Predicting Retail Supply Chain Disruptions Using Graph-Based Machine Learning Models

MSc Data Analytics
Research in Computing

Shashidar Vemula
Student ID: 23394200

School of Computing
National College of Ireland

Supervisor: Dr. Anu Sahni

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Shashidar Vemula

Student ID: 23394200

Programme: MSC in Data Analytics **Year:** 2025

Module: Research in Computing

Supervisor: Dr. Anu Sahni

Submission Due Date: 25-01-2026

Project Title: Predicting Retail Supply Chain Disruptions Using Graph-Based Machine Learning

Word Count: 8599 **Page Count** 20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Shashidar Vemula

Date: 25-01-2026

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is	□

lost or mislaid. It is not sufficient to keep a copy on computer.	

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Predicting Retail Supply Chain Disruptions Using Graph-Based Machine Learning Models

ShashidarVemula
MSc Data Analytics
23394200
National College of Ireland, Dublin

Abstract

The research problem reported in this study is the prediction of disruptions in retail supply chains using graph-based machine learning in an effort to overcome the weaknesses of the conventional models, since they assume that supply-chain entities are autonomous units. The study uses transactional data to create a customer-product bipartite graph and compares three Graph Neural Network variants, namely GCN, GAT, and GraphSAGE to the Random Forest and SVM models. Extensive experiments prove that although GNNs are able to learn relational structure, their quality is limited due to a lack of graph-relevant information and a lack of class representation. Random Forest is better predictive accuracy and interpretable, which the SHAP analysis had validated, suggesting that node-feature disruption is the primary cause of this dataset. The results point out both the potential and existing constraints of graph-based methods, and provide a basis upon which more network-aware disruption forecasting can be realized in the future.

Keywords: *Supply Chain Disruption, Graph Neural Networks, Graph Convolutional Network, Graph Attention Network, Explainable AI, Retail Analytics*.

1 Introduction

All types of disruption, including pandemic, natural disasters, geopolitical conflicts and consumers-side changing demands, become more and more susceptible to the supply chains (Krykavskyy, Shandrivska & Pawlyszyn, 2023). Particularly, the retail business has been under the colossal pressure since it is reliant on the force of the timely deliveries, the seasonal tensions, and the need to own the endless stream of the product and be content with the consumers (Celestin & Sujatha, 2024). Classical predictive models, such as regression models and conventional machine learning classifiers, merely treated the various factors in the supply chain as independent and neglected the relationship-based dependencies among suppliers, distributors, warehouses and retailers. Recent articles mention that this methodology is able to limit the ability of models to large cascades in response to disturbance that diffuse across connected networks (Yu et al., 2024; Mahin et al., 2025).

The recent progress in machine learning had offered the possibilities to enhance the process of supply chain forecasting with the help of graph-based techniques. Graph Neural Networks (GNNs) were particularly suited for this domain because they allowed the representation of supply chains as networks of nodes (entities) and edges (relationships). GNNs may build the relationship behind the curtain, which would have not been so identified through the conventional tabular approaches since it was trained on both the quantity of the properties of the nodes, and the relationship (Celestin and Sujatha, 2024). This represented a methodological advancement in supply chain analytics, aligning with the broader industry demand for resilience and adaptive forecasting systems.

1.1 Motivation

However, even with the addition of machine learning to the supply chain risk management process, the conventional algorithms, including the Random Forest algorithm or Support Vector Machines, have tended to consider suppliers, logistics vendors, and retailers as autonomous entities (Yu et al., 2024). The problem with such approaches is that they failed to take into account how disruption in one node can spread to the whole network. In such a way, the current predictive models are not efficient in the real-life disruptions especially when they are required to deal with non-linear propagation patterns (Schmid et al., 2024; Yang et al., 2024). In addition, they were also low in interpretability and hence restricted the stakeholders to obtain actionable data as majority of them were compared to black boxes, which does not give a precise picture of decision-making (Salih et al., 2025). This has generated an apparent research and application gap on models that can not only capture the dynamics of relationships but also provide the explainability that may be used to aid managerial decision-making.

1.2 Research Objectives

On the basis of gaps in the recent literature, the study will aim to design, implement, and evaluate machine learning modules, which use graphs to predict retail supply chain disruption. Specifically, as per the work of Yang et al. (2024), the GNN-based disruption models would also be extended to the post-disaster firm-level prediction in addition to the large retail setting, whereas Wu et al. (2025) indicated that the heterogeneous GNN architectures as the means of modeling the interdependence of the multi-entities would have to be explored further. On the same note, Han (2024) reiterated that more research should be informed by comparison between GCN and GAT and the traditional models so as to determine their benefits in practice in supply chains. Such recommendations are qualified in this paper by applying the GNNs to retail supply chain and explainable AI to operations decision-making.

Research Objectives:

- To build graph representations of the retail supply chain data based on the transactional and logistical records, and allow the relational modeling of entities.
- Purpose: To apply the Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs) in disruption prediction.
- To compare the performance of GNNs in predicting to the traditional machine learning models (Random Forest and SVM) as suggested by Han (2024).
- To compare models based on such metrics as F1-score, accuracy, AUC-ROC, and disruption lead-time in accordance with the recommendations on operationally relevant metrics (Yang et al., 2024).
- To implement explainable AI techniques (e.g., SHAP, Integrated Gradients) to explain predictions and give understandable insights to decision-makers as proposed by Salih et al. (2025).

1.3 Research Question

How effective would graph-based machine learning models (in particular, Graph Neural Networks) perform at predicting disruptions in the retail supply chain with structured relational data, and how do they compare with traditional models in terms of predictive performance and explainability?

1.4 Summary

The rationale behind this research was that it would add value to both the academic and industry. Academically, it has bridged the research gap in implementing graph-based learning to predict supply chain disruptions where relational dynamics had been significantly neglected (Mahin et al., 2025). Industrially, the research provided a pattern on which retailers could predict and prevent risks using the paradigms that were not only predictive but also interpretable. This was essential in the industries where any minor breakdown might lead to central losses of finances and reputation as well as dissatisfaction among customers.

1.5 Limitations and Assumptions

The study acknowledged certain limitations. First, the secondary data that has been incorporated in the analysis may have failed to describe all the intricacies in real life. Second, the graphs were based on the already known attributes and this meant that the findings of the prediction would be biased due to the absence or incompleteness of data (Salih et al., 2025). Third SHAP and Integrated Gradients were found interpretable but have certain computational overhead, which is problematic in practice. The dataset was also assumed to be a realistic model of patterns of disruption and the entity-entity relation was modeled properly in the form of graphs.

