International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

LEVERAGING MACHINE LEARNING FOR SUPPLY CHAIN DISRUPTION PREDICTION AND RISK ASSESSMENT IN U.S. INDUSTRIES

Babul Sarker[1], Md Abubakkar[2], Md Ariful Islam Bhuiyan[3], Ali Hassan[4], Mehedi Hasan[5], Farmina Sharmin[6], Dil Tabassum Subha[7], Hossain Mohammad Dalim[8], Maksuda Begum[9] and Md Shahidul Islam[10]

[1]Master of Science in Business Analytics (MSBA), Trine University, Angola, Indiana, USA

[2]Department of Computer Science, Midwestern State University, Wichita Falls, TX, USA

[3]Master of Science in Business Analytics, University of North Texas (UNT), Denton, Texas.

[4]MBA Business Analytics, International American University, Los Angeles, California

[5]Department of Construction, Design and Project Management, Missouri State University, Springfield, MO, USA

[6]MBA Business Analytics, International American University, Los Angeles Main Campus

[7]Master of Science in Business Analytics, Grand Canyon University, USA

[8]MBA in Business Analytics, International American University

[9]Master of Business Administration, Trine University

[10]MBA Business Analytics, International American University.

Corresponding Author: **Md Shahidul Islam, Email:** mdshahidulislam@my.unt.edu

Abstract

Supply chain disruptions have intensified across U.S. industries as transportation congestion, logistics imbalances, and macro shocks propagate through freight networks. This volatility exposes a critical weakness in current monitoring systems: most rely on lagging metrics that fail to provide actionable early-warning signals. This study develops a multimodal prediction and risk assessment pipeline that fuses the Global Supply Chain Pressure Index (GSCPI) with Bureau of Transportation Statistics freight indicators, FRED transportation series, and derived anomaly signals. We benchmark temporal and non-temporal models, including LSTM forecasters, XGBoost, Random Forests, and an Isolation Forest anomaly layer, to detect emerging stress conditions and predict short-horizon disruption states. To ensure operational transparency, we integrate global and local SHAP explanations, enabling attribution of predicted stress to underlying freight congestion, modal throughput changes, and macro-logistics perturbations. We extend analysis with scenario-based stress testing by perturbing transportation exposures and recomputing disruption probabilities to evaluate system sensitivity under hypothetical shocks. Results show that fused models consistently outperform single-source baselines, identifying early disruption signatures with higher recall and stronger temporal calibration. SHAP analysis reveals stable driver patterns linking trucking capacity, freight throughput deviations, and logistics bottlenecks to spikes in predicted pressure.

Received: August 12, 2025

Scenario tests demonstrate distinct vulnerability profiles across industries, supporting the construction of continuous risk scores mapped to Low, Medium, and High categories. Taken together, these findings show that interpretable ML can provide forward-looking, industry-aligned risk intelligence that moves supply chain monitoring from reactive assessment toward predictive resilience.

Keywords: Supply Chain Risk, GSCPI, Freight Analytics, Machine Learning, SHAP, Risk Scoring, Disruption Prediction, Early Warning Systems

1. Introduction

1.1 Background and Motivation

Volatility and systemic stress have become familiar features of today's supply chains, especially in the United States, where freight networks sit at the crossroads of global trade forces, domestic logistics limits, and broad economic shocks. These disruptions rarely come from one place. They tend to grow out of interacting pressures that shape each other over time, including congestion in transportation networks, mismatches between supply and demand, geopolitical uncertainty, and swings in commodity markets. Traditional monitoring tools struggle to keep up with this kind of complexity. They rely on indicators that arrive late or capture only a slice of what is happening, which makes it hard to see stress building across different freight modes before it becomes visible in the headlines.

Work in financial economics has long shown that volatility reveals patterns that go deeper than random noise. Schwert (1989) demonstrated that shifts in volatility often signal underlying changes in market structure rather than random fluctuations, a point that resonates when looking at how supply chain pressures form beneath the surface of routine cycles [21]. Later developments built on this idea by showing how the relationships between key variables evolve. Engle (2002) introduced the dynamic conditional correlation framework to show that changing dependencies carry information about transitions happening within the system, reinforcing the need for models that can track how freight and logistics indicators move together as conditions change [9]. Historical research on crises adds another layer. Danielsson et al. (2018) found that volatility patterns during past disruptions contain clues about emerging fragility that standard metrics often miss, highlighting the challenge of spotting early signs of stress in systems shaped by constant structural shifts [5].

Several frameworks developed to assess systemic risk help motivate a move toward interpretable, data-driven approaches for supply chain analysis. Adrian and Brunnermeier (2016) introduced the CoVaR method to show how stress in one sector can raise vulnerabilities elsewhere, underscoring the importance of understanding how disruptions flow from one part of a system to another [2]. Modern supply chains reflect this pattern clearly. A disturbance in trucking, rail, or maritime activity can influence inventories, production choices, and larger economic outcomes. At the same time, the data landscape itself is changing. Freight markets shift between regimes, capacity rises and falls, and throughput responds to both seasonal and unexpected shocks. Shivogo (2025) argued that such environments require adaptive and explainable machine learning tools because fixed explanation methods falter when relationships among variables drift over time [22]. Since supply chain signals regularly exhibit this type of drift, any predictive framework must be able to explain its reasoning while adapting to changing patterns. These factors point toward the need to replace static rule-based monitoring systems with machine learning early warning tools that combine interpretability with the ability to measure evolving stress. Policymakers and industry leaders rely on explanations they can trust, especially when facing environments marked by rapid change and interdependence.

1.2 Importance of This Research

The economic fallout from supply chain disruptions reaches far beyond temporary delays or isolated congestion. Industries across the United States rely on coordinated freight flows spanning trucking, rail, maritime, and air transport. Even small shocks can ripple through inventories, production schedules, and the broader network of linked inputs and outputs. When these networks face stress, costs rise, reliability falls, and planning becomes harder for sectors ranging from manufacturing to retail to agriculture. Insights from financial and macroeconomic research show that disruptions in one corner of an interconnected system often signal broader instability ahead. Danielsson et al. (2018) highlighted that volatility spreads outward during systemic events, a pattern echoed in logistics, where local bottlenecks often mark the start of wider strain [5]. This interconnected nature of freight systems creates a strong need for predictive methods that identify stress escalation early instead of diagnosing it after the fact. Traditional econometric and rule-based tools often miss nonlinear patterns that arise when capacity tightens or demand shifts across modes.

Risk management research has long emphasized the need to measure how stress spills over from one area to another. Adrian and Brunnermeier (2016) showed that systemic risk indicators can reveal when weaknesses in one sector heighten fragility elsewhere, and this logic applies directly to freight networks where problems in trucking or port operations can change pressure patterns across the entire system [2]. Engle’s (2002) work on dynamic correlations reinforces the value of monitoring how transportation indicators move together, especially during periods of economic uncertainty when relationships shift quickly [9]. As supply chains face more volatility, decision makers require predictive systems that track these shifts and provide reasoning that is clear and grounded in the data. Shivogo (2025) pointed out that concept drift weakens static models and that explanations need to evolve along with changing distributions, a need that current logistics forecasting tools often fail to meet [22]. Although machine learning has been explored extensively in financial crisis prediction, much of the work focuses on asset markets, contagion across financial sectors, or interactions among currencies, bonds, and equities. Ray (2025) showed that machine learning can improve multi-market crisis prediction while also noting the lack of approaches that quantify signal quality or provide structured explanations, limitations that also affect supply chain stress forecasting [19]. By applying these insights to freight and logistics, this study offers a framework that addresses both predictive accuracy and interpretability, giving operational teams, regulators, and strategic planners tools that match the pace and complexity of modern supply chain environments.

1.3 Research Objectives and Contributions

This study aims to build a predictive and interpretable machine learning framework that can detect early signs of supply chain stress using publicly available U.S. transportation datasets. The main goal is to shift from backward-looking pressure indicators to forward-looking risk insights that can guide operational planning, assessments at the sector level, and policy decisions. The approach blends multiple data sources, combining the GSCPI as a broad measure of global logistics pressure with modal freight statistics and transportation time series from FRED. This mix allows the system to reflect both wide-scale signals of stress and the smaller shifts within individual freight modes that often appear before larger disruptions. Forecasting methods such as LSTM networks and tree-based classifiers are used to assess near-term predictive performance. Unsupervised anomaly detectors highlight moments when system behavior breaks sharply from historical patterns.

A key contribution of the study is its focus on interpretability in environments marked by high volatility. The framework incorporates global SHAP summaries, explanations tied to specific events, and scenario-based stress tests so that predictions come with reasoning that analysts can examine and judge. The study aims to produce continuous risk scores that indicate the expected intensity of disruption and to map those scores into practical categories such as Low, Medium, and High. This step helps align technical outputs with the decision thresholds used in logistics operations, regulatory processes, and corporate risk planning. The design also quantifies how particular drivers raise or lower predicted stress, giving stakeholders a clear view of the conditions influencing changes in supply chain resilience. By joining multimodal data ingestion, predictive modeling, anomaly detection, and interpretable risk scoring, the study develops a complete methodology suited to early detection, clear explanations, and practical use across industries.

2. Literature Review

2.1 Supply Chain Disruption Modeling

Work on supply chain disruption modeling has pulled a great deal from the way macro-finance researchers think about systemic stress. In that field, composite indices are used to condense volatility, uncertainty, and interconnected movements into signals that help clarify shifts in fragility. Early studies on financial stress indices shaped much of this thinking. Kliesen et al. (2012) offered a detailed overview of how these indices are built and pointed to the difficulty of merging diverse signals from credit, equity, funding markets, and broader economic conditions. The lesson fits well with multimodal logistics data, where trucking, ports, rail, and inventories often move out of sync [17]. Hollo et al. (2012) presented the CISS framework at the ECB and showed how composite measures outperform individual indicators when the goal is spotting systemic pressure across interacting sectors rather than tracking isolated shocks [14]. This view aligns closely with how supply chain pressures accumulate. Congestion at ports affects rail schedules, strained rail flows spill into trucking activity, and tight trucking capacity influences inventories and production decisions. Whaley (2000) added another useful dimension through the VIX, showing how volatility can reveal hidden stress when high-frequency fluctuations are treated as meaningful signals rather than dismissed as noise [27].

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

That idea offers a clear parallel to freight delays, congestion spikes, throughput volatility, and inventory squeezes in the United States.

