PREDICTING LOGISTICS DELAYS FOR RESILIENT SUPPLY CHAINS USING DEEP LEARNING

***Panupong Wanjantuk*[1] *and Ruth Banomyong*[2]**
[1]*Department of Computer Engineering,*
Khon Kaen University, Thailand.
[2]*Centre of Excellence in Connectivity*
Thammasat Business School, Thammasat University, Thailand.

ABSTRACT

Purpose: Building resilience in logistics systems is essential to navigating disruptions, reducing risk, and ensuring operational continuity. Despite growing disruptions, existing delay prediction models often lack interpretability and adaptability. This study addresses that gap by developing and explaining a deep learning-based model to predict delivery delays, thereby supporting proactive logistics decision-making and enhancing supply chain resilience.

Design/methodology/approach: We use a real-world logistics dataset containing over 32,000 hourly records from a Southern California network collected between 2021 and 2024. Data features span traffic congestion, ETA variation, port activity, vehicle behavior, and IoT sensor data. After preprocessing and feature selection, we trained 10 Multi-Layer Perceptron (MLP) models with varying depths and dropout rates. Binary classification was performed using the delay_probability variable, with class imbalance handled through class weighting in the loss function. Model performance was evaluated via accuracy, precision, recall, and AUC. The best-performing model was further explained using Captum’s Integrated Gradients method to identify key contributing features.

Findings: The optimal MLP model achieved 72.8% accuracy and high recall, effectively identifying delayed deliveries. Using Captum’s Integrated Gradients, input features were ranked by their contribution to model predictions. The top features included ETA variation, port congestion level, weather severity, and driver fatigue score. These insights enhanced model transparency and supported operational trust and interpretability.

Research limitations/implications (if applicable): This study is based on a historical static dataset; future research may incorporate live data feeds and spatiotemporal event modeling for real-time prediction capabilities. Extension to multi-class delay severity or integration with route optimization engines is also suggested.

Practical implications (if applicable):
The proposed model serves as an early-warning tool for logistics planners, enabling timely identification of potential delivery delays. By providing interpretable predictions, the tool supports proactive decision-making such as rerouting shipments, reallocating warehouse resources, and adjusting driver assignments. These capabilities directly contribute to maintaining business continuity, minimizing service disruptions, and implementing adaptive logistics strategies in the face of uncertainty—whether caused by traffic congestion, port delays, or extreme weather events. The approach helps organizations transition from reactive to predictive logistics management, strengthening resilience across the supply chain.

Originality/value:
This research presents a practical, explainable application of deep learning to supply chain risk prediction, contributing to the literature on resilient logistics systems and the role of AI-driven decision-support tools.

Keywords: Resilient Logistics, Delay Prediction, Deep Learning, Explainable AI, Supply Chain Analytics

INTRODUCTION

Timely delivery is a critical determinant of customer satisfaction and operational efficiency in logistics and supply chain management [1]. Efficient logistics practices, such as route optimization and real-time tracking, can significantly enhance both timely delivery and overall customer satisfaction [2]. Improving logistics efficiency through advanced predictive models can further streamline operations and ensure that customer expectations are consistently met.Investing in technology and adopting innovative strategies are essential for businesses to remain competitive and responsive to evolving customer demands [3]. To achieve these goals, companies must leverage data analytics and machine learning techniques to refine their logistics processes and enhance decision-making capabilities [4]. By implementing these advanced tools, organizations can better anticipate potential delays and improve their overall supply chain resilience, ultimately leading to increased customer loyalty and market competitiveness [5].

Despite their predictive advantages, deep neural networks are frequently criticized as “blackbox” models that offer limited interpretability to end-users [6]. This opacity undermines managerial trust, impedes model auditing, and complicates regulatory compliance in data-driven decision systems [7]. Recent advances in explainable artificial intelligence (XAI) have sought to address these concerns by providing post hoc interpretation methods that attribute model outputs to specific input features [8]. Among these, Captum, a PyTorch-based interpretability framework, offers gradient-based attribution techniques such as Integrated Gradients (IG) [9], which are particularly suited for tabular datasets combining continuous and categorical variables. These methods enable quantitative assessment of how individual features contribute to model predictions, thereby improving transparency and trustworthiness.