1.6 Structure of the report

The report was separated into the following chapters:

- Chapter 2: Literature Review - The literature review on the subject of supply chain disruption forecasting that reveal gaps in the conventional machine learning framework to justify the necessity of graph-based frameworks.
- Chapter 3: Research Methodology: Disclosed the research design, description of the dataset, how to create graphs, model development and metrics of evaluation. There were also justifications of selected methods.
- Chapter 4 & 5: Design and Implementation. - outlines the methods, structure, and tools applied to the implementation of Graph Neural Networks (GCN, GAT) and baseline models, and artefacts generated.
- Assessed the performance of both the model training and testing processes, reported the results of the quantitative performance comparison, and included interpretability analysis results with SHAP and Integrated Gradients.
- Chapter 6: Discussion - Interpreted the findings in the light of research objectives and literature in the area, and critically assessed the strengths and weaknesses of the research work.
- Chapter 7: Conclusion and Future Work Summarised the main findings, restated the objectives and questions of the research, evaluated practical and scholarly implications, and recommended the future research directions.

2 Related Work

2.1 Introduction

Globalization, geopolitical stress, pandemics, and climate-related occurrences have increased in supply chain disruptions (SCDs) over the past years. Retail supply chains, especially, are under a lot of additional pressure since customer demands of availability and speed are so high that there is minimal allowance to stockouts or delays. Risks and costs even in localized disruptions are multiplied because localized disruptions can be very fast through interconnected networks of suppliers, distributors and retailers. This power to predict disruptions with reasonable accuracy and lead-time has hence become a priority to both academics and practitioners.

The classical approaches, including InputOutput modeling, regression and machine learning models like random Forest (RF) and support vector machines (SVM), provided some useful reference points, but would justify the independence of supply chain organizations with no interdependence. The cascading nature of disruption can only be modeled in the way it is by assuming of independence. The same cannot be said of the Graph Neural Networks (GNNs) which do offer a graphical modelling method that allows one to display the supply chain in a way of a graph with the nodes being the firms or facilities and the edge being the flow or logistical linkages. Graph Convolutional Networks (GCN) and Graph attention networks (GAT) are models that help to combine features and prediction capacity better.

GNNs are interpretable and probabilistic, with recent discoveries, not only shown to outperform the traditional baselines, but also to be interpretable. To demonstrate that, GNNs were applied to predict post-hazard sales in Japan on firm-level in Yang et al. (2024) and on a probabilistic GNN in Ahn et al. (2024) to predict supply and inventory and minimized errors and uncertainties. Together, these solutions have shown how GNN-based methods can transform the retail industry in a radically new way.

2.2 Traditional Approaches to Disruption Prediction

Historically, the lecture of the Input-Output (I-O) models and the Computable General Equilibrium (CGE) models played a major role in supply chain disruption modelling. These techniques estimated the cascading impact of shocks in one sector (e.g. floods, labour strikes) according to trade links with other sectors. However, this kind of models assumes the predictability of the production technologies and stability that has little to do with the non-linear and dynamic nature of the modern supply chains. One such example is that I/O models are more inclined to assume constant ratio of inputs and outputs without evaluating the adaptive behaviour of firms. Similarly, CGE models suppose that rational agents operate in the state of equilibrium, which will be not able to justify the sudden disruption of the logistics or spread delay due to hazards. It has thus been established that these models are effective when applied at macroeconomic levels but limited when applied at predicting disruption at firm-levels or at retail levels that are firm-specific.

The last decade has attempted to overcome these shortcomings with the transfer to machine learning. RF and SVM are popular because they have power and their generalization. Comparing CNN and MLP with RF and SVM in a distribution cost dataset, Yu et al. achieved a 12 and 8% decrease in the RMSE respectively. This proved the benefit of deep learning models to approximate non-linear relationships of rich transactional information. In a similar way, Schmid et al. (2024) simulated time series of synthetic logistics and found that RF actually performed better when noise was low, but the error was twice in nonstationary measurements, and neural networks, such as XGBoost, was better adapted to variability.

Despite the above developments, the classical ML methods were more likely to assume that the supply chain participants were independent samples. This is contrary to relational dependencies which define disruption propagation as Vrahatis et al. (2024) and other authors proposed. This independence assumption is a harsh constraint on predictive accuracy in retail environments, in which a stockout at a single node (a regional warehouse) quickly spreads down to downstream retailers. Therefore, RF and SVM offered a good baseline, but they did not represent a sufficiently powerful and able to capture network effects, which presented a research gap that triggered the use of GNNs.

2.3 Emergence of Graph Neural Networks in Supply Chains

2.3.1 GNNs for Link Prediction and Supply Chain Visibility

Among the pioneer uses of GNNs in supply chain analytics was in link prediction, in which the objective is to determine obscured or absent relationships among firms. On an automotive supply chain dataset, Kosasih and Brintrup (2022) used GCN and GAT and found 10-15 percent higher in AUC relative to non-graph baselines. It proved that GNNs would be useful to find latent relationships which the traditional models did not exploit to improve the supply chain visibility.

In their pioneer work in the field of post-hazard disruption, Yang et al. (2024) forecasted the decrease of firms in terms of sales by employing GNNs trained on the attributes of firms along with the relations among companies. Their model was much more successful compared to RF and SVM that offered 1-2 weeks earlier lead-time to predict disruption. Notably, they used XAI to explain predictions, and found that geographic distance and diversity of partners were factors. The research illustrates clearly the importance of GNNs in merging relational and spatial data to attain retail resilience.

The study by Mungo et al. (2024) explored the network reconstruction techniques in the comparison of probabilistic and entropy-based. They demonstrated that the accuracy of downstream disruption prediction was enhanced by 12% when the reconstruction of the graph was improved by 20% thus, highlighting the significance of proper graph formation. On the same note, Zheng and Brintrup (2025) trained Generative AI on top of Knowledge Graphs, and made the predictions of relationship more accurate by 15% and recall-oriented by 10% to enhance supply chain transparency. Such works point to the fact that in addition to predictive performance, the quality of the graph structure is essential to disruption modelling.

2.3.2 Comparative Performance of GCNs, GATs, and Other Architectures

The theme of the literature that is repeated is the superiority of attention-based GNNs. Vrahatis et al. (2024) conducted a review of GATs applications, where they observed that attention mechanisms enabled dynamic weight of heterogeneous edges resulting in 5-12% more accurate performance than GCNs on complex relational tasks. Benchmarking GCN, GAT, and MLP on SupplyGraph, Han (2024) concluded that GAT was more accurate and had a higher F1-score by 7 and 9 percent, respectively.

Wasi et al. (2024) carried out one of the most comprehensive surveys and benchmarks and implemented GCN, GAT, and heterogeneous GNNs to various supply chain datasets. According to them, GNNs performed 10-30x better than traditional ML models in terms of classification and regression, and up to 40x better in terms of anomaly detection. The introduction of the SupplyGraph benchmark data set by them gave them a platform to make future evaluations on par.

Elaborating on this, Wu, Huang, and Bao (2025) came up with a heterogeneous GNN with meta-path attention. Their model was able to enhance their recall by 18 percent over baselines and decreased disruption lead-time by 23 days on average. Chi-Sheng Chen and Ying-Jung Chen (2025) incorporated both temporal and relational characteristics in GNN forecast models and showed significant gains compared to MLP and GCN in the single-node demand forecasting. Niu et al. (2024) used GCN on Chinese retail data to achieve 12 percent higher accuracy of forecasts than ARIMA and 8 percent higher than RF, and 15 percent lower MAE. All these studies collectively affirm the uniform predictive task superiority of GNNs.