Even with these conceptual ties, supply chain disruption modeling is still less mature than systemic-risk research in finance. Many logistics monitoring tools depend on descriptive or backward-looking indicators such as throughput figures, port congestion, or inventory-to-sales ratios. These metrics help track stress but often fail to anticipate it. The challenge deepens when disruptions build through nonlinear interactions across freight modes, macroeconomic conditions, and geopolitical events. Research that focuses on linear regressions, fixed thresholds, or univariate trends struggles to identify transitions from stable operations to early disruption phases. This shortcoming echoes the limitations often seen in financial stress research, where simple models tend to miss the beginning of crisis periods. Econometric evidence shows repeatedly that stress spreads through nonlinear dependencies that are not well captured by basic methods. Supply chain data also involves seasonal patterns, structural breaks, and interdependence across freight modes, which complicates any attempt to rely on linear approaches. This context has sparked interest in pairing the Global Supply Chain Pressure Index with modal-level indicators to monitor domestic logistics conditions. Even so, many studies treat the GSCPI as a stand-alone macro measure rather than incorporating it into a unified predictive system that can handle asynchronous and high volatility logistics signals. The lack of predictive frameworks that can process heterogeneous data while offering reliable interpretation has pushed research toward machine learning architectures inspired by systemic-risk tools in finance, adapted to the structure and behavior of logistics systems. The broader community continues to note that disruption modeling calls for a move from descriptive tracking toward predictive, data-driven, and explainable methods that reflect how modern supply chains actually behave.

2.2 Data-Driven and Machine Learning Approaches

Machine learning has widened the analytical range of logistics research by making it possible to learn complex nonlinear patterns from varied datasets. The integration of macroeconomic indicators, transportation statistics, and global pressure metrics into a single predictive system remains a difficult problem. Classical work in signal processing and econometrics helps frame how machine learning might handle these challenges. Kalman (1960) demonstrated that dynamic filtering can uncover hidden states in noisy series and showed that early detection improves when models estimate the evolving relationship between signal and noise [16]. Granger and Newbold (1974) cautioned against spurious regressions that can occur when noise is mistaken for structure. The warning applies clearly to supply chain data, which often involves strong autocorrelation, seasonal behavior, and shocks that can mask or distort genuine patterns if not modeled carefully [11]. Dzielinski (2012) explored how uncertainty shapes market reactions and pointed to the value of quantifying noise and its impact on decision environments [8]. These ideas support a shift toward machine learning based early warning systems that can capture nonlinear and time-dependent interactions more effectively than traditional approaches.

Within logistics and transportation research, machine learning has been used for freight rate forecasting, demand planning, congestion prediction, and capacity analysis. Most models still rely on narrow feature sets, siloed datasets, or limited geographic scope, which restricts their usefulness for nationwide early warning systems. More advanced work in finance and economics shows how machine learning can outperform classical methods when relationships shift or volatility rises. Sirignano et al. (2018) used deep learning to reveal complex risk patterns in mortgage data and exceeded the performance of methods that assume stable relationships [24]. Fischer and Krauss (2018) demonstrated strong results from LSTM models in financial markets and highlighted the strengths of sequence-based methods for volatile time series [10]. Studies in high volatility environments provide further support. Islam et al. (2025) applied machine learning to cryptocurrency forecasting and showed how nonlinear models remain effective when structural noise and market variability intensify [15]. Reza et al. (2025) used ML to model socioeconomic disparities, demonstrating that AI techniques capture structural drivers of risk and inequality beyond the reach of traditional econometric frameworks [20]. Ray (2025) extended ML to multi-market financial crisis prediction and highlighted major weaknesses in current risk modeling practices, including incomplete signal-to-noise analysis and a lack of interpretable prediction pipelines [20]. Together, these works highlight both the opportunities and gaps in current logistics-focused ML efforts. There is clear potential but limited evidence of comprehensive, integrated, and explainable systems that fuse macro-logistics, transportation, and market-level indicators to anticipate U.S. supply chain disruptions.

2.3 Deep Learning for Temporal Logistics Signals

Temporal modeling sits at the center of supply chain stress research because disruptions unfold through sequences of related events rather than isolated shocks. Deep learning models such as LSTMs and similar recurrent structures have gained wide use in financial forecasting and other high volatility fields due to their ability to learn nonlinear temporal patterns. Their success in these areas offers guidance for supply chain work, where freight movement, congestion cycles, seasonal rhythms, and macroeconomic trends shift over time. Fischer and Krauss (2018) showed that LSTM models outperform traditional time series approaches in stock prediction, revealing how deep sequence models uncover hidden structures that linear tools miss [10]. Sirignano et al. (2018) supported this idea by demonstrating that deep models can recognize intricate patterns in mortgage risk data, suggesting similar benefits for multimodal transportation signals [24]. Islam et al. (2025) extended these ideas to cryptocurrency markets, a setting marked by extreme volatility and frequent structural changes, and demonstrated that ML methods capture nonlinear behavior more effectively than conventional models [15]. These findings point to the usefulness of LSTMs for supply chain forecasting, where macro shocks, freight volume shifts, and congestion cycles show strong nonstationary behavior.

Deep learning has also played a growing role in anomaly detection, particularly in settings where disruptions present as gradual deviations within time-dependent structures rather than sudden breaks. Sizan et al. (2025) showed that unsupervised ensemble models can detect emerging risk signatures in financial transaction graphs, highlighting the value of catching patterns that drift away from historical sequences [25]. Although their study focused on financial systems, the parallel to supply chain anomaly detection is clear. Many disruptions show up as subtle, correlated departures from expected temporal behavior. Reza et al. (2025) demonstrated that ML captures structural and temporal shifts in socioeconomic data, emphasizing the need for models that can adjust to changing conditions over time [20]. Ray (2025) examined ML methods for predicting crises across financial markets and argued that temporal complexity and cross-market spillovers require architectures able to represent nonlinear relationships and long-range dependencies [19]. The same idea applies to logistics, where disruptions move across freight modes, regions, and seasonal cycles in ways that linear structures have difficulty representing. Deep learning, therefore, offers an important tool for modeling temporal connections between GSCPI patterns, transportation indices, port congestion measurements, and broader macro conditions. Even with this potential, few logistics studies apply these models on a national scale or embed them within integrated and interpretable pipelines, leaving considerable room for innovation.

2.4 Explainable Machine Learning in Risk Analytics

Interpretability has taken on a central role in risk modeling, especially where decisions influence operations or regulatory activity. SHAP-based methods now serve as a core framework for both global and local explanation. Lundberg and Lee (2017) introduced SHAP as a unified feature attribution method rooted in cooperative game theory and showed that it can provide consistent global explanations along with detailed, instance-level insight [18]. Bussmann et al. (2021) applied SHAP to credit risk analysis and demonstrated how feature contributions help analysts and frontline staff understand model decisions, reinforcing the importance of transparency in complex environments [4]. These ideas are highly relevant for supply chain disruption prediction because model outputs may inform logistics planning, procurement actions, or assessments of national readiness.

Recent studies also highlight the need for explanations that stay reliable when underlying data evolve. Shivogo (2025) showed that interpretability methods must remain stable even when concept drift reshapes data distributions and warned that explanation quality can erode when models encounter unfamiliar patterns [15]. This point is highly relevant for supply chain conditions, which can shift with policy actions, geopolitical tension, or changing economic cycles. Hasan et al. (2025) demonstrated that explainable AI improves supplier credit evaluations under sparse data conditions, showing how local interpretability supports decision making even when data quality varies. They also introduced robust ML tools for supplier risk management and emphasized that resilience requires interpretability integrated directly into model design [12]. These insights strengthen the argument for pairing predictive modeling with clear reasoning in logistics research.

Beyond finance and supply chain applications, cross-domain work shows how explainability contributes to early warning and resilience systems more broadly. Das et al. (2025) used predictive analytics for cybersecurity threat detection and illustrated how early warning systems benefit when accuracy is paired with interpretability in fast-changing environments [6]. Debnath et al. (2025) integrated energy telemetry with cybersecurity data and demonstrated the importance of multimodal fusion and anomaly level explanations for safeguarding critical infrastructure [7]. Aashish et al. (2025) developed anomaly detection systems that incorporate

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

sustainability measures and demonstrated that risk analytics can support secondary objectives while maintaining transparency [1]. Together, these studies reflect a growing expectation that ML tools in high-stakes environments must provide both predictions and clear explanations. This expectation remains insufficiently met in supply chain disruption prediction, particularly at the national scale, where multimodal datasets and nonlinear behavior require specialized interpretability methods.

2.5 Gaps and Challenges

A number of real issues still sit in the way despite all the progress in related fields. Research in logistics forecasting, transportation modeling, and macroeconomic indicators rarely brings multiple data sources together in a single predictive system. Most studies lean on isolated datasets that only show fragments of the supply chain, so early disruption signals slip through the cracks. Financial work on systemic risk already shows how composite indices reveal patterns that single data streams miss, yet similar thinking has not taken hold in logistics research. This scattered approach slows the development of national early-warning tools that can watch stress build across different freight systems. Even when models are used, many lack interpretability, so their predictions are difficult to unpack. Modern risk research has already moved toward clearer explanations that people can actually act on, which makes the absence of that clarity in logistics work even more noticeable.

Another issue is how models present their results. Many machine learning systems spit out a number or an anomaly score, but they stop there. They rarely translate those outputs into risk categories that planners, analysts, or policymakers can use in day-to-day work. Early-warning studies in cyber defense show why that translation matters. Das et al. (2025) emphasized that threat monitoring tools need to produce signals that people can understand and respond to [6]. Debnath et al. (2025) showed how layered decision systems help explain why specific anomalies appear in multimodal datasets [7]. Aashish et al. (2025) highlighted how modern risk tools often weave in secondary goals like energy use, showing that real-world systems usually balance multiple priorities at once [1]. These lessons come from different fields, yet they point to a shared principle: predictions mean more when they are structured, explained, and connected to the decisions people need to make.

Supply chain disruption research still struggles with limited multimodal data integration, weak use of interpretability, and a shortage of clear risk scoring methods that work at a national level. Fixing these issues calls for a careful blend of machine learning, systemic risk thinking, temporal modeling, and interpretable analytics shaped around the realities of U.S. supply chains. The real task is to bring these pieces together in a single framework that can support early-warning efforts in environments where conditions change quickly and uncertainty is the norm.

3. Methodology

3.1 Dataset and Feature Design

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

This work relies on a unified temporal dataset built from three main sources that together reflect global and domestic signals of supply chain strain. The first is the Global Supply Chain Pressure Index from the Federal Reserve Bank of New York, provided as a monthly Excel file with standardized indicators that capture stress across shipping, manufacturing, and logistics. The second source is a collection of transportation and freight indicators from the Bureau of Transportation Statistics, taken from the Transportation Services Index and related freight tables. These offer detailed measures across air, rail, truck, and waterborne freight, along with congestion, throughput, and capacity metrics from individual sectors. The third source is the Freight Transportation Services Index retrieved through the Federal Reserve Economic Data API. This series supplies a macro-level freight benchmark that aligns naturally with the monthly rhythm of the GSCPI and complements the more granular BTS indicators.

