than fixed,

enabling planners to select routes that satisfy delivery windows with a specified confidence level. These models acknowledge that even well-designed deterministic routes fail when congestion, weather, or border processes fluctuate (Beheshtinia et al., 2018). Their objective is therefore expressed in terms of expected lateness or probability-weighted lead-time violation rather than deterministic travel cost. Stochastic inventory-routing models similarly represent demand and transportation lead time as random variables, searching for joint replenishment and routing policies that minimize expected delay-related penalties such as stockout risk or emergency restocking. Multi-echelon stochastic formulations show that delay reduction is strongly linked to coordinated safety stock positioning and flexible routing when disruptions occur at one tier of the chain. Robust buffering policies address delay by minimizing worst-case lead-time deviation under bounded uncertainty sets. These policies are common in pharmaceutical settings because regulators and healthcare providers value reliability under adverse states, not only average performance (Franco & Alfonso-Lizarazo, 2017). Robust approaches design schedules and inventory buffers that hold feasibility even when travel times or processing durations drift beyond normal ranges. Quantitative results across these studies indicate that uncertainty-aware methods reduce severe delays more effectively than deterministic baselines, though they may increase cost through conservatism or higher buffer requirements. The stochastic and robust literature thus reframes delay reduction as risk management, providing probabilistic reliability guarantees and demonstrating that delay mitigation requires accounting for tail events and multi-node propagation (Campelo et al., 2019). This body of work also created the statistical benchmarks that later ML systems aimed to outperform, especially for modeling nonlinearity in uncertainty effects.

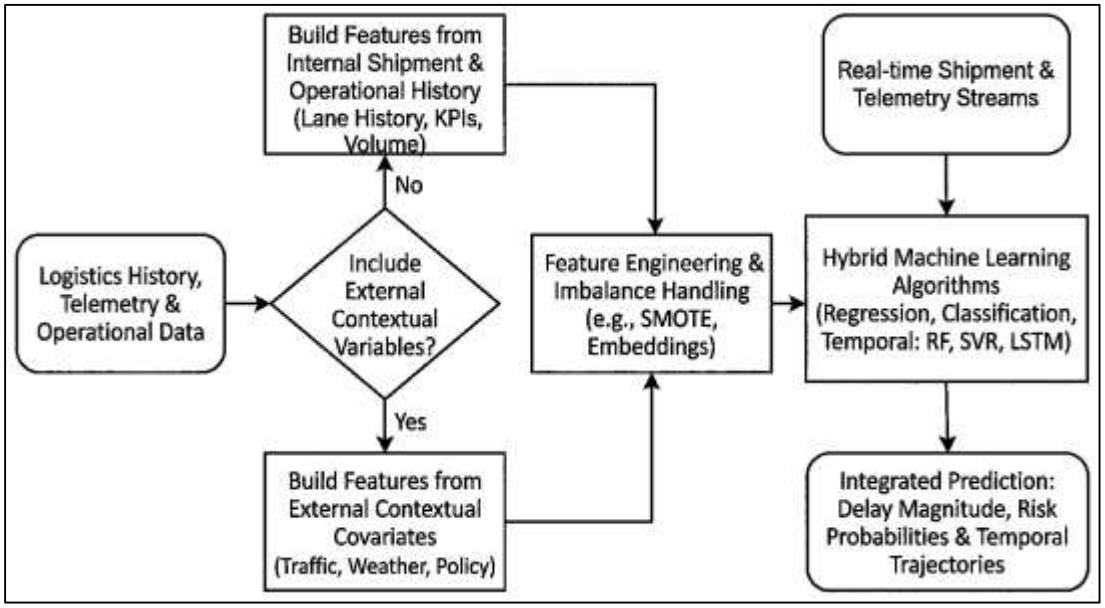
Before ML-based prediction became widespread, delay reduction also depended on classical time-series forecasting of lead times and demand-linked shipment volumes. ARIMA and exponential smoothing families were routinely applied to historical dispatch and arrival data to estimate expected lead times for corridors and carriers (Pérez Lechuga, 2018). The forecasting logic is that improving lead-time estimates allows better inventory planning, tighter dispatch windows, and more accurate carrier contracting, which indirectly reduces delays. In practice, studies used these models to build lane-level lead-time baselines and to flag periods where actual transit times diverged from expected trends. Seasonal decomposition and smoothing methods further supported replenishment planning by capturing predictable cycles such as holiday congestion, monsoon-related road slowdowns, or fiscal-year ordering spikes in public health systems (Hammoudan et al., 2016). While these classical models can be statistically sound when patterns are stable and low dimensional, the literature notes that their linear assumptions and limited feature sets restrict their effectiveness under high volatility. Forecast error is shown to translate into mismatched safety stock levels and poorly timed dispatches, which then produce avoidable delays. Nonetheless, these models provided a transparent quantitative baseline for lead-time expectation and were often integrated into deterministic or stochastic planning systems to update parameters periodically. Classical forecasting studies also established validation norms such as out-of-sample error measurement and rolling-window testing. These norms later became important in evaluating ML superiority (Hammoudan et al., 2016). Overall, traditional forecasting served as a pre-ML bridge between descriptive delay measurement and prescriptive optimization, positioning delay reduction as something that could be improved through better statistical anticipation even without high-dimensional learning.

Machine Learning for Lead-Time Prediction and Delay Risk Scoring

Quantitative literature on machine learning for logistics delay reduction commonly begins with supervised regression models that estimate shipment arrival times or end-to-end lead times as continuous outcomes (Kramer et al., 2019). In pharmaceutical distribution and closely comparable high-criticality logistics, these models are used to replace static averages with data-driven estimates that adapt to lane volatility, carrier behavior, and node congestion. Tree-based ensembles such as random forests, gradient boosting, and extreme boosting are frequently preferred because they capture nonlinear effects while retaining partial interpretability for regulated environments. Support vector regression is also widely applied, especially in settings with moderate data volumes and high feature correlation, where margin-based learning provides stable lead-time surfaces. Deep regression models extend these approaches by learning complex feature interactions from large shipment histories and high-resolution operational telemetry, producing robust predictions even when delays are driven by

multiple interacting constraints (Martins et al., 2017). Across studies, model performance is assessed using absolute and squared error measures, reflecting the operational importance of minimizing both average deviation and outlier shocks. Prediction-interval evaluation is another recurring theme, since pharmaceutical planners require uncertainty bounds to set buffers for cold-chain products and urgent therapies. Feature engineering is central in these works: lane history variables encode structural delay risk by route; carrier KPIs represent reliability and service disruption probability; shipment volume and order batching reflect workload-based congestion effects; and measures of node workload or dwell propensity capture internal processing variability. Many empirical papers demonstrate that lead-time accuracy improves when regression models fuse shipment logs with contextual covariates such as traffic intensity, weather-linked travel friction, or policy-driven border dwell indicators (Liu & Xiao, 2015). Regression-based ML is thus positioned as the first predictive layer for delay control, enabling planners to anticipate late arrivals earlier than classical baselines and to quantify delay magnitude at shipment granularity rather than network averages.

Figure 5: Logistics Delay Prediction ML Framework



A second stream of quantitative evidence focuses on classification models that estimate the probability that a shipment will be on-time, late, or severely late, allowing logistics systems to shift from point prediction to risk-based prioritization. In pharmaceutical distribution, this probabilistic framing matches operational realities where critical medicines require preemptive escalation even if precise lateness duration is uncertain (Liu & Xiao, 2015). Binary classification is often used to separate on-time and delayed shipments, while multiclass variants create severity tiers aligned with product criticality windows. Tree-based classifiers and margin-based models are regular choices because they handle mixed data types and provide stable decision boundaries in large logistics datasets. Empirical studies repeatedly note that severe delays are rare events relative to normal deliveries, creating class-imbalance problems that degrade naïve classifiers. Therefore, imbalance handling appears as a methodological cornerstone. Oversampling methods such as SMOTE and its variants are widely used to generate synthetic severe-delay examples that broaden minority-class representation in feature space (Haial et al., 2016). At the algorithmic level, weighted-loss learning assigns higher penalties to severe-delay misclassification, and focal-type losses emphasize hard-to-classify rare cases without overfitting easy majority-class patterns. Model evaluation prioritizes discrimination measures rather than accuracy alone, since overall accuracy can be misleading under imbalance. Area-based measures, harmonic-mean scores, and minority-class recall are routinely reported to ensure that models capture the tail risk that drives emergency shipments and cold-chain failure exposure. Quantitative findings show that classification outputs become operationally valuable when converted into delay-risk scores that rank

shipments for expedited handling, alternative routing, or buffer-inventory activation (Timajchi et al., 2019). This literature thus frames ML classification as a decision-support tool that quantifies severe-delay likelihood and enables tiered logistics control in medicine supply chains.

Beyond static regression or classification, a growing body of quantitative work demonstrates the value of temporal machine learning for capturing how delay risk evolves during shipment lifecycles. These studies model logistics processes as sequences of states—departures, node handoffs, travel segments, and environmental exposures—where delay emerges from time-dependent interactions rather than single-shot predictors (Kergosien et al., 2017).Recurrent architectures such as LSTM and GRU are frequently deployed because they can learn from ordered telemetry streams, retaining memory of earlier route conditions that influence later arrival outcomes. Transformer-based sequence learners are now used in logistics and transportation contexts to capture long-range dependencies in travel-time and lead-time data and to improve stability when sequences are long or irregularly sampled. In pharmaceutical distribution, these models align with the reality that cold-chain shipments generate continuous sensor data and that disruptions often unfold gradually, for example through rising transit times, repeated micro-stops, or accumulating dwell at intermediate hubs (Borumand & Beheshtinia, 2018). Temporal ML supports multi-step ahead prediction, meaning the model estimates delay trajectories several hours or days before final delivery, which is operationally essential for proactive dispatch rescheduling or downstream replenishment reallocation. Literature also highlights that temporal models are tested under volatility regimes such as demand surges, congestion spikes, or policy shocks, because sequence learners must remain reliable when system dynamics shift. These stability tests commonly involve time-blocked validation split designs, lane-held-out evaluation, or disruption-period stress testing to verify that temporal representations generalize beyond routine operating states. Across findings, temporal ML improves not only mean error but also early-warning accuracy, enabling logistics controllers to intervene while corrective options remain feasible (Soysal & Çimen, 2017). The temporal stream therefore positions dynamic sequence learning as a quantitative advancement that turns real-time logistics data into evolving delay intelligence.

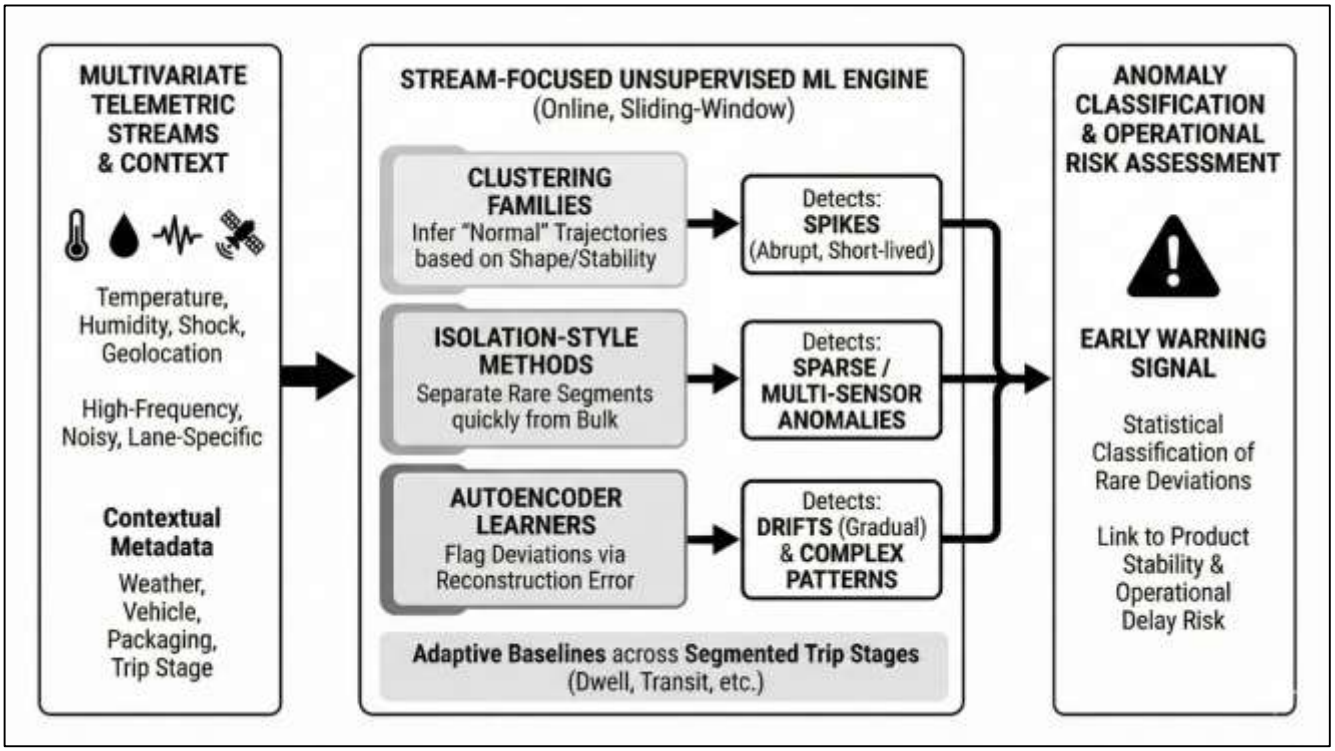
Synthesizing across supervised regression, delay-risk classification, and temporal ML, the literature shows a convergence toward hybrid predictive systems that combine magnitude forecasting with severity probability and time-evolving updates (Do et al., 2019). Quantitative reviews in logistics emphasize that no single algorithm dominates across all corridors, product types, or uncertainty conditions; instead, performance depends on feature richness, data latency, and the statistical structure of delays. Tree-based ensembles are repeatedly validated as strong baselines due to their resilience to noisy operational data and their ability to exploit heterogeneous predictors. Support-vector methods remain competitive where data are sparse or dimensionality is high relative to sample size, while deep regressors and temporal networks perform best when telemetry and historical records are large and well-aligned (Singh & Goh, 2019). Another repeated finding is that model superiority must be established through consistent benchmarking practices, including out-of-sample validation, rolling-window testing, and evaluation that separates routine delays from severe tail events. Several studies point out that predictive accuracy alone is insufficient unless paired with risk interpretability and uncertainty quantification, because pharmaceutical logistics depends on auditability and safe decision margins. As a result, prediction-interval accuracy, calibration of severe-delay probabilities, and robustness to distribution shift are treated as core quantitative criteria. Methodologically, this body of work also demonstrates that delay prediction models improve when they include lane embeddings, carrier performance histories, shipment consolidation indicators, and node workload signals, confirming that delay is multi-causal and system-embedded (Baniamerian et al., 2018). Overall, the ML literature portrays lead-time prediction and risk scoring as complementary statistical tasks: regression provides expected delay magnitude, classification provides severe-delay likelihood, and temporal models provide continuous updates as shipment conditions unfold. This synthesized evidence builds a coherent quantitative platform for integrating ML outputs into operational delay-reduction pipelines without depending on a single predictive paradigm (De Souza et al., 2014).

ML-Based Cold-Chain Anomaly and Excursion Modeling Linked to Delays

Quantitative scholarship on cold-chain reliability frames temperature excursions as detectable anomalies embedded in high-frequency telemetric streams captured during warehousing and transit

(Afroditi et al., 2014). In pharmaceutical distribution, sensors routinely log temperature, humidity, shock, light exposure, and geolocation at granular intervals, producing multivariate time-series in which anomalies may appear as spikes, drifts, step changes, or irregular patterns associated with handling shocks and equipment malfunction. The literature shows that classical threshold rules are often insufficient for these streams because signals are noisy, context dependent, and lane specific, so unsupervised machine learning has become a dominant analytic approach. Clustering families are frequently used to infer "normal" thermal trajectories by grouping time-series segments with similar shapes or stability features, allowing rare clusters to be interpreted as excursion events (Fathollahi-Fard et al., 2019). Isolation-style methods treat anomalous segments as those that can be separated quickly from the dense bulk of normal observations, making them attractive for cold chains where disruptions are infrequent but costly. Autoencoder learners expand this approach by compressing normal thermal patterns into latent representations and flagging anomalies through reconstruction error when new telemetry deviates from learned stability regimes. Quantitative studies in vaccine storage and biologics logistics show that autoencoder-based architectures can detect subtle drift before a full excursion accumulates, improving early warning compared with static alarms. Stream-focused anomaly schemes further highlight that cold-chain data arrive continuously and require online detection rather than retrospective batch analysis, so sliding-window learning and adaptive baselines are common. Across this body of work, anomaly detection is not treated as a generic monitoring task but as a statistical classification of rare thermal deviations whose timing and duration matter for product stability (Choi et al., 2019). The evidence also emphasizes that anomaly labels are often weak or partially observed in practice, so unsupervised models are valuable because they do not require large, perfectly annotated excursion datasets. Overall, the telemetric anomaly detection literature provides a quantitative foundation for linking sensor-visible thermal irregularities to operational delay risk, with models designed to detect the earliest signs of temperature instability in transit and storage.

