

Quantum-Enhanced Anomaly Detection for Supply Chain Disruption Risk in Renewable Energy Component Procurement: An Empirical Comparison of Variational Quantum Circuits and Classical Machine Learning Baselines

Syada Nusrat Nahar, Umme Jahara Simki, Mirza Golam Karim Chowdhury,

Md Adnan Chowdhury Ifti, Prithula Sarker

North South University

syada.nahar.253@northsouth.edu, umme.simki@northsouth.edu, mirza.chowdhury@northsouth.edu, adnan.ifti@northsouth.edu, prithula.sarker.252@northsouth.edu

Abstract

Renewable energy supply chains — dependent on geographically concentrated raw materials such as lithium, cobalt, rare earths, and polysilicon — are increasingly exposed to procurement disruption risk from geopolitical shocks, commodity price volatility, and demand shifts tied to climate policy. This study develops and empirically tests a hybrid quantum-classical anomaly detection framework using a Variational Quantum Circuit (VQC) to flag early-warning disruption signals in renewable energy procurement data, and benchmarks it against three classical baselines: Isolation Forest, One-Class SVM, and an LSTM Autoencoder. A monthly panel (2008–2025, $n = 216$) combining the Caldara–Iacoviello Geopolitical Risk Index, World Bank Pink Sheet commodity prices, and IRENA renewable capacity data was engineered into 147 rolling-window features and reduced via PCA to six components for a 6-qubit angle-encoded circuit, implemented independently in PennyLane and Qiskit. Using three documented macro-disruption windows (2008–09 Global Financial Crisis, COVID-19 onset, and the 2022 Russia–Ukraine invasion) as ground truth, the VQC was evaluated against the classical baselines on precision, recall, F1-score, false-positive rate, per-typology recall, noise robustness, and computational cost. Contrary to the study's initial hypotheses, Isolation Forest achieved the highest F1-score (0.593) among all five models, modestly outperforming both VQC implementations ($F1 = 0.556$). The VQC showed no measurable robustness advantage over Isolation Forest at any tested noise level (0–50% of feature standard deviation) and required approximately 1,230 times longer to train than Isolation Forest for a comparable or weaker result. All models, including the VQC, failed to detect the COVID-19-era logistics disruption window. These findings suggest that, on this dataset and circuit design, near-term quantum machine learning does not yet offer a measurable advantage over well-tuned classical anomaly detectors for renewable energy procurement risk monitoring, and the paper discusses the practical and methodological implications of this null result for procurement risk management and future QML research.

Keywords: Quantum Machine Learning; Variational Quantum Circuit; Supply Chain Risk Management; Anomaly Detection; Renewable Energy Procurement; Isolation Forest

1. Introduction and Background

The global transition toward renewable energy has triggered a rapid expansion of supply chains for solar panels, wind turbines, lithium-ion batteries, and associated components. Unlike traditional energy supply chains, renewable energy procurement is characterized by a high dependency on geographically concentrated raw material sources, including lithium, cobalt, rare earth elements, and polysilicon, many of which are subject to geopolitical risk, trade policy volatility, and climate-related logistical disruptions (Hund et al., 2020).

Disruption events in these supply chains carry cascading consequences: project delays, cost overruns, energy-transition setbacks, and financial penalties for developers. The COVID-19 pandemic exposed serious vulnerabilities in global solar and wind component supply chains, and geopolitical shocks such as the 2022 Russia-Ukraine invasion further destabilized energy-commodity markets. As renewable deployment accelerates under frameworks such as the Paris Agreement and UN SDG 7, resilient procurement risk management has become increasingly important (Ivanov, 2021).

Classical anomaly detection methods — Isolation Forest, One-Class SVM, and LSTM-based autoencoders — have been applied to inventory and logistics anomalies in other domains, but their capacity to capture the high-dimensional, non-linear risk signatures of renewable energy procurement has not yet been systematically benchmarked. Quantum Machine Learning (QML), and Variational Quantum Circuits (VQC) in particular, have been proposed as a promising direction for high-dimensional pattern recognition through quantum feature mapping and entanglement-based encoding (Biamonte et al., 2017). However, empirical applications of QML to supply chain disruption risk remain scarce, and, to this author's knowledge, none has yet targeted the renewable energy procurement domain specifically.

To address this gap, the study it builds a hybrid quantum-classical anomaly detection pipeline for renewable energy procurement risk and empirically compares it, under the same controlled experimental conditions, against three widely used classical baselines. Rather than assuming a quantum advantage, the study is designed as a genuine test of that assumption, and reports the result — favourable or not — transparently.

2. Problem Statement

Despite growing awareness of supply chain vulnerability in the renewable energy sector, existing disruption detection systems suffer from three limitations:

Classical ML models (LSTM, Isolation Forest) may lack sufficient sensitivity to early-stage, multidimensional disruption signals in sparse procurement datasets.

Current frameworks are predominantly reactive — they identify disruptions after they manifest — rather than proactive risk detection tools.

Quantum Machine Learning methods, despite theoretical promise for high-dimensional anomaly detection, have not been empirically tested against classical baselines on renewable energy supply chain procurement risk data.

The study therefore asks: Can a hybrid Quantum-Classical anomaly detection framework outperform classical baselines in identifying early-stage procurement disruption risk signals in renewable energy supply chains — and, if not, what does that tell us about the current practical readiness of QML for this application?

3. Research Objectives

To design a hybrid Quantum-Classical anomaly detection pipeline integrating a Variational Quantum Circuit (VQC) with a classical data pre-processing and feature extraction layer.