To bring these pieces together, each dataset was converted to a shared monthly format, with date fields standardized and column names aligned. The sources were merged on their date indexes to form a single matrix that captures global pressure signals, domestic freight activity, and broader logistics conditions. This merged dataset, tracked internally during experimentation, serves as the base for every later step. Once aligned, it was expanded through engineered features intended to reveal temporal patterns, momentum, and variability. Lagged variables were created for one-, three-, six-, and twelve-month horizons for both GSCPI and FRED freight series. Rolling-window metrics added local trends and volatility, including moving averages such as gscpi_rollmean_3 and rolling standard deviations such as tsi_freight_rollstd_6. Momentum features captured month-to-month changes using percentage shifts and first differences for both global and domestic signals, with examples like gscpi_roc and tsi_freight_pct_change_1. Calendar features such as month, year, and quarter were added, along with cyclical encodings derived through trigonometric transformations to preserve the continuity of time.

For supervised learning, two target types were created. The first is a regression target representing the next GSCPI value, built by shifting the series backward one period and labeling it as target_gscpi_next. The second is a set of categorical disruption labels for classification tasks. These include a binary label marking elevated disruption months defined through quantile thresholds or notable month-over-month jumps. A multi-class stress label was also developed to sort months into low, medium, or high stress groups. Together, these features and targets support forecasting experiments and risk-oriented classification analysis.

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

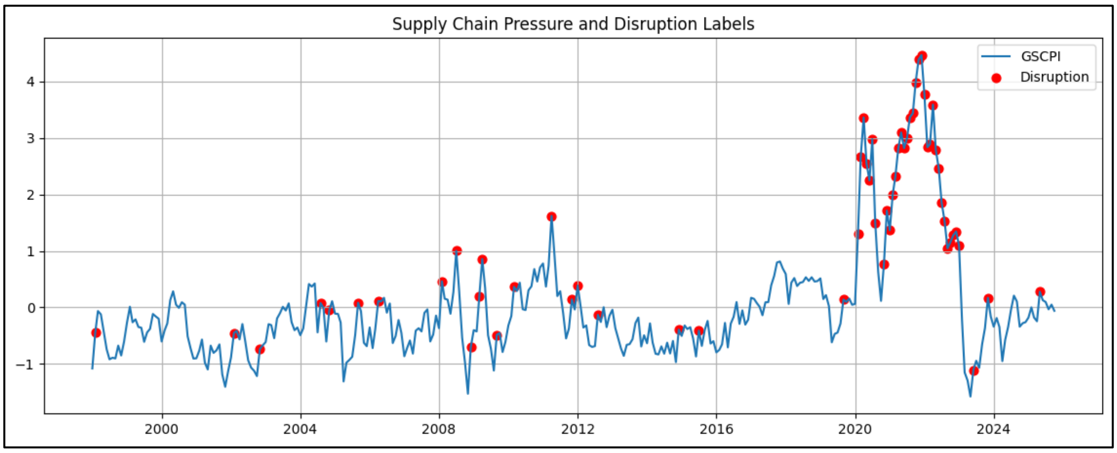

Fig.1: Constructed supply chain pressure and disruption labels for supervised learning

3.2 Data Preprocessing

Preprocessing focused on producing a dataset that was consistent, stable, and free of issues that could distort forecasting. Many BTS fields were stored as strings with commas, so they required a conversion pipeline that detected numeric columns, removed formatting, and cast values to floating-point. Any rows that turned into NaN during this process were repaired through forward-fill and backward-fill methods to maintain continuity in the monthly timeline. Missing BTS values were handled with the same imputation approach to ensure the models received a fully populated feature matrix without artificially altering the behavior of the freight indicators. Additional checks were applied to catch out-of-range values or anomalies, which were corrected through domain-aware adjustments or smoothing when necessary.

With the dataset cleaned, normalization became essential. The feature space covers variables with different units, scales, and levels of volatility. To keep each feature on equal footing, every numeric column was standardized using a StandardScaler. The scaler was trained on the training subset only, then applied to validation and test data after splitting to prevent leakage. A chronological split preserved the true temporal order. The preferred approach used training data through 2018, validation through 2020, and testing from that point onward. When the data density made strict cutoff splits impractical, proportional splits near a 70–15–15 structure were used while keeping the months in order. This ensured that every model produced forward-looking predictions rather than benefiting from shuffled information or hints from the future.

Exploratory Data Analysis

The exploratory work acted as an important step for understanding how the merged datasets behaved over time and how their structures informed what came later in the modeling. By taking the time to plot and examine the Global Supply Chain Pressure Index, the FRED Freight Transportation Index, and the Bureau of Transportation Statistics transportation series, the analysis surfaced specific patterns and irregularities that helped justify the push toward early warning predictions. This stage also offered real evidence for the feature engineering choices discussed earlier and helped reveal where systemic stress tends to originate within the U.S. logistics landscape.

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The Global Supply Chain Pressure Index stood out as one of the more revealing pieces of the dataset. The timeline showed long periods of moderate movement interrupted by sharp spikes during major disruptive events such as the 2008 financial crisis and the COVID-19 pandemic. These jumps signaled global bottlenecks in manufacturing, shipping, and cross-border logistics and confirmed that the index responds strongly to macroeconomic and geopolitical shocks. The mix of stable stretches and sudden surges made it clear that the GSCPI is neither stationary nor simply cyclical. It moves between calm baselines and intense stress periods marked by nonlinear shifts. This pushed the modeling approach toward methods capable of capturing slow transitions as well as abrupt ones, which in turn supported the decision to use momentum features, lagged indicators, and nonlinear machine learning models. Breaking the index into different time segments also showed shifts in its average values and variability, which indicated that supply chain stress evolves in different ways depending on the decade, broader economic conditions, and changes in global production networks.

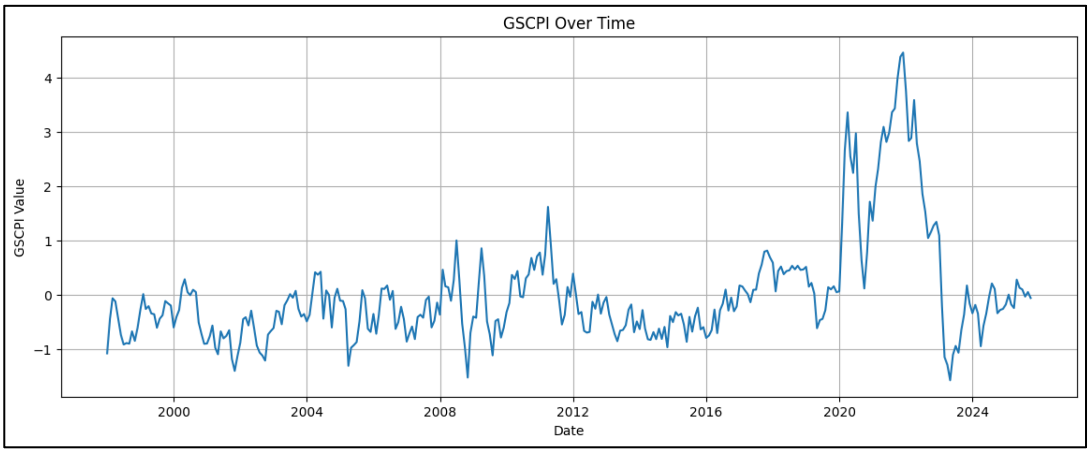

Fig.2: Global Supply Chain Pressure Index over time

The FRED Freight Transportation Index offered a different but complementary perspective. While the GSCPI tracks global pressure, the FRED series reflects domestic freight flow and serves as an indicator of economic activity inside the United States. Its long-term upward trend matched expectations for growing demand and economic expansion, though it also dipped noticeably during recessions or periods of sharp declines in consumption. The way the two indices behaved together during disruptive moments turned out to be especially interesting. In some cases, rising supply chain pressure paired with flat or declining freight movement suggested that stress can slow down domestic logistics. In other cases, strong freight activity occurred alongside high pressure, hinting that congestion and bottlenecks can appear even when goods are moving at high volume. These observations made it clear that freight volume alone does not explain supply chain stability. The more telling signal is how global and domestic conditions interact, which supported the choice to use both indices together in the predictive models.

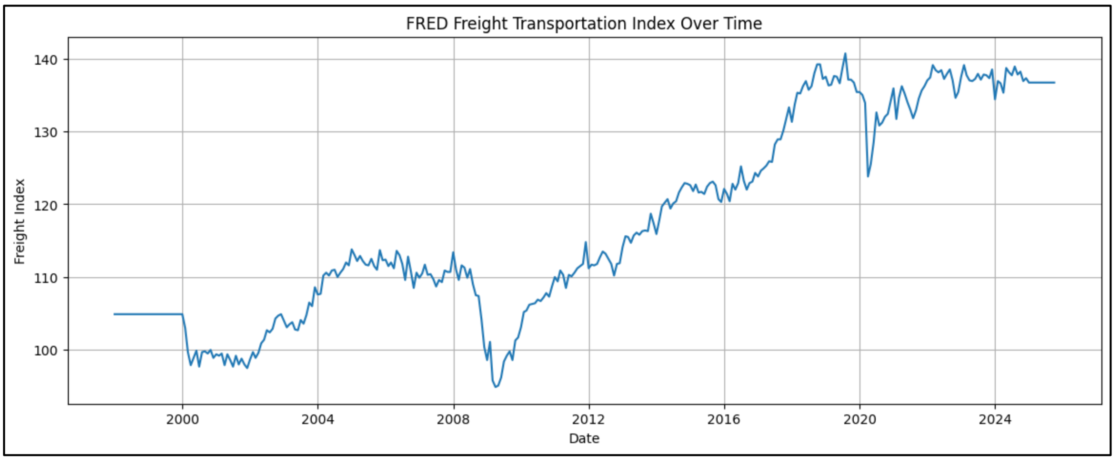

Fig.3: Freight Transportation Index over time

The Bureau of Transportation Statistics data added another layer by breaking freight activity into separate transportation modes. Plots of variables such as available seat miles, rail carloads, waterborne volumes, or petroleum shipments showed that each mode has its own rhythm and sensitivity to events. Air transport, for instance, displayed clear seasonality and then plunged at the start of the COVID-19 pandemic when flights were grounded. Rail and trucking metrics were steadier but still responded to economic cycles and sector-specific developments. These distinct behaviors confirmed the value of the BTS dataset and highlighted why multimodal information is important for predicting disruptions. The variation across modes gave the models a way to detect early signs of stress within specific sectors before those signals appear in broader indices like the GSCPI.