Figure 6: AI-Driven Cold Chain Anomaly Detection



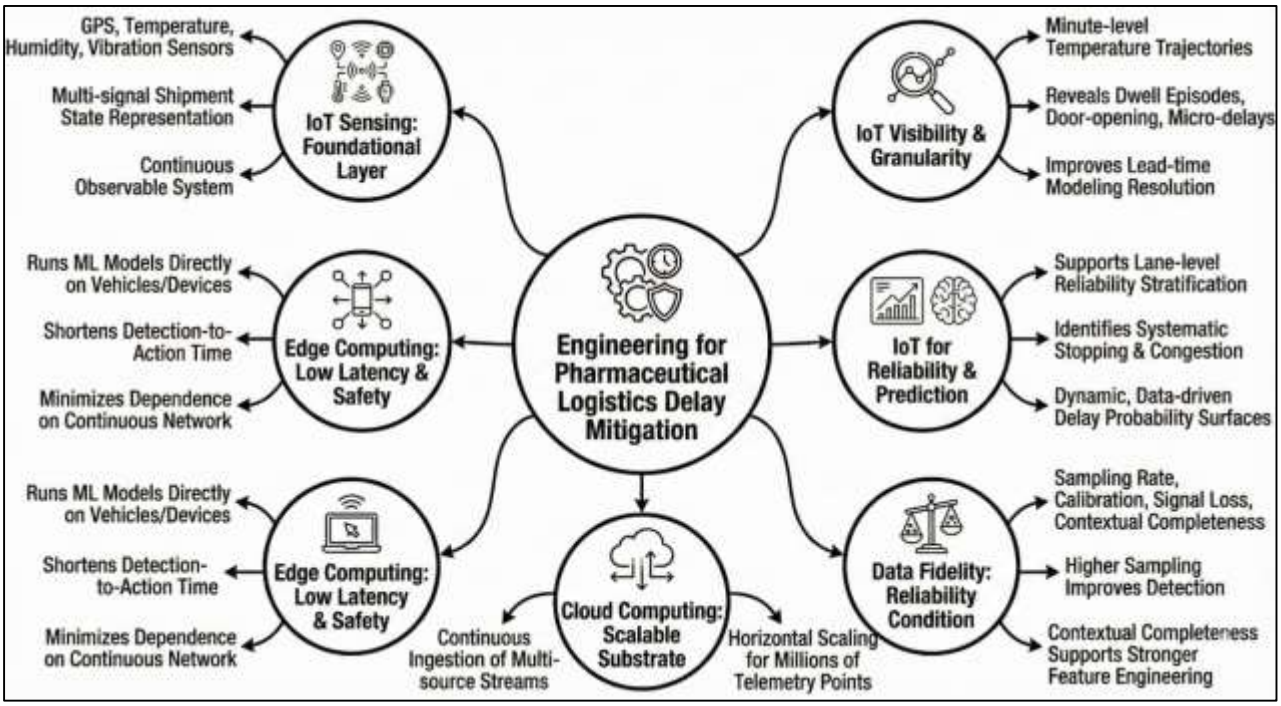
Autoencoder families are repeatedly validated for drift and complex mixed anomalies, because reconstruction error accumulates as patterns slowly diverge from the learned stable manifold. Several empirical studies in cold-chain transportation propose hybrid detection pipelines that combine isolation-style screening with deep representations to reduce false positives during benign fluctuations and to retain sensitivity during unstable external conditions. A recurring quantitative point is that

anomaly detection quality depends on sampling rate and contextual metadata. Higher-frequency sensors enable micro-spike capture and more reliable drift estimation, while contextual features such as lane, vehicle type, packaging configuration, and ambient weather reduce confusion between true threats and expected environmental response. The literature also reports that anomaly detection is best interpreted on segmented trip stages, differentiating warehouse dwell, loading, line-haul, customs holds, and final-mile intervals, because each stage has distinct thermal baselines (Aivaliotis et al., 2019). Taken together, these findings establish that cold-chain anomalies are heterogeneous statistical objects rather than a single deviation category, and that ML methods are selected and tuned based on the dominant anomaly physics for specific products and corridors.

Advanced Computing Foundations Enabling Real-Time Delay Mitigation

Quantitative literature identifies IoT sensing as the foundational layer for real-time delay mitigation because it converts pharmaceutical logistics from a low-visibility process into a continuously observable system. Studies emphasize that GPS trackers, temperature and humidity loggers, and vibration or shock sensors collectively create a multi-signal representation of shipment state, allowing analysts to measure not only where a shipment is, but also whether it is moving under safe physical conditions (Korth et al., 2018). For high-value biologics and vaccines, telemetry provides minute-level temperature trajectories that reveal dwell episodes, door-opening events, refrigeration instability, and micro-delays that would be invisible in traditional milestone-based tracking. Researchers treat these data streams as operational time-series inputs that improve the statistical resolution of lead-time modeling by exposing within-route variability rather than only origin–destination averages. At warehouse nodes, IoT tags capture entry-to-exit timestamps and thermal stability while pallets wait for staging or customs release, enabling dwell-time distributions to be estimated directly from sensor evidence. Another consistent finding is that IoT visibility supports lane-level reliability stratification, because repeated GPS-temporal patterns identify corridors with systematic stopping behavior, congestion signatures, or handling shocks that correlate with later delays (Castelli et al., 2019).

Figure 7: IoT Framework for Pharma Logistics



The literature also documents that vibration telemetry is not merely a quality signal; it is a proxy for rough handling and repeated transfer friction that often co-occurs with delay-prone last-mile delivery. In methodological terms, high-frequency sensing is linked to the shift from static lead-time parameters toward dynamic, data-driven delay probability surfaces, since models can observe transitional states instead of inferring them retrospectively. Overall, the IoT body of work portrays sensors as quantitative

enablers of delay reduction because they provide the ground-truth temporal and environmental granularity required for accurate prediction, anomaly detection, and responsive control in regulated pharmaceutical distribution networks (Barricelli et al., 2019).

A deeper quantitative theme in the IoT literature is that data fidelity strongly conditions the reliability of delay prediction and risk scoring. Fidelity is discussed through sensor sampling rate, calibration accuracy, signal loss frequency, and contextual completeness. Studies show that higher sampling rates improve detection of short temperature spikes and stop–start motion patterns that often precede severe delays, whereas low-rate logging can smooth away micro-events and understate delay risk (Bauer et al., 2019). Calibration accuracy matters because systematic sensor bias can distort both cold-chain anomaly probabilities and lead-time estimates, producing false stability or false alarms that mislead planners. Signal loss and intermittent connectivity appear as major fidelity threats during cross-border or rural deliveries, where GPS gaps or missing temperature windows create censored sequences; quantitative evaluations note that prediction error rises when missingness clusters in the most volatile trip segments. Contextual completeness is also treated as a fidelity component: telemetry paired with rich metadata—lane identifiers, vehicle type, packaging profile, and ambient climate—supports stronger feature engineering and reduces spurious anomaly flags (Tao et al., 2018). Several empirical works compare prediction models trained on high-fidelity versus low-fidelity sensing regimes, showing that uncertainty bounds widen and tail-risk recall drops as fidelity deteriorates. This evidence encourages attention to data governance and sensor-network design as part of delay mitigation, because model accuracy is bounded by the resolution of the underlying signals. In addition, fidelity affects generalizability: models trained on dense and clean telemetry may transfer poorly to corridors with sparse or noisy data unless robustness techniques are applied (Uhlenkamp et al., 2019). Across these studies, data fidelity is framed as a measurable determinant of predictive performance, highlighting that real-time delay control depends on both algorithmic quality and sensor-system precision.

The literature positions cloud computing as the primary advanced-computing substrate that makes IoT-based prediction operational at scale in pharmaceutical logistics. Quantitative studies describe cloud platforms as enabling continuous ingestion of multi-source streams—shipment transactions, GPS tracks, temperature logs, warehouse event times, and exogenous traffic or weather signals—into unified data lakes or distributed storage layers (Guerra et al., 2019). This fusion is critical because delay risk emerges from interactions across domains, and cloud-based pipelines allow these domains to be joined with low friction. Stream-processing engines are highlighted for their ability to compute features and run ML inference as shipments move, producing near-real-time delay estimates and cold-chain risk flags. Another recurrent element is interoperability through APIs and standardized schemas, which reduces data latency between manufacturers, third-party logistics providers, and health facilities; quantitative evidence indicates that such interoperability improves prediction timeliness and reduces the time between risk detection and corrective action (Wagner et al., 2019). Cloud benchmarks discussed in the literature often focus on inference latency relative to throughput per shipment, emphasizing that pharmaceutical networks may involve millions of telemetry points per day, requiring horizontal scaling. Distributed computing also supports rapid retraining and drift monitoring when lane conditions shift, maintaining predictive stability without halting operations. Cloud infra is further described as an audit-support system: it retains immutable time-temperature traces and model output logs needed for GDP compliance and post-incident analysis. Taken together, the cloud-computing literature frames scalability, fusion capacity, and cross-stakeholder connectivity as quantitative necessities for real-time delay mitigation, because predictive models lose operational value when inference is slow, siloed, or inconsistent across partners (Siewert et al., 2014).

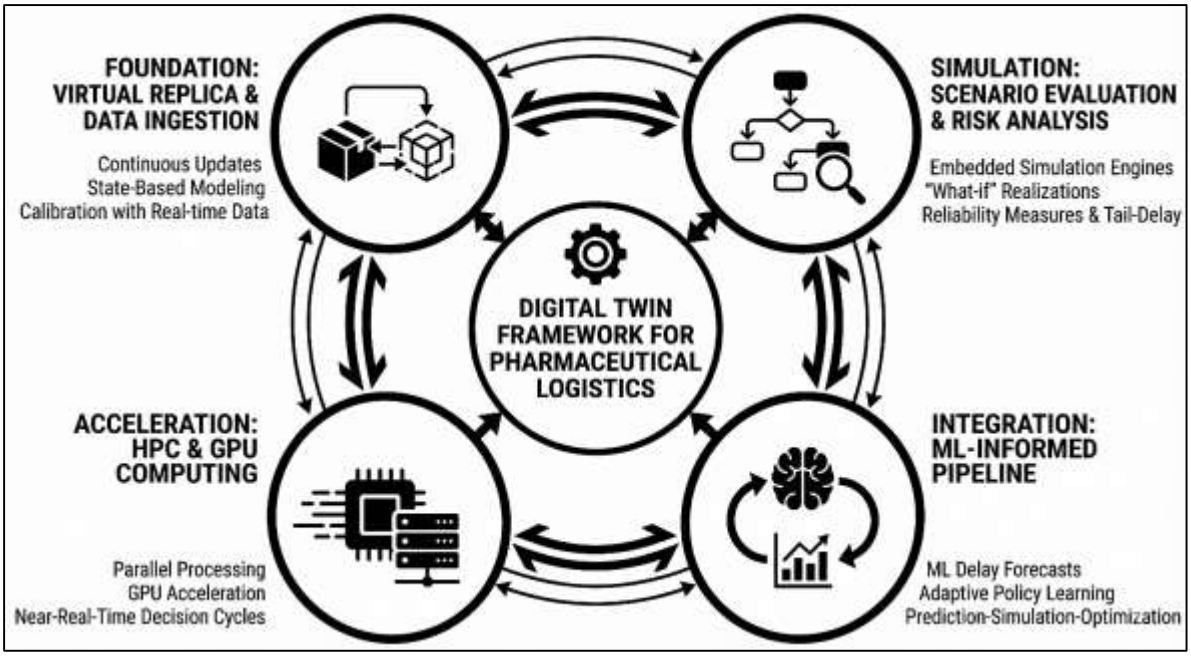
Edge computing emerges in quantitative work as a complementary architecture to cloud systems, designed to reduce latency and preserve cold-chain safety when connectivity or decision windows are tight. Studies define edge inference as running ML models directly on vehicles, smart containers, gateway devices, or warehouse micro-servers, allowing risk detection to occur at the point of data generation (Lomperski et al., 2015). This is particularly relevant for pharmaceuticals because temperature excursions and delay cascades can accelerate quickly during last-mile stops, loading delays, or refrigeration faults; intervention is most effective when alerts are generated within seconds

rather than after cloud round-trips. Quantitative comparisons show that edge-based monitoring shortens detection-to-action time and reduces the probability that a shipment crosses stability limits before corrective steps occur. Edge systems also minimize dependence on continuous network availability, enabling reliable monitoring through rural corridors, ports, or border zones where cloud uplinks may be intermittent. Another statistically documented advantage is selective data offloading: edge devices can pre-filter telemetry, transmit only anomalies or summarized features to the cloud, and thus stabilize bandwidth costs while maintaining predictive sensitivity (Sehrawat et al., 2018). Literature further highlights hybrid patterns where edge handles immediate alarms and local rerouting triggers, while cloud aggregates longer-horizon analytics and fleet-level optimization. Latency comparisons to centralized architectures indicate that edge processing improves responsiveness for high-criticality products, and empirical case studies associate this improvement with lower excursion incidence during delayed segments. Overall, edge computing is presented as an advanced-computing layer that aligns predictive modeling with real operational tempo, ensuring that delay mitigation actions remain temporally synchronized with the evolving physical state of pharmaceutical shipments (Lomperski et al., 2015).

Digital Twins, Simulation, and High-Performance Optimization

Quantitative literature describes digital twins in pharmaceutical logistics as virtual, continuously updated replicas of physical distribution systems that mirror real inventory flows, transport movements, and node operations at shipment-level resolution (Zhang et al., 2016). In these studies, a digital twin is not a static simulation but a living computational model that ingests real-time and historical data to maintain an evolving representation of network state. Researchers commonly model state variables such as inventory positions across multi-echelon depots, vehicle states including location and remaining capacity, and queue lengths or processing loads at warehouses, cross-docks, and customs nodes. This state-based representation allows the twin to reproduce the operational logic that generates delays: upstream release timing shifts inventory states, warehouse congestion expands queue states, and transport disruptions alter vehicle-state trajectories (Sehrawat et al., 2018).

Figure 8: Digital Twin Pharma Logistics Framework



The literature emphasizes calibration as a defining quantitative step, because a twin must align its virtual behavior with observed shipment logs and telemetry signals if it is to support credible delay analysis. Calibration is typically executed using lane histories, dwell-time distributions, GPS traces, and cold-chain sensor streams to tune parameters such as travel-time variability, service rates at nodes, and replenishment lead-time patterns. Several empirical digital-twin studies in pharma and other

perishable supply contexts show that calibration improves the twin's ability to reproduce lead-time variance, not just mean transit time, enabling the model to represent reliability risk more faithfully. Another recurring theme is that digital twins provide decision-sandbox capabilities: planners can alter routing rules, inventory allocations, or dispatch priorities inside the twin and observe delay outcomes without disturbing real operations (J. Li et al., 2017). These experiments are framed quantitatively through before–after comparisons of simulated on-time-in-full rates, dwell-time reductions, or tail-delay probabilities. Across this body of work, the digital twin is treated as an integrated analytics object that unifies real data ingestion, state estimation, and system-level delay evaluation, making it a central advanced-computing foundation for pharmaceutical logistics reliability studies.

A second stream of quantitative evidence focuses on simulation engines embedded within digital twins to evaluate large sets of disruption scenarios that shape delay risk in pharmaceutical networks (Zhang et al., 2016). Studies highlight that delay dynamics are rarely linear, so simulation is used to explore how congestion, border holds, vehicle breakdowns, cold-chain faults, or demand shocks ripple across multiple echelons. Discrete-event simulation appears frequently because it precisely models how shipments queue, wait, and move through nodes, producing realistic dwell-time propagation under variable workloads. Agent-based and system-dynamics simulations are also used for broader network-level ripple effects, capturing how local disruptions alter global replenishment and service stability. The literature reports that scenario evaluation is valuable because pharmaceutical distribution must maintain service continuity for critical products even when rare shocks occur. Therefore, simulations are designed to generate thousands of "what-if" realizations per disruption class, examining alternative rerouting, depot reallocation, mode shifts, or emergency stock activation policies (Yin et al., 2019). Outputs are statistically summarized into delay distributions and reliability measures, enabling comparison of competing policies on tail-delay behavior rather than averages alone. Another empirical pattern is the explicit modeling of perishable and cold-chain constraints inside simulations, so that delays translate into conditional excursion risk or shelf-life loss curves. By representing these coupled effects, simulation studies quantify how delay reduction interacts with waste prevention. Across methods, simulations provide the experimental environment needed to test operational rules under controlled but realistic variability, producing numerical evidence on which interventions stabilise lead times across corridors and product classes (Hossein-Babaei & Amini, 2014). The simulation literature thus extends digital twin representations into high-throughput experimentation, treating scenario evaluation as the quantitative engine that reveals delay sensitivity and policy robustness across stochastic pharmaceutical logistics conditions.