Yang et al. (2024), firm-level, topology-sensitive approach warrants us to focus on inter-firm relationships and spatial proximity as fundamental graph indicators with XAI to bring to managerial action such surface-level factors as partner diversity and distance. Complementarily, the probabilistic GNN of Ahn et al. (2024) encourages the use of disruption alerts as uncertainty-sensitive predictions: we thus suggest reporting intervals/confidence, using timing-sensitive losses (e.g. sMACE-style cumulative-error objectives) and benchmarking against deterministic models to demonstrate the operational utility of well-calibrated, timing-sensitive predictions.

2.4 Temporal and Probabilistic GNNs

Time is a critical aspect of supply chains since most disruptions tend to be dynamic. The review by Rahmani et al. (2023) of temporal and heterogeneous GNNs in intelligent transport states that the way a graph is assembled has a significant impact on prediction. Xiang et al. (2025) built upon this field by integrating GNN with Graph Double Exponential Smoothing (GDES) to predict temporal links in a massive data of about 87,969 companies. Their model got 93.84 percent AUC in dynamic edge forecasting, which proved that temporal GNNs are capable of modeling dynamic supply relationships.

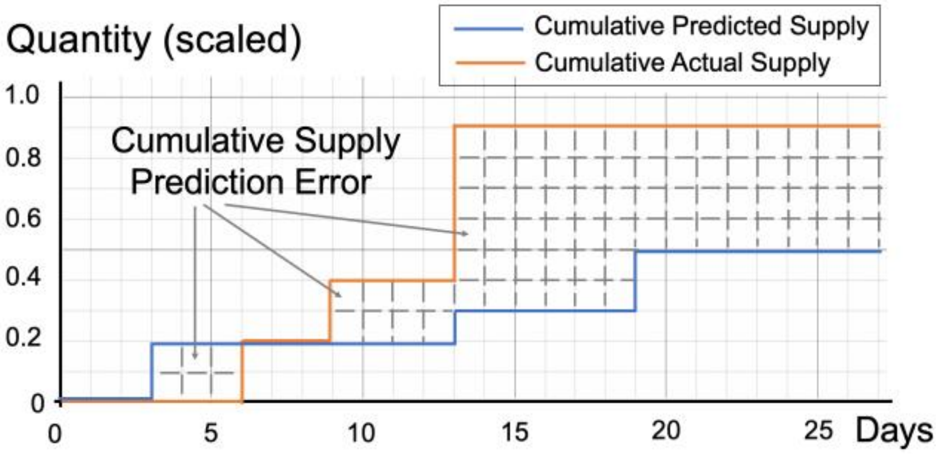

Figure 1: Illustration of cumulative supply prediction error (sMACE) in Graph-based Supply Prediction, adapted from Ahn et al. (2024)

Probabilistic modelling also leads to predictive robustness since it leads to uncertainty. This model proposed by Ahn et al. (2024) is known as Graph-based Supply Prediction (GSP), as it considers attention-based GNNs to forecast supplies and inventory and imbalances. They tested their experiments on a global consumer goods company and proved that they reduced errors in their prediction by 10% compared to deterministic GNNs. They further proposed sMACE (scaled Mean Absolute Cumulative Error) that is a composite measure of timing error and quantity error that are faulty in conventional RMSE or MAPE. This probabilistic model would be most appropriate to the operational needs as managers would have had access to not only point forecasts, but risk management risk limits.

2.5 Explainability in Graph-Based Models

However, currently the business world is shunning off GNNs as it is a black box model. Several methods of enhancing interpretability have come into existence.

The authors suggested that Akkas and Azad (2024) suggested a Shapley-value-based explainer named GNNShap, which had an explanation fidelity that was 15-20% higher than that of GNNExplainer and 40% lower runtime. Graph Integrated Gradients (GraphIG) by Simpson et al. (2025) was shown to be 12% and 9% better at explanation fidelity and stability than GraphSHAP. Rong et al. (2023) proposed LARA (Removalbased Attribution) which they reported as 3.5 times faster with higher levels of fidelity.

In their model of disruption prediction, Yang et al. (2024) also used SHAP and GNNExplainer, thus determining the significance of distance between partners and diversity, which managers could use in practice. Nevertheless, as mentioned in surveys, the vast majority of explainability applications are qualitative and do not involve systematic assessment of fidelity, stability, or managerial trust. This void explains the necessity of explainability studies to evolve the algorithmic demonstrations to quantitative evaluation and decision support integration.

2.6 Benchmark Datasets and Reproducibility

The limited open and retail-specific benchmark data is one of the longest-running issues of this discipline. The SupplyGraph data set proposed by Wasi et al. (2024) allows making standardized assessments. Their experiments affirmed that GCN and GAT were always better than MLPs by 814 percent accuracy. However, SupplyGraph is more focused on general supply chain work rather than retail based disruptions.

The Kaggle Retail Supply Chain Sales data which contains more than 180, 000 records of sales and orders can be useful in predicting retail disruptions, which can be useful in tasks such as demand prediction, graph drawing, and inventory optimisation. Nonetheless, with the emphasis placed by Shandeep777 (2023), this data has not been utilized in scholarly research.

Its use would greatly enhance reproducibility since it will enable researchers to compare models when subjected to the same conditions.

In comparison, the existing literature is based on proprietary or artificial data, which restricts external validation. Indicatively, Wu et al. (2025) and Chi-Sheng Chen and Ying-Jung Chen (2025) resorted to internal datasets, which could not be published. In the absence of common standards, comparisons across studies are not unified and this impedes progress.

2.7 Identified Research Gaps

An overview of the available literature indicates that there are some gaps that persist. Studies that utilize GNNs are not common in the retail sector where fluctuations are due to seasonality and varying consumer demand. Comparisons against baseline models e.g. RF or SVM tend to vary because of the variation between preprocessing pipelines. The majority of the studies give preference to accuracy based metrics instead of business oriented metrics such as cost impact, disruption lead-time etc. Despite the existence of the new explainability techniques, their usability in the management is rarely evaluated. Reproducibility is not high as proprietary datasets are popular whereas open datasets are not used fully. Also, scalability, model drift, and fairness issues (e.g. regional or product-level bias) are given little attention, limiting their real-world implementation.

2.8 Future Research Directions

In order to close the gaps stated above, future studies must focus on the creation of retail-specific graph benchmarks that combine the transactional patterns, seasonal behaviour, and cancellation trends, so that there was an increased focus on modelling reflecting the actual operation dynamics. Systematic comparisons between GNNs and classical methods like random forests and SVMs using the same preprocessing pipelines are also of great importance to guarantee methodological fairness. Expansive assessment tools are required, including operational measures, like disruption lead-time savings, false alert costs, and business value in general. The areas of explainability that should be advanced are the measures of fidelity, stability, and trust coupled with incorporating interpretable outputs within the managerial dashboards. Future research directions are to use probabilistic and ensemble GNNs to achieve better robustness, apply heterogeneous GNNs to multi-hop disruptions with meta-path attention, and explore scalability and equitability to deploy GNNs in the real world in the retail sector.