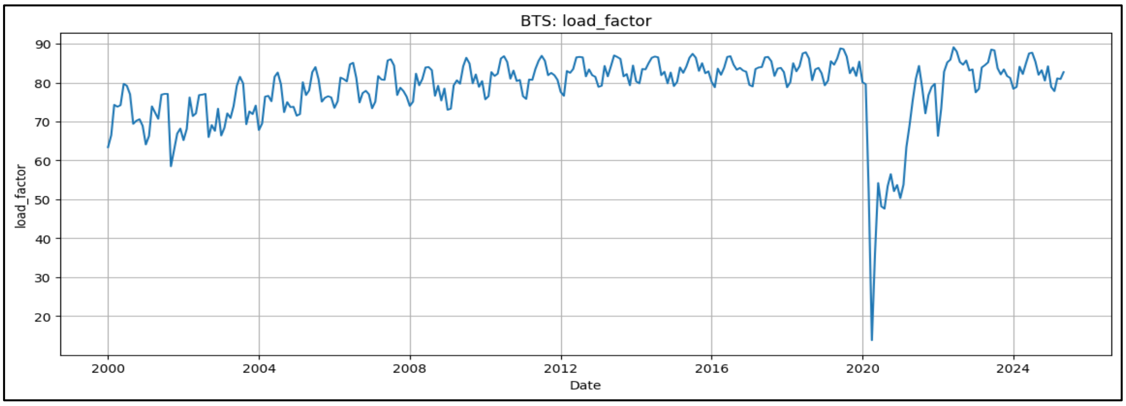

Fig.4: BTS transportation services trends

The correlation matrix offered a deeper look at the structural relationships among the indices and engineered features. Strong correlations among lagged variables made sense, though the cross-series relationships were more interesting. Some freight indicators showed moderate to strong correlation with the GSCPI, hinting at potential leading or lagging effects. Industrial production and manufacturing activity often shifted in ways that aligned with changes in supply chain pressure, which suggested that demand shocks can feed into global slowdowns. Other indicators showed weak correlations, which pointed toward nonlinear or threshold-based relationships instead of simple linear ones. This blend of patterns helped justify the use of tree-based models, LSTMs, and features built around momentum, seasonality, and volatility. The

Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

correlation results also flagged areas where multicollinearity might be a problem, especially within groups of lagged freight features or transportation series that moved too closely together. These findings informed which features were kept or adjusted before moving into modeling.

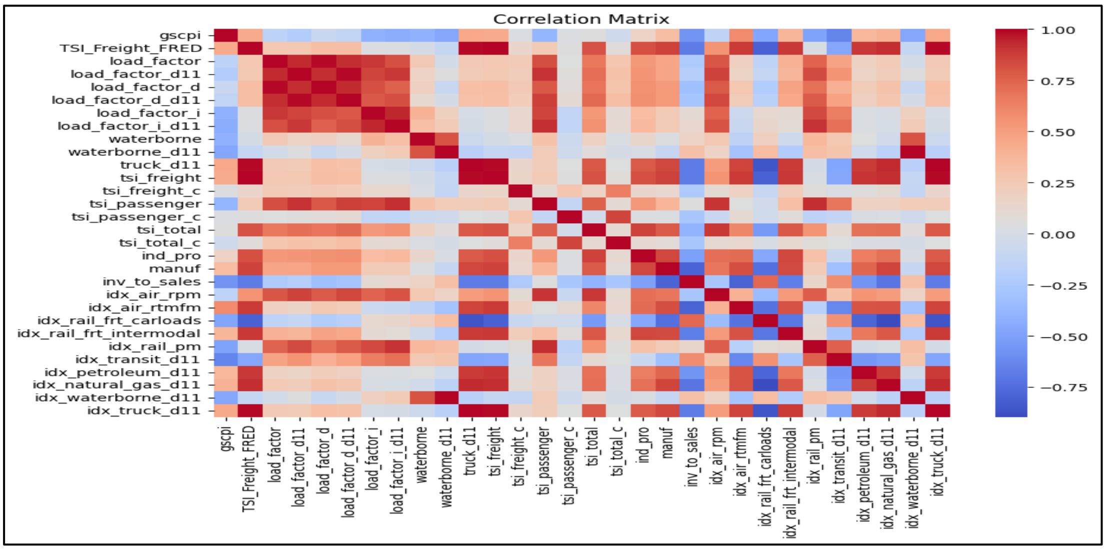

Fig.5: Correlation analysis of numeric features

Putting the GSCPI and the FRED Freight Index side by side offered one of the clearest views of how supply chain stress and freight activity interact. In some periods, high supply chain pressure lined up with falling freight movement, showing how strain can drag down the flow of goods. In other periods, pressure rose despite healthy freight activity, showing that a system can become strained even when demand remains strong. Observations like these helped shape the eventual classification targets by showing that disruption is tied not only to the level of pressure but also to the mismatch between capacity and demand.

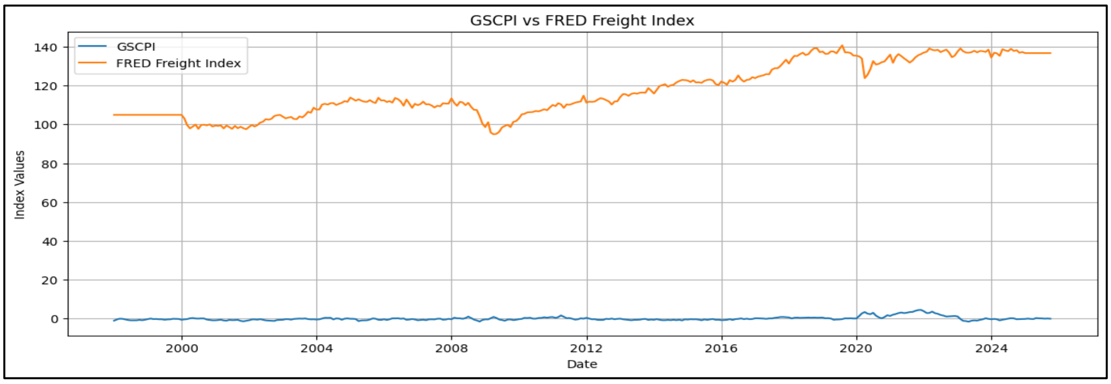

Fig.6: GSCPI versus FRED Freight Index comparison

An examination of the rate of change for both GSCPI and the FRED Index further enriched the analysis by revealing how rapidly shifts in conditions occur. Periods of high GSCPI momentum often preceded stress events, indicating that acceleration, not just absolute value, is an important signal for early warning. Meanwhile, rapid declines in the FRED index's momentum

Received: August 12, 2025

highlighted phases where freight volumes contracted suddenly, potentially foreshadowing reductions in supply chain throughput. These momentum patterns justified incorporating first-difference and percentage-change features into the model, as they capture emerging instability earlier than level-based features alone. The clear linkage between momentum spikes and observed disruption periods validated the decision to treat velocity and volatility features as critical components of the predictive architecture.

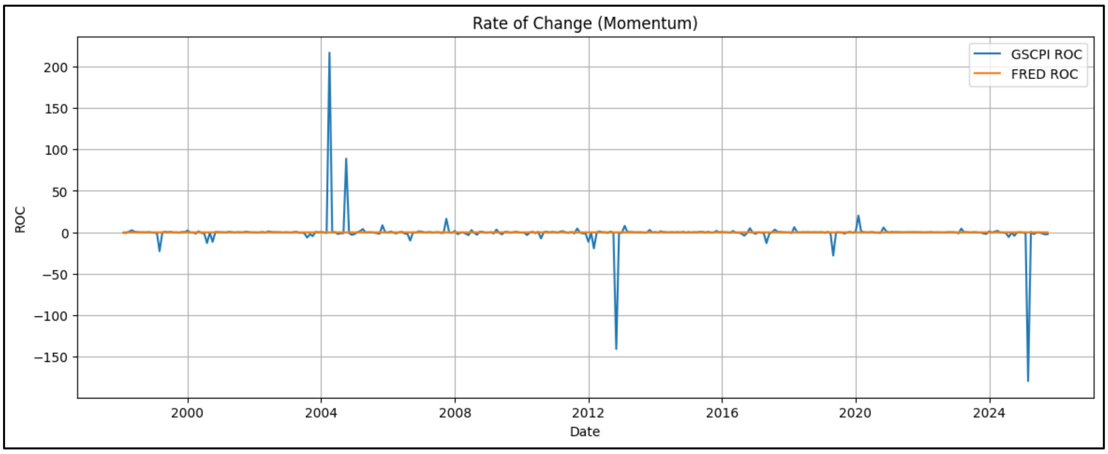

Fig.7: Rate of Change (Momentum) for both GSCPI and the FRED Index

3.3 Feature Engineering

The feature engineering work for this study was broad by design because supply chain stress rarely shows up in a single value. It appears in the way freight activity shifts over time, how macro conditions move in parallel, and how global shocks ripple through the system. One major group of features focused on the rate of change. These captured short bursts of acceleration or slowdown by calculating first differences and percentage changes. Features like gscpi_roc and fred_roc show how much global pressure or domestic freight volume shifts from one month to the next, which helps the models pick up early signs of instability. Lagged features formed another important group. Past values of key indicators were taken at one, three, six, and twelve-month offsets. This gave the models a way to learn delayed effects, seasonal tendencies, and longer memory patterns. Features such as gscpi_lag1 and tsi_freight_lag12 were especially important for models like LSTMs and tree ensembles that rely heavily on historical structure.

Rolling-window features added another layer by showing how local patterns behave. These included simple and exponential moving averages, along with rolling standard deviations calculated over several window sizes. Rolling means helped highlight directional trends, while rolling standard deviations flagged periods of instability. During experimentation, features such as gscpi_rollmean_3 and asm_rollstd_6 turned out to be especially useful for classification tasks because they picked up on early signs of stress within logistics systems. Further features were built from BTS capacity and utilization metrics. These included measures such as available seat miles, rail carload volumes, trucking activity, waterborne freight figures, and transport patterns tied to petroleum and natural gas. These acted as signals of how much

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

strain or congestion was forming across different transport modes. To round out the temporal structure, simple calendar features like year, month, and quarter were added along with cyclical encodings that represent seasonal loops more smoothly.