To collect, pre-process, and curate a multi-source renewable energy procurement dataset combining supply, price, and geopolitical risk indicators.

To empirically evaluate the proposed QML framework against three classical baselines — Isolation Forest, LSTM Autoencoder, and One-Class SVM — across key performance metrics.

To analyse quantum advantage conditions: identifying the data characteristics and disruption typologies under which VQC does (or does not) outperform classical methods.

To derive actionable implications for supply chain risk management practitioners in the renewable energy sector.

4. Research Questions

RQ1: Does a VQC-based anomaly detection model achieve higher disruption detection accuracy than classical ML baselines on renewable energy procurement data?

RQ2: Under what data conditions (noise level, dimensionality) does the quantum model demonstrate measurable performance advantages?

RQ3: What are the practical trade-offs between detection accuracy, false-positive rate, and computational cost for quantum versus classical approaches?

RQ4: Which procurement risk typologies (demand shock, geopolitical shock, logistics disruption, supplier failure) are most effectively captured by the quantum model versus classical models?

5. Hypotheses

Consistent with the theoretical promise of QML reviewed in Section 6.3, the study was designed to test the following directional hypotheses before any experiments were run:

H1: The VQC-based model will achieve an equal or higher F1-score than all three classical baselines on the labelled disruption dataset.

H2: The VQC's relative performance advantage over classical models will increase as Gaussian noise is added to the input features (i.e., the VQC will degrade more slowly than Isolation Forest under noise).

H3: The VQC will show more balanced recall across disruption typologies (demand shock, geopolitical shock, logistics disruption) than the classical baselines, which are expected to specialize in specific anomaly shapes.

Section 9 (Discussion) reports whether each hypothesis was supported, partially supported, or rejected by the empirical results.

6. Literature Review

6.1 Supply Chain Disruption Risk in Renewable Energy

The renewable energy supply chain literature has identified several structural vulnerabilities. The World Bank's mineral-intensity analysis shows that clean energy technologies are substantially more mineral-intensive than their fossil-fuel counterparts, and that graphite, lithium, and cobalt production will need to expand sharply to meet climate targets — concentrating procurement risk in a small number of supplier geographies (Hund et al., 2020). Ivanov (2021) formalizes the related concept of supply chain viability, arguing that resilience research must move from reactive disruption response toward proactive, adaptation-oriented risk management — a framing this study adopts directly by treating anomaly detection as an early-warning rather than after-the-fact diagnostic tool.

More recent work has pushed this proactive framing further by combining AI with geospatial and network-level risk data. Adegboye et al. (2025) propose an AI-driven framework for green energy storage supply chains that fuses satellite imagery, trade data, geopolitical risk indices, and climate hazard models, using graph neural networks and probabilistic risk modeling to flag vulnerable nodes in lithium supply routes across Southeast Asia and sub-Saharan Africa before disruptions materialize. This is a close conceptual parallel to the present study's own attempt to use geopolitical and commodity indicators as leading

indicators of renewable energy procurement risk, though it relies on graph-structured supplier network data rather than the macro time-series panel used here. Similarly, Iyer and Raji's (2026) literature review of AI-powered digital twin systems in global logistics — synthesizing 142 peer-reviewed papers and 20 industry reports from 2019–2026 — finds that digital twins combined with reinforcement and deep learning improve resilience through real-time monitoring and scenario simulation, but also flags limited empirical validation and difficulty integrating geopolitical data as open challenges; both issues are directly relevant to the present study's own limitations (Section 10). A separate strand of work models geopolitical risk propagation at the macro-network level: a study combining graph theory, random forest, and neural networks on WTO and customs data shows how trade sanctions and international conflict propagate business-failure risk across countries and sectors (Zhang et al., 2025), reinforcing the choice of the Caldara–Iacoviello Geopolitical Risk Index (Caldara & Iacoviello, 2022) as a relevant leading indicator for this study's feature set.

6.2 Anomaly Detection and Predictive Analytics in Supply Chain Management

Classical anomaly detection in supply chains has progressed through statistical process control, rule-based systems, and machine learning. Isolation Forest — the tree-based ensemble method used as this study's primary classical baseline — was originally proposed by F. T. Liu et al. (2008) and has since become one of the most widely deployed unsupervised anomaly detectors precisely because it requires no labeled training data and scales well to high-dimensional data, both properties this study's disruption-detection setting also requires. Its applicability to supply chain data specifically is not merely assumed: Moradi et al. (2025) apply Isolation Forest as the unsupervised pre-filtering stage of a two-phase fraud-detection pipeline on the real-world DataCo Smart Supply Chain Dataset, reporting an F1-score of 0.817 and a false-positive rate below 3% — direct empirical evidence that classical, non-quantum anomaly detection already performs strongly on real supply chain data when a well-defined ground truth is available. A separate hybrid framework applied to 100 raw-material suppliers of a Nigerian beverage company similarly pairs Isolation Forest with a Local Outlier Factor model downstream of K-Means clustering and a Random Forest performance classifier, and finds that Isolation Forest and LOF together surface 14 high-risk suppliers, roughly half of whom sat in the "mid-performance" tier and would have been missed by conventional supplier scoring alone (Ehichoya & Oshodin, 2025). Together, these are the classical benchmarks this study's VQC is measured against, rather than an untested or arbitrary comparator: the study asks whether QML can improve on a method already shown, elsewhere in the literature, to be effective for this problem class.