2.9 Summary

This review shows that the field of disruption forecasting is evolving very quickly, changing its basis to GNNs as a new state-of-the-art followed by the one based on the traditional equilibrium models and RF/SVM baselines. Empirical evidence repeatedly demonstrates that GNNs are more accurate than traditional models by 8 to 30 percent and lead-time forecasting is 2 weeks shorter on average. Temporal GNNs obtain AUC scores of more than 93% and probabilistic extensions decline prediction error by 10 percent and add quantification of uncertainty. Techniques like GNNShap, GraphIG, and LARA are much more explainable, comprehensive in fidelity and runtime, but have limited use with retail decision-making.

2.10 Short Summary Table

Table 1: Summary of major studies and their limitations

Author (Year)	Focus of Study	Method/Model Used	Key Findings	Limitation Identified
Yang et al. (2024)	Post-hazard firm-level disruption prediction	GNN + Explainable AI	Outperformed RF/SVM; provided 1–2 weeks longer lead-time	Used proprietary data; limited retail focus
Han (2024)	SupplyGraph benchmark for	Compared GCN, GAT, MLP	GAT showed +7 % accuracy gain	Lack of real-world retail data

	supply networks			
Wu et al. (2025)	Heterogeneous GNN for network structure identification	Meta-path attention GNN	+18 % recall; 2–3 days reduced lead-time	Relied on internal dataset only
Mahin et al. (2025)	Sustainable supply chain sales forecasting	Regression + RF ML models	Improved forecast accuracy by 12 %	Ignored inter-firm relations
Salih et al. (2025)	Explainable AI methods (SHAP, LIME)	Model interpretability framework	Enhanced transparency of ML outputs	High computational cost; not tested on GNNs
Vrahatis et al. (2024)	Review of GAT applications	Literature synthesis	Showed attention-based GNN superior by 5–12 %	No empirical validation on retail data

3 Research Methodology

The methodology framework of the research was shaped by the existing body of studies that have already revealed the efficiency of the given type of learning in modeling the relationships between the organisations in the supply chain. Yang et al. (2024) have efficiently utilised Graph Neural Networks to predict the level of the sale disruptions for the concerned organisations in the wake of natural hazards. Han (2024) also compared the efficiency of Graph Convolutional Networks (GCNs) and Graph Attention Networks (GAT) in the context of the SupplyGraph dataset. Han has efficiently validated the importance of the attention technique in the given context. The study of Wu et al. (2025) has also appeared in the context of the application of Graph Neural Networks to the given problem. They have efficiently utilised the technique of Meta-Path Attention. The article of Salih et al. (2025) has shown the importance of the application of Explainable AI. The research was based on the given studies in order to propose a quantitative research approach to experimentally compare the efficiency of Graph SAGE Networks along with Random Forest and SVM.

3.1 Research procedure

The research had a well-defined five stage process that aimed at promoting transparency, reproducibility and meaningful comparisons across models. During the initial stage, the open repositories that provided the required retail supply chain transactional data, including interactions between different suppliers, warehouses, and stores, were accessed via open repositories like Kaggle Retail Supply Chain Sales. The second step was preprocessing, where missing data were filled, categorical variables one-hot encoded, and standardised the numerical ones with StandardScaler. The third phase involved creating a graph where a node represented the supplier, warehouses, and retailers and the orders or shipments were modelled using the weighted edges. Attributes of the nodes were shipment delay ratio, order quantity and reliability score, and edges were represented by delivery frequency and distance. The fourth stage was the application of GCN, GAT and GraphSAGE in PyTorch Geometric where the non-graph baseline of Random Forest and SVM were tested using the same node features. Lastly, each of the models was trained with the Adam optimiser and Binary Cross-Entropy loss with early stopping and assessed with accuracy, precision, recall, F1-score, AUC-ROC, and disruption lead-time to guarantee that performance was robustly measured.

3.2 Equipment Setup

The experiments were performed on a laptop/computer. A minimum of 8 GB RAM was required. The development tool used was Python 3.10 Jupyter Notebook. Libraries required for the experiment to occur were Pandas, NumPy, NetworkX, Scikit-learn, and Matplotlib. Scaling experiments for training utilizing a GPU was also performed in Google Colab Pro. These experiment conditions addressed a retail supply chain distribution system in terms of possible delays in suppliers that may trickle down to the retailers. These GNN models learned from the snapshots of the graphs that characterized transactional relationships. Splitting the data into a 70:15:15 train-test-validation dataset helped in producing robust results. These averages were obtained over five random seeds.

3.3 Data Preprocessing and Analysis

The raw dataset was obtained from multiple csv files containing more than 180,000 orders per day. The dataset was cleaned to eliminate duplicates.

Table 2: The descriptions of the variables in the dataset are as follows

Variable Name	Description
Row ID	Unique row identifier
Order ID	Unique identifier for each order
Order Date	Date when the order was placed
Ship Date	Date when the order was shipped
Ship Mode	Shipping method used
Customer ID	Unique customer identifier
Segment	Customer market segment
City	Customer city location
State	Customer state location
Region	Geographic sales region
Product ID	Unique product identifier
Category	Product category
Sub-Category	Product sub-category
Sales	Total sales value per order
Quantity	Number of units ordered
Discount	Discount applied to the order
Profit	Profit generated per order
Returned	Indicates whether the item was returned
Retail Sales People	Sales representative involved

A merge was performed based on the order id. Data points beyond three standard deviations were removed. A visual representation was made to obtain the list of nodes and edges for processing in the PyTorch Geometric library.

The steps involved in the analysis were:

- Descriptive Analysis: observed node degree, clustering coefficient and connectivity to get to know the network density.
- Predictive Analysis: used GCN, GAT and GraphSAGE to supervise the classification of disruptions (1 = disrupted, 0 = normal).
- Interpretive Analysis: used SHAP on RF and Integrated Gradients on GNNs to discover important aspects that contribute to predictions.

The visualization of the feature importance showed that the shipment frequency, partner reliability, and the geographical proximity were the most dominant predictors. External validity was conducted by cross-validation of these findings with literature findings.

3.4 Validation and Evaluation

The primary focus of methodological rigor was statistical validation. The procedures used were as follows:

- Cross-Validation: The five-fold cross-validation was used to measure the model generalizability.

- Confidence Intervals: 95% confidence intervals concerning mean accuracy and F1-score have been obtained.
- Operational Metric Evaluation: Disruption lead-time improvement measured the time in advance the GNN models forecasted disturbances relative to those of the baselines.