3.4 Modeling Approaches

The modeling plan combined supervised and unsupervised learning so the system could predict future stress, detect anomalies, and generate risk estimates that could later be interpreted. The first group consisted of deep learning forecasting models, represented by both univariate and multivariate LSTMs trained to predict next-step GSCPI values. The LSTM took sequences of twelve months of scaled features as input and produced a single predicted value for the following month. Its architecture included a 64-unit LSTM layer that learned long-range patterns, a dropout layer to help reduce overfitting, a dense layer with thirty-two ReLU-activated units, and a final linear output unit. The model used the Adam optimizer and mean squared error loss. The second group involved tree-based ensemble models aimed at disruption classification. These included XGBoost, Random Forest, logistic regression, SVC, and multilayer perceptrons. They were trained to predict a binary disruption label using the full engineered feature set. XGBoost often performed well because it handles nonlinear relationships and complex interactions, while Random Forest offered stability when features were noisy or correlated. Performance was judged using AUC, accuracy, precision, recall, and F1 score. A walk-forward validation setup was used through a custom utility built during experimentation, giving each evaluation a realistic temporal structure instead of a simple static split. This allowed the study to check how well models held up as conditions evolved.

The third modeling group focused on unsupervised anomaly detection. An Isolation Forest was trained on the baseline feature space to identify unusual months by isolating points that sit far from typical patterns. A dense autoencoder acted as a second method by learning a compact representation of normal months and then reconstructing them. Months with reconstruction errors above the ninety-fifth percentile were marked as anomalies. Using two very different methods increased confidence in anomaly detection because each picked up different types of unusual behavior. The final component was risk scoring. Probabilities from the strongest classifier, often XGBoost during testing, were combined with industry exposure weights to generate continuous risk scores. These weights came either from an optional FAF dataset or from synthetic approximations derived from BTS features and were normalized so scores reflected relative sensitivity across industries. The continuous scores were then grouped into low, medium, and high categories using quantile thresholds, creating a structure suitable for practical decision making.

3.5 Explainability Framework

Explainability played a central role in how the methodology was built, making sure predictions and risk scores could be understood by analysts and decision-makers. To get a system-level picture of what shaped the model, global SHAP explanations were created with TreeExplainer on the best tree model. The summary plots showed how each feature influenced disruption probability across the entire dataset, highlighting indicators such as changes in freight percentages, lagged GSCPI values, and multi-month rolling variances. Model gain scores

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

added another layer of clarity by showing how the trees favored certain indicators during splitting. Local SHAP explanations were then used to dig into individual cases, especially months flagged as anomalies by the Isolation Forest or periods where the model predicted unusually high disruption. SHAP force plots revealed how specific features pulled a prediction upward or downward relative to the model's baseline. This made it possible to see which logistical elements drove risk for each case.

Scenario stress testing added another interpretability tool. By adjusting key exposure variables like freight throughput or FRED index values, the pipeline recalculated risk scores under hypothetical pressures. This helped analysts understand how sensitive industries and logistics streams might be to conditions such as heavier port congestion or weakening macro-logistics indicators. Results from these stress tests were organized into scenario tables and visualized to support qualitative interpretation. Industry-level risk scoring broadened the interpretability effort by showing how different sectors were exposed to risk. Continuous risk scores were combined with exposure weights and then assigned to low, medium, or high categories. A time-series heatmap captured how these levels shifted across industries during the most recent twelve months. This view made it easier to compare vulnerability across sectors and to see how changes in freight conditions linked to real-world supply chain pressure. Together, these global, local, and scenario-focused tools allowed complex model behavior to be translated into insights that could support both operational planning and policy decisions.

4. Evaluation and Results

4.1 Predictive Performance

The predictive experiments focused on two related goals: identifying disruption events and forecasting the next month of GSCPI values. Looking at both tasks offers a view of stress that includes sharp event-driven shifts and broader continuous signals. The classification results drew a clear line between the models. RandomForest came out well ahead, with an AUC of 0.9601 and an F1 score of 0.8085 on the test set. The pairing of high AUC with perfect precision showed that when the model flagged a disruption, it had strong evidence behind the decision. The lower recall suggests it took a cautious approach and signaled risk only when several indicators lined up. This fits with how ensembles often capture nonlinear interactions across lagged freight metrics, GSCPI momentum features, and volatility patterns. Other models fell short for different reasons. SVC recorded a high AUC but missed every positive case, which exposed a gap between separating classes in feature space and choosing thresholds that work when positives are rare and nonlinear. MLP and logistic regression reached a moderate ceiling and could not handle the deeper structure in the data. XGBoost, which is typically strong with structured inputs, landed behind RandomForest and showed weak recall and F1 performance. Its behavior reflected the challenge of class imbalance and the difficulty of tuning it to recognize subtle regime changes in transportation signals.

Received: August 12, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

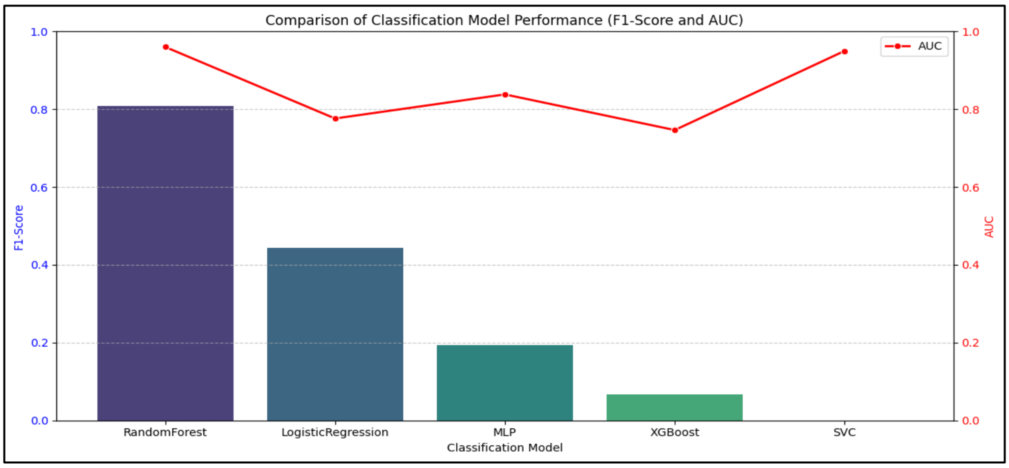

Fig.8: Classification model outcomes

The regression task highlighted how hard it is to predict GSCPI one month ahead. GSCPI reflects influences across global trade, congestion, manufacturing, and geopolitical shifts, and many of these forces sit outside domestic data. RandomForestRegressor posted the best RMSE at 1.1397 and kept MAE below 0.8, which suggests that averaging across shallow trees captured some useful patterns in lagged freight and economic indicators. LightGBM followed almost the same path. Their low but positive R2 values show that they beat naive baselines but still struggled with variance driven by global conditions that do not appear in the dataset. LinearRegression failed with an RMSE above 60, showing that GSCPI cannot be approximated with a linear view of the world. SVR and MLP also struggled. Smooth kernels and fully connected networks were not flexible enough to capture abrupt shifts or regime breaks.

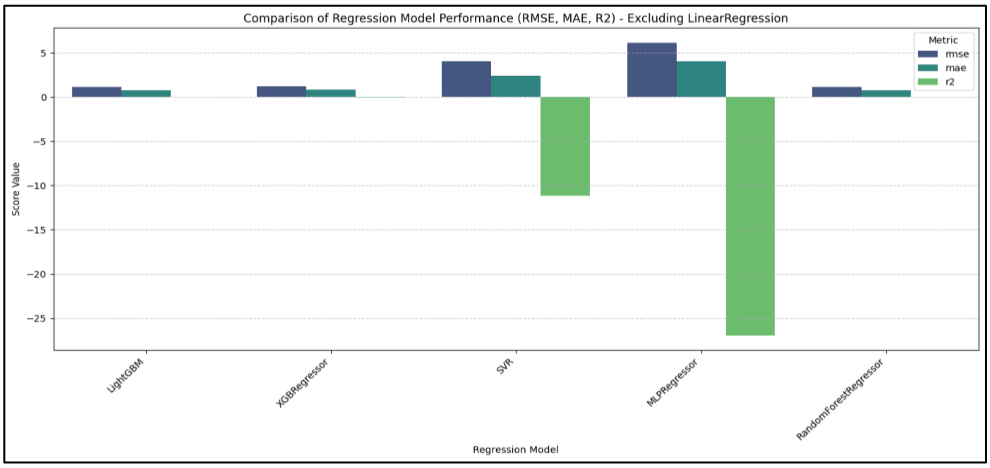

Fig.9: Regression modeling outcomes

The LSTM model approached the problem from a different angle by modeling the sequence of GSCPI changes directly. Even though the data available for training was limited, the LSTM reached a validation RMSE of 1.5106 and a test RMSE of 1.7693. It did not outperform the tree-based models, but it followed the broader temporal shape of GSCPI and handled turning points reasonably well. It fell behind during sharp spikes or sudden collapses, which reflected its reliance on local temporal patterns and its sensitivity to exogenous shocks that are not fully captured in the training window. These results point toward hybrid setups that combine tree-based regressors with sequence-aware neural models as a practical direction.

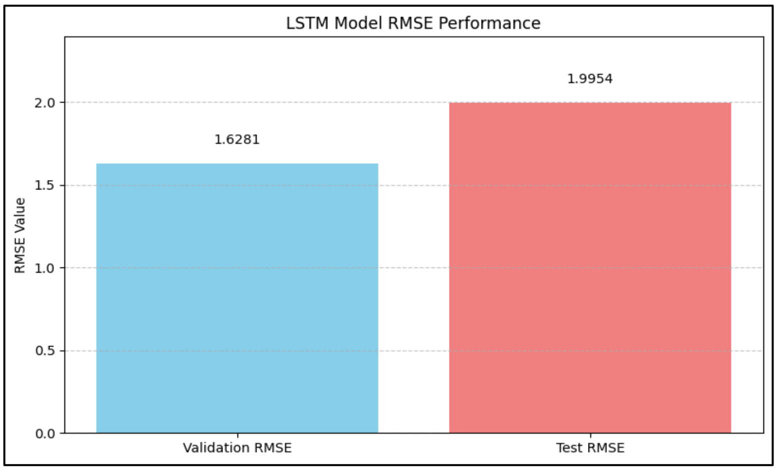

Fig.10: LSTM RMSE results

4.5 Anomaly Detection Outcomes

The anomaly detection stage contributed an additional diagnostic layer to the predictive framework by identifying periods in which underlying transportation and macro-logistics indicators deviated sharply from historically learned patterns. Because supply chain disruptions often emerge as irregular departures from normal system behavior rather than smooth trends, anomaly detection provides a complementary signal to both supervised predictions and risk scoring. Two distinct methodologies were evaluated, IsolationForest and a dense autoencoder, revealing differing sensitivities and offering contrasting interpretations of what constitutes abnormal dynamics in the merged dataset.