The other two classical baselines used here also have direct precedent in the supply chain literature. Nguyen et al. (2021) combine LSTM-based multivariate forecasting with an LSTM Autoencoder plus One-Class SVM anomaly detector, evaluated on real-world fashion retail sales data, and find that the

LSTM Autoencoder approach outperforms earlier LSTM-only anomaly detection methods — directly supporting this study's choice to include both the LSTM Autoencoder and One-Class SVM as baselines, and offering a point of comparison for why the two methods might be expected to behave differently on procurement-style time series. At a broader level, a review of ML-driven anomaly detection in 5G-enabled industrial automation and supply chains surveys supervised, unsupervised, deep, federated, reinforcement, and graph-based learning approaches for detecting cyberattacks, fraud, counterfeiting, and operational anomalies, and concludes that ML-based anomaly detection is foundational to building resilient, trustworthy supply chains, while also flagging interpretability and heterogeneous-data-integration as persistent open challenges (James, 2022) — challenges this study also encounters in interpreting VQC anomaly scores (Section 10). Anomaly detection has also been embedded directly into broader risk-scoring pipelines: Lin and Lu (2026) apply a two-stage shared-resource Data Envelopment Analysis plus machine learning framework to NVIDIA's semiconductor supply chain, using anomaly detection to flag unusual supplier behaviour alongside predictive analytics for supplier risk forecasting, and a blockchain-integrated framework for interconnected renewable energy microgrids (L. Liu et al., 2026) uses anomaly detection to identify abnormal consumer behaviour and prevent fraud within a stochastic, MILP-optimized energy trading system — the latter a rare example of anomaly detection applied within a renewable-energy-specific operational context, even though its focus (microgrid energy trading) differs from this study's procurement-risk focus. At a broader systems level, a proposed AI/ML framework for supply chain resilience during national emergencies — pandemics, disasters, cyberattacks, geopolitical conflict — integrates IoT, satellite, and social media data with reinforcement learning and graph neural networks for real-time emergency threat detection (Khan et al., 2022), and Tandon et al. (2025) report over 95% accuracy and F1-score for a Bidirectional LSTM model on real-time anomaly detection in IoT-enabled supply chain networks, outperforming Decision Tree, SVM, and XGBoost baselines as well as a standard LSTM model on IoT network-attack data — evidence, from a different application context, that LSTM-family architectures remain competitive against a broader set of classical alternatives, consistent with this study's own finding that the LSTM Autoencoder was the closest classical competitor to Isolation Forest (Section 8.2).

Looking beyond global supply chains, regional research bodies such as Supply Chain Insider have documented a parallel trend toward technology-driven forecasting and risk management. Hasan et al. (2022) apply blockchain-enabled demand forecasting to FMCG logistics in Bangladesh, while Haque et al. (2022) examine blockchain adoption for e-commerce supply chain visibility in the same market — both framing improved data infrastructure as a precondition for better anomaly and demand signal detection. More recently, Mahmud et al. (2024) apply ensemble machine learning models to multivariate demand forecasting in agricultural supply chains, illustrating that classical and ensemble ML methods remain a practical and effective option for supply chain forecasting problems in this kind of research setting. Sikder

et al. (2023) and Noha et al. (2023) extend this technology-adoption literature to circular-economy and Industry 5.0 contexts, reinforcing that the supply chain research community is actively testing emerging computational paradigms — but, notably, none of this body of work, including the additional sources reviewed above, has yet tested quantum computational methods against classical anomaly detectors on renewable energy procurement data specifically.

6.3 Why Benchmark Quantum Against Classical Models at All?

One reasonable question about the study design is: given that classical anomaly detectors are already established and effective (Section 6.2), why introduce a quantum model rather than simply refining the classical approach further? The rationale is methodological and is grounded in the QML literature rather than in an assumption made by this study. Bowles et al. (2024) conducted the largest systematic benchmarking study to date of quantum machine learning against classical baselines — 12 quantum models across 160 datasets built from 6 classification tasks — and found that out-of-the-box classical machine learning models outperform quantum classifiers overall, and that removing entanglement from a quantum model often performs as well as or better than the entangled version, suggesting "quantumness" is often not the decisive ingredient on small-scale problems. Importantly, Bowles et al. (2024) also document a strong publication bias in the QML literature: of 55 relevant papers published on arXiv until December 2023 containing both "quantum machine learning" and "outperform" in the title or abstract, about 40% claim that a quantum model outperforms a classical one, while only 3 papers, or about 4%, find that a quantum model does not outperform a classical one. This asymmetry is itself evidence that direct, transparently reported QML-versus-classical comparisons — including ones that find no quantum advantage, as this study does — are systematically under-represented in the literature and therefore of independent scientific value, not merely a null result to be minimized. This study is designed in that spirit: it treats "does QML help here" as a genuinely open, falsifiable question against classical baselines that are themselves independently validated on supply chain data (Moradi et al., 2025; Ehichoya & Oshodin, 2025), rather than starting with the assumption that a quantum advantage must exist.

6.4 Quantum Machine Learning for Anomaly Detection

Quantum Machine Learning leverages superposition, entanglement, and interference to encode and process data in Hilbert space. Variational Quantum Circuits (VQC) are hybrid classical-quantum models in which parameterized quantum gates are optimized through classical gradient-based methods. Biamonte et al. (2017) formalized the theoretical basis for QML and catalogued early evidence for the ways in which quantum computers may outperform classical computers on machine learning tasks. Subsequent applied work has explored VQC-based classification on small benchmark datasets, generally in domains far removed from supply chain analytics. The intersection of QML and renewable energy supply chain risk

management therefore remains a largely unaddressed empirical question — the gap this study targets directly.