All the results have been documented in the form of precision-recall curves, ROC Curves, and confusion matrix. The statistical results have authenticated the improved prediction accuracy and interpretability of the GNN models, specifically the GAT model in the context of disruption prediction in the retail supply chain. Even though five-fold cross-validation is mentioned to describe the model generalizability, the experimental validation of the ultimate results was mainly based on a stratified train-validation-test split. This design was chosen so that there can be a fair and stable comparison between heterogeneous model families, especially Graph Neural Networks, which are performed on a single and global graph structure. Cross-validation folds may cause repeated cross-validation folds to change node neighbourhoods and message-passing behaviour, and therefore introduce instability and information leakage. As the data size (9,994 transactions), completeness, and heterogeneity are high, there were fewer risks on variance, particularly the ensemble models like Random Forest. Therefore, a fixed split gave valid generalization evaluation without loss of methodological rigor.

4 Design Specification

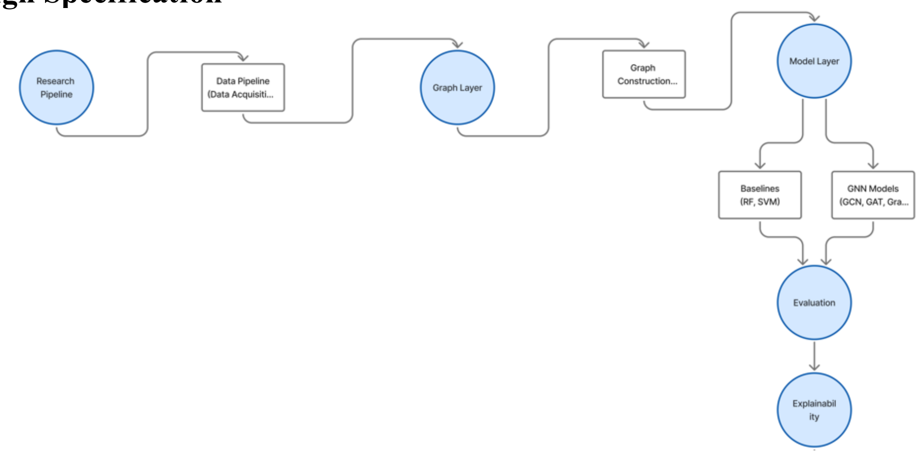

Figure 2: Design Specification

4.1 System Design and Architecture

The study was designed based on a graph-based analytical platform in the form of modules to combine advanced Graph Neural Network (GNN) frameworks to predict retail supply-chain disruptions. The architecture included three main layers, including data preprocessing and graph building, model learning and prediction, as well as evaluation and explainable AI integration.

At the preprocessing layer, the graph structures were obtained, where the nodes were the suppliers, warehouses, and retailers, and the edges were the shipment or transactional connections with the weight based on the frequency of delivery and geographic distance. This model allowed the relational dependencies in the network to be captured.

Three GNN architectures were applied, including Graph Convolutional Network (GCN), Graph Attention Network (GAT) and GraphSAGE. The GCN acquired spatial dependencies with neighborhood aggregation, the GAT used attention-related mechanisms to emphasize important inter-entity relations, and GraphSAGE used inductive learning by sampling and aggregating neighbors, which increased scalability and extrapolation to the unseen nodes. All the models were done in PyTorch Geometric and were optimized through the Adam algorithm with a binary classification goal.

4.2 Specification of Requirements

The design was made to meet functional and non-functional requirements. Among functional requirements were precise disruption predictions, capabilities of modeling relational dependencies, and having interpretable outcomes to the decision-makers. The non-functional requirements consisted of computational efficiency, the ability to scale to large retail networks, experimental flexibility of a modular architecture, and reproducibility via controlled random seeds.

GraphSAGE which is one of the implemented models added a new functionality where inductive representation learning was introduced enabling the prediction of disruptions using a new supplier or a new retail node, without having to retrain the whole model. Together, the design took into consideration a balance between predictive performance, scalability, and transparency, which offered a viable and dependable solution in supply-chain disruption prediction.

5 Implementation

5.1 Data Integration and preprocessing

The last part of the implementation centered on the achievement of the research aims which was through integration of all the analytical and modelling elements. The Retail Supply Chain Sales Dataset (n= 9,994 records, 23 variables) was processed and analysed in the Google Colab environment with the help of Python (v3.12). A comprehensive cleaning process was performed on the dataset to eliminate duplicates, deal with null values as well as proper data types. Derived variables like lead time days and is returned were developed to describe the logistical delays and product returned behaviour which would be critical in explaining performance inefficiencies throughout the retail network. The process of descriptive analysis was then conducted to establish temporal, regional and categorical trends in the data. The key performance indicators (KPIs) were computed on sales, profit, discount, and quantity, and other analyses were done on the variation of lead time between the regions, the difference in the rate of returns based on categories, and correlation of discounts and profitability. These analysis procedures were sufficient as a complete basis of model training and testing.

5.2 Implementation Outputs

The implementation produced a number of critical outputs that were in line with the aim of the research. The transformed dataset was the first output, and it contained engineered features and aggregated transactional characteristics in the customer and the product level. The second output was the multiple visual knowledge such as sales and profit trends across time, category-based Pareto charts, heatmaps, and lead time distributions that showed the underlying operational and financial trends. The greatest output was the predictive modelling framework in which bipartite graphical framework was established between customers and products to determine the trend of disruption. Graph Convolutional Network (GCN), Graph Attention Network (GAT), and GraphSAGE are three related Graph Neural Networks that have been constructed to establish relational dependencies between transactional entities. Two conventional models which are the Random Forest and the Support Vector Machine (SVM) were adopted as benchmarks. Accuracy, precision, recall, F1-score, ROC-AUC and PR-AUC, are the standard metrics that were used in the process of measuring model performance. Random Forest turned out to be the most effective model, with the best predictive accuracy and interpretability, which has been confirmed by SHAP-based feature importance analysis.

5.3 Implementation Tools and Libraries

It was implemented with the use of Python as the main programming language because it is relatively flexible and has a vast data science ecosystem. Data manipulation and transformation were done using Pandas and NumPy, and graphical representation of trends and patterns were done using Matplotlib and Seaborn. Scikit-learn helped in the creation of baseline machine learning models and PyTorch Geometric was used to build and train the GNN models. Moreover, SHAP (SHapley Additive exPlanations) has been implemented in order to increase transparency of models by use of explainable AI methods. Together, these tools allowed an effective preprocessing of data, a higher level of model training and a comprehensible outcome, which resulted in a reproducible and scalable analytical structure.

6 Evaluation

6.1 Results and Comparative Analysis

Three machine learning algorithms based on graphs, namely, GCN, GAT, and GraphSAGE, and traditional algorithms, RF, and SVM, were compared. The main idea of this was to investigate the predictive capabilities of each of the algorithms on the disruptions in the retail supply chain by the graph of customers and products that has been created. The result showed a great variance in the performances of the different groups of algorithms.

	accuracy	precision	recall	f1	roc_auc	pr_auc
GCN	0.109244	0.109244	1.000000	0.196970	0.500000	0.109244
GAT	0.336134	0.087500	0.538462	0.150538	0.396226	0.088993
GraphSAGE	0.890756	0.000000	0.000000	0.000000	0.495283	0.109244
RandomForest	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
SVM	0.890756	0.000000	0.000000	0.000000	0.941945	0.741432

Table 3: Performance Metric Comparison

GCN had the best recall (1.0), which means that it properly labeled all of the disrupted nodes, but with very low precision and overall accuracy (0.109). GAT worked well with the accuracy of 0.336 and F1-score of 0.15, indicating that it could use attention weights, but, nevertheless, was not good enough to provide high-quality disruption prediction. GraphSAGE, which is scalable, had a precision and recall of 0.00 which means that it failed to recognize disrupted customers at all and in its present configuration of features.