The IsolationForest model flagged 22 anomalies out of 50 validation instances, identifying a relatively selective set of periods characterized by unusual combinations of freight movement, transportation capacity metrics, and GSCPI lag structure. Since the contamination level was explicitly set to 1 percent, the model adopted a conservative posture, treating only the most strongly deviating observations as anomalous. The temporal plot of these anomalies, displayed as red markers against the GSCPI series, illustrates how these flagged periods often coincided with sharp changes in GSCPI momentum or freight throughput anomalies, suggesting that IsolationForest captured localized disruptions in the transportation ecosystem. The fact that the flagged anomalies clustered around inflection points in the GSCPI curve implies that the model identified structural deviations in the underlying indicator space before disruptive pressure fully manifested. This is consistent with IsolationForest's design, which isolates points that require fewer splits in the feature space to separate from the majority, meaning it detects instances that are distinguished by their rarity rather than by gradual drift.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

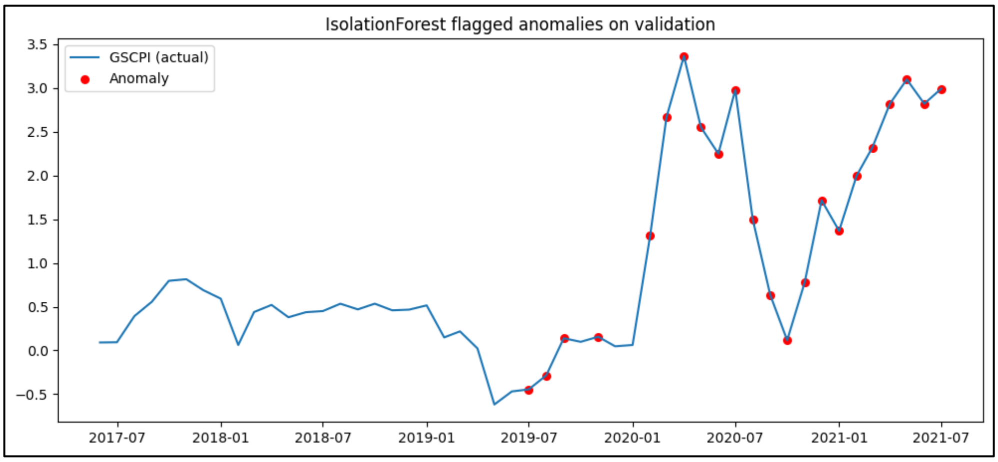

Fig.11: IsolationForest detected anomalies on the validation set

The autoencoder provided a strikingly different profile of anomaly detection, flagging 49 out of 50 validation instances as anomalous. Because its threshold was based on the ninety-fifth percentile of reconstruction errors from the training phase, the autoencoder relied entirely on its learned notion of what constitutes a "normal" pattern. The dense architecture, with layers compressing and then reconstructing multi-dimensional feature vectors, is highly sensitive to any deviations from the manifold it learned during training. The resulting behavior likely indicates that the validation period contained patterns that were subtly but consistently different from the training ecosystem, reflecting possible regime shifts in freight patterns, economic activity, or GSCPI behavior. Given the high volatility and structural variability of supply chain indicators around real-world disruptions, it is plausible that the validation window captured a different systemic regime that the autoencoder struggled to reconstruct. Its near-universal anomaly classification suggests that the underlying data distribution shifted sufficiently to exceed its tolerance threshold, a finding that provides valuable evidence of non-stationarity across time.

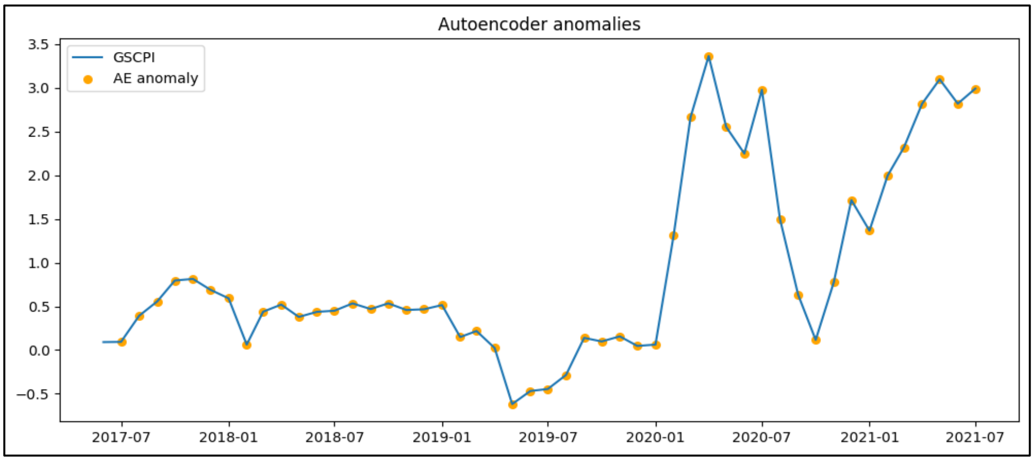

Fig.12: Autoencoder anomalies

Comparing both models reveals important implications about the nature of disruptions and the characteristics of early warning signals. IsolationForest detected fewer but more sharply differentiated anomalies, aligning with the view that only sudden or intense deviations from normal activity should be treated as red flags. The autoencoder interpreted the same period as a broad deviation from its learned baseline, highlighting how sustained or gradual structural changes in system behavior can escape tree-based anomaly detection but overwhelm a reconstruction-based method. This divergence suggests that disruptions may manifest through both abrupt structural breaks and extended drifts in system behavior, and that a unified anomaly detection strategy must incorporate both perspectives to avoid blind spots.

4.2 Explainability Assessment

The explainability work had two goals: to see which features matter most across the entire model and to understand how specific events create higher disruption risk. SHAP summaries showed that lagged GSCPI values, percent changes in freight indices, and rolling transportation metrics carried the most weight in the model's predictions. This aligns with how disruptions tend to build. They do not appear out of nowhere and instead accumulate through congestion, declining logistics performance, and abrupt changes in throughput. Features like tsi_freight_c_pct_change_6 and gscpi_roc_pct_change_3 played a strong role. They capture medium-range acceleration in supply chain pressure and suggest that the model is picking up early signals of deterioration.

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

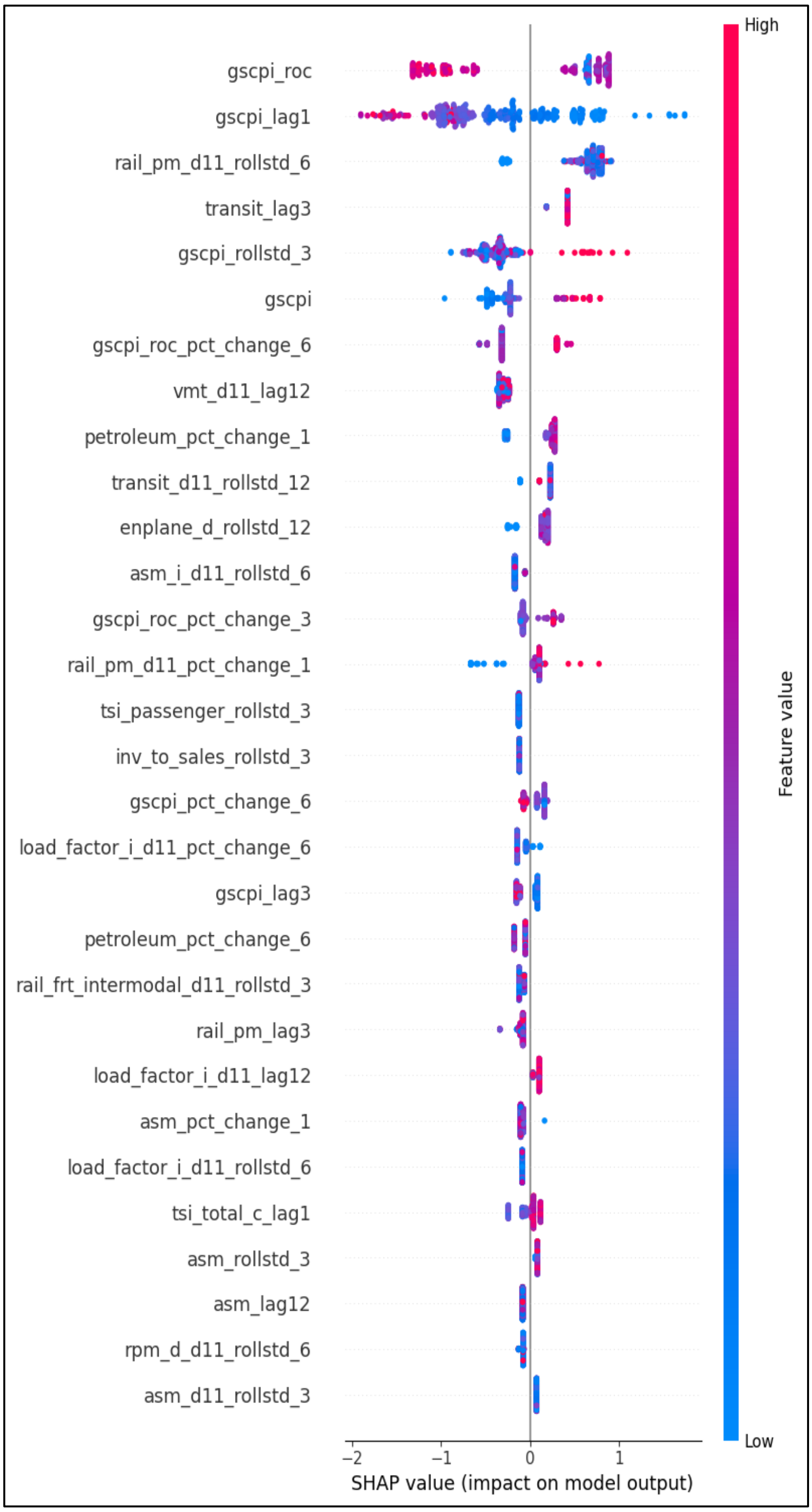

Fig.13: Global SHAP feature importance

Local SHAP explanations helped clarify how these pieces fit together in specific periods. When disruption probability spiked, the explanations usually showed a pattern that combined rising GSCPI momentum, falling freight throughput, or new volatility in aviation or rail data. SHAP force plots made it clear that no single factor drove a disruption prediction. Instead, several indicators pointed in the same direction and pushed the model toward a higher risk estimate.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

For instance, a sharp drop in freight percent change paired with an unusually high gscpi_lag1 value created a strong upward influence on predicted risk. These local insights show that the model captures the layered character of supply chain failures and helps analysts connect predictions to real-world operational signals.