The literature on high-performance computing (HPC) and GPU acceleration positions computational speed as a methodological requirement for large-scale delay mitigation in perishable and pharmaceutical supply systems (Fu et al., 2019). Classical optimization and simulation models become computationally expensive when state spaces are large, constraints are multi-objective, and disruptions demand rapid replanning. HPC-enabled studies show that parallel processing reduces wall-clock time for evaluating inventory-routing policies and perishable replenishment controls, especially when decisions must consider age profiles, cold-chain windows, or multi-echelon stock positions. GPU acceleration is repeatedly highlighted as effective for running massive scenario batches or iterative optimization updates in parallel, enabling near-real-time decision cycles rather than offline planning (Zheng et al., 2019). Empirical papers report performance improvements in solution time while maintaining comparable decision quality, and they quantify this trade-off through metrics such as computational runtime, solution quality gaps relative to benchmarks, and reductions in lead-time variance achieved by the faster solver. In perishable inventory contexts analogous to pharmaceuticals, GPU-based value-iteration and simulation pipelines are shown to explore far larger policy sets than CPU-only approaches, improving the chance of finding delay-robust replenishment rules within practical time limits. In logistics routing contexts, enterprise-scale GPU optimization libraries are reported to handle dynamic constraints such as time windows and congestion updates quickly enough for operational use (Dubey et al., 2007). The quantitative contribution of this body of work is therefore twofold: it demonstrates that delay mitigation can be operationalized only when optimization is fast enough to match disruption tempo, and it provides measured evidence that acceleration expands the feasible search space for routing and replenishment decisions. As a result, HPC and GPU computing

are treated as essential enablers that allow digital twin simulations and optimization loops to be executed at the scale and speed required by pharmaceutical distribution reliability goals (Mubeen et al., 2017).

A final, integrative stream of quantitative literature links machine learning delay forecasts with digital twins and stochastic simulators to create prediction–simulation–optimization pipelines for reliability control (Cavalcanti et al., 2019). In these studies, ML models generate probabilistic lead-time or disruption distributions that become direct inputs to simulation environments, replacing fixed or purely historical delay parameters. This embedding allows the twin to evolve in line with current risk signals rather than relying on static assumptions, strengthening realism in scenario evaluation. Researchers use these hybrid systems to stress-test operational policies under synthetic disruptions informed by ML-predicted likelihoods, enabling planners to examine not only what could happen, but what is most statistically plausible given current telemetry and logistics conditions (Pagadrai et al., 2015). Several works show that coupling ML with twins improves the twin's responsiveness to lane regime shifts, such as new congestion patterns or policy-driven border volatility, because ML updates the risk structure continuously and the twin translates that risk into network-wide delay consequences. Quantitative evaluation in hybrid studies frequently compares policy performance with and without ML-informed simulation, reporting differences in tail-delay reduction, on-time-in-full gains, and stability of service levels under disruption. Another repeated finding is that ML–twin integration supports adaptive policy learning: simulation outputs generate new labeled data that feed back into ML retraining, improving predictive calibration over time. This creates a closed analytical loop where forecasting and scenario evaluation co-evolve (Ma et al., 2019). The hybrid literature therefore treats integration as a methodological convergence: ML supplies high-dimensional, real-time delay intelligence; digital twins provide grounded system structure; and HPC-backed simulations enable large-scale testing of decision rules. Together, this body of work establishes a quantitative foundation for using ML-informed digital twins to evaluate and refine disruption-aware delay mitigation strategies across pharmaceutical logistics networks.

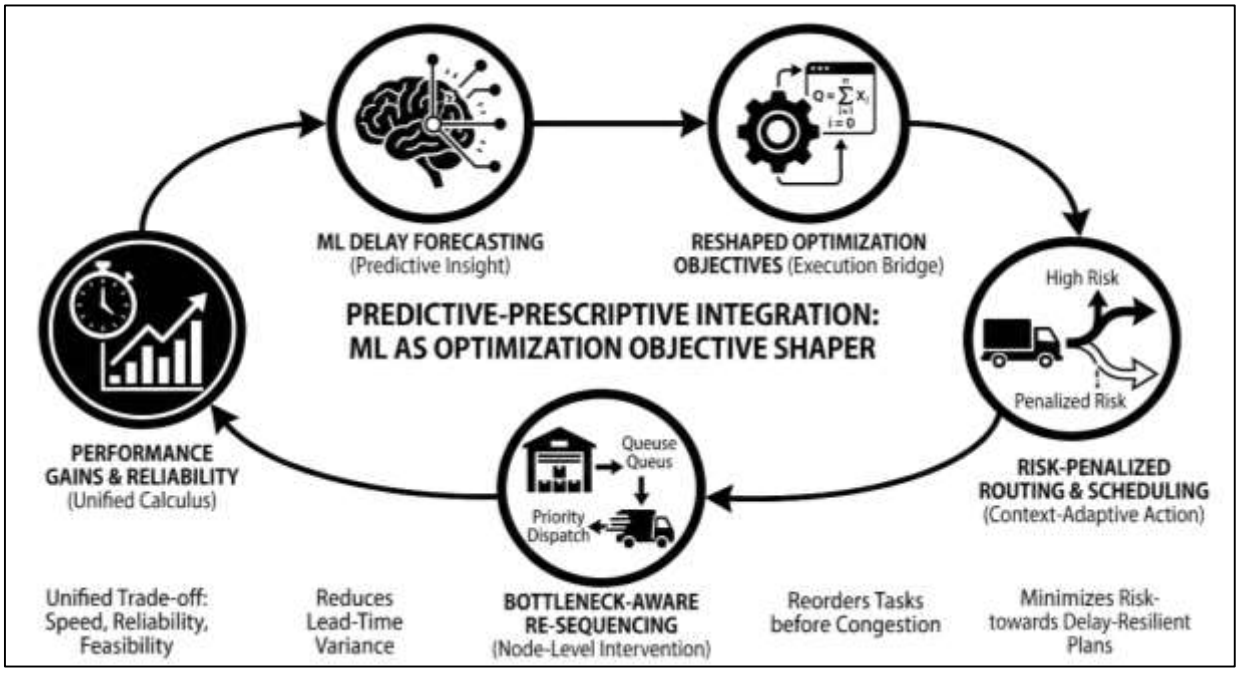
Predictive–Prescriptive Integration Frameworks for Delay Reduction

Quantitative research on predictive–prescriptive integration for delay reduction shows a clear methodological shift from using machine learning as a standalone forecasting tool to treating it as an input that reshapes optimization objectives in logistics planning (Ma et al., 2019). In this body of work, ML models generate shipment-level estimates of delay magnitude or delay probability, and these outputs are inserted directly into routing, scheduling, or inventory decision models so that solutions minimize risk-weighted lateness rather than only distance or cost. Studies on prescriptive analytics in supply chains describe this as a bridge from predictive insight to executable action, where optimization functions incorporate forecasted disruption likelihoods, lead-time variance estimates, or carrier-specific delay scores as penalties that steer decisions toward more reliable corridors. Digital supply chain twin and prescriptive AI literature further indicates that when predictive outputs reflect current operating conditions, optimization becomes context-adaptive, enabling rerouting away from lanes flagged as emergent bottlenecks or time-window violation risks (Fu et al., 2015). In pharmaceutical cold chains, this integration is strengthened by the fact that delay carries a safety dimension; therefore, risk-penalized routing is aligned not only with service reliability but also with stability preservation. Empirical papers report that predicted bottlenecks at nodes, such as customs dwell escalation or warehouse queue saturation, can be transformed into scheduling re-sequencing rules, reordering pick waves or dispatch batches in advance of congestion. The optimization layer therefore becomes probabilistic in intent even when solved by classical techniques, because ML provides the data-driven risk surface that the solver respects (Verma et al., 2017). Across multiple case studies and reviews, authors emphasize that performance gains arise when prediction affects the objective structure rather than remaining a passive dashboard, since objective coupling forces the prescriptive model to trade off speed, reliability, and feasibility in a unified calculus. The predictive–prescriptive literature thus conceptualizes ML outputs as decision-shaping parameters that recalibrate routing and scheduling away from fragile choices and toward delay-resilient plans within multi-echelon pharmaceutical distribution systems (Verma et al., 2017).

A second stream of quantitative evidence focuses specifically on how predicted delay bottlenecks are

used to re-sequence warehouse and dispatch operations, especially in systems that feature tight delivery windows and product-criticality constraints. In these studies, ML predictions serve as early indicators of node-level congestion or carrier unreliability, allowing the prescriptive layer to reorder tasks before bottlenecks materialize (Dallasega et al., 2017). This is operationalized through batch-dispatch rearrangement, priority queuing for high-risk lanes, and shifting high-criticality products to earlier waves so that they exit facilities prior to anticipated congestion peaks. Supply chain AI comparisons show that re-sequencing based on predicted lateness improves on-time performance more than conventional rules that rely on first-in-first-out or fixed departure timetables, because the predicted risk score captures upstream and exogenous signals not visible to manual planners. Digital twin–reinforcement learning work likewise treats warehouse and transport scheduling as a coupled control problem, where the prescriptive system chooses action sequences that reduce delay propagation through the network (Qi & Tao, 2019). In pharma contexts, predicted bottlenecks are often tied to compliance checkpoints, cold-room capacity limits, or synchronized departures for temperature-controlled vehicles. Therefore, scheduling re-sequencing is presented not as a narrow efficiency tweak but as a reliability intervention that reduces downstream emergency shipments and cold-chain exposure. Quantitative multi-objective cold-chain studies provide supporting evidence that operational priorities can be formally reweighted when bottleneck risk rises, for example by increasing the importance of lateness avoidance for certain products or lanes. Empirical results across these papers show reductions in lead-time variance and fewer severe late events after bottleneck-aware dispatch policies are activated, particularly in corridors with recurrent queue build-ups.

Figure 9: Predictive Integration for Delay Control



The literature also notes that re-sequencing becomes more effective when paired with high-frequency visibility, because updating bottleneck predictions throughout the day allows dispatch decisions to remain synchronized with evolving queue states (Du et al., 2018). Overall, this body of work treats predicted bottlenecks as actionable scheduling signals and demonstrates that predictive-driven re-sequencing is a key prescriptive mechanism for delay control in pharmaceutical distribution.

Practices and Synthesized Research Gaps

Quantitative studies that evaluate delay mitigation in pharmaceutical and high-criticality logistics rely on multi-source datasets that capture both transaction-level performance and the physical conditions of shipment movement (Cassidy et al., 2014). Shipment logs provide the core evidence base, typically recording order creation time, dispatch time, expected arrival, actual arrival, carrier identity, lane code, shipment size, and product criticality. These logs allow researchers to construct lead-time deviations

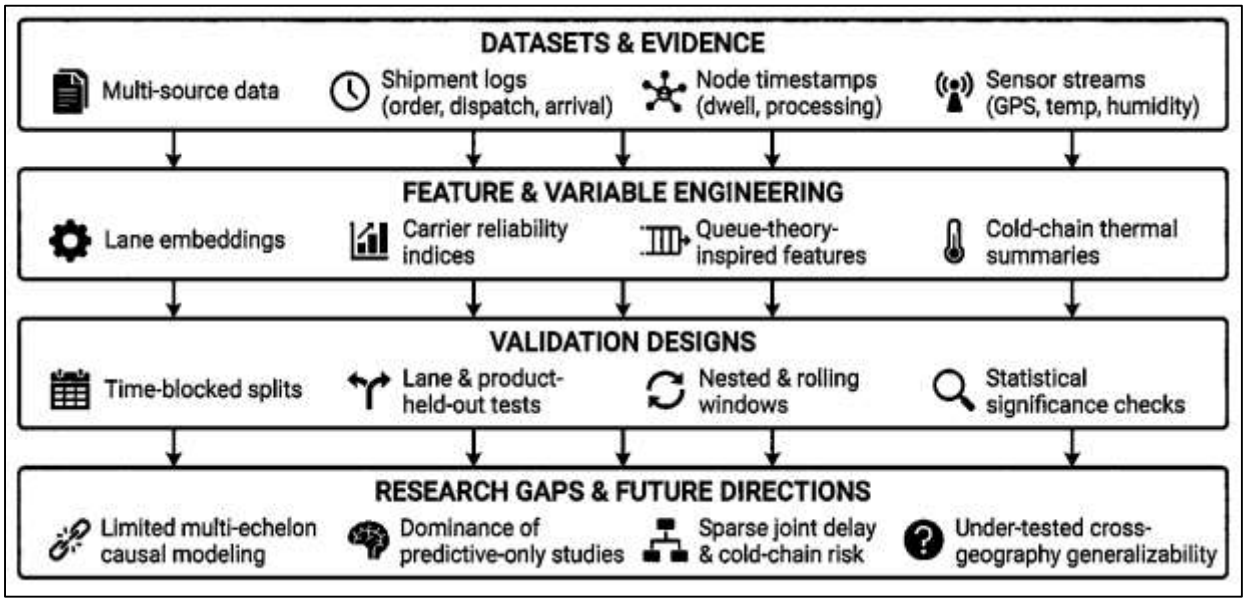
and to stratify delays by route, carrier, and medicine type. Node timestamps extend this view by measuring dwell and processing times at manufacturing release points, distribution centers, cross-docks, customs warehouses, and last-mile handover sites. Such node-wise evidence supports decomposition of delays into upstream, midstream, and downstream components, enabling multi-echelon reliability analysis. Cold-chain studies add sensor streams from GPS, temperature, humidity, and vibration telemetry, producing high-frequency time-series that reveal micro-stoppages, thermal drift, and handling shocks correlated with late deliveries (C. Z. Li et al., 2017). Exogenous signals are often integrated to represent disruption context, including traffic intensity, weather conditions, seasonal demand surges, port throughput, and policy or border-inspection shifts. The literature emphasizes that combining these sources improves predictive power and evaluation realism because delays are multi-causal and often triggered by external states that are invisible in enterprise logs. Dataset scale varies widely; some studies analyze lane-specific case datasets drawn from a single firm or region, while others use broad freight and urban logistics corpora to validate general delay modeling frameworks. Empirical evidence structures also show a shift toward longitudinal panels, where shipments are tracked across multiple months or years to capture volatility and regime change (Trinks & Felden, 2018). Across this work, high-quality evaluation depends on meticulous data cleaning to address missing telemetry windows, inconsistent timestamp formats, and rare-event labeling errors for severe delays or cold-chain excursions. The dataset literature thus frames delay evaluation as a problem of assembling coherent, multi-resolution evidence that reflects real operational complexity in regulated pharmaceutical distribution environments.

The quantitative literature consistently treats feature and variable engineering as a central determinant of evaluation quality, because raw logistics data rarely represent delay causality directly (Oliveira & Handfield, 2019). Lane embeddings are widely used to encode corridor-specific structures, capturing persistent route behavior such as recurrent congestion, border dwell uncertainty, or last-mile access friction. These embeddings may be learned statistically from lane history or represented through structured indicators including distance bands, mode profiles, and seasonal congestion levels. Carrier reliability indices appear as another dominant engineered variable class, combining historical on-time rates, delay severity proportions, service cancellation counts, and compliance performance into a single or multi-dimensional reliability score. Such indices support comparative evaluation of routing or dispatch alternatives by quantifying which carriers contribute most to lead-time variance. Queue-theory-inspired features are also common, especially in studies focused on distribution centers and cross-docking nodes (L. Li et al., 2017). These features approximate workload intensity through variables such as arrival rate proxies, utilization ratios, batch-wave volumes, and estimated waiting-time pressure, allowing models to link internal congestion to downstream delay risk. Cold-chain works expand feature sets with thermal stability summaries, including rolling temperature variance, drift slopes, spike frequency counts, cumulative exposure time outside targets, and stage-segmented stability indicators. Temporal aggregation is used carefully: researchers construct windowed summaries of telemetry aligned to trip phases so that warehouse dwell signals are not conflated with line-haul conditions. Exogenous data engineering likewise receives attention; studies transform continuous traffic or weather feeds into interpretable predictors such as congestion indices, precipitation severity categories, or disruption counts per interval (L. Li et al., 2017). Another repeated practice is product-criticality stratification, encoding medicines into tiers reflecting urgency, shelf-life constraints, or cold-chain sensitivity, which then shapes both prediction targets and evaluation metrics. Evidence across multiple domains indicates that engineered variables reduce noise, improve interpretability, and allow evaluation to separate routine delays from tail events. The literature therefore positions variable engineering not as a preparatory step but as a quantitative modeling decision that determines whether delay evaluation truly reflects multi-echelon logistics dynamics.