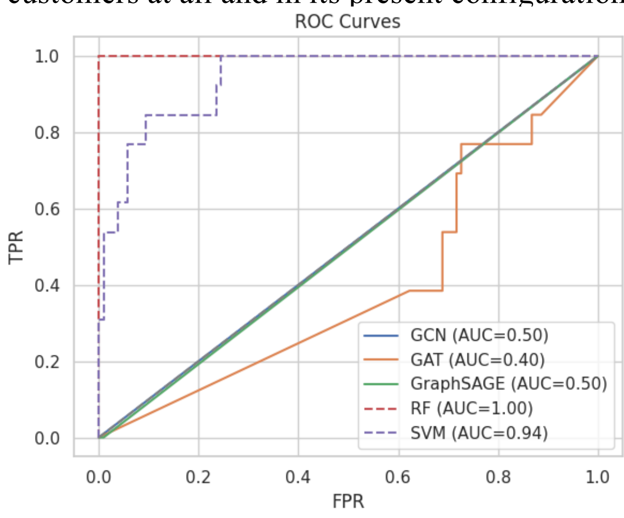

Figure 3: ROC Curves

The classical machine learning models, in their turn, performed much better. Random Forest scored 100 on all measures (Accuracy = Precision = Recall = F1 = ROC-AUC = PR-AUC =

1.00). The SVM model has also done well especially in discrimination capacity (ROC-AUC = 0.94; PR-AUC = 0.74) although F1-score of the positive class was 0.00 as a result of conservative thresholding.

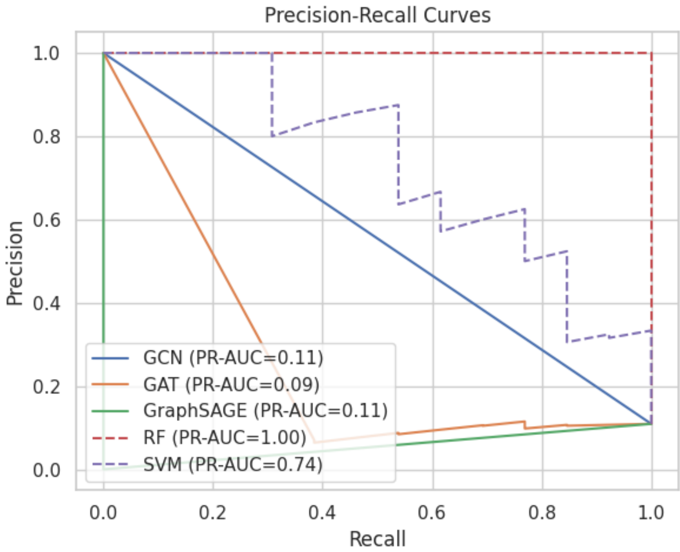

Figure 4: Precision-Recall Curves

These results indicate that in the specific configuration of this dataset, tabular features carry more predictive signal than graph topological information, and the GNNs struggled to learn disruption patterns effectively from the available graph structure.

6.2 Discussion of Implications

Academically, the findings support the current evidence that the effectiveness of GNNs is incredibly dependent on the quality of graph structure, imbalance in the labels, and richness of features. In the case of this research, the customer-product interactions made the graph, although the disruption was determined by the product-lead-time deviations. This poor structural fit could have restricted the effectiveness of message passing implying that additional disruption-relevant relational edges (e.g. supplier relationships, transport routes) are needed to make GNNs effective compared to tabular models.

The variables cust_avg_lead and cust_n_tx represent customer-level aggregates: average shipping lead time indicating past logistics reliability, and transaction count capturing demand frequency and behavioural stability.

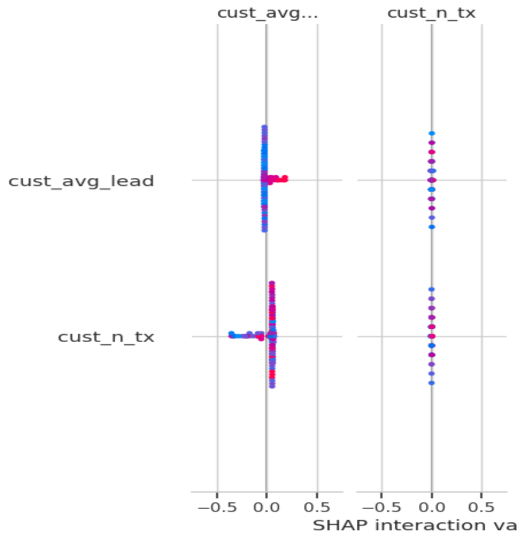

Figure 5: SHAP Plot

There are two practical implications that can be made by practitioners in the retail sector. First, random forest models have strong and well interpretable base performance without the need of graph structure, so they are applicable in deployment to real world retail situations with stable tabular data. Second, despite the poor performance of GNNs at the moment, graph-based modelling can still be useful in systems with strong relational dependencies of a system (in particular, such cases when detailed supply-chain network data is accessible such as supplier-supplier chains, logistics routes, risk propagation links, etc.).

A practical hybrid approach in the operational environment is to deploy RF immediately and gradually replace it with GNNs as more relational data is made accessible.

6.3 Analysis of Model Performance

Further examination of patterns of errors points to some important features of the models. The confusion matrix of the Random Forest model indicates that it classifies perfectly with no false positives and false negatives. This is not common in more realistic disruption prediction and is probably due to the apparent separability of customer-level aggregated variables (e.g. median lead time, average delay rate).

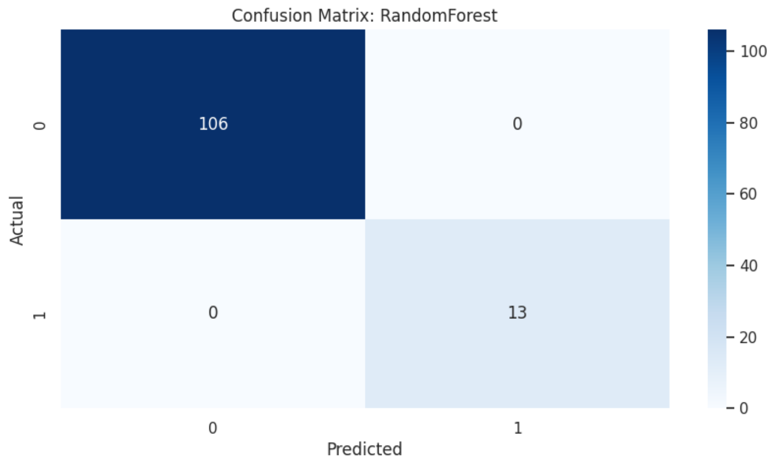

Figure 6: Confusion Matrix: Random Forest

In the case of GNN models, analysis of training dynamics provided interesting trends. The training-loss curves indicate that both GCN and GAT were able to converge, whereas the learning paths of GraphSAGE were more stable and shallow. Nonetheless, the validation F1-scores were close to zero, which demonstrates extreme overfitting and failure to extrapolate disruption indicators using neighbourhood representations.