6.5 Research Gap

No published study, to this author's knowledge, has applied VQC-based anomaly detection to renewable energy procurement disruption risk, nor benchmarked it direct against classical baselines that are themselves independently validated on real supply chain data, under matched experimental conditions (same features, same ground truth, same evaluation protocol). This study addresses that gap and, importantly, treats the outcome as an open empirical question rather than a foregone conclusion — a distinction that matters given the results reported in Section 8.

7. Research Methodology

7.1 Research Design

A quantitative, computational-experimental design was used rather than a survey-based approach because the research questions focus on measurable model performance. The research questions (Section 4) ask whether a specific class of algorithm — a Variational Quantum Circuit — can out-detect established classical models on a clearly defined and measurable task: flagging documented macro-disruption months in historical procurement-relevant data. Because this is an algorithm-performance question, controlled experiments on historical data provide a direct way to evaluate it; a practitioner questionnaire alone could not establish whether no amount of surveying procurement managers about their current risk-detection practices would establish whether a VQC actually outperforms Isolation Forest on this data. A hybrid quantum-classical anomaly detection pipeline was therefore implemented and evaluated through controlled computational experiments, comparing five model variants — Isolation Forest, One-Class SVM, LSTM Autoencoder, VQC (PennyLane), and VQC (Qiskit) — on an identical, held-fixed feature set and identical ground-truth labels. A complementary primary-research component (a short-structured questionnaire for renewable energy procurement managers) was scoped for this study and is presented in Section 11.4 as a logical next step; it was not administered within this paper's timeline, and the results reported here are computational only. This was a deliberate scope decision rather than an overlooked research gap.

7.2 Data Sources

Table 1. Raw data sources merged into the monthly analysis panel.

Dataset	Source	Variables used	Coverage
Geopolitical Risk Index	Caldara & Iacoviello (2022)	GPR, GPRT (threats), GPRA (acts)	Monthly, 1960–2025
Commodity prices	World Bank Pink Sheet (World Bank, n.d.)	Crude oil (avg./Brent/Dubai/WTI), coal, aluminium, iron ore, copper, lead, tin, nickel, zinc, platinum, silver	Monthly, 1960–2025
Renewable capacity	IRENA (n.d.)	Installed capacity (MW): solar PV, onshore wind, offshore wind	Annual, forward-filled monthly

The three sources were merged on a common monthly date key into a single panel of 792 monthly observations spanning January 1960 to December 2025 and 22 raw variables. After feature engineering and the removal of rows with insufficient rolling-window history, the final modelling sample covers 216 consecutive months from January 2008 to December 2025 — the period for which all three-source series overlap with sufficient density.

7.3 Feature Engineering and Dimensionality Reduction

For each of the 22 raw variables, three rolling-window statistics (3-, 6-, and 12-month rolling mean and standard deviation) were computed in addition to the raw monthly value, yielding 147 engineered features. Features were standardized (zero mean, unit variance) and reduced via Principal Component Analysis (PCA) to six components — chosen to match a 6-qubit angle encoding for the quantum circuit, consistent with near-term simulator and hardware constraints on circuit width.

7.4 Variational Quantum Circuit Architecture

The VQC encodes each 6-dimensional feature vector via per-qubit RY angle encoding, followed by four variational layers of parameterized RY/RZ rotations and ring-topology CNOT entanglement. The circuit is trained as an autoencoder-style reconstruction model: the expectation value of the Pauli-Z operator on each qubit is compared against the original (angle-scaled) input, and the mean squared reconstruction error is used both as the training loss and, at inference time, as the anomaly score (higher error implies a less "normal" pattern). The circuit was optimized with the Adam optimizer (learning rate 0.05) for 60 epochs on a subsample of approximately 60 training points, for computational tractability on a classical simulator.

To rule out a framework-specific artifact, the identical circuit design was independently re-implemented in Qiskit using a finite-difference gradient estimator and a state vector estimator backend, rather than PennyLane's native auto differentiation. The two implementations were trained and scored separately; their outputs (Spearman rank correlation, reported in Section 8.1) serve as a cross-framework consistency check on the circuit design.

7.5 Classical Baseline Models

Three classical baselines were trained on the same six PCA features:

Isolation Forest (scikit-learn; contamination = 0.05) — ensemble-based anomaly detection via random recursive partitioning.

One-Class SVM (scikit-learn; RBF kernel, $\nu = 0.1$) — kernel-based boundary estimation.

LSTM Autoencoder (PyTorch; hidden size 16, latent size 8, 6-month input sequences, 50 training epochs) — sequence reconstruction-error anomaly detector.

7.6 Ground-Truth Labelling and Its Limits

In the absence of firm-level supplier-failure records, three well-documented macro-level disruption windows were used as ground truth: the 2008–09 Global Financial Crisis (Demand Shock, 10 months), the COVID-19 onset of February–June 2020 (Logistics Disruption, 5 months), and the Russia-Ukraine invasion of February–August 2022 (Geopolitical Shock, 7 months) — 22 of 216 months (10.2%) labeled positive. This is an explicit limitation: the proposal's originally intended fourth typology, Supplier Failure, requires firm-level procurement records that were not available for this study and could not be labeled. This is stated here as a limitation rather than silently dropped, and is revisited in Section 10.