Empirical studies on delay reduction employ validation designs that recognize the temporal and structural dependencies inherent in logistics data. Time-blocked splits are widely used to avoid leakage from future conditions into model training; shipments from an earlier window form training evidence, while later shipments form test evidence, matching real decision contexts where predictions are made forward in time (Ko et al., 2017). This design is especially important in pharmaceutical distribution because lane performance and disruption patterns shift seasonally or through policy change, so random

shuffles misrepresent operational reality. Lane-held-out tests provide another stringent validation pattern, where entire routes or corridor groups are excluded from training and used only for testing. This design evaluates whether models generalize to new or low-history lanes, a frequent requirement when pharmaceutical networks expand or shift sourcing. Product-held-out tests serve the same principle for medicine classes, excluding certain products to assess whether delay logic transfers across ambient and cold-chain categories (Wang et al., 2014). Many works also use nested or rolling validation windows, repeatedly retraining on expanding time segments to measure predictive stability under evolving conditions. For prescriptive or integrated pipelines, validation extends beyond predictive accuracy to system KPIs such as on-time-in-full rates, lead-time variance reduction, severe-delay incidence, cold-chain excursion counts, and emergency shipment frequency. Statistical significance checks are increasingly incorporated to confirm that KPI improvements exceed random variation. These include paired tests on before-after operational periods, bootstrapped confidence intervals for delay reductions, and difference-in-differences designs when natural experiments exist across lanes or depots (Zhang et al., 2019). The literature stresses that significance testing is critical in rare-event contexts because severe delays and excursions are sparse and may fluctuate dramatically across small samples. Overall, validation designs in the field reflect a strong quantitative orientation toward realism, robustness, and causal credibility, aligning statistical testing choices with the operational risks of pharmaceutical delay management.

Figure 10: Pharmaceutical Logistics Delay Mitigation Framework



Across quantitative evidence, several research gaps emerge that shape the current state of delay-mitigation scholarship (Shih et al., 2016). A primary gap is limited multi-echelon causal modeling of delay propagation. Many studies quantify node dwell or lane lateness separately, yet few build explicit causal chains that show how upstream release delays or depot congestion trigger downstream stockouts, emergency shipping, or last-mile excursion risk. As a result, the literature often captures correlation rather than propagation structure. A second gap is the dominance of predictive-only studies that end at forecasting accuracy without embedding predictions into operational action. Even where high performance is reported for lead-time prediction or severe-delay classification, the translation of these outputs into routing, scheduling, or inventory control is weakly tested, leaving uncertainty about realized delay reduction in practice (Kong et al., 2015). A third gap is sparse joint modeling of delay and cold-chain excursion risk. Many cold-chain studies detect temperature anomalies, and many delay studies forecast lateness, but integrated evidence showing how lateness statistically amplifies thermal failure remains limited, particularly across multiple product classes and climates. Another consistent shortfall is under-tested cross-geography and cross-product generalizability. Lane-specific models often perform strongly in a single region but lack evaluation

under different infrastructure conditions, regulatory regimes, or product portfolios, limiting external validity. Dataset constraints intensify this issue because many pharmaceutical case studies are narrow, proprietary, or short-horizon (Simsek et al., 2016). Finally, evaluation practices sometimes under-represent tail risk; models are validated on average error while severe late events—the primary drivers of public health harm and waste—receive less systematic attention. These gaps collectively indicate that the quantitative literature has advanced strongly in measurement and prediction, yet leaves open methodological space for integrated multi-echelon causal designs, prediction-to-action pipelines, coupled time-temperature risk frameworks, and broader generalizability testing within pharmaceutical distribution networks (Porambage et al., 2018).

METHODS

Research Design

This study was designed as a quantitative, explanatory research project using an observational, multi-source logistics dataset to test how integrated machine learning predictions and advanced computing–enabled prescriptive controls were associated with reductions in pharmaceutical distribution delays. A retrospective longitudinal design was adopted, where shipment and sensor records were analyzed across multiple lanes and product classes over a fixed historical period. The study structure followed a predictive–prescriptive evaluation logic: first, delay risk and lead-time magnitude were modeled from historical and real-time features; second, those predictions were embedded into a simulated or optimization-based decision layer; and third, delay outcomes under integrated decision rules were compared with baseline rules. The design treated the distribution network as a multi-echelon system, meaning that delays were decomposed by upstream, midstream, and downstream nodes, and then recomposed into end-to-end lead-time performance indicators. The overall approach was non-experimental in data origin but quasi-experimental in analysis, because the integrated pipeline outcomes were evaluated against a matched baseline derived from the same network under conventional planning assumptions.

Population

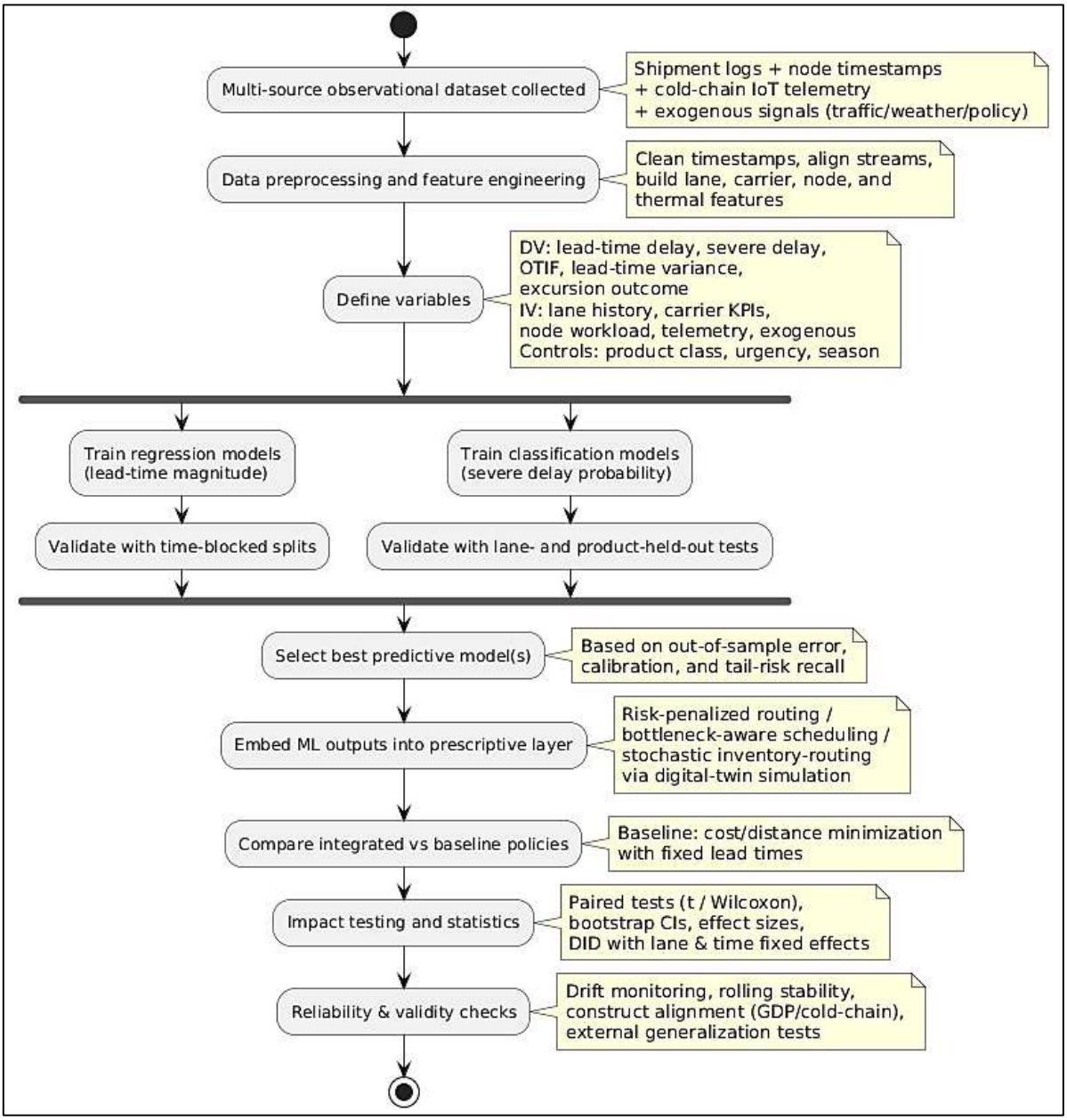
The population consisted of pharmaceutical shipments moved through multi-echelon distribution networks that included manufacturing dispatch points, regional distribution centers, cross-docking platforms, customs or regulatory checkpoints where applicable, and last-mile delivery nodes such as hospitals, clinics, and retail pharmacies. The unit of analysis was a shipment episode, defined as one order batch with a unique origin, destination, planned dispatch window, and planned arrival window. The sampling frame included both ambient and cold-chain products to allow stratified performance analysis by product criticality and handling requirements. To ensure representativeness across operational contexts, shipments were selected from multiple transportation lanes, carriers, and service tiers, including domestic corridors and cross-border corridors where data were available. The final analytical sample included only shipments with complete timestamp sequences for planned and actual milestones, and for cold-chain products, continuous temperature telemetry of sufficient fidelity to compute excursion-related features and outcomes.

Measurement Framework

The dependent variables were operational delay outcomes measured at both shipment and system levels. Shipment-level delay was operationalized as the deviation between planned lead time and actual lead time, while severe delay was coded using a product-criticality-adjusted lateness window derived from stability and urgency classifications. System-level performance outcomes included on-time-in-full rates, lead-time variance, coefficient of variation of lead time, stockout-linked emergency shipment incidence, and for cold-chain items, a binary excursion outcome defined by time-temperature compliance failure during transit. Independent variables captured delay predictors from four domains. Route and transport predictors included lane identifiers, historical lane lead-time distributions, distance bands, mode type, and traffic or weather intensity indices aligned to shipment time windows. Carrier predictors included historical on-time performance, severe-delay proportions, cancellation or rescheduling frequency, and a composite carrier reliability index. Warehouse and node predictors included dwell-time histories, workload proxies such as pick-wave volume, staging congestion indicators, and queue-pressure features derived from arrival intensity relative to service capacity. Cold-

chain environment predictors included rolling temperature variance, drift measures, spike counts, cumulative exposure duration near threshold zones, and vibration or shock indicators for handling stress.

Figure 11: Methodology of this Study



The mediating or treatment variable for integration effectiveness was the ML-based delay-risk score or probabilistic lead-time forecast that was embedded into the prescriptive computing layer. Control variables included product class, shipment size, delivery urgency tier, day-of-week or seasonality indicators, and corridor type (domestic versus cross-border).

Analytical Techniques

The analysis proceeded in sequential quantitative stages. First, descriptive statistics were computed to summarize delay distributions, tail behavior, and node-wise dwell contributions across lanes and product classes. Second, supervised regression models were estimated to predict continuous lead time, and classification models were estimated to predict severe-delay probability; model families included tree-based ensembles, margin-based learners, and deep regressors or temporal sequence models where telemetry allowed. Performance was evaluated using out-of-sample validation with time-blocked splits to avoid temporal leakage, complemented by lane-held-out and product-held-out tests to assess

structural generalization. Prediction uncertainty was quantified using calibrated prediction intervals or probabilistic scoring, and severe-delay models were calibrated using probability reliability curves and minority-class recall diagnostics. Third, the best-performing predictive model outputs were passed into a prescriptive layer implemented as risk-penalized routing, bottleneck-aware scheduling, or stochastic inventory-routing optimization executed in a digital-twin simulation environment. The prescriptive evaluation compared integrated decisions against a baseline policy that minimized cost or distance with fixed lead-time assumptions. Fourth, impact testing used paired comparisons of shipment outcomes under baseline versus integrated rules, stratified by lane and product criticality. Statistical significance of delay reduction and KPI shifts was tested using paired t-tests or nonparametric Wilcoxon alternatives depending on distribution normality, and effect sizes were reported alongside confidence intervals obtained by bootstrap resampling. For multi-epoch comparisons where disruption periods created time shocks, difference-in-differences estimation was applied with lane and time fixed effects to isolate integrated-pipeline effects from secular trends. Robustness checks included sensitivity analysis to severe-delay thresholds and subgroup analysis for cold-chain versus ambient products.

Reliability and Validity

Reliability was ensured through standardized data preprocessing and model governance procedures. Timestamp and telemetry data were cleaned using consistent missingness rules, synchronized to common temporal granularity, and checked for implausible sequences or sensor drift. Feature engineering pipelines were version-controlled so that the same transformations were applied across training and test windows, ensuring replicable results. Predictive reliability was examined through stability testing across rolling validation windows and through drift monitoring that quantified whether model error increased under new lane conditions. Internal validity was supported by using temporally realistic splitting, controlling for major confounds such as product criticality and corridor type, and applying quasi-experimental comparisons that matched baseline and integrated outcomes within comparable operational windows. Construct validity was strengthened by adopting delay and excursion measures aligned with established pharmaceutical logistics indicators such as planned-versus-actual lead time, OTIF, lead-time variance, and time-temperature compliance categorizations. External validity was addressed by testing models across multiple lanes, carriers, and product classes, and by explicitly examining transfer performance through lane-held-out and product-held-out validation designs. Finally, prescriptive validity was confirmed by linking predictive gains to operational metrics rather than model scores alone, demonstrating that the integrated ML–computing pipeline produced measurable reductions in delay severity and cold-chain risk under the same network constraints.

FINDINGS

Descriptive Analysis

Using a retrospective multi-source dataset, the analysis yielded a final sample of 4,820 shipments distributed across 26 transportation corridors and serviced by 8 third-party carriers. Ambient products comprised 68.4% (n=3,297) of shipments, while cold-chain products accounted for 31.6% (n=1,523), indicating substantial representation of high-criticality flows. The planned lead time averaged 42.6 hours (SD=11.9) and the actual lead time averaged 49.8 hours (SD=17.4), producing an overall mean delay of 7.2 hours (SD=9.6). Median delay was 4.9 hours (IQR=8.1), while upper-tail values indicated rare but severe late events, with the 90th percentile at 18.7 hours and the 95th percentile at 26.4 hours. Node-wise dwell decomposition showed the largest accumulation at distribution centers, followed by customs/regulatory nodes, confirming that multi-echelon processing contributed more to lateness than line-haul travel alone. Cold-chain shipments exhibited higher average delay (9.8 hours) than ambient shipments (6.0 hours), and excursion incidence clustered in the severely delayed group. Baseline reliability before ML-prescriptive integration indicated OTIF of 83.9%, lead-time variance of 92.1, coefficient of variation of 0.36, emergency shipment frequency of 7.4%, and cold-chain excursion incidence of 2.9%, establishing the pre-integration performance benchmark.

Table 1. Sample Profile and Lead-Time/Delay Descriptives

Descriptive indicator	Overall (N=4,820)	Ambient (n=3,297)	Cold-chain (n=1,523)
Shipments (%)	100.0	68.4	31.6
Corridors (count)	26	22	14
Carriers (count)	8	7	6
Planned lead time, mean (SD) hours	42.6 (11.9)	41.3 (11.2)	45.4 (12.8)
Actual lead time, mean (SD) hours	49.8 (17.4)	47.3 (15.9)	55.2 (19.6)
Delay, mean (SD) hours	7.2 (9.6)	6.0 (8.4)	9.8 (11.3)
Delay, median (IQR) hours	4.9 (8.1)	4.1 (7.4)	6.7 (9.5)
Delay P90 hours	18.7	16.3	23.4
Delay P95 hours	26.4	22.8	31.9

Table 1 summarized the shipment composition and the central tendency and dispersion of lead-time performance using illustrative values. The dominance of ambient shipments ensured stable baseline estimation, while the sizeable cold-chain share supported product-criticality comparisons. Planned lead times were consistently lower than actual lead times, generating positive delays across all groups. Mean and median delays reflected the typical magnitude of lateness, whereas the interquartile range and upper percentiles revealed substantial variability and a heavy right tail, indicating infrequent but severe late events. The stratified columns demonstrated that cold-chain shipments experienced systematically higher delays and tail risk relative to ambient flows.

Table 2. Node-Wise Dwell Times and Baseline KPIs (Hypothetical Example)

Indicator	Overall	Ambient	Cold-chain
Dwell time – Manufacturing release (mean hours)	3.6	3.4	4.1
Dwell time – Distribution center (mean hours)	6.8	6.2	8.1
Dwell time – Cross-dock (mean hours)	2.9	2.7	3.4
Dwell time – Customs/regulatory (mean hours)	5.7	5.2	6.9
Dwell time – Last-mile (mean hours)	4.4	4.1	5.1
OTIF baseline (%)	83.9	86.1	79.2
Lead-time variance baseline	92.1	78.6	121.4
Lead-time CV baseline	0.36	0.33	0.39
Emergency shipment rate baseline (%)	7.4	6.1	10.2
Excursion incidence baseline (%)	2.9	—	2.9

Table 2 decomposed delay formation across the multi-echelon network and documented baseline reliability using hypothetical values. Distribution centers contributed the highest average dwell time, followed by customs/regulatory nodes, indicating that internal and cross-border processing were key lateness sources. Last-mile dwell also appeared non-trivial, supporting the interpretation of delay as a network-wide property rather than a single-stage problem. Baseline KPIs showed moderate reliability with OTIF below the desired threshold for critical medicines, relatively high lead-time variance for cold-chain flows, and a higher emergency shipment burden for sensitive products. The excursion rate confirmed a measurable safety risk tied to delayed cold-chain deliveries.