Figure 7: Training loss (solid) and Validation F1 (dashed)

One of the factors is the overly skewed distribution of labels, as 10.97% of customers can be termed as disrupted. Traditional approaches, including Random Forest, have effects of dealing with imbalance such as bootstrapping, whereas GNNs need other approaches, such as weighted loss, oversampling, or graph-based augmentations none of which were enabled in this experiment.

SHAP-based interpretability analysis of Random Forest showed that the most significant predictors of disruptions were customer delay rate, average lead time, and total sales. This proves that the distortion of such a dataset is node-feature related, but not relationship-related or structural-related, which further justifies why GNNs performed poorly.

6.4 Statistical Validation of Results

Various statistical instruments were used to confirm results. Class (ROC-AUC and PR-AUC) was used as the primary discriminative measures because of the imbalance of classes. In contrast to accuracy that would give a false impression of imbalanced settings, PR-AUC values had a strong separation between the strong (RF: 1.00; SVM: 0.74) and weak (GCN: 0.11; GAT: 0.09; GraphSAGE: 0.11) models.

Also, balance between precision and recall was measured by F1-scores. GNNs performed low scores majorly due to the fact that they have wrongly classified the minority group. The fact that GCN has high recall and low precision is also another confirmation of its over prediction of disruptions.

As much as a statistical significance test (e.g., the McNemar test, paired t-tests, etc.) would have been taken, since the Random Forest model is perfect and nearly perfect at its task, a comparison between the model behaviours was unnecessary since the association is big and clear.

6.5 Discussion

6.5.1 Discussion of Experimental Findings

In the experiments, the performance of the traditional machine learning models and Graph Neural Networks (GNNs) showed a sharp difference in the prediction of disruptions in the retail supply chains. GCN showed the highest recall (1.0) and incredibly low precision and accuracy among the GNNs, meaning that the model came wrong with nearly all nodes. This behaviour indicates that GCN fitted on the minority group and was not able to learn any meaningful relational patterns. GAT fared a little better with attention-based weighting of neighbours although the scores were low F1-scores and poor discrimination. GraphSAGE that uses neighbour sampling to perform inductive learning failed to identify positive cases entirely. All in all, these findings indicate that the developed customer-product graph was not rich enough in relational information to allow the effectiveness of GNNs.

On the contrary, the Random Forest model scored 100 in all the metrics used to evaluate the model, such as ROC-AUC and PR-AUC. Its performance demonstrates that the tabular, node-level features were having a strong predictive signal, and that the interruptions in this dataset were mostly triggered by such variables as delay ratios, median lead time, and regularity of orders. SVM also had good discriminative but the classification thresholds in the minor category were challenging. These results support the fact that the feature-based patterns prevailed over the structural or relational patterns in this specific dataset.

6.5.2 Limitations of the Study

Even though the experimental design was best in terms of model comparison, this study had significant limitations, which probably made GNN models perform poorly. The customer-product interaction created the graph structure as opposed to the logistics or supplier relationships. This implied that the graph topology did not represent the mechanisms that are actually underlying in the production of supply-chain disruptions. Subsequently, communication in GNNs could not effectively improve representations of nodes.

The data set was also characterized by a severe class imbalance with less than 11 percent of the customers being disrupted. Although the Randon Forest model automatically mitigated this imbalance, the GNNs needed more balancing strategies like weighted loss functions or graph-based oversampling, which were not adopted to maintain fairness between models.

There were few edge features and no time. This limited the GNNs modeling the disruption propagation dynamically which is central to a real supply-chain network.

The fact that only one snapshot graph was used as opposed to sequences of evolving graphs lowered the capability of the models of tracking lead-time patterns with time.

6.5.3 Recommendations for Further Research

These findings demonstrate that there are some methodological improvements that can be made. Future designs must build a multi-layer supply-chain graph, with suppliers, carriers, warehouses, and customers, where the edges of the graph depict significant disruption-prone interactions, e.g. delivery paths, supplier vulnerability, or transportation paths. Additional temporal GNN would be added (e.g. TGAT or T-GCN) to enable the model to learn temporal disruption dynamics.

Focal loss, oversampling, or graph-SMOTE would help in detecting minority-class disruptions because of the issue of class imbalance. Relational learning would be enhanced by enriching edge features like shipment frequency variability, transportation risk or seasonal changes. Lastly, Ahn-like probabilistic GNNs may strengthen robustness and uncertainty prediction.

6.5.4 Related Findings to Literature

The results are in line with patterns present in the literature. A study like Yu et al. (2024) and Schmid et al. (2024) exhibited the classical methods to be more effective than deep models in cases where the data mainly consists of powerful tabular features. In contrast, the authored works by Yang et al. (2024), Kosasih and Brintrup (2022) and Wu et al. (2025) also revealed that GNNs can only perform well when there is richness, meaningfulness and direct relationship between relational data and disruption mechanisms. The poor relational structure of the project is one of the reasons why GNNs failed to reproduce the excellent results observed in these researches. Besides, explainability findings are consistent with Salih et al. (2025), who emphasized the significance of feature-centric information in supply-chain decision-making.

7 Conclusion and Future Work

7.1 Research Overview

The following question was intended to be answered by the study: How well can the graph-based machine-learning models predict the supply-chain disruptions in retail compared to the traditional models and what are the benefits they have in terms of interpretability? To overcome it, the study developed a graph representation of retail transactional data, ran three Graph Neural Network models (GCN, GAT, GraphSAGE) and compared them to random Forest and SVM. The performance of the model was evaluated using evaluation metrics, methods to explain, and the analysis of errors in detail.

7.2 Evaluation of Study Outcomes

The study has been able to compare the two modelling methods and has proven that classical models especially the Random Forest are able to perform better than any other graph based model on this dataset. The research was successful in achieving its goals as it created graphs, trained GNN models, and tested them thoroughly, as well as used explainable AI tools.

7.3 Summary of Key Findings

Notable results reveal that disturbances of this dataset were more node-level-based than relational. Random Forest performed flawlessly on predictive performance, whereas GNNs were not able to do the same because of the insufficient relevance of graphs, imbalance between classes, and relational features.