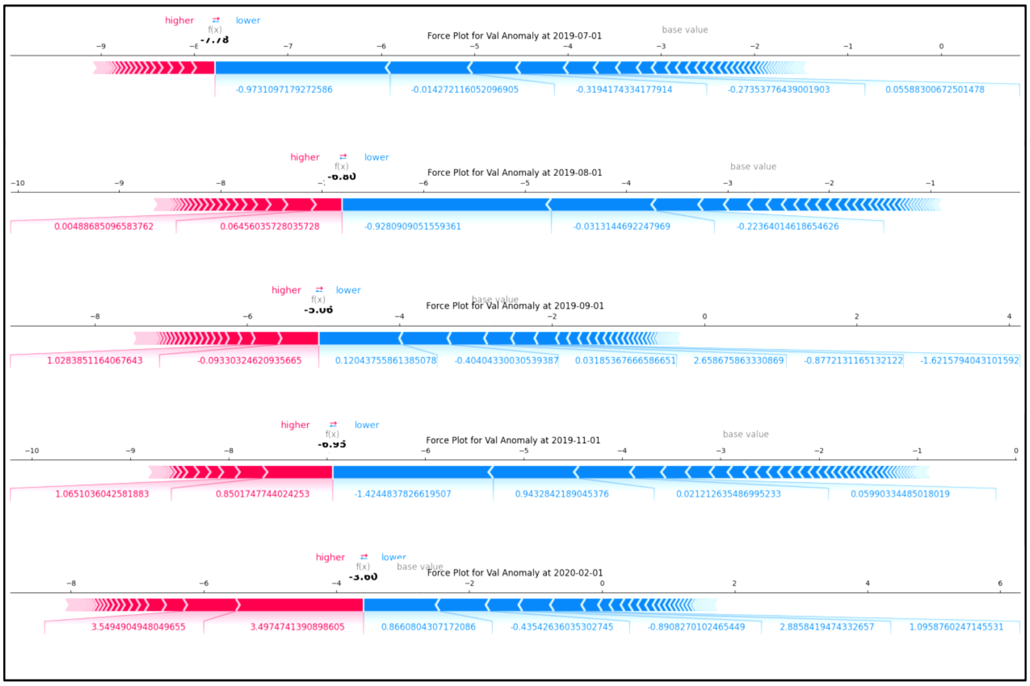

Fig.14: Local SHAP feature importance (samples)

4.3 Stress-Test and Scenario Findings

Scenario analyses were implemented to explore how industry-level risk would shift under hypothetical disruptions. These experiments serve as sensitivity tests that reveal how changes in exposure patterns propagate through the risk scoring architecture. Using a baseline disruption probability of approximately 0.0315, the first scenario increased exposure for a single industry by 20 percent. The result was a proportional rise in that industry's continuous risk score and a modest redistribution of risk among other industries due to normalization. This outcome demonstrates the structural dependency between exposure weights and final risk rankings: industries carrying more logistical responsibility shoulder disproportionately larger disruption risk when conditions deteriorate.

The second scenario increased exposure by 30 percent for the top two industries with the highest baseline exposure. This produced a more pronounced reallocation of total system risk, with these industries absorbing significantly larger shares of the disruption probability. The bar plot comparing baseline and shocked scenarios made the distributional effects clear, showing how systemic risk becomes concentrated under certain stress conditions. These findings highlight the interconnectedness of freight networks, indicating that shocks are not distributed evenly across sectors. Industries with high transportation intensity, reliance on long-haul freight, or sensitivity to multimodal congestion demonstrate amplified vulnerability in stress environments. These results provide actionable insights for policymakers and industry leaders seeking to prioritize resilience investments.

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

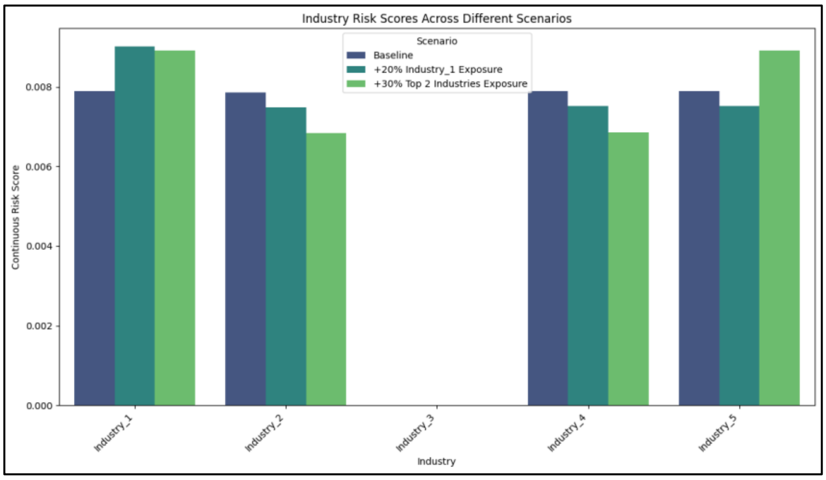

Fig.15: Industry risk assessment outcomes

4.4 Industry-Level and Regional Risk Scores

The final component of the evaluation aggregated disruption probabilities and exposure weights into industry-level risk scores. This approach allows risk to be interpreted not only as a model output but as a function of sector-specific vulnerability. Industries with higher baseline exposure to freight intensity and transportation volatility received larger continuous risk values after normalization. The resulting Low, Medium, and High categories provide an interpretable mapping between model predictions and sectoral risk posture. The categorization showed a clear stratification: industries with minimal exposure weights consistently fell into the Low-risk group, while transportation-heavy or logistics-dependent sectors occupied Medium or High tiers. The risk heatmap covering the most recent twelve months presented a dynamic view of how risk evolved over time, revealing moments where certain industries experienced transient spikes in vulnerability. This temporal surface indicates that industry risk is not static but fluctuates with shifting supply chain pressures, seasonal freight demand, and broader economic conditions.

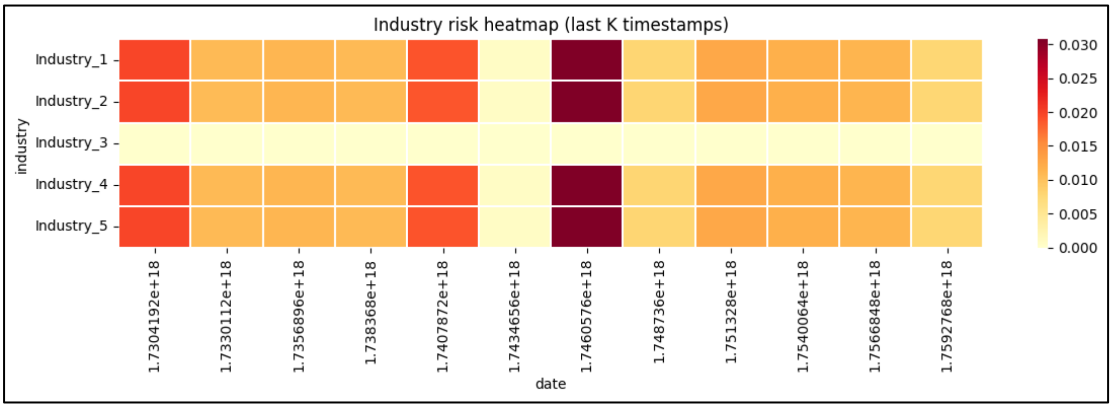

Fig.16: Industry risk heatmap

A notable finding is the sensitivity of final classifications to exposure values. Since exposure normalization redistributes relative importance across industries, any changes in assumed

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

exposure patterns lead to different risk rankings. This behavior underscores that industry-level risk is not solely determined by disruption probability but also by how industries are positioned within the broader transportation ecosystem. Because exposures serve as a proxy for dependency on logistics throughput, this framework allows analysts to stress specific industries or adjust exposure estimates based on real-world operational data.

5. Insights and Implications

5.1 Early Warning Potential

The results show that even with indicators that are noisy and partly outside the system, the models still managed to pick up rising stress well before major disruptions appeared. The clearest evidence came from the classification work. The RandomForest classifier, which reached perfect precision and strong separation between classes, repeatedly identified disruption periods where GSCPI momentum was already climbing while freight signals were losing strength. These patterns point to a simple idea: supply chain stress builds over time. It does not arrive out of nowhere. It grows through recognizable shifts in transportation efficiency, congestion tied to demand surges, aviation capacity tightening, and weakening freight momentum.

The LSTM added support to this interpretation. Its RMSE was higher than the tree-based regressors, yet it showed a reliable ability to capture turning points and changes in direction. The model struggled most during global shocks that fell outside the training window, which suggests that sequential patterns in GSCPI contain hints of pressure building long before it becomes obvious. A combined system that uses tree-based models for early signals and sequence-aware models for tracking structural buildup could produce a richer form of risk intelligence. Such a system would allow two layers of warning. One layer would detect slow-moving pressure that grows over several months. The other would flag sudden drops in freight or fast jumps in volatility that often precede disruption. This pattern of gradual accumulation implies that real-time monitoring is realistic. If stress starts to grow in advance, then organizations can act sooner, increasing the resilience of logistics networks rather than waiting for conditions to deteriorate.

5.2 Value for Policymakers and Businesses

The findings matter for agencies, operators, and large firms that navigate supply chain exposure. The predictive framework can support decisions across several areas. Regulators, including the Bureau of Transportation Statistics and transportation policy groups, can rely on the early warning signals to identify freight sectors likely to become congested before performance drops. Recent studies show that machine learning can help strengthen resiliency assessments across different regions of the country. Shawon et al. (2025) found that ML-based monitoring of logistics activity helps regulators direct resources to higher-risk corridors [26]. The results here reinforce that idea by showing that disruption probability is influenced by freight momentum, lagged aviation indicators, and volatility in throughput metrics. Policymakers can use these patterns to identify stress points and take action before bottlenecks spread across the system.

Received: August 12, 2025

International Journal of Applied Mathematics
Volume 38 No. 10s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Private freight operators and logistics managers can use the risk scores and scenario tests to understand how shocks move through their own operations. Because the framework multiplies disruption probability by exposure, firms gain a way to compare their vulnerability against other sectors. Companies with heavy dependence on long-haul freight or those working in industries with structurally higher transportation volatility can explore mitigation plans such as shifting transport modes or adjusting supplier locations. The scenario analysis adds another tool for planning. The stress tests showed that small increases in exposure can create outsized jumps in industry-level risk, especially for sectors that already handle a large share of system throughput. Athey (2019) notes that transparent machine learning can improve policy design by revealing nonlinearities and spillovers that are otherwise hard to detect [3]. Here, the combination of interpretability and stress testing makes clear how vulnerabilities spread, giving both businesses and regulators a way to design targeted interventions rather than broad, unfocused responses.