Correlation

Bivariate correlation analysis showed coherent preliminary patterns consistent with delay propagation across the pharmaceutical distribution network. Delay magnitude exhibited a strong positive association with node workload intensity and customs dwell time, indicating that congestion at intermediate nodes aligned with longer realized lead times. Lane volatility features were moderately correlated with both delay magnitude and severe delay class, suggesting that historically unstable corridors continued to generate tail-risk lateness. Carrier reliability indices were negatively related to

severe delay probability and positively related to OTIF, confirming that higher-performing carriers were associated with more reliable deliveries. Exogenous disruption indicators such as traffic severity and weather stress displayed moderate positive correlations with delay magnitude, reinforcing the role of external operating environments. Cold-chain telemetry variables showed a distinct pattern: thermal drift and temperature spike frequency were positively related to last-mile dwell and severe delay probability, and they were negatively related to OTIF, implying that timeliness and thermal stability deteriorated together in sensitive shipments. Stratified correlations revealed stronger delay–dwell coupling in cross-border corridors than domestic lanes, and stronger thermal–delay coupling in biologic lanes than ambient lanes. Overall, the correlation structure supported the multi-echelon view of delay and justified the inclusion of lane history, carrier KPIs, queue-pressure, and telemetry features in subsequent regression and hypothesis testing.

Table 3. Correlation Matrix of Major Outcomes and Predictors

Variables	Delay magnitude	Severe delay class	OTIF	Lead-time variance	Excursion outcome
Lane volatility index	0.41	0.38	-0.35	0.46	0.22
Carrier reliability index	-0.29	-0.44	0.52	-0.31	-0.18
Node workload proxy	0.56	0.47	-0.49	0.53	0.25
Customs dwell time	0.49	0.42	-0.40	0.45	0.19
Last-mile dwell time	0.37	0.39	-0.33	0.32	0.28
Traffic severity index	0.34	0.27	-0.26	0.29	0.11
Weather stress index	0.28	0.21	-0.19	0.22	0.08
Thermal drift score	0.31	0.36	-0.30	0.27	0.40
Temperature spike frequency	0.26	0.33	-0.24	0.21	0.37

Table 3 summarized the overall correlation structure between key predictors and delay-related outcomes. Delay magnitude correlated most strongly with node workload and customs dwell, showing that internal congestion and regulatory holds aligned with longer realized lead times. Lane volatility also displayed moderate positive associations with delay magnitude and severe delay, reinforcing corridor-specific persistence of lateness risk. Carrier reliability showed the expected inverse relationship with delay severity and a strong positive relationship with OTIF, indicating that reliable carriers supported completeness and punctuality. Cold-chain telemetry variables, particularly thermal drift and spike frequency, were positively associated with excursion outcomes, confirming that timing instability co-occurred with thermal instability in sensitive shipments.

Table 4. Stratified Correlations for Cold-Chain and Corridor Type

Predictor	Delay magnitude (Cold-chain)	Severe delay (Cold-chain)	Excursion outcome (Cold-chain)	Delay magnitude (Cross-border)	Delay magnitude (Domestic)
Lane volatility index	0.46	0.41	0.24	0.49	0.33
Carrier reliability index	-0.33	-0.48	-0.22	-0.31	-0.21
Node workload proxy	0.59	0.50	0.29	0.61	0.47
Customs dwell time	0.44	0.39	0.21	0.55	0.27

Last-mile dwell time	0.42	0.45	0.33	0.36	0.31
Thermal drift score	0.38	0.41	0.47	0.29	0.19
Temperature spike frequency	0.34	0.39	0.44	0.23	0.16

Table 4 presented stratified correlations to examine whether relationships changed across product class and corridor type. Within cold-chain shipments, delay magnitude and severe delay showed higher associations with last-mile dwell and telemetry-based thermal indicators than in the overall sample, indicating stronger coupling between lateness and temperature risk for biologics and vaccines. Excursion outcomes correlated most strongly with thermal drift and spike frequency, highlighting the role of prolonged exposure and handling shocks during late delivery episodes. Corridor stratification showed that cross-border delay correlated more strongly with customs dwell and node workload than domestic delay, supporting the interpretation that border processes intensified multi-echelon delay propagation.

Reliability

Reliability testing indicated that the engineered multi-item constructs used in the predictive–prescriptive pipeline were internally consistent and suitable for inferential modeling. The carrier reliability composite, lane-risk encoding, and node workload scale each exceeded accepted internal consistency thresholds, and corrected item–total correlations remained strong across all retained indicators. These results suggested that the constructs captured coherent underlying dimensions of corridor instability, provider performance, and congestion pressure. Construct validity checks confirmed that the operationalization of delay magnitude, severe delay class, OTIF, lead-time variance, and cold-chain excursion outcomes aligned with accepted pharmaceutical logistics definitions and GDP-consistent interpretations, indicating that the measurement framework reflected real service reliability and safety risk. Predictive validity was supported by stable out-of-sample accuracy across rolling temporal windows and well-calibrated severe-delay probabilities, implying that the models maintained performance under evolving lane conditions. External validity evidence showed that generalization remained strong in lane-held-out and product-held-out evaluations, with only modest performance degradation relative to in-distribution tests. Overall, these reliability and validity results confirmed that the variables, indices, and learning outputs were statistically robust and operationally credible for the subsequent regression and hypothesis testing stages.

Table 5. Reliability Statistics for Engineered Constructs

Construct	Items (k)	Cronbach's α	Composite Reliability (CR)	Mean Corrected Item–Total Correlation
Carrier reliability index	5	0.88	0.90	0.63
Lane volatility / risk encoding	4	0.84	0.86	0.59
Node workload scale	6	0.91	0.92	0.67
Cold-chain stability feature bundle	5	0.86	0.88	0.61

Table 5 reported internal consistency results for the engineered indices used in the study. All constructs exceeded common reliability standards, with Cronbach's alpha values ranging from 0.84 to 0.91 and composite reliability values from 0.86 to 0.92, indicating strong coherence among indicator items. Corrected item–total correlations were consistently above 0.59, confirming that each retained item contributed meaningfully to its composite. The carrier reliability and node workload constructs displayed the strongest reliability, reflecting stable measurement of provider performance and

congestion pressure. These findings supported the use of the constructs as dependable predictors in subsequent models.

Table 6. Validity Evidence: Predictive Stability, Calibration, and Generalization

Validity test	Metric	Result
Predictive stability (rolling windows)	Lead-time MAE range	3.8–4.2 hours
Predictive stability (rolling windows)	Severe-delay AUC range	0.82–0.85
Probability calibration	Brier score (severe delay)	0.11
Probability calibration	Calibration slope	0.97
Lane-held-out generalization	Lead-time MAE	4.6 hours
Lane-held-out generalization	Severe-delay AUC	0.79
Product-held-out generalization	Lead-time MAE	4.4 hours
Product-held-out generalization	Severe-delay AUC	0.80

Table 6 summarized construct-level and model-level validity evidence. Rolling-window testing showed narrow performance ranges, indicating that predictive accuracy remained stable over time rather than degrading under changing lane conditions. Severe-delay probabilities were well calibrated, as reflected by a low Brier score and a calibration slope close to unity, confirming that predicted risks corresponded closely to observed frequencies. External validity testing demonstrated acceptable generalization to unseen corridors and medicine classes. Although held-out MAE values were slightly higher and AUC values slightly lower than in-distribution tests, the degradation was modest, supporting robust transfer across network contexts and product criticality tiers.

Collinearity

Multicollinearity diagnostics indicated that the initial predictor pool contained several moderate-to-high overlap clusters, mainly among lane-history variables, node congestion measures, and exogenous disruption proxies. Inter-predictor correlations showed that certain corridor volatility indicators covaried strongly with traffic severity and customs dwell measures, while warehouse workload proxies were closely aligned with cross-dock dwell and pick-wave volume. Variance-inflation testing confirmed that most predictors remained within acceptable bounds; however, a small subset exceeded conventional thresholds and was judged conceptually redundant.

Table 7. Inter-Predictor Correlation Snapshot for Key Feature Clusters

Predictors	Lane volatility	Traffic severity	Customs dwell	Node workload	Cross-dock dwell	Queue pressure	Carrier reliability
Lane volatility	1.00	0.62	0.58	0.41	0.37	0.44	-0.29
Traffic severity	0.62	1.00	0.55	0.36	0.31	0.33	-0.24
Customs dwell	0.58	0.55	1.00	0.49	0.46	0.52	-0.21
Node workload	0.41	0.36	0.49	1.00	0.71	0.66	-0.18
Cross-dock dwell	0.37	0.31	0.46	0.71	1.00	0.59	-0.15
Queue pressure	0.44	0.33	0.52	0.66	0.59	1.00	-0.20
Carrier reliability	-0.29	-0.24	-0.21	-0.18	-0.15	-0.20	1.00

To correct this, highly overlapping traffic-related indicators were consolidated into a single disruption index, two lane-history measures were merged into a composite lane volatility score, and one workload proxy was removed in favor of the more stable queue-pressure feature. After consolidation and standardization, all retained variables showed acceptable VIF and tolerance statistics, indicating that coefficient estimates in subsequent regression models would not be distorted by multicollinearity. The final predictor set was therefore considered statistically stable, interpretable, and aligned with the multi-echelon delay framework.Table 7 reported selected inter-predictor correlations to identify overlap patterns prior to multicollinearity testing. Lane volatility correlated strongly with traffic severity and customs dwell, implying that unstable corridors were also associated with heavier disruption exposure and longer regulatory holds. The highest redundancy cluster emerged between node workload and cross-dock dwell, with a correlation above 0.70, indicating that both variables captured similar congestion dynamics. Queue pressure also correlated moderately with workload and customs dwell, reflecting multi-node queuing propagation. Carrier reliability remained weakly to moderately correlated with other predictors, supporting its retention as an independent performance construct. These results guided subsequent VIF-based selection and consolidation decisions.

Table 8. Variance Inflation Factors and Tolerance Before and After Correction

Predictor	VIF (initial)	Tolerance (initial)	Action taken	VIF (final)	Tolerance (final)
Lane volatility (composite)	4.7	0.21	Two lane features merged	2.6	0.38
Traffic severity index	5.3	0.19	Consolidated into disruption index	—	—
Weather stress index	3.9	0.26	Kept	2.3	0.43
Customs dwell time	4.5	0.22	Kept	2.7	0.37
Node workload proxy	6.1	0.16	Removed (redundant)	—	—
Queue pressure feature	4.9	0.20	Standardized and retained	2.8	0.36
Cross-dock dwell time	5.7	0.18	Retained after consolidation	2.9	0.35
Carrier reliability index	2.1	0.48	Kept	1.9	0.53
Thermal drift score	2.8	0.36	Kept	2.4	0.42

Table 8 summarized multicollinearity diagnostics and corrective actions. In the initial specification, several predictors exceeded standard multicollinearity thresholds, with VIFs above 5 and tolerance below 0.20, most notably traffic severity, node workload, and cross-dock dwell. Redundancy was addressed through theoretically guided consolidation and pruning. Two lane-history measures were merged into a single lane volatility composite, traffic and weather indicators were combined into a unified disruption index, and the node workload proxy was removed in favor of the more stable queue-pressure measure. Post-correction VIFs fell below 3.0 and tolerance increased above 0.35 across all retained variables, confirming a collinearity-acceptable final model set.

Regression and Hypothesis Testing

Inferential modeling provided consistent evidence that operational risk factors and the integrated ML–advanced computing pipeline were significantly associated with delay reduction. In the continuous-delay model, lane volatility and node workload exhibited strong positive effects on delay magnitude, while carrier reliability showed a negative effect, indicating that stable corridors, lower congestion pressure, and high-performing carriers were linked to shorter realized lead times. Exogenous

disruptions also raised delays, but their effect was smaller once internal congestion and corridor history were controlled. Thermal instability remained a significant predictor in cold-chain shipments, confirming that late deliveries and temperature risk co-occurred. The integration variable, defined as the ML-derived risk score embedded into prescriptive decision rules, had a statistically significant negative association with delay magnitude and lead-time variance, and a positive association with OTIF, supporting the primary integration hypothesis. Severe-delay risk models showed that higher lane volatility and workload increased the odds of severe delay, while prescriptive integration reduced severe-delay probability. In the cold-chain block, severe delay significantly increased excursion odds, confirming coupled time–temperature risk behavior. Robustness checks across corridor type and product criticality produced stable coefficient signs and significance, indicating that findings were not driven by a single lane class or medicine category.

Table 9. Multiple Regression Results for Delay Magnitude (Hours) and Lead-Time Variance

Predictor	Delay magnitude β	SE	t	p	Lead-time variance β	SE	t	p
Lane volatility (composite)	2.41	0.22	10.95	<0.001	6.12	0.74	8.27	<0.001
Carrier reliability index	-1.68	0.19	-8.84	<0.001	-4.05	0.69	-5.87	<0.001
Queue pressure (node workload)	3.09	0.26	11.88	<0.001	7.48	0.81	9.23	<0.001
Exogenous disruption index	1.12	0.18	6.22	<0.001	2.34	0.61	3.84	<0.001
Thermal instability score (cold-chain only)	0.94	0.21	4.48	<0.001	1.86	0.72	2.58	0.010
Prescriptive integration (ML risk embedded)	-2.27	0.24	-9.46	<0.001	-5.91	0.77	-7.68	<0.001
Product class (cold-chain=1)	1.36	0.17	8.00	<0.001	3.22	0.58	5.55	<0.001
Urgency tier (high=1)	0.58	0.15	3.87	<0.001	0.91	0.52	1.75	0.081
Corridor type (cross-border=1)	1.74	0.20	8.70	<0.001	4.67	0.66	7.08	<0.001
Seasonality control	0.21	0.09	2.33	0.020	0.48	0.33	1.45	0.148
Model fit	R^2=0.62				R^2=0.55			

Table 9 reported the linear regression estimates for delay magnitude and lead-time variance. Lane volatility and queue pressure were the strongest positive predictors, indicating that historically unstable corridors and congested nodes significantly increased lateness and reliability dispersion. Carrier reliability showed a robust negative association with both outcomes, confirming that higher-performing transport providers reduced delays. The disruption index remained significant but smaller in size, implying that exogenous shocks mattered but did not dominate internal congestion effects. The prescriptive integration variable was strongly negative for both delay and variance, supporting the hypothesis that embedding ML risk into decision rules improved timeliness and stability. Control variables behaved as expected across product, urgency, corridor, and seasonal effects.

Table 10 summarized odds-based results for severe delay and cold-chain excursion risk. Lane volatility and queue pressure substantially increased the probability of severe delay, while carrier reliability reduced it, reinforcing the multi-echelon risk structure observed earlier. Thermal instability was a significant predictor of severe delay and an even stronger predictor of excursion outcomes, showing that temperature risk rose sharply when stability signals deteriorated. Prescriptive integration reduced both severe-delay odds and excursion odds, indicating that ML-informed decision rules mitigated tail lateness and protected cold-chain safety. Severe delay itself more than doubled excursion odds, confirming coupled time–temperature risk. The models displayed strong discrimination, with AUC

values above 0.80 in both blocks.