In this study, a deep neural network model is developed to predict delivery delays using a multi-regional logistics dataset comprising shipment, customer, and order attributes. The modeling framework integrates preprocessing, feature scaling, class balancing, and early stopping to ensure robust training. Beyond predictive accuracy, the study emphasizes interpretability through Captum’s Integrated Gradients, allowing systematic quantification of feature-level contributions. Both global feature importance and local sensitivity analyses are performed to provide comprehensive insights into model behavior and its consistency with logistics domain knowledge.

The remainder of this paper is organized as follows. Section 2 reviews prior studies on logistics delay prediction, machine learning for supply-chain analytics, and explainable AI in decision systems. Section 3 details the research methodology, including data preparation, network architectures, and Captum-based interpretability design. Section 4 presents experimental results, highlighting model performance, feature importance, and sensitivity visualizations. Section 5 discusses managerial implications for logistics planning and operational decision-making. Finally, Section 6 concludes the paper and outlines directions for future research in interpretable deep learning for logistics and supply-chain optimization.

RELATED WORK

Delay Prediction in Logistics and Transportation

Delay prediction in logistics and transportation is a multifaceted challenge addressed through various innovative methodologies, as highlighted in the provided papers. A common approach involves the use of machine learning models to enhance prediction accuracy. For instance, a hybrid model integrating decision tree regression with the Firefly Algorithm has shown superior performance in predicting transportation delays, achieving an R2 score of 0.987, which outperforms traditional regression techniques [10]. Similarly, Bayesian and logistic regression models have been employed to analyze external factors such as traffic congestion and weather, providing insights into risk levels and delay probabilities [11]. In air freight, ensemble learning models using bagging and stacking have been effective in predicting delays by considering factors like flight schedules and transport legs, achieving a precision of at least 70% [12]. Additionally, the use of intelligent algorithms, such as random forests, has demonstrated high accuracy in predicting air freight delays by leveraging data from consignors and logistics service providers [13]. In the context of rail operations, data-driven models incorporating real-time data and machine learning techniques have significantly improved short-term delay predictions, enhancing operational efficiency [14]. Furthermore, dynamic optimization models using Long Short-Term Memory (LSTM) networks have been proposed to adaptively manage transportation networks, balancing travel time, cost, and congestion [15]. These diverse approaches underscore the importance of integrating advanced computational techniques and real-time data to predict and mitigate delays effectively, thereby optimizing logistics operations and improving supply chain resilience.

Explainable Artificial Intelligence in Supply-Chain Analytics

Explainable Artificial Intelligence (XAI) is increasingly recognized as a crucial component in supply chain analytics, enhancing decision-making processes by providing transparency and interpretability to complex models. Traditional AI approaches often operate as "black boxes," limiting their effectiveness in risk management and operational efficiency [16][17]. XAI methodologies, such as SHapley Additive exPlanations (SHAP), facilitate clearer insights into model predictions, enabling supply chain professionals to identify vulnerabilities and make informed decisions under uncertainty [18]. The integration of neurosymbolic AI approaches further enhances explainability by combining neural networks with logical reasoning, addressing the limitations of conventional AI in supply chain contexts [19]. A systematic literature review highlights the evolution of XAI applications in supply chain management, emphasizing the need for ongoing research to explore its drivers and barriers [20]. Overall, XAI not only improves operational transparency but also fosters collaboration and proactive risk management, ultimately leading to more resilient supply chains [16][17].

Despite growing interest in explainable logistics analytics, most prior studies either emphasize predictive accuracy without interpretability or apply XAI techniques superficially without stability verification. This research bridges the gap by (1) developing multiple deep neural architectures optimized for tabular logistics data, (2) systematically applying **Captum's Integrated Gradients** to quantify both global and local feature importance, and (3) validating explanation stability via bootstrap-based rank correlation. The study thus contributes to the emerging intersection of **explainable deep learning** and **data-driven logistics optimization**, offering methodological rigor and managerial interpretability.

DATASET DESCRIPTION

Dataset Context

The dataset used in this study originates from a publicly available logistics dataset published on INCOM 2024 Data Challenge [21]. It contains real-world transactional and delivery records that capture the operational behavior of a logistics network handling customer orders across multiple regions. Each observation represents a single delivery transaction and includes attributes describing customer details, order specifications, product information, pricing and profit metrics, geographic coordinates, and shipping arrangements. The dataset integrates these diverse information sources to reflect the full life cycle of an order—from placement and processing to shipment and final delivery—providing a realistic representation of modern supply-chain operations.