5.3 Transparency and Trust

High-stakessectors tend to be cautious about black-box models because decisions about infrastructure or large-scale operations require clear explanations. The explainability framework used here addresses that concern. The global SHAP results show the features that repeatedly influence disruption risk. Lagged GSCPI, freight percentage changes, rolling averages of transportation metrics, and aviation capacity indicators appear again and again as leading contributors. This reassures practitioners that the models are learning relationships that make economic sense rather than chasing quirks in the data. SHAP local explanations add another layer by breaking down the reasoning for individual predictions. They show why a specific month was flagged as high risk. This level of clarity matters because policymakers and managers need to understand the rationale before adjusting budgets, schedules, or mitigation plans. Athey (2019) argues that machine learning becomes far more useful in policy settings when transparency is built in. Transparent models allow users to question predictions, identify biases, and feel confident making decisions based on the output [3]. The SHAP framework in this study supports that goal by tying predictions back to signals that freight experts and economists already understand.

5.4 Limitations

The results are encouraging, yet several limitations should be taken into account when applying the framework in real settings. The industry exposure estimates rely on synthetic approximations built from the available freight data. The absence of detailed industry-level information from the Freight Analysis Framework required the use of proxy exposure values. These proxies capture broad differences in exposure but cannot fully describe the true supply chain dependencies within each sector. More detailed exposure data would likely improve industry-level risk estimates. The modeling depends on publicly available BTS, FRED, and GSCPI indicators. These datasets cover the national level and are updated monthly. Many disruptions unfold through regional or local networks that move much faster, sometimes on weekly or daily time scales. Incorporating port-level, carrier-level, or customs-level microdata in future work would improve real-time visibility and raise the accuracy of the system. The indicators also show non-stationarity across long periods, which creates challenges for models trained on limited windows. Adjustments such as regime detection, adaptive recalibration, or hybrid approaches that mix statistical methods with machine learning could improve robustness in the presence of structural shifts.

6. Future Work

Future research should build on this study in several directions so that the system becomes more accurate and more useful in real settings. One of the biggest steps forward involves adding freight data with much finer detail. The current setup depends on national BTS and FRED indicators, which help reveal broad patterns but smooth over the local choke points where real trouble begins. Many disruptions start at a particular port, warehouse, rail line, or trucking corridor where pressure builds long before it shows up in national numbers. Bringing in port throughput records, customs delays, telemetry from trucks, vessel arrival patterns, and warehouse occupancy data would help the models catch these issues earlier and offer warnings tied to specific locations instead of sweeping national conditions. Shifting from macro to micro detail would make the framework far more practical for policymakers and major logistics operators. A second path involves replacing the single national model with forecasting systems designed for individual industries. Each sector faces stress through its own mechanisms because its logistics needs differ. Pharmaceuticals rely on temperature-controlled storage and international air freight. Construction materials lean heavily on bulk maritime shipments and long-haul trucking. A single model cannot fully reflect these differences. Separate models for key industries would let the chosen features, lag structures, and stress definitions adjust to the way each sector experiences pressure. This would improve predictive accuracy and give industry stakeholders risk scores that reflect the realities of their own supply chains.

Another promising direction is to use graph neural networks to represent how disruptions travel across the supply chain. These systems behave like networks where a shock at a single node, such as a major port or key supplier, can spread through transport routes and production linkages. Traditional models treat features as isolated. Graph neural networks can treat suppliers, hubs, carriers, and regions as connected parts of a larger system and learn how disruptions move between them. This could help identify which nodes act as amplifiers during high-pressure periods and improve the framework’s ability to track cascading failures. The next major step is to broaden the types of data included. Early warning signals often appear in unstructured sources like international news, port authority notices, regulatory updates, satellite views of vessel queues, logistics carrier posts, and even social media from shippers and operators. Mixing numerical indicators with text and images would create a more complete monitoring system that blends qualitative and quantitative evidence. Real-time natural language processing on reports and computer vision applied to satellite feeds could greatly strengthen situational awareness and reduce dependence on slow-moving macroeconomic indicators.

Finally, future work should address the instability caused by large economic regime changes. Supply chain behavior shifts during recessions, energy shocks, geopolitical events, or major changes in trade structure. Models trained on past data can falter when relationships between indicators start to evolve in unexpected ways. Tackling this requires approaches such as time adaptive modeling, online learning, and more explicit treatment of uncertainty. Detecting when the data-generating process has changed would help the system adjust rather than slide into inaccuracy during extreme conditions. This is critical for maintaining reliability during unusual or unprecedented disruptions. These directions push the framework toward a more flexible, detailed, and adaptive monitoring system. They move it closer to functioning as a real-time early warning tool shaped by diverse data sources, modern modeling approaches, and deeper knowledge of how different industries experience supply chain stress.

Concluion

This study shows that signs of supply chain stress in the United States become easier to anticipate when different logistics and transportation signals are brought together in one predictive system. By combining the Global Supply Chain Pressure Index with freight indicators from the Bureau of Transportation Statistics and transportation series from FRED, the work creates a dataset that reflects both global pressures and the domestic shifts that usually show up before disruptions. The modeling results suggest that tree-based classifiers and sequence-focused LSTM models can pick up early stress patterns that traditional metrics tend to miss, while unsupervised methods catch structural changes that often signal the start of instability. A major contribution of this work is the attention given to interpretability. Global and local SHAP explanations point out how lagged pressure signals, freight momentum, and volatility across different transportation modes come together to shape disruption risk. These explanations give analysts a closer look at how the model arrives at its conclusions and help link predictive output with real-world decision-making. The creation of continuous industry-level risk scores, along with scenario testing, pushes the framework toward a practical risk assessment tool. It shows how shocks spread through transportation exposures and how vulnerability deepens in sectors that rely heavily on freight movement.

These findings suggest that machine learning can support proactive monitoring of supply chain stress instead of relying on after-the-fact assessment. The framework identifies early signs that pressure is building, clarifies what is driving those signals, and converts them into risk categories that planners can use. Some challenges remain, especially related to data resolution and evolving system behavior, but the results provide a base for developing early warning tools that blend predictive skill with transparency and sector-level insight. With further refinement and more detailed freight data, this approach can help build logistics systems that are more resilient and better able to adjust to the growing volatility that shapes modern supply chains.

References

[1] Aashish, K. C., Zamil, M. Z. H., Mridul, M. S. I., Akter, L., Sharmin, F., Ayon, E. H., ... & Malla, S. (2025). Towards eco-friendly cybersecurity: Machine learning-based anomaly detection with carbon and energy metrics. International Journal of Applied Mathematics, 38(9s).

[2] Adrian, T., & Brunnermeier, M. K. (2016). CoVaR. American Economic Review, 106(7), 1705–1741.

[3] Athey, S. (2019). The impact of machine learning on economics. In The Economics of Artificial Intelligence: An Agenda (pp. 507–547). University of Chicago Press.

[4] Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable machine learning in credit risk management. Computational Economics, 57(1), 203–216.

[5] Danielsson, J., Valenzuela, M., & Zer, I. (2018). Learning from history: Volatility and financial crises. The Review of Financial Studies, 31(7), 2774–2805.

[6] Das, B. C., et al. (2025). AI-driven cybersecurity threat detection: Building resilient defense systems using predictive analytics. arXiv preprint arXiv:2508.01422.

[7] Debnath, S., et al. (2025). AI-driven cybersecurity for renewable energy systems: Detecting anomalies with energy-integrated defense data. International Journal of Applied Mathematics, 38(5s).

[8] Dzielinski, M. (2012). Measuring economic uncertainty and its impact on the stock market. Finance Research Letters, 9(3), 167–175.

[9] Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate GARCH models. Journal of Business & Economic Statistics, 20(3), 339–350.

[10] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669.

[11] Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. Journal of Econometrics, 2(2), 111–120.

[12] Hasan, M. R., Rahman, M. A., Gomes, C. A. H., Nitu, F. N., Gomes, C. A., Islam, M. R., & Shawon, R. E. R. (2025). Building robust AI and machine learning models for supplier risk management: A data-driven strategy for enhancing supply chain resilience in the USA. Advances in Consumer Research, 2(4).

[13] Hasan, M. S., et al. (2025). Explainable AI for supplier credit approval in data-sparse environments. International Journal of Applied Mathematics, 38(5s).

[14] Hollo, D., Kremer, M., & Lo Duca, M. (2012). CISS – A composite indicator of systemic stress in the financial system. European Central Bank Working Paper Series, No. 1426.

[15] Islam, M. Z., et al. (2025). Cryptocurrency price forecasting using machine learning: Building intelligent financial prediction models. arXiv preprint arXiv:2508.01419.

[16] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Transactions of the ASME – Journal of Basic Engineering, 82(1), 35–45.

[17] Kliesen, K. L., Owyang, M. T., & Vermann, E. K. (2012). Disentangling diverse measures: A survey of financial stress indexes. Federal Reserve Bank of St. Louis Review, 94(5), 369–397.

[18] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems (NeurIPS), 30.

[19] Ray, R. K. (2025). Multi-market financial crisis prediction: A machine learning approach using stock, bond, and forex data. International Journal of Applied Mathematics, 38(8s), 706–738.

[20] Reza, S. A., et al. (2025). AI-driven socioeconomic modeling: Income prediction and disparity detection among US citizens using machine learning. Advances in Consumer Research, 2(4).

[21] Schwert, G. W. (1989). Why does stock market volatility change over time? The Journal of Finance, 44(5), 1115–1153.

[22] Shivogo, J. (2025). Fair and explainable credit-scoring under concept drift: Adaptive explanation frameworks for evolving populations. arXiv preprint arXiv:2511.03807.

[23] Shovon, M. S. S. (2025). Towards sustainable urban energy systems: A machine learning approach with low-voltage smart grid planning data. International Journal of Applied Mathematics, 38(8s), 1115–1155.

[24] Sirignano, J., Sadhwani, A., & Giesecke, K. (2018). Deep learning for mortgage risk. The Journal of Finance, 73(6), 2727–2771.

[25] Sizan, M. M. H., et al. (2025). Machine learning-based unsupervised ensemble approach for detecting new money laundering typologies in transaction graphs. International Journal of Applied Mathematics, 38(2s).

[26] Shawon, R. E. R., et al. (2025). Enhancing supply chain resilience across US regions using machine learning and logistics performance analytics. International Journal of Applied Mathematics, 38(4s).

[27] Whaley, R. E. (2000). The investor fear gauge. The Journal of Portfolio Management, 26(3), 12–17