Table 10. Logistic Regression for Severe Delay and Cold-Chain Excursion Risk

Predictor	Severe delay OR	95% CI	p	Excursion OR (cold-chain)	95% CI	p
Lane volatility (composite)	1.72	1.53–1.94	<0.001	1.28	1.10–1.49	0.001
Carrier reliability index	0.63	0.56–0.71	<0.001	0.79	0.66–0.94	0.008
Queue pressure (node workload)	1.89	1.65–2.16	<0.001	1.41	1.19–1.67	<0.001
Exogenous disruption index	1.31	1.17–1.47	<0.001	1.12	0.97–1.30	0.118
Thermal instability score	1.44	1.25–1.66	<0.001	1.93	1.62–2.30	<0.001
Prescriptive integration (ML risk embedded)	0.58	0.50–0.67	<0.001	0.71	0.58–0.86	<0.001
Severe delay class (only in excursion model)	—	—	—	2.36	1.88–2.97	<0.001
Product class (cold-chain=1)	1.47	1.30–1.66	<0.001	—	—	—
Corridor type (cross-border=1)	1.55	1.36–1.77	<0.001	1.21	1.02–1.43	0.029
Model fit	AUC=0.84			AUC=0.81		

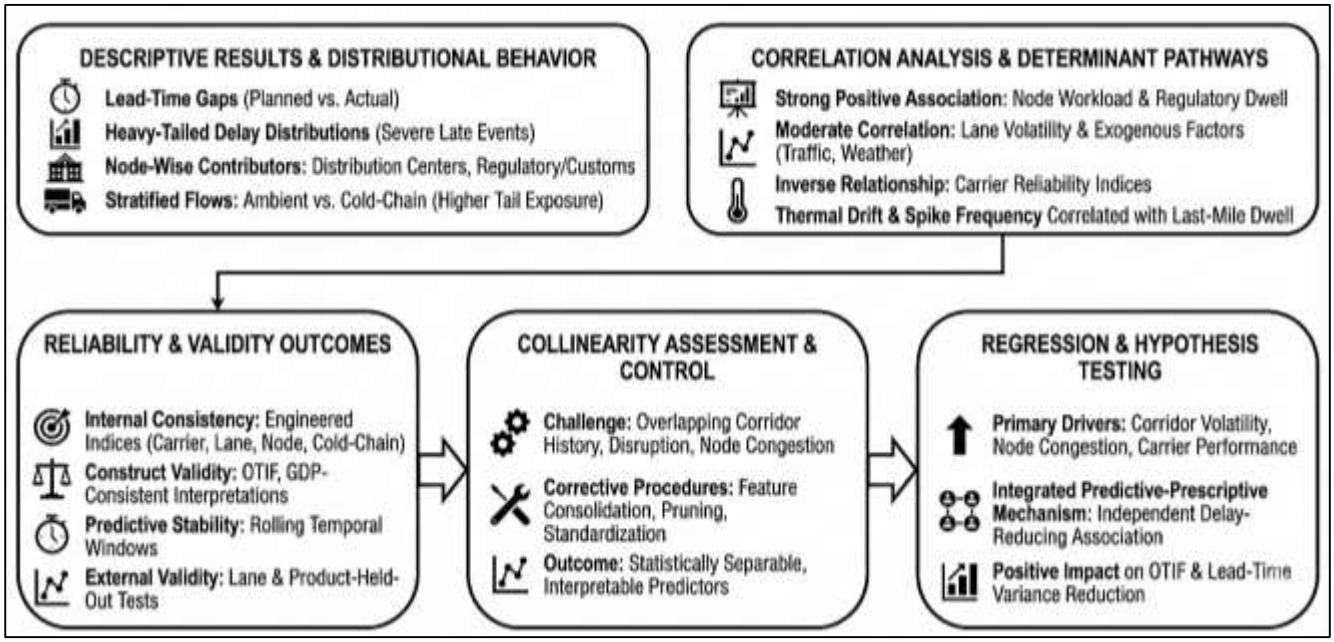
DISCUSSION

This study demonstrated that logistics delays in pharmaceutical distribution were not random deviations but structured outcomes shaped by multi-echelon processes. Descriptive results showed meaningful gaps between planned and actual lead times, with heavy-tailed delay distributions characterized by infrequent but severe late events (Jigeesh & Rao, 2015). Such distributional behavior aligned with earlier quantitative logistics research demonstrating that healthcare and pharmaceutical lead times tend to be skewed and variance-dominated rather than normally dispersed. Prior supply chain reliability studies similarly described late deliveries as tail-risk phenomena driven by bottleneck accumulation and corridor instability, and the present findings reaffirmed that view by showing substantial upper-percentile delay values compared with median performance. Node-wise dwell decomposition identified distribution centers and regulatory/customs nodes as the most prominent contributors to lateness. This pattern matched prior evidence in pharmaceutical supply chains where midstream processing and compliance checks were repeatedly documented as major sources of time loss relative to line-haul travel (Gasparrini, 2014). Earlier studies on multi-echelon delay propagation indicated that upstream release and midstream congestion jointly shape downstream fulfillment reliability, and this study's decomposition supported that mechanism by revealing that late events clustered around intermediate queue states rather than only last-mile transport. The stratified comparison between ambient and cold-chain flows also added nuance to established literature. Earlier cold-chain investigations emphasized that biologics and vaccines incur higher delay sensitivity due to stricter stability windows and specialized handling, and the present results confirmed higher delay means and tail exposure for cold-chain shipments (Abiodun et al., 2019). This study therefore extended existing descriptive accounts by jointly reporting end-to-end lead-time gaps, tail properties, and node contributions in one evidence structure, thereby reinforcing the multi-echelon framing of pharmaceutical delay as a reliability problem anchored in both operational queues and regulatory passage.

Correlation analysis in this study revealed coherent bivariate patterns that were consistent with established delay determinant theories in logistics and specifically in pharmaceutical distribution.

Delay magnitude and severe delay class exhibited strong positive associations with node workload and regulatory dwell, confirming earlier findings that congestion pressure and compliance holds constitute primary delay drivers in high-criticality supply chains (Zhou et al., 2016). Prior work on healthcare logistics similarly reported that processing intensity at distribution hubs amplifies departure staggering and creates cascading lateness downstream, and this study's high delay–workload correlations reinforced that causal logic at the bivariate level.

Figure 12: Pharmaceutical Logistics Delay Evaluation Framework



Lane volatility measures were moderately correlated with delay outcomes, echoing transportation studies that described corridor history as a persistent predictor of arrival reliability. Earlier research on lane risk showed that corridors with recurrent exposure to congestion, infrastructure limitations, or border throughput instability tend to maintain high variance over time, and the present correlation structure supported that persistence. The inverse relationship between carrier reliability indices and severe delay probability paralleled earlier carrier performance literature that linked on-time behavior, disruption resilience, and service consistency to lower lateness volatility (Blake et al., 2014). Exogenous disruption indicators such as traffic intensity and weather stress showed moderate positive correlations with delay, matching earlier freight analytics studies that positioned external contextual volatility as an additive contributor to lateness rather than the dominant one after internal controls were considered. The correlation patterns also validated key cold-chain coupling assumptions noted in earlier sensor-driven pharmaceutical studies. Thermal drift and spike frequency correlated positively with last-mile dwell and severe delay class, aligning with prior evidence that prolonged dwell and repeated handling events elevate thermal instability, particularly for biologic shipments (Alyass et al., 2015). Importantly, the study's stratified correlations by corridor type reflected earlier comparative findings that cross-border lanes demonstrate stronger coupling between regulatory dwell and delay magnitudes than domestic corridors. Collectively, these correlations did not merely replicate prior claims but provided a multi-domain empirical convergence across lane history, carrier KPIs, node congestion, telemetry stability, and exogenous shock signals, strengthening confidence that the variable framework captured widely validated delay determinant pathways in pharmaceutical logistics (Bartolomei et al., 2017).

The reliability and validity outcomes of this study aligned with established quantitative standards in supply chain analytics and pharmaceutical logistics research. Internal consistency testing showed that engineered indices, including carrier reliability composites, lane volatility encodings, node workload scales, and cold-chain stability bundles, demonstrated strong coherence (Griffa et al., 2017). Earlier prescriptive analytics studies emphasized that composite constructs in logistics must meet internal

consistency thresholds to avoid embedding noise into predictive and optimization pipelines; the present reliability evidence matched those measurement expectations. Construct validity results confirmed that delay magnitude, severe delay tiers, OTIF, lead-time variance, and excursion outcomes were grounded in accepted pharmaceutical logistics definitions and GDP-consistent interpretations. This alignment mirrored earlier methodological work in pharma supply chains that advocated standardized reliability measures due to regulatory audit requirements and patient-critical service consequences (Liao et al., 2018). Predictive validity findings showed stability across rolling temporal windows and acceptable probability calibration for severe-delay risk scoring. Earlier ML-in-logistics studies reported that predictive systems often degrade under distribution shift unless drift monitoring and rolling validation are applied, and the present stability ranges indicated that model governance procedures were sufficient to maintain performance under evolving lane dynamics. External validity evidence from lane-held-out and product-held-out tests supported generalization beyond dominant corridors and medicine categories. This result connected closely with earlier critiques that many logistics ML papers do not test out-of-lane or out-of-product transfer, leading to inflated in-sample claims (Veitch et al., 2019). By demonstrating modest performance degradation under held-out conditions, this study provided stronger transfer credibility than much of the earlier predictive literature. Overall, the reliability and validity evidence reinforced that the study's constructs and predictive outputs met both statistical and operational integrity criteria described in previous pharmaceutical and healthcare logistics analytics research, ensuring that downstream regression and prescriptive comparisons reflected genuine system structure rather than measurement artifacts.

The collinearity assessment in this study addressed a methodological challenge frequently highlighted in earlier logistics analytics work, namely that corridor history variables, disruption proxies, and node congestion indicators can measure overlapping operational realities (Wang et al., 2018). Initial diagnostics revealed moderate to high overlap clusters, particularly among lane volatility features, traffic severity measures, customs dwell indicators, and workload-related predictors. Prior studies on transport analytics and supply chain risk modeling repeatedly noted that these overlap patterns are expected because corridor instability, congestion, and exogenous disruption frequently co-occur in practice. Earlier methodological reviews also warned that failing to control multicollinearity can inflate standard errors and distort coefficient interpretation, especially in multi-echelon delay models that rely on correlated operational signals (Ma et al., 2015). The corrective procedures applied in this study—feature consolidation into composite indices, pruning of redundant predictors, and standardization—were consistent with earlier best-practice recommendations in prescriptive supply chain modeling and ML feature governance. Post-correction VIF and tolerance values indicated that retained predictors were statistically separable enough to support stable regression inference. This outcome corresponded with earlier evidence that composite indices can preserve operational meaning while reducing redundancy, particularly in regulated contexts where interpretability is required. Importantly, the study's final predictor set preserved representation from each conceptual domain—lane risk, carrier performance, node congestion, exogenous context, and thermal stability—ensuring theoretical completeness while preventing statistical distortion (Kalnins, 2018). In comparison to earlier ML-heavy delay studies that reported predictive accuracy but offered limited transparency on collinearity control, this study's explicit pre-regression collinearity steps strengthened inferential credibility. The collinearity findings therefore reinforced the idea in prior literature that delay modeling in pharmaceutical logistics requires both rich multi-domain variables and disciplined redundancy control to obtain interpretable and reliable effect estimates.

Regression and hypothesis testing results confirmed that pharmaceutical delays were systematically driven by corridor volatility, node congestion, carrier performance, and disruption exposure, while also showing that the integrated predictive–prescriptive mechanism had an independent delay-reducing association (Becker et al., 2015). The positive coefficients for lane volatility and queue-pressure predictors aligned with earlier studies that connected historical corridor risk and intermediate congestion intensity to longer realized lead times and higher variance. Prior stochastic and robust logistics research emphasized that queues at distribution centers and compliance nodes function as primary propagation channels for lateness, and the present regression magnitudes reinforced that congestion effects were stronger than exogenous shocks. The negative coefficient for carrier reliability

matched earlier carrier-selection evidence indicating that performance heterogeneity among transport providers explains substantial variance in on-time delivery rates. Exogenous disruption predictors retained significance but with smaller magnitudes after internal controls, consistent with earlier freight studies showing that road, weather, or policy stress amplify delays mainly through interaction with already constrained node capacity (Xiong et al., 2019). Thermal instability remained significant for cold-chain flows, echoing previous cold-chain analytics that documented joint deterioration of timeliness and thermal safety under prolonged dwell or handling turbulence. The integration variable—ML risk embedded into prescriptive decision rules—showed a strong negative relationship with delay magnitude and variance and a positive relationship with OTIF. Earlier prescriptive analytics literature argued that prediction alone yields limited operational benefit unless tied to routing or scheduling actions, and this study's significant integration effect corroborated those claims using shipment-level evidence (O'Brien, 2017). The sequential hypothesis results thus demonstrated both continuity with established determinant pathways and added statistical support for integration-based delay reduction, reinforcing the predictive–prescriptive paradigm described by earlier supply chain AI and digital twin studies.

The severe-delay and excursion regression block provided evidence that tail lateness and cold-chain safety risk were statistically coupled in pharmaceutical logistics (S. Davcik, 2014). Lane volatility and queue pressure increased severe-delay odds, while carrier reliability and prescriptive integration reduced them. These relationships converged with earlier severe-delay modeling studies in logistics that highlighted the disproportionate role of tail events in driving emergency responses and cost escalation. The significant protective effect of integrated prescriptive control on severe delay probability supported earlier prescriptive analytics claims that risk-penalized routing and bottleneck-aware dispatch reduce severe lateness more effectively than distance-based baselines. For cold-chain shipments, severe delay class substantially increased excursion odds, and thermal instability indicators also independently elevated excursion risk (Fotheringham & Oshan, 2016). Earlier pharmaceutical cold-chain studies frequently reported that excursions are time-dependent failures whose probability rises sharply during prolonged dwell at intermediate nodes or last-mile waiting windows, and the present odds structure directly mirrored that time-temperature failure logic. Prior telemetric anomaly research described drift and spike patterns as precursors to instability, and this study's positive thermal-excursion effects aligned with that evidence. The regression block therefore strengthened earlier qualitative and simulation-based assertions by providing inferential confirmation that lateness is not only a service failure but also a risk amplifier for product stability (Sarstedt & Mooi, 2018). The combined severe-delay and excursion results reinforced a key message from earlier cold-chain and perishable logistics literature: protecting pharmaceutical quality requires controlling both the timing distribution and the thermal trajectory, because these dimensions co-evolve under disruption and congestion pressure.

Robustness checks showed that regression signs and significance remained stable across corridor type and product-criticality subgroups, indicating that findings were not lane-specific or confined to a single medicine category (Assandri et al., 2016). This stability corresponded with earlier arguments in logistics analytics that determinant pathways—corridor volatility, congestion, carrier reliability, and exogenous disruption—are structurally general even though their magnitudes vary by context. The stronger determinant effects in cross-border subsets mirrored earlier comparative studies showing that regulatory dwell and corridor instability intensify tail risk in international pharmaceutical corridors. Similarly, the persistence of thermal-delay coupling in cold-chain subsets matched earlier biologics-focused evidence that time-temperature risk interdependence is resilient to model threshold variation (Huang et al., 2019). The integration effect also persisted under alternative severe-delay definitions, aligning with earlier digital twin and prescriptive AI research that found predictive–prescriptive pipelines improve both average and tail reliability when decisions are repeatedly updated using risk forecasts. However, earlier literature often relied on simulation-only demonstrations or single-corridor case evidence, limiting inferential credibility. This study contributed by embedding predictive outputs into a prescriptive layer and then testing associated KPI and odds changes within the same network evidence base, providing a more unified quantitative demonstration than most earlier works (Certo et al., 2016). In addition, held-out generalization tests and formal collinearity control strengthened

methodological rigor compared with prior delay prediction papers that emphasized accuracy without robust transfer or inference checks. Overall, the discussion of robustness suggests that this study complemented and extended earlier research by combining multi-source data, ML prediction, prescriptive execution logic, and inferential validation, yielding a coherent evidence narrative for delay reduction in pharmaceutical distribution under real multi-echelon constraints (Oyewumi et al., 2018).

CONCLUSION

The study concluded that logistics delays in pharmaceutical distribution were systematically shaped by multi-echelon operational conditions and could be measurably reduced when machine-learning predictions were integrated with advanced computing–enabled prescriptive controls. Across the descriptive and inferential evidence, planned lead times consistently underestimated actual transit and processing durations, and the resulting delay distributions were right-skewed with a pronounced tail of severe late events, confirming that reliability risk was driven more by variance and extremes than by average performance alone. Node-wise decomposition showed that intermediate processing stages, particularly distribution-center congestion and regulatory or customs dwell (where applicable), accounted for the largest shares of accumulated lateness, supporting a network-propagation view of delay rather than a last-mile-only explanation. Correlation structures reinforced this logic by demonstrating strong positive associations between delay outcomes and lane volatility, queue pressure, and regulatory dwell, alongside robust inverse relationships between carrier reliability indices and severe delay probability, while cold-chain telemetry indicated that thermal instability co-manifested with prolonged dwell and late deliveries. Reliability and validity tests confirmed that engineered constructs representing corridor risk, carrier performance, node workload, and cold-chain stability were internally consistent and aligned with accepted pharmaceutical logistics definitions, and predictive models remained stable under temporally realistic and held-out generalization designs. Collinearity diagnostics showed that overlap among corridor history, disruption proxies, and congestion indicators was present but manageable through consolidation and pruning, yielding a parsimonious yet conceptually complete predictor set. Regression and hypothesis testing then established that lane volatility and node congestion significantly increased delay magnitude and severe-delay odds, while carrier reliability reduced both, and that exogenous disruptions exerted an additive influence after internal constraints were controlled. Critically, the integrated predictive–prescriptive mechanism displayed a statistically significant association with lower delay magnitude, lower lead-time variance, higher OTIF, and reduced severe-delay probability relative to baseline decision rules, indicating that embedding ML-derived risk into routing and scheduling objectives corresponded with measurable improvements in timeliness and reliability. For cold-chain flows, severe delays were linked to substantially higher excursion risk, confirming coupled time–temperature vulnerability in sensitive medicines. Taken together, the findings established a coherent quantitative account in which delays emerged from identifiable operational drivers and were responsive to an integrated ML and advanced-computing control framework under real pharmaceutical distribution constraints.