7.7 Evaluation Protocol

For each model, the 15% of months with the highest anomaly score (32 of 216 months) were flagged as predicted anomalies — a common convention in unsupervised anomaly-detection benchmarking that avoids selecting a single, model-specific threshold. Models were then compared on: (a) Spearman rank correlation of anomaly scores across models; (b) Precision, Recall, and F1-score against the labeled ground truth; (c) False Positive Rate; (d) per-typology recall across the three labeled crisis windows; (e) robustness to synthetic Gaussian noise added to the input features at five levels (0%, 10%, 20%, 30%, and 50% of each feature's standard deviation), comparing VQC and Isolation Forest F1-score degradation; and (f) wall-clock training and inference time.

7.8 Implementation Tools

PennyLane (Bergholm et al., 2022) and Qiskit (Qiskit contributors, n.d.) for quantum circuit simulation and optimization; scikit-learn (Pedregosa et al., 2011) and PyTorch for classical baselines and preprocessing; pandas for data engineering; all experiments run on classical simulators (no physical quantum hardware was used, which is itself a limitation discussed in Section 10).

8. Results

8.1 Descriptive Comparison of Anomaly Scores

Figure 1 plots each model's anomaly score, normalized to its own 0–100% range for visual comparability, across the full 2008–2025 sample. All five models show a pronounced joint spike around the 2008–09 financial crisis and around March 2022 (Russia-Ukraine invasion), visually confirming that each model is sensitive to genuine macro-disruption periods. Outside these windows, the One-Class SVM in particular shows persistently elevated and noisy scores throughout 2011–2019, suggesting a higher baseline false-alarm tendency that is quantified formally in Section 8.3.

Figure 1. Normalized (0–100%) anomaly scores for all five models, January 2008 – December 2025. Vertical line marks March 2022 (Russia-Ukraine invasion).

8.2 Precision, Recall, and F1-Score (RQ1)

Table 2. Precision, Recall, and F1-Score for all five models against the three-window ground truth, flagging the top 15% (32/216) highest-scoring months per model. Ranked by F1-score (highest first).

Model	Precision	Recall	F1-Score	Months Flagged
Isolation Forest	0.500	0.727	0.593	32
LSTM Autoencoder	0.471	0.727	0.571	32
VQC (PennyLane)	0.469	0.682	0.556	32
VQC (Qiskit)	0.469	0.682	0.556	32
One-Class SVM	0.375	0.545	0.444	32

Isolation Forest achieved the highest F1-score (0.593), narrowly ahead of the LSTM Autoencoder (0.571) and both VQC implementations (0.556 each). The One-Class SVM trailed all other models (F1 = 0.444). Both VQC variants produced identical Precision, Recall, and F1 values, which is expected given their shared circuit design and near-identical anomaly score rankings; this cross-framework agreement supports the internal validity of the circuit implementation, even though the absolute performance level fell short of the classical Isolation Forest baseline.

8.3 False Positive Rate (RQ3)

Table 3. False Positive Rate (FPR) per model - a proxy for operational "alert fatigue" in a procurement risk monitoring context. Ranked lowest (best) to highest.

Model	FPR	True Positives	False Positives
Isolation Forest	0.082	16	16

Model	FPR	True Positives	False Positives
VQC (PennyLane)	0.088	15	17
VQC (Qiskit)	0.088	15	17
LSTM Autoencoder	0.093	16	18
One-Class SVM	0.103	12	20

Isolation Forest again ranked best (FPR = 0.082), with both VQC implementations close behind (0.088) and the LSTM Autoencoder and One-Class SVM trailing. The gap between the best and worst FPR (0.082 vs. 0.103) is modest in absolute terms, indicating that all five models operate in a broadly similar false-alarm regime once a common 15%-flagging threshold is applied.

8.4 Per-Typology Recall (RQ4)

Table 4. Recall by disruption typology, showing which kind of crisis each model is best suited to catch.

Disruption Type (Event)	Months	Isolation Forest	One-Class SVM	LSTM Autoencoder	VQC (PennyLane)	VQC (Qiskit)
Demand Shock (2008–09 GFC)	10	9/10 (90%)	6/10 (60%)	10/10 (100%)	9/10 (90%)	8/10 (80%)
Logistics Disruption (COVID-19)	5	0/5 (0%)	1/5 (20%)	0/5 (0%)	0/5 (0%)	0/5 (0%)
Geopolitical Shock (Russia-Ukraine)	7	7/7 (100%)	5/7 (71%)	6/7 (86%)	6/7 (86%)	7/7 (100%)

The proposal's originally intended fourth typology, Supplier Failure, is not represented: the study's ground truth is limited to macro-level economic and geopolitical events, and no firm-level supplier data was available to construct that label. This limitation is stated explicitly rather than being omitted.

Three patterns are particularly clear. First, the LSTM Autoencoder and VQC (Qiskit) tied for the best Demand Shock / Geopolitical Shock coverage among their respective categories (100% each), while the One-Class SVM underperformed across both typologies. Second, and most strikingly, the COVID-19-era Logistics Disruption window was essentially undetected by every model tested — four of five models (including both VQC variants) flagged 0 of 5 months, and the best performer (One-Class SVM) caught only 1 of 5 (20%). This is a substantive negative finding: none of the tested models, quantum or classical, is currently well suited to detecting this particular disruption typology using macro-level commodity, geopolitical, and capacity data alone. Third, the VQC (Qiskit) implementation matched or exceeded Isolation Forest on the Geopolitical Shock typology specifically (100% vs. 100%), suggesting any typology-specific quantum advantage, if it exists at all in this data, is narrow and inconsistent rather than general.

8.5 Noise Robustness (RQ2)

To test whether the VQC's relative standing changes under noisier data — the central question behind RQ2 — Isolation Forest and the VQC (PennyLane) were retrained from scratch at five Gaussian noise levels (0%, 10%, 20%, 30%, and 50% of each feature's standard deviation), holding all other settings fixed.