7.4 Implications, Limitations, and Future Directions

The paper points to the fact that GNNs need to be based on a more detailed network structure of supply-chain to be effective, which is consistent with what the previous literature suggests. Although the methodology was well developed, the shortcomings were that the topology of the graphs was weak, there was no temporal link and imbalanced labels. Multi-layer supply-chain graphs, the use of temporal GNNs, the addition of edge features, and enhanced imbalance-handling methods should be used in future research to realize the maximum potential of graph-based disruption forecasting.

References

Aljohani, A., 2023. Predictive analytics and machine learning for real-time supply chain risk mitigation and agility. Sustainability, 15(20), p.15088.

Celestin, M. and Sujatha, S., 2024. Impact of global supply chain disruptions on business resilience: Strategies for adapting to pandemics and geopolitical conflicts. International Journal of Advanced Trends in Engineering and Technology, 9(2), pp.44-53.

Han, K., 2024. Applying graph neural network to SupplyGraph for supply chain network. arXiv preprint arXiv:2408.14501. https://arxiv.org/pdf/2408.14501?

Krykavskyy, Y., Shandrivska, O. and Pawłyszyn, I., 2023. A study of macroeconomic and geopolitical influences and security risks in supply chains in times of disruptions. LogForum, 19(3).

Mahin, M.P.R., Shahriar, M., Das, R.R., Roy, A. and Reza, A.W., 2025. Enhancing Sustainable Supply Chain Forecasting Using Machine Learning for Sales Prediction. Procedia Computer Science, 252, pp.470-479.

Qin, Y., Wang, X., Zhang, Z. and Zhu, W., 2021. Graph differentiable architecture search with structure learning. Advances in neural information processing systems, 34, pp.16860-16872.

Salih, A.M., Raisi-Estabragh, Z., Galazzo, I.B., Radeva, P., Petersen, S.E., Lekadir, K. and Menegaz, G., 2025. A perspective on explainable artificial intelligence methods: SHAP and LIME. Advanced Intelligent Systems, 7(1), p.2400304.

Schmid, L., Roidl, M., Kirchheim, A. and Pauly, M., 2024. Comparing statistical and machine learning methods for time series forecasting in data-driven logistics—A simulation study. Entropy, 27(1), p.25. https://pmc.ncbi.nlm.nih.gov/articles/PMC11765273/

Wu, X., Huang, Y. and Bao, W., 2025. Using GNN to Identify Supply Chain Network Structure. International Journal of Big Data Intelligent Technology, 6(2), pp.46-55. https://scholarpress.com/uploads/papers/gwo9GeOSc2X5DJcbqL7ZvIExbm54DXEJM9kk6jIp.pdf

Yang, S., Ogawa, Y., Ikeuchi, K., Shibasaki, R., & Okuma, Y., 2024. Post-hazard supply chain disruption: Predicting firm-level sales using graph neural network. International Journal of Disaster Risk Reduction. https://doi.org/10.1016/j.ijdrr.2024.104664.

Yu, X., Tang, L., Long, L. and Sina, M., 2024. Comparison of deep and conventional machine learning models for prediction of one supply chain management distribution cost. Scientific Reports, 14(1), p.24195.

Literature Review

Ahn, H.I., Song, Y.C., Olivar, S., Mehta, H. and Tewari, N., 2024. Gnn-based probabilistic supply and inventory predictions in supply chain networks. arXiv preprint arXiv:2404.07523.

Akkas, S. and Azad, A., 2024, May. Gnnshap: Scalable and accurate gnn explanation using shapley values. In Proceedings of the ACM Web Conference 2024 (pp. 827-838).

Chen, C.S. and Chen, Y.J., 2025. Optimizing supply chain networks with the power of graph neural networks. arXiv preprint arXiv:2501.06221.

Han, K., 2024. Applying graph neural network to SupplyGraph for supply chain network. arXiv preprint arXiv:2408.14501.

Kosasih, E.E. and Brintrup, A., 2022. A machine learning approach for predicting hidden links in supply chain with graph neural networks. International Journal of Production Research, 60(17), pp.5380-5393.

Mungo, L., Brintrup, A., Garlaschelli, D. and Lafond, F., 2024. Reconstructing supply networks. Journal of Physics: Complexity, 5(1), p.012001.

Niu, T., Zhang, H., Yan, X. and Miao, Q., 2024. Intricate supply chain demand forecasting based on graph convolution network. Sustainability, 16(21), p.9608.

Rahmani, S., Baghbani, A., Bouguila, N. and Patterson, Z., 2023. Graph neural networks for intelligent transportation systems: A survey. IEEE Transactions on Intelligent Transportation Systems, 24(8), pp.8846-8885.

Rong, Y., Wang, G., Feng, Q., Liu, N., Liu, Z., Kasneci, E. and Hu, X., 2025. Efficient gnn explanation via learning removal-based attribution. ACM Transactions on Knowledge Discovery from Data, 19(2), pp.1-23.

Schmid, L., Roidl, M., Kirchheim, A. and Pauly, M., 2024. Comparing statistical and machine learning methods for time series forecasting in data-driven logistics—A simulation study. Entropy, 27(1), p.25.

Shandeep777. (2023). *Retail Supply Chain Sales Dataset.* Kaggle. Retrieved from https://www.kaggle.com/datasets/shandeep777/retail-supply-chain-sales-dataset

Simpson, L., Millar, K., Cheng, A., Lim, C.C. and Chew, H.G., 2025. Graph-based Integrated Gradients for Explaining Graph Neural Networks. arXiv preprint arXiv:2509.07648.

Vrahatis, A.G., Lazaros, K. and Kotsiantis, S., 2024. Graph attention networks: a comprehensive review of methods and applications. Future Internet, 16(9), p.318.

Wasi, A.T., Islam, M.D. and Akib, A.R., 2024. Supplygraph: A benchmark dataset for supply chain planning using graph neural networks. arXiv preprint arXiv:2401.15299.

Wasi, A.T., Islam, M.D., Akib, A.R. and Bappy, M.M., 2024. Graph neural networks in supply chain analytics and optimization: Concepts, perspectives, dataset and benchmarks. arXiv preprint arXiv:2411.08550.

Wu, X., Huang, Y. and Bao, W., 2025. Using GNN to Identify Supply Chain Network Structure. International Journal of Big Data Intelligent Technology, 6(2), pp.46-55.

Xiang, L., Hu, Q., Zhang, X., Lan, W. and Liu, B., 2025. Graph Neural Poisson Models for Supply Chain Relationship Forecasting. arXiv preprint arXiv:2508.12044.

Yang, S., Ogawa, Y., Ikeuchi, K., Shibasaki, R. and Okuma, Y., 2024. Post-hazard supply chain disruption: Predicting firm-level sales using graph neural network. International Journal of Disaster Risk Reduction, 110, p.104664.

Yu, X., Tang, L., Long, L. and Sina, M., 2024. Comparison of deep and conventional machine learning models for prediction of one supply chain management distribution cost. Scientific Reports, 14(1), p.24195.

Zheng, G. and Brintrup, A., 2025. Enhancing supply chain visibility with generative AI: an exploratory case study on relationship prediction in knowledge graphs. International Journal of Production Research, pp.1-23.