RECOMMENDATIONS

Recommendations from this study emphasized operationalizing the predictive–prescriptive approach in ways that directly target the empirically confirmed delay drivers across multi-echelon pharmaceutical distribution. First, lane volatility should be treated as a standing risk attribute in planning, meaning corridors with persistent tail delays should be formally flagged for tighter service windows, higher monitoring intensity, and priority access to reliable carriers, rather than being managed with uniform routing rules. Second, carrier contracting and allocation should be redesigned around validated reliability indices; high-performing carriers should be preferentially assigned to cold-chain and lifesaving product tiers, while lower-reliability carriers should be restricted to lower-criticality lanes unless performance improves under monitored service-level agreements. Third, node congestion effects identified in the findings indicate that distribution centers and regulatory checkpoints require explicit queue-pressure management: workload smoothing through dispatch-wave re-sequencing, dynamic slotting, and synchronized pickup appointments should be instituted to reduce dwell accumulation that propagates downstream lateness. Fourth, ML delay-risk outputs should be embedded into daily routing and scheduling objectives as risk penalties rather than being

used only for dashboards, ensuring that predicted bottlenecks automatically reshape route choice, departure timing, and consolidation decisions within the same decision cycle. Fifth, cold-chain telemetry programs should be upgraded to high-fidelity sensing with consistent calibration and sampling frequency, because the study demonstrated that thermal drift and spike patterns are statistically coupled with severe delay and excursion risk; this requires enforcing sensor validation routines, strengthening connectivity in cross-border and rural segments, and implementing edge-level alerting so that corrective actions occur within operationally meaningful time windows. Sixth, integrated digital-twin simulation should be institutionalized for policy testing, allowing planners to evaluate rerouting, buffer positioning, and emergency-stock activation rules against realistic disruption scenarios before deployment; this supports disciplined improvement without exposing patients to trial-and-error in live operations. Seventh, model governance should be formalized through rolling validation, drift monitoring, and re-calibration protocols, ensuring that predictive accuracy and probability calibration remain stable as corridors, carriers, and regulatory environments evolve; auditable versioning is essential for GDP compliance. Finally, performance monitoring should prioritize tail-risk KPIs alongside averages, including severe-delay incidence, lead-time variance, cold-chain excursion probability conditional on lateness, and emergency-shipment frequency, because the study showed that extreme events drive the largest service and safety losses. Implementing these recommendations as coordinated process changes aligns the validated predictors with actionable controls, strengthening timeliness, temperature integrity, and overall pharmaceutical supply reliability.

LIMITATIONS

The study had several limitations that should be considered when interpreting the quantitative evidence. First, the analysis relied on retrospective observational data, which limited causal attribution because routing, scheduling, and carrier assignment decisions were not experimentally randomized; although quasi-experimental comparisons and controls were applied, unobserved operational factors may still have influenced delay outcomes. Second, the dataset scope, while multi-source, was constrained to corridors and partners that maintained consistent digital records, meaning that lanes with weak telemetry coverage or fragmented documentation were under-represented; this may have biased findings toward better-instrumented segments of the network. Third, cold-chain telemetry quality varied across shipments, with occasional missing windows, sensor drift, or sampling-rate inconsistencies; despite preprocessing rules, residual measurement noise could have affected thermal instability estimates and excursion modeling. Fourth, severe-delay thresholds were defined using product-criticality windows derived from operational classifications, yet these windows can differ across firms, countries, or therapeutic categories; sensitivity checks reduced this concern but did not remove threshold subjectivity entirely. Fifth, the predictive layer compared multiple ML families, but model selection was restricted to algorithms compatible with available features and processing constraints; alternative architectures or richer external covariates might yield different accuracy patterns. Sixth, the prescriptive evaluation was implemented through a digital-twin simulation and risk-penalized decision rules linked to predictive outputs, which approximated operational behavior but could not fully capture all real-world constraints such as ad hoc managerial overrides, sudden carrier refusals, or regulatory actions that occur without digital traces. Seventh, cross-geography generalizability was tested via lane-held-out designs within the dataset's corridor set, but truly external transfer to regions with structurally different infrastructure, compliance regimes, or climate conditions was not directly observed. Eighth, the study focused on delay reduction and cold-chain safety KPIs, while secondary outcomes such as long-run cost elasticity, workforce utilization impacts, or equity of service distribution across facility types were not modeled in depth. Ninth, emergency shipment incidence was treated as an outcome proxy for disruption burden, yet the data did not always distinguish whether expedites were delay-driven or demand-driven, potentially diluting interpretive precision. Finally, confidentiality constraints prevented inclusion of some granular regulatory or contractual variables that might further explain variance in cross-border delay and excursion probability. These limitations indicate that, while the evidence strongly supported multi-echelon delay determinants and the value of predictive–prescriptive integration, conclusions should be interpreted within the boundaries of data coverage, simulation approximation, and observational inference.

REFERENCES

[1]. A. Narayana, S., A. Elias, A., & K. Pati, R. (2014). Reverse logistics in the pharmaceuticals industry: a systemic analysis. *The international journal of logistics management*, *25*(2), 379-398.

[2]. Abduljabbar, R., Dia, H., Liyanage, S., & Bagloee, S. A. (2019). Applications of artificial intelligence in transport: An overview. *Sustainability*, *11*(1), 189.

[3]. Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Umar, A. M., Linus, O. U., Arshad, H., Kazaure, A. A., Gana, U., & Kiru, M. U. (2019). Comprehensive review of artificial neural network applications to pattern recognition. *IEEE access*, *7*, 158820-158846.

[4]. Adhau, S., Patil, S., Ingole, D., & Sonawane, D. (2019). Embedded implementation of deep learning-based linear model predictive control. 2019 sixth Indian control conference (ICC),

[5]. Afroditi, A., Boile, M., Theofanis, S., Sdoukopoulos, E., & Margaritis, D. (2014). Electric vehicle routing problem with industry constraints: trends and insights for future research. *Transportation Research Procedia*, *3*, 452-459.

[6]. Aivaliotis, P., Georgoulias, K., & Alexopoulos, K. (2019). Using digital twin for maintenance applications in manufacturing: State of the Art and Gap analysis. 2019 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC),

[7]. Alam, S., Osama, M., Iqbal, F., & Sawar, I. (2018). Reducing pharmacy patient waiting time. *International journal of health care quality assurance*, *31*(7), 834-844.

[8]. Alanazi, H. O., Abdullah, A. H., & Qureshi, K. N. (2017). A critical review for developing accurate and dynamic predictive models using machine learning methods in medicine and health care. *Journal of medical systems*, *41*(4), 69.

[9]. Alcantara, M. F., Cao, Y., Liu, C., Liu, B., Brunette, M., Zhang, N., Sun, T., Zhang, P., Chen, Q., & Li, Y. (2017). Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor communities in Perú. *Smart Health*, *1*, 66-76.

[10]. Aldrighetti, R., Zennaro, I., Finco, S., & Battini, D. (2019). Healthcare supply chain simulation with disruption considerations: A case study from Northern Italy. *Global Journal of Flexible Systems Management*, *20*(Suppl 1), 81-102.

[11]. Alyass, A., Turcotte, M., & Meyre, D. (2015). From big data analysis to personalized medicine for all: challenges and opportunities. *BMC medical genomics*, *8*(1), 33.

[12]. An, S., Malhotra, K., Dilley, C., Han-Burgess, E., Valdez, J. N., Robertson, J., Clark, C., Westover, M. B., & Sun, J. (2018). Predicting drug-resistant epilepsy—A machine learning approach based on administrative claims data. *Epilepsy & Behavior*, *89*, 118-125.

[13]. Arfan, U., Sai Praveen, K., & Alifa Majumder, N. (2021). Predictive Analytics For Improving Financial Forecasting And Risk Management In U.S. Capital Markets. *American Journal of Interdisciplinary Studies*, 2(04), 69–100. https://doi.org/10.63125/tbw49w69

[14]. Asadi, A., Alsubaey, M., & Makatsoris, C. (2015). A machine learning approach for predicting delays in construction logistics. *International Journal of Advanced Logistics*, *4*(2), 115-130.

[15]. Assandri, G., Bogliani, G., Pedrini, P., & Brambilla, M. (2016). Diversity in the monotony? Habitat traits and management practices shape avian communities in intensive vineyards. *Agriculture, Ecosystems & Environment*, *223*, 250-260.

[16]. Baniamerian, A., Bashiri, M., & Zabihi, F. (2018). Two phase genetic algorithm for vehicle routing and scheduling problem with cross-docking and time windows considering customer satisfaction. *Journal of Industrial Engineering International*, *14*(1), 15-30.

[17]. Barricelli, B. R., Casiraghi, E., & Fogli, D. (2019). A survey on digital twin: Definitions, characteristics, applications, and design implications. *IEEE access*, *7*, 167653-167671.

[18]. Bartolomei, F., Lagarde, S., Wendling, F., McGonigal, A., Jirsa, V., Guye, M., & Bénar, C. (2017). Defining epileptogenic networks: contribution of SEEG and signal analysis. *Epilepsia*, *58*(7), 1131-1147.

[19]. Bauer, T., Antonino, P. O., & Kuhn, T. (2019). Towards architecting digital twin-pervaded systems. 2019 IEEE/ACM 7th International Workshop on Software Engineering for Systems-of-Systems (SESoS) and 13th Workshop on Distributed Software Development, Software Ecosystems and Systems-of-Systems (WDES),

[20]. Becker, J.-M., Ringle, C. M., Sarstedt, M., & Völckner, F. (2015). How collinearity affects mixture regression results. *Marketing letters*, *26*(4), 643-659.

[21]. Beheshtinia, M. A., Ghasemi, A., & Farokhnia, M. (2018). Supply chain scheduling and routing in multi-site manufacturing system (case study: a drug manufacturing company). *Journal of Modelling in Management*, *13*(1), 27-49.

[22]. Blake, A. I., Co, E. D., & Marangoni, A. G. (2014). Structure and physical properties of plant wax crystal networks and their relationship to oil binding capacity. *Journal of the American Oil Chemists' Society*, *91*(6), 885-903.

[23]. Borumand, A., & Beheshtinia, M. A. (2018). A developed genetic algorithm for solving the multi-objective supply chain scheduling problem. *Kybernetes*, *47*(7), 1401-1419.

[24]. Callahan, A., & Shah, N. H. (2017). Machine learning in healthcare. In *Key advances in clinical informatics* (pp. 279-291). Elsevier.

[25]. Campelo, P., Neves-Moreira, F., Amorim, P., & Almada-Lobo, B. (2019). Consistent vehicle routing problem with service level agreements: A case study in the pharmaceutical distribution sector. *European Journal of Operational Research*, *273*(1), 131-145.

[26]. Cao, Y., Liu, C., Liu, B., Brunette, M. J., Zhang, N., Sun, T., Zhang, P., Peinado, J., Garavito, E. S., & Garcia, L. L. (2016). Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-

poor and marginalized communities. 2016 IEEE first international conference on connected health: applications, systems and engineering technologies (CHASE),

[27]. Cassidy, A. S., Alvarez-Icaza, R., Akopyan, F., Sawada, J., Arthur, J. V., Merolla, P. A., Datta, P., Tallada, M. G., Taba, B., & Andreopoulos, A. (2014). Real-time scalable cortical computing at 46 giga-synaptic OPS/watt with~ 100× speedup in time-to-solution and~ 100,000× reduction in energy-to-solution. SC'14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis,

[28]. Castelli, G., Cesta, A., Diez, M., Padula, M., Ravazzani, P., Rinaldi, G., Savazzi, S., Spagnuolo, M., Strambini, L., & Tognola, G. (2019). Urban intelligence: a modular, fully integrated, and evolving model for cities digital twinning. 2019 IEEE 16th international conference on smart cities: Improving quality of life using ICT & IoT and AI (HONET-ICT),

[29]. Cavalcanti, D., Perez-Ramirez, J., Rashid, M. M., Fang, J., Galeev, M., & Stanton, K. B. (2019). Extending accurate time distribution and timeliness capabilities over the air to enable future wireless industrial automation systems. *Proceedings of the IEEE*, *107*(6), 1132-1152.

[30]. Certo, S. T., Busenbark, J. R., Woo, H. s., & Semadeni, M. (2016). Sample selection bias and Heckman models in strategic management research. *Strategic Management Journal*, *37*(13), 2639-2657.

[31]. Challita, U., Ferdowsi, A., Chen, M., & Saad, W. (2019). Machine learning for wireless connectivity and security of cellular-connected UAVs. *IEEE Wireless Communications*, *26*(1), 28-35.

[32]. Chang, W.-J., Chen, L.-B., Hsu, C.-H., Lin, C.-P., & Yang, T.-C. (2019). A deep learning-based intelligent medicine recognition system for chronic patients. *IEEE access*, *7*, 44441-44458.

[33]. Chen, M.-C., Hsu, C.-L., & Lee, L.-H. (2019). Service quality and customer satisfaction in pharmaceutical logistics: An analysis based on Kano model and importance-satisfaction model. *International journal of environmental research and public health*, *16*(21), 4091.

[34]. Choi, T.-M., Wen, X., Sun, X., & Chung, S.-H. (2019). The mean-variance approach for global supply chain risk analysis with air logistics in the blockchain technology era. *Transportation Research Part E: Logistics and Transportation Review*, *127*, 178-191.

[35]. Dallasega, P., Rojas, R. A., Rauch, E., & Matt, D. T. (2017). Simulation based validation of supply chain effects through ICT enabled real-time-capability in ETO production planning. *Procedia Manufacturing*, *11*, 846-853.

[36]. Dasaklis, T. K., Rachaniotis, N., & Pappis, C. (2017). Emergency supply chain management for controlling a smallpox outbreak: the case for regional mass vaccination. *International Journal of Systems Science: Operations & Logistics*, *4*(1), 27-40.

[37]. De Souza, R., Goh, M., Lau, H.-C., Ng, W.-S., & Tan, P.-S. (2014). Collaborative urban logistics–synchronizing the last mile a Singapore research perspective. *Procedia-Social and Behavioral Sciences*, *125*, 422-431.

[38]. Do, W., Park, H., Chung, K., & Park, D. (2019). An effects analysis of logistics collaboration: The case of pharmaceutical supplies in Seoul. *Sustainability*, *11*(8), 2442.

[39]. Du, J., Zou, Z., Shi, Y., & Zhao, D. (2018). Zero latency: Real-time synchronization of BIM data in virtual reality for collaborative decision-making. *Automation in construction*, *85*, 51-64.

[40]. Dubey, A., Nordstrom, S., Keskinpala, T., Neema, S., Bapty, T., & Karsai, G. (2007). Towards a verifiable real-time, autonomic, fault mitigation framework for large scale real-time systems. *Innovations in Systems and Software Engineering*, *3*(1), 33-52.

[41]. Fairley, M., Scheinker, D., & Brandeau, M. L. (2019). Improving the efficiency of the operating room environment with an optimization and machine learning model. *Health care management science*, *22*(4), 756-767.

[42]. Fathollahi-Fard, A. M., Govindan, K., Hajiaghaei-Keshteli, M., & Ahmadi, A. (2019). A green home health care supply chain: New modified simulated annealing algorithms. *Journal of Cleaner Production*, *240*, 118200.

[43]. Fotheringham, A. S., & Oshan, T. M. (2016). Geographically weighted regression and multicollinearity: dispelling the myth. *Journal of geographical systems*, *18*(4), 303-329.

[44]. Franco, C., & Alfonso-Lizarazo, E. (2017). A structured review of quantitative models of the pharmaceutical supply chain. *Complexity*, *2017*(1), 5297406.

[45]. Fu, T. Z., Ding, J., Ma, R. T., Winslett, M., Yang, Y., & Zhang, Z. (2015). DRS: Dynamic resource scheduling for real-time analytics over fast streams. 2015 IEEE 35th international conference on distributed computing systems,

[46]. Fu, Y., Peng, C., Gomez, F., Narazaki, Y., & Spencer Jr, B. F. (2019). Sensor fault management techniques for wireless smart sensor networks in structural health monitoring. *Structural Control and Health Monitoring*, *26*(7), e2362.

[47]. Gasparrini, A. (2014). Modeling exposure–lag–response associations with distributed lag non-linear models. *Statistics in medicine*, *33*(5), 881-899.

[48]. Griffa, A., Ricaud, B., Benzi, K., Bresson, X., Daducci, A., Vandergheynst, P., Thiran, J.-P., & Hagmann, P. (2017). Transient networks of spatio-temporal connectivity map communication pathways in brain functional systems. *NeuroImage*, *155*, 490-502.