Table 5. F1-score of Isolation Forest vs. VQC (PennyLane) across synthetic noise levels. A positive "VQC advantage" would indicate the VQC outperforming Isolation Forest at that noise level; all observed values are negative.

Noise Level	Isolation Forest F1	VQC F1	VQC Advantage
0%	0.593	0.519	−0.074
10%	0.593	0.556	−0.037
20%	0.593	0.556	−0.037
30%	0.593	0.556	−0.037
50%	0.556	0.519	−0.037

At every noise level tested, Isolation Forest's F1-score met or exceeded the VQC's, and the gap did not narrow as noise increased — if anything, Isolation Forest's F1-score stayed exactly flat (0.593) from 0% to 30% noise, only degrading at the highest tested noise level (50%, F1 = 0.556), at which point it still matched the VQC's noise-free F1-score. No noise level produced a positive VQC advantage.

8.6 Computational Cost (RQ3)

Table 6. Wall-clock training and inference time on the 216-month, 6-feature dataset (classical simulator; PennyLane default. Qubit device). Qiskit VQC timing was not separately benchmarked here, as its finite-difference gradient estimator is already known from Section 7.4 to be substantially slower than PennyLane's native auto differentiation for the same circuit.

Model	Training Time	Inference Time (all 216 months)
One-Class SVM	0.00s	0.001s
Isolation Forest	0.07s	0.004s
LSTM Autoencoder	0.19s	0.001s
VQC (PennyLane)	86.11s	1.519s

The difference in computational cost is substantial: the VQC (PennyLane) required 86.11 seconds to train, versus 0.07 seconds for Isolation Forest — approximately 1,230 times longer — for a lower F1-score. Inference time shows a similar gap (1.519s vs. 0.004s for 216 months of scoring). Even accounting for the small dataset size used here, this cost asymmetry is an important practical consideration for any near-term deployment decision, independent of the accuracy comparison in Section 8.2.

9. Discussion

9.1 RQ1: Does the VQC outperform classical baselines?

No. Isolation Forest achieved the highest F1-score (0.593) of all five models tested, with both VQC implementations trailing at 0.556 — a 3.7-percentage-point gap in F1 and a similarly modest gap in Precision and Recall (Table 2). H1 is therefore rejected: on this dataset, feature set, and circuit design, the VQC does not match, let alone exceed, the best classical baseline. Notably, the VQC did outperform the One-Class SVM, so the result is not that quantum methods perform worse overall — only that they do not lead.

This result is not unique to the present implementation; it is consistent with the broader pattern documented by Bowles et al. (2024) in their large-scale, 160-dataset benchmarking study, which found that out-of-the-box classical models outperform quantum classifiers overall (Section 6.3). Reading alongside that literature, this study's result is better understood as one more data point in a systematically under-reported category of finding — QML tested fairly against classical baselines and found not to lead — rather than as evidence that the VQC implementation here was flawed. The classical baselines it lost to are themselves independently validated on real supply chain data by Moradi et al. (2025) and by the Isolation Forest/LOF supplier-risk study reviewed in Section 6.2 (Ehichoya & Oshodin, 2025), which strengthens rather than weakens the credibility of this comparison: the VQC was tested against genuinely strong, externally validated classical benchmarks, not a weak strawman.

9.2 RQ2: Under what conditions does the VQC show an advantage?

No such conditions emerged within the tested range. The noise-robustness sweep (Table 5) shows Isolation Forest's F1-score meeting or exceeding the VQC's at every one of the five noise levels tested, with no narrowing trend as noise increased. H2, which predicted the VQC would degrade more slowly than Isolation Forest under noise, is rejected within the tested range (0–50% of feature standard deviation). This does not rule out a quantum advantage at noise levels, dimensionalities, or data volumes outside this study's scope — only that none emerged within it.

9.3 RQ3: What is the practical accuracy/cost trade-offs?

For this dataset, the trade-off clearly favours the classical approach: it is both less accurate (lower F1) and dramatically more expensive to train (roughly 1,230× the wall-clock time of Isolation Forest) than the best classical baseline, with a marginally higher false-positive rate as well (Table 3). For a procurement risk team operating on a monthly monitoring cadence, this combination offers no practical justification for adopting the VQC over Isolation Forest at this stage.

9.4 RQ4: Which typologies does each model capture best?

Recall varies meaningfully by typology and by model (Table 4), but not in a way that favours the VQC specifically: the LSTM Autoencoder and VQC (Qiskit) each achieved standout recall on one typology (Demand Shock and Geopolitical Shock, respectively), while all models — quantum and classical alike — essentially failed on the Logistics Disruption (COVID-19) typology. H3, which predicted more balanced VQC recall across typologies, is not supported: the VQC's typology profile is uneven in the same way the classical models' profiles are, and it shares the same blind spot-on logistics disruptions. This is arguably the study's most practically important negative finding: a supply chain risk team relying on any of these five models, in their current form, would have missed the COVID-19-era logistics shock in this dataset.

10. Limitations

Macro-level ground truth only. The three labeled crisis windows (22 of 216 months) are macro-economic/geopolitical events, not firm-level supplier failures; the Supplier Failure typology from the original proposal could not be tested for lack of firm-level data.

Small positive-class sample. With only 22 positive months, precision/recall estimates carry wide uncertainty; no confidence intervals or statistical significance tests were computed, and results should be read as descriptive rather than inferential.