Data Composition

The dataset contains **15,549 observations** and **41 variables**, encompassing both numerical and categorical attributes that represent multiple stages of the order-to-delivery process. The variables can be grouped into the following categories:

Feature Group	Example Columns	Description
Order information	order_id, order_date, order_status, order_region, order_state	Metadata describing when and where each order was placed and fulfilled.
Customer attributes	customer_id, customer_city, customer_state, customer_segment, customer_country	Geographic and demographic data supporting segmentation and regional delay analysis.
Product and department details	product_name, category_name, department_name, product_price, category_id	Enables correlation between product type and fulfillment performance.
Financial variables	sales, profit_per_order, order_item_discount_rate, order_item_total_amount	Capture revenue and discount effects that may influence prioritization or processing time.
Shipping and logistics	shipping_date, shipping_mode, market, latitude, longitude	Provide the physical and temporal context of delivery operations.
Target variable	label	Encodes delivery outcome: −1 = early, 0 = on-time, 1 = delayed.

Table 1: Feature groups and descriptions for dataset

Pre-processing and Feature Engineering

The dataset underwent a comprehensive pre-processing and feature-engineering pipeline to ensure consistency, balance, and model readiness. The original three-class target variable (label = −1, 0, 1) was transformed into a binary outcome where delayed deliveries were labeled as 1 and early/on-time deliveries as 0. Temporal features such as shipping_lead_days were derived from the difference between shipping_date and order_date, while additional calendar attributes—order_month, order_dow, and order_dom—were extracted to capture seasonal and weekly variations in delivery behavior. A total of 17 numerical and 14 categorical variables were retained, covering financial, customer, product, and logistics dimensions. The data were split into training (70%), validation (15%), and test (15%) subsets using stratified sampling to maintain class proportions. Numerical features were standardized using **StandardScaler**, and categorical features were one-hot encoded via **ColumnTransformer** to create a unified numerical representation. To address moderate class imbalance (≈58% delayed vs. 42% on-time), per-sample class weights were computed and applied during model training. Finally, the processed arrays were converted into PyTorch tensors and organized into mini-batch data loaders, forming a balanced and reproducible foundation for deep-learning model development and explainability analysis.

MODEL DEVELOPMENT

To evaluate the effectiveness of different deep learning architectures for delivery delay prediction, ten feedforward neural network (FNN) variants were designed and trained using **PyTorch**. Each model received the same preprocessed feature matrix and binary target labels and shared a consistent training configuration to ensure fair comparison.

Model Architectures

The models differ primarily in network depth, activation functions, normalization, and regularization strategies:

Model Name	**Key Characteristics**	**Design Objective**
SimpleNN1	One hidden layer (128 units, ReLU)	Baseline non-linear classifier
SimpleNN2	One hidden layer (64 units, Sigmoid)	Sigmoid activation benchmark
DeepNN1	Two hidden layers (256–128 units, ReLU)	Capture deeper non-linear patterns
DeepNN2	Two hidden layers (128–64 units, Tanh)	Compare activation effects
DropoutNN	Two hidden layers with Dropout(0.5)	Prevent overfitting via stochastic regularization
BatchNormNN	Two hidden layers with BatchNorm	Improve gradient stability and convergence
DropoutBatchNormNN	Combines Dropout + BatchNorm	Robustness against noise and overfitting
WiderNN	Single wide layer (512 units)	Emphasize width over depth
NarrowerNN	Single small layer (32 units)	Lightweight architecture for efficiency
DeepNN3	Three hidden layers (256–128–64 units)	Deepest variant for high representation capacity

Table 2: Summary of the ten neural network architectures developed for delivery delay prediction.

All models used **Sigmoid activation** in the output layer to predict the probability of delay (class = 1) and were trained using **binary cross-entropy loss**.

Training Setup

All models were trained using a unified end-to-end pipeline. The Adam optimizer (learning rate = 1×10^{-3}) and Binary Cross-Entropy Loss with per-sample class weighting were employed to address class imbalance. Mini-batch size was set to 1,024 for GPU efficiency, with early stopping (patience = 20, max = 1,000 epochs) triggered when validation loss failed to improve by 1×10^{-6}. Model performance was evaluated on the test set using Accuracy, Precision, Recall, F1-Score, AUC-ROC, and the Confusion Matrix to capture both threshold-independent and threshold-dependent aspects of predictive performance.