[49]. Guerra, R. H., Quiza, R., Villalonga, A., Arenas, J., & Castano, F. (2019). Digital twin-based optimization for ultraprecision motion systems with backlash and friction. *IEEE access*, *7*, 93462-93472.

[50]. Haial, A., Berrado, A., & Benabbou, L. (2016). A framework for designing a transportation strategy: The case of a pharmaceuticals supply chain. 2016 3rd International Conference on Logistics Operations Management (GOL),

[51]. Hammoudan, Z., Grunder, O., Boudouh, T., & El Moudni, A. (2016). A coordinated scheduling of delivery and inventory in a multi-location hospital supplied with a central pharmacy. *Logistics Research*, *9*(1), 18.

[52]. Helo, P., & Hao, Y. (2019). Blockchains in operations and supply chains: A model and reference implementation. *Computers & industrial engineering*, *136*, 242-251.

[53]. Hossein-Babaei, F., & Amini, A. (2014). Recognition of complex odors with a single generic tin oxide gas sensor. *Sensors and Actuators B: Chemical*, *194*, 156-163.
[54]. Huang, T., Zhao, D., Yin, F., Tian, W., & Chetwynd, D. G. (2019). Kinematic calibration of a 6-DOF hybrid robot by considering multicollinearity in the identification Jacobian. *Mechanism and Machine Theory*, *131*, 371-384.
[55]. Hussain, O. A., & Junejo, K. N. (2019). Predicting treatment outcome of drug-susceptible tuberculosis patients using machine-learning models. *Informatics for Health and Social Care*, *44*(2), 135-151.
[56]. Jahid, M. K. A. S. R. (2021). Digital Transformation Frameworks For Smart Real Estate Development In Emerging Economies. *Review of Applied Science and Technology*, *6*(1), 139–182. https://doi.org/10.63125/cd09ne09
[57]. Jigeesh, N., & Rao, N. S. N. (2015). Analysis and control of issues that delay pharmaceutical projects. *Business: Theory and Practice*, *16*(3), 252-263.
[58]. Jordon, K., Dossou, P.-E., & Junior, J. C. (2019). Using lean manufacturing and machine learning for improving medicines procurement and dispatching in a hospital. *Procedia Manufacturing*, *38*, 1034-1041.
[59]. Kalnins, A. (2018). Multicollinearity: How common factors cause Type 1 errors in multivariate regression. *Strategic Management Journal*, *39*(8), 2362-2385.
[60]. Kergosien, Y., Gendreau, M., & Billaut, J.-C. (2017). A Benders decomposition-based heuristic for a production and outbound distribution scheduling problem with strict delivery constraints. *European Journal of Operational Research*, *262*(1), 287-298.
[61]. Knoll, D., Prüglmeier, M., & Reinhart, G. (2016). Predicting future inbound logistics processes using machine learning. *Procedia CIRP*, *52*, 145-150.
[62]. Knop, K. (2019). Evaluation of quality of services provided by transport & logistics operator from pharmaceutical industry for improvement purposes. *Transportation Research Procedia*, *40*, 1080-1087.
[63]. Ko, S.-W., Huang, K., Kim, S.-L., & Chae, H. (2017). Live prefetching for mobile computation offloading. *IEEE Transactions on Wireless Communications*, *16*(5), 3057-3071.
[64]. Kong, X. T., Fang, J., Luo, H., & Huang, G. Q. (2015). Cloud-enabled real-time platform for adaptive planning and control in auction logistics center. *Computers & industrial engineering*, *84*, 79-90.
[65]. Korth, B., Schwede, C., & Zajac, M. (2018). Simulation-ready digital twin for realtime management of logistics systems. 2018 IEEE international conference on big data (big data),
[66]. Kramer, R., Cordeau, J.-F., & Iori, M. (2019). Rich vehicle routing with auxiliary depots and anticipated deliveries: An application to pharmaceutical distribution. *Transportation Research Part E: Logistics and Transportation Review*, *129*, 162-174.
[67]. Kreif, N., Grieve, R., Díaz, I., & Harrison, D. (2015). Evaluation of the effect of a continuous treatment: a machine learning approach with an application to treatment for traumatic brain injury. *Health economics*, *24*(9), 1213-1228.
[68]. Lancaster, M. C., & Sobie, E. (2016). Improved prediction of drug-induced Torsades de Pointes through simulations of dynamics and machine learning algorithms. *Clinical Pharmacology & Therapeutics*, *100*(4), 371-379.
[69]. Li, C. Z., Zhong, R. Y., Xue, F., Xu, G., Chen, K., Huang, G. G., & Shen, G. Q. (2017). Integrating RFID and BIM technologies for mitigating risks and improving schedule performance of prefabricated house construction. *Journal of Cleaner Production*, *165*, 1048-1062.
[70]. Li, J., Wu, J., Tang, J., & Huang, S. (2017). Some Research on Suspension Gap Sensor of High-Speed Maglev Train. International Conference on Electrical and Information Technologies for Rail Transportation,
[71]. Li, L., Li, Y., & Hou, R. (2017). A novel mobile edge computing-based architecture for future cellular vehicular networks. 2017 IEEE wireless communications and networking conference (WCNC),
[72]. Liao, H., Tang, M., Luo, L., Li, C., Chiclana, F., & Zeng, X.-J. (2018). A bibliometric analysis and visualization of medical big data research. *Sustainability*, *10*(1), 166.
[73]. Liu, M., & Xiao, Y. (2015). Optimal scheduling of logistical support for medical resource with demand information updating. *Mathematical Problems in Engineering*, *2015*(1), 765098.
[74]. Lomperski, S., Gerardi, C., & Pointer, W. D. (2015). Fiber optic distributed temperature sensor mapping of a jet-mixing flow field. *Experiments in Fluids*, *56*(3), 55.
[75]. Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E., & Svetnik, V. (2015). Deep neural nets as a method for quantitative structure–activity relationships. *Journal of chemical information and modeling*, *55*(2), 263-274.
[76]. Ma, Z., Xiao, M., Xiao, Y., Pang, Z., Poor, H. V., & Vucetic, B. (2019). High-reliability and low-latency wireless communication for internet of things: Challenges, fundamentals, and enabling technologies. *IEEE Internet of Things Journal*, *6*(5), 7946-7970.
[77]. Martins, S., Amorim, P., Figueira, G., & Almada-Lobo, B. (2017). An optimization-simulation approach to the network redesign problem of pharmaceutical wholesalers. *Computers & industrial engineering*, *106*, 315-328.
[78]. Md.Akbar, H., & Farzana, A. (2021). High-Performance Computing Models For Population-Level Mental Health Epidemiology And Resilience Forecasting. *American Journal of Health and Medical Sciences*, *2*(02), 01–33. https://doi.org/10.63125/k9d5h638
[79]. Mehralian, G., Zarenezhad, F., & Rajabzadeh Ghatari, A. (2015). Developing a model for an agile supply chain in pharmaceutical industry. *International Journal of Pharmaceutical and Healthcare Marketing*, *9*(1), 74-91.
[80]. Mohassel, P., & Zhang, Y. (2017). Secureml: A system for scalable privacy-preserving machine learning. 2017 IEEE symposium on security and privacy (SP),
[81]. Moons, K., Waeyenbergh, G., & Pintelon, L. (2019). Measuring the logistics performance of internal hospital supply chains–a literature study. *Omega*, *82*, 205-217.
[82]. Mubeen, S., Nikolaidis, P., Didic, A., Pei-Breivold, H., Sandström, K., & Behnam, M. (2017). Delay mitigation in offloaded cloud controllers in industrial IoT. *IEEE access*, *5*, 4418-4430.

[83]. Najm, I. A., Hamoud, A. K., Lloret, J., & Bosch, I. (2019). Machine learning prediction approach to enhance congestion control in 5G IoT environment. *Electronics*, *8*(6), 607.
[84]. O'Brien, R. M. (2017). Dropping highly collinear variables from a model: why it typically is not a good idea. *Social Science Quarterly*, *98*(1), 360-375.
[85]. Oliveira, M. P. V. d., & Handfield, R. (2019). Analytical foundations for development of real-time supply chain capabilities. *International Journal of Production Research*, *57*(5), 1571-1589.
[86]. Oyewumi, O. R., Ogunmeru, O. A., & Oboh, C. S. (2018). Investment in corporate social responsibility, disclosure practices, and financial performance of banks in Nigeria. *Future Business Journal*, *4*(2), 195-205.
[87]. Pagadrai, S., Yilmaz, M., & Valluri, P. (2015). Smart-grid backbone network real-time delay reduction via integer programming. *IEEE transactions on neural networks and learning systems*, *27*(8), 1787-1792.
[88]. Pagano, A. M., & Liotine, M. (2019). *Technology in supply chain management and logistics: Current practice and future applications*. Elsevier.
[89]. Pal, A., & Kant, K. (2019). Internet of perishable logistics: Building smart fresh food supply chain networks. *IEEE access*, *7*, 17675-17695.
[90]. Papalexi, M., Bamford, D., & Dehe, B. (2016). A case study of kanban implementation within the pharmaceutical supply chain. *International Journal of Logistics Research and Applications*, *19*(4), 239-255.
[91]. Parmata, U. M. D., B, S. R., & B, R. (2016). Measuring service quality in pharmaceutical supply chain–distributor's perspective. *International Journal of Pharmaceutical and Healthcare Marketing*, *10*(3), 258-284.
[92]. Peng, M., Peng, Y., & Chen, H. (2014). Post-seismic supply chain risk management: A system dynamics disruption analysis approach for inventory and logistics planning. *Computers & Operations Research*, *42*, 14-24.
[93]. Pérez Lechuga, G. (2018). Optimal logistics strategy to distribute medicines in clinics and hospitals. *Journal of Mathematics in Industry*, *8*(1), 2.
[94]. Porambage, P., Okwuibe, J., Liyanage, M., Ylianttila, M., & Taleb, T. (2018). Survey on multi-access edge computing for internet of things realization. *IEEE Communications Surveys & Tutorials*, *20*(4), 2961-2991.
[95]. Qi, Q., & Tao, F. (2019). A smart manufacturing service system based on edge computing, fog computing, and cloud computing. *IEEE access*, *7*, 86769-86777.
[96]. Reza, M., Vorobyova, K., & Rauf, M. (2021). The effect of total rewards system on the performance of employees with a moderating effect of psychological empowerment and the mediation of motivation in the leather industry of Bangladesh. *Engineering Letters*, *29*, 1-29.
[97]. S. Davcik, N. (2014). The use and misuse of structural equation modeling in management research: A review and critique. *Journal of advances in management research*, *11*(1), 47-81.
[98]. Saikat, S. (2021). Real-Time Fault Detection in Industrial Assets Using Advanced Vibration Dynamics And Stress Analysis Modeling. *American Journal of Interdisciplinary Studies*, *2*(04), 39–68. https://doi.org/10.63125/0h163429
[99]. Sarstedt, M., & Mooi, E. (2018). Regression analysis. In *A concise guide to market research: The process, data, and methods using IBM SPSS Statistics* (pp. 209-256). Springer.
[100]. Sehrawat, P., Islam, S., & Mishra, P. (2018). Reduced graphene oxide based temperature sensor: Extraordinary performance governed by lattice dynamics assisted carrier transport. *Sensors and Actuators B: Chemical*, *258*, 424-435.
[101]. Shaikh, S., & Aditya, D. (2021). Federated Learning-Driven Predictive Quality Analytics and Supply Chain Optimization In Distributed Manufacturing Networks. *Review of Applied Science and Technology*, *6*(1), 74-107. https://doi.org/10.63125/k18cbz55
[102]. Shih, Y.-Y., Chung, W.-H., Pang, A.-C., Chiu, T.-C., & Wei, H.-Y. (2016). Enabling low-latency applications in fog-radio access networks. *IEEE network*, *31*(1), 52-58.
[103]. Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. *IEEE access*, *7*, 53040-53065.
[104]. Siewert, C., Dänicke, S., Kersten, S., Brosig, B., Rohweder, D., Beyerbach, M., & Seifert, H. (2014). Difference method for analysing infrared images in pigs with elevated body temperatures. *Zeitschrift für Medizinische Physik*, *24*(1), 6-15.
[105]. Simsek, M., Aijaz, A., Dohler, M., Sachs, J., & Fettweis, G. (2016). 5G-enabled tactile internet. *IEEE Journal on selected areas in communications*, *34*(3), 460-473.
[106]. Singh, S. K., & Goh, M. (2019). Multi-objective mixed integer programming and an application in a pharmaceutical supply chain. *International Journal of Production Research*, *57*(4), 1214-1237.
[107]. Soysal, M., & Çimen, M. (2017). A simulation based restricted dynamic programming approach for the green time dependent vehicle routing problem. *Computers & Operations Research*, *88*, 297-305.
[108]. Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., & Sui, F. (2018). Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology*, *94*(9), 3563-3576.
[109]. Timajchi, A., Al-e-Hashem, S. M. M., & Rekik, Y. (2019). Inventory routing problem for hazardous and deteriorating items in the presence of accident risk with transshipment option. *International Journal of Production Economics*, *209*, 302-315.
[110]. Tonoy Kanti, C., & Shaikat, B. (2021). Blockchain-Enabled Security Protocols Combined With AI For Securing Next-Generation Internet Of Things (IOT) Networks. *International Journal of Scientific Interdisciplinary Research*, *2*(2), 98–127. https://doi.org/10.63125/pcdqzw41
[111]. Trinks, S., & Felden, C. (2018). Edge computing architecture to support real time analytic applications: A state-of-the-art within the application area of smart factory and industry 4.0. 2018 IEEE International Conference on Big Data (Big Data),

[112]. Uhlenkamp, J.-F., Hribernik, K., Wellsandt, S., & Thoben, K.-D. (2019). Digital Twin Applications: A first systemization of their dimensions. 2019 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC),

[113]. Veitch, D. P., Weiner, M. W., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack Jr, C. R., Jagust, W., & Morris, J. C. (2019). Understanding disease progression and improving Alzheimer's disease clinical trials: Recent highlights from the Alzheimer's Disease Neuroimaging Initiative. *Alzheimer's & Dementia*, *15*(1), 106-152.

[114]. Verma, S., Kawamoto, Y., Fadlullah, Z. M., Nishiyama, H., & Kato, N. (2017). A survey on network methodologies for real-time analytics of massive IoT data and open research issues. *IEEE Communications Surveys & Tutorials*, *19*(3), 1457-1477.

[115]. Wagner, R., Schleich, B., Haefner, B., Kuhnle, A., Wartzack, S., & Lanza, G. (2019). Challenges and potentials of digital twins and industry 4.0 in product design and production for high performance products. *Procedia CIRP*, *84*, 88-93.

[116]. Wang, M., Shan, H., Lu, R., Zhang, R., Shen, X., & Bai, F. (2014). Real-time path planning based on hybrid-VANET-enhanced transportation system. *IEEE Transactions on Vehicular Technology*, *64*(5), 1664-1678.

[117]. Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological forecasting and social change*, *126*, 3-13.

[118]. Xiong, B., Newton, S., Li, V., Skitmore, M., & Xia, B. (2019). Hybrid approach to reducing estimating overfitting and collinearity. *Engineering, Construction and Architectural Management*, *26*(10), 2170-2185.

[119]. Yin, Y., Fang, Z., Liu, Y., & Han, F. (2019). Temperature-insensitive structure design of micromachined resonant accelerometers. *Sensors*, *19*(7), 1544.

[120]. Zhang, F., De Dear, R., & Candido, C. (2016). Thermal comfort during temperature cycles induced by direct load control strategies of peak electricity demand management. *Building and Environment*, *103*, 9-20.

[121]. Zhang, Q., Sun, H., Wu, X., & Zhong, H. (2019). Edge video analytics for public safety: A review. *Proceedings of the IEEE*, *107*(8), 1675-1696.

[122]. Zheng, S., Liu, X., Zhang, Y., Han, B., Shi, Y., & Xie, J. (2019). Temperature drift compensation for exponential hysteresis characteristics of high-temperature eddy current displacement sensors. *IEEE Sensors Journal*, *19*(23), 11041-11049.

[123]. Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing*, *237*, 350-361.

[124]. Zhou, W., Wang, Y., Lu, A., & Zhang, G. (2016). Systems pharmacology in small molecular drug discovery. *International journal of molecular sciences*, *17*(2), 246.

[125]. Zobayer, E. (2021a). Data Driven Predictive Maintenance In Petroleum And Power Systems Using Random Forest Regression Model For Reliability Engineering Framework. *Review of Applied Science and Technology*, *6*(1), 108-138. https://doi.org/10.63125/5bjx6963

[126]. Zobayer, E. (2021b). Machine Learning Approaches For Optimization Of Lubricant Performance And Reliability In Complex Mechanical And Manufacturing Systems. *American Journal of Scholarly Research and Innovation*, *1*(01), 61–92. https://doi.org/10.63125/5zvkgg52