Simulator-only quantum experiments. All VQC results were obtained on noiseless classical simulators (PennyLane default. Qubit; Qiskit StatevectorEstimator); no physical quantum hardware was used, so results do not reflect real-device noise, decoherence, or NISQ-era error rates.

Reduced VQC training subset. For computational tractability, the VQC was trained on a ~60-point subsample (PennyLane) or ~20-point subsample (Qiskit finite-difference) rather than the full 216-month series; this may understate the VQC's achievable performance relative to the classical baselines, which were trained on the full sample.

Single circuit design and hyperparameter setting. Only one VQC architecture (4 layers, RY/RZ + ring CNOT, angle encoding) and one hyperparameter configuration were tested; a broader architecture/hyperparameter search might change the comparative result.

Threshold convention. The top-15%-flagged convention used for Precision/Recall/F1 is a common but arbitrary choice; results may shift under a different flagging threshold or a threshold tuned per model.

11. Practical Implications for Procurement Risk Management

These results provide practical guidance for a renewable energy procurement or supply chain risk team, beyond the model-selection recommendation itself.

11.1 A minimum-viable monthly monitoring routine

A procurement risk desk could operationalize this study's pipeline directly: refresh the Geopolitical Risk Index, World Bank commodity prices, and IRENA capacity figures monthly, recompute the six PCA features, and re-score with Isolation Forest (sub-second retraining makes monthly refresh trivial). A month scoring in the top 15% historically should trigger a manual review — not an automatic supplier action — focused on the specific commodities driving the score (e.g., a spike in polysilicon-adjacent metals or a jump in the Geopolitical Risk Index for a key sourcing country).

11.2 Where this pipeline should not be trusted alone

Section 8.4's finding that every model missed the COVID-19-era logistics window has a direct procurement implication: macro commodity and geopolitical indicators are not a substitute for shipment-level visibility. Teams whose primary exposure is logistics/freight risk (port congestion, container availability, carrier delays) should treat this model family as a complement to, not a replacement for, freight-rate indices, port-throughput data, or direct supplier/carrier check-ins.

11.3 A cost-aware technology roadmap

Given the $\sim 1,230\times$ training-time and modest accuracy gap documented in Section 8.6, organizations should not adopt VQC-based monitoring operationally today. A reasonable roadmap is: (1) deploy Isolation Forest now for monthly monitoring; (2) revisit VQC-based methods only if training costs fall substantially (e.g., via better gradient-estimation methods or access to real quantum hardware) or if a future study demonstrates a genuine accuracy advantage on a typology this study could not test, such as firm-level supplier failure.

11.4 Proposed primary-research extension (not administered in this study)

To connect these computational findings to how procurement teams actually behave, this study scoped — but did not administer, given time constraints — a short structured questionnaire for renewable energy procurement and supply chain professionals in Bangladesh, covering: (a) which of the disruption typologies in Table 4 they currently monitor and how; (b) whether they would act on a monthly anomaly flag of the kind this pipeline produces, and what would make them trust or distrust it; and (c) what data sources (beyond commodity prices and geopolitical indices) they consider indispensable for disruption

early-warning. A structured instrument covering these items was drafted for this purpose but is not included here, as it was not administered within this study's timeline; a sample of 15–25 responses from procurement/supply chain managers at renewable energy developers or EPC contractors would be sufficient for an indicative (not statistically representative) qualitative validation of the computational findings above.

12. Conclusion

This study set out to test — not assume — whether a Variational Quantum Circuit could outperform classical anomaly detectors on renewable energy supply chain disruption risk. Using a matched experimental design across five models, three ground-truth crisis windows, five noise levels, and a direct compute-cost comparison, the evidence points consistently in one direction: on this dataset and circuit design, the VQC does not outperform the best classical baseline (Isolation Forest) on detection accuracy, does not show a robustness advantage under added noise, and is roughly three orders of magnitude more expensive to train. All models share a common and policy-relevant blind spot on COVID-19-era logistics disruption. These are honest null and negative results relative to the study's initial hypotheses, and the paper reports them as such rather than reframing the outcome after the fact. This does not close the door on QML for supply chain risk management — it narrows the open question to specific conditions (real quantum hardware, larger training samples, alternative circuit designs, or different data typologies, including the node-level and network-level risk framings explored in Adegboye et al. (2025) and Iyer and Raji (2026)) that this study did not have the scope to test, and which future work should target directly. Future work should also administer the practitioner questionnaire proposed in Section 11.4, to test whether procurement professionals' lived experience of these four disruption typologies aligns with, or diverges from, the purely computational picture presented here.

References

- Adegboye, O., Arowosegbe, O. B., & Prosper, O. (2025). AI-optimized supply chain mapping for green energy storage systems: Predictive risk modeling under geopolitical and climate shocks. *International Journal of Research Publication and Reviews*, 6(5), 63–86. <https://doi.org/10.55248/gengpi.6.0525.1801>
- Bergholm, V., Izaac, J., Schuld, M., Gogolin, C., Ahmed, S., Ajith, V., Alam, M. S., Alonso-Linaje, G., AkashNarayanan, B., Asadi, A., Arrazola, J. M., Azad, U., Banning, S., Blank, C., Bromley, T. R., Cordier, B. A., Ceroni, J., Delgado, A., Di Matteo, O., ... Killoran, N. (2022).