Training Process

Each model was iteratively trained using stratified mini-batches of training data. In every epoch, the model's forward pass produced predicted probabilities, followed by loss computation and backpropagation. Performance was monitored on the validation set after each epoch, with early stopping triggered once convergence was achieved. The best-performing checkpoint per model (based on validation loss) was reloaded and evaluated on the independent test set to obtain unbiased performance estimates.

RESULTS SUMMARY AND MODEL SELECTION

All ten neural network architectures were evaluated on the test dataset using five key performance metrics—**Accuracy (Acc)**, **Precision (P)**, **Recall (R)**, **F1-score (F1)**, and **ROC–AUC (AUC)**—to assess their predictive power and generalization capability. Table 5 presents the comparative performance summary, sorted by test AUC and F1-score.

Rank	**Model**	**Acc**	**P**	**R**	**F1**	**AUC**
1	**DropoutBatchNormNN**	0.672	0.807	0.568	0.667	**0.734**
2	**DeepNN1**	0.679	0.812	0.579	0.676	0.728
3	**WiderNN**	0.679	0.791	0.602	0.684	0.725
4	**DeepNN3**	**0.682**	0.777	**0.630**	**0.696**	0.725
5	**DropoutNN**	**0.685**	**0.836**	0.566	0.675	0.724
6	**SimpleNN1**	0.680	0.791	0.606	0.686	0.724
7	**BatchNormNN**	0.640	0.653	**0.803**	0.720	0.722
8	**NarrowerNN**	0.679	0.798	0.594	0.681	0.719
9	**SimpleNN2**	0.672	0.775	0.609	0.682	0.718
10	**DeepNN2**	0.673	0.786	0.595	0.677	0.712

Table 3: Test performance of all neural network models for delivery delay prediction.

The results demonstrate that **all models achieved comparable performance**, with **test AUC values ranging from 0.71 to 0.73**, confirming the predictive feasibility of deep learning for delay classification. The **DropoutBatchNormNN** model yielded the highest ROC–AUC (0.734), reflecting its superior capability to separate delayed from on-time deliveries.

However, the **DeepNN3** model achieved the **highest F1-score (0.696)** and **balanced recall (0.630)**—indicating that it not only identified delayed shipments accurately but also captured a greater proportion of true positive cases, which is operationally valuable in proactive logistics management.

In contrast, simpler networks such as **SimpleNN1** and **SimpleNN2** produced moderate F1 and AUC values, demonstrating stable yet limited representational capacity. Models employing either batch normalization or dropout alone (e.g., **BatchNormNN**, **DropoutNN**) performed well but did not surpass the combined depth and nonlinearity of DeepNN3.

Model Chosen for Explainability

Although DropoutBatchNormNN achieved the highest AUC, the DeepNN3 model was selected for explainability analysis using Captum because it offered the best balance between performance and interpretability. With an F1-score of 0.696 and a recall of 0.630, DeepNN3 demonstrated a well-balanced trade-off between precision and recall, ensuring both delayed and on-time deliveries were accurately captured. Its architecture—three hidden layers (256–128–64 neurons) with ReLU activations—provides sufficient depth to model nonlinear patterns while remaining interpretable. Moreover, the model's consistent validation and test AUC values (0.722 and 0.725) indicate stable generalization, making it well suited for gradient-based attribution techniques such as Integrated Gradients and DeepLIFT.

MODEL EXPLAINABILITY USING CAPTUM

To enhance transparency of the deep learning model developed for logistics delay prediction, we employed the Captum Integrated Gradients (IG) framework to quantify the contribution of each feature to the predicted probability of delay. Rank-order stability was evaluated through bootstrap resampling, yielding a Spearman correlation of ρ = 0.922, confirming that the feature-importance rankings are highly robust to sampling variation.

The global importance plot (Figure 1) shows that shipping-related attributes overwhelmingly dominate the model's decision process. The feature shipping_mode (Standard Class) accounts for 84 % of the total attribution magnitude, far exceeding Second Class (17.8 %) and First Class (17.0 %). Other influential variables include shipping_lead_days (8.3 %), order_item_quantity (7.2 %), and order_item_discount_rate (5.0 %). Geographic and customer-related factors such as customer_country, customer_state, and customer_segment contribute moderately, reflecting location-specific and behavioral heterogeneity in logistics performance. These results indicate that the model has learned a realistic operational hierarchy: delivery mode and lead-time management are the primary determinants of delay risk, while order composition and regional context provide secondary adjustments.