PennyLane: Automatic differentiation of hybrid quantum-classical computations. Xanadu Quantum Technologies. <https://pennylane.ai>

Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., & Lloyd, S. (2017). Quantum machine learning. *Nature*, 549(7671), 195–202. <https://doi.org/10.1038/nature23474>

Bowles, J., Ahmed, S., & Schuld, M. (2024). Better than classical? The subtle art of benchmarking quantum machine learning models. *arXiv*. <https://doi.org/10.48550/arXiv.2403.07059>

Caldara, D., & Iacoviello, M. (2022). Measuring geopolitical risk. *American Economic Review*, 112(4), 1194–1225. <https://doi.org/10.1257/aer.20191823>

Ehichoya, M., & Oshodin, O. D. (2025). A hybrid machine learning approach to supplier segmentation and anomaly detection for proactive circular supply chain risk management. *NIPES – Advances in Engineering Design Technology*, 7(4), 45–61. <https://doi.org/10.5281/zenodo.17857462>

Hasan, N., Ferdouse, N., Khan, T., Adiba, A. F., & Shams, K. A. I. (2022). Implementing blockchain in the supply chain of FMCG companies for better demand forecasting and logistics operation. *Supply Chain Insider*, 8(1).

Haque, A. Z., Shuvo, A. A., Hossain, M. M., & Ebrahim, J. A. (2022). A scenario of adopting blockchain technology in supply chain: A study of e-commerce in Bangladesh. *Supply Chain Insider*, 8(1).

Hund, K., La Porta, D., Fabregas, T. P., Laing, T., & Drexhage, J. (2020). Minerals for climate action: The mineral intensity of the clean energy transition. World Bank.

International Renewable Energy Agency. (n.d.). Renewable capacity statistics (Solar photovoltaic, Onshore and Offshore wind energy). IRENA. Retrieved August 28, 2026, from <https://www.irena.org/Data>

Ivanov, D. (2021). Supply chain viability and the COVID-19 pandemic: A conceptual and formal generalisation of four major adaptation strategies. *International Journal of Production Research*, 59(12), 3535–3552. <https://doi.org/10.1080/00207543.2021.1890852>

Iyer, S. S., & Raji, B. (2026). Geopolitical risk simulation using AI-powered digital twin systems in global logistics: A critical literature review. *International Journal of Scientific Research and Management (IJSRM)*, 14(5), 10621–10634. <https://doi.org/10.18535/ijrm/v14i05.em05>

James, U. U. (2022). Machine learning-driven anomaly detection for supply chain integrity in 5G industrial automation systems. *International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET)*, 9(2), 542–566. <https://doi.org/10.32628/IJSRSET22549>

Khan, R. S., Sirazy, M. R. M., Das, R., & Rahman, S. (2022). An AI and ML-enabled framework for proactive risk mitigation and resilience optimization in global supply chains during national emergencies. *Sage Science Review of Applied Machine Learning*.

Lin, S.-W., & Lu, W.-M. (2026). A strategic framework for supplier performance assessment and anomaly detection in high-tech supply chains: Integrating shared-resource DEA with machine learning. *IEEE Transactions on Engineering Management*, 73, 2470–2484. <https://doi.org/10.1109/TEM.2026.3675293>

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. In *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM)* (pp. 413–422). IEEE. <https://doi.org/10.1109/ICDM.2008.17>

Liu, L., Cheng, Y., Baziar, A., & Basheer, S. (2026). Novel financial, economic, and management framework for sustainable energy supply chain transformation: Consumer behavior anomaly detection and risk analytics. *IEEE Transactions on Consumer Electronics*. Advance online publication. <https://doi.org/10.1109/TCE.2026.3685891>

Mahmud, F. B., Rouf, M. J. B., & Mahin, M. M. I. (2024). Optimization of rice supply chain through multivariate forecasting using ensemble learning models. *Supply Chain Insider*, 14(1).

Moradi, F., Tarif, M., & Homaei, M. (2025). Semi-supervised supply chain fraud detection with unsupervised pre-filtering. *arXiv*. <https://doi.org/10.48550/arXiv.2508.06574>

Nguyen, H. D., Tran, K. P., Thomassey, S., & Hamad, M. (2021). Forecasting and anomaly detection approaches using LSTM and LSTM autoencoder techniques with the applications in supply chain management. *International Journal of Information Management*, 57, Article 102282. <https://doi.org/10.1016/j.ijinfomgt.2020.102282>

Noha, B. J., Ali, A., Shah, P. K., Tamang, P., & Khan, N. J. (2023). Circular economy practices in supply chains of Bangladesh. *Supply Chain Insider*, 11(1).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Qiskit contributors. (n.d.). Qiskit: An open-source framework for quantum computing. IBM. Retrieved August 28, 2026, from <https://www.ibm.com/quantum/qiskit>

Sikder, M. R., Saif, S. J. A., Nadim, S. A., Tarik, K. I. B., & Islam, R. (2023). Integrating circular economy and Industry 5.0 for sustainable supply chain management in the food industry. *Supply Chain Insider*, 9(1).

Tandon, R., Thomas, J., Vedi, K. V., & Gupta, S. (2025). Prediction and anomaly detection methods based on artificial intelligence techniques in supply chain management. In *Proceedings of the 2025 International Conference on Information, Implementation, and Innovation in Technology (I2ITCON)*. Pune, India.

World Bank. (n.d.). Commodity markets - "Pink Sheet" monthly price data. World Bank. Retrieved August 28, 2026, from <https://www.worldbank.org/en/research/commodity-markets>

Zhang, Y., Sánchez Arnau, E., & Sánchez Pérez, E. A. (2025). Impact of geopolitical and international trade dynamics on corporate vulnerability and insolvency risk: A graph-based approach. *Information*, 16(7), Article 525. <https://doi.org/10.3390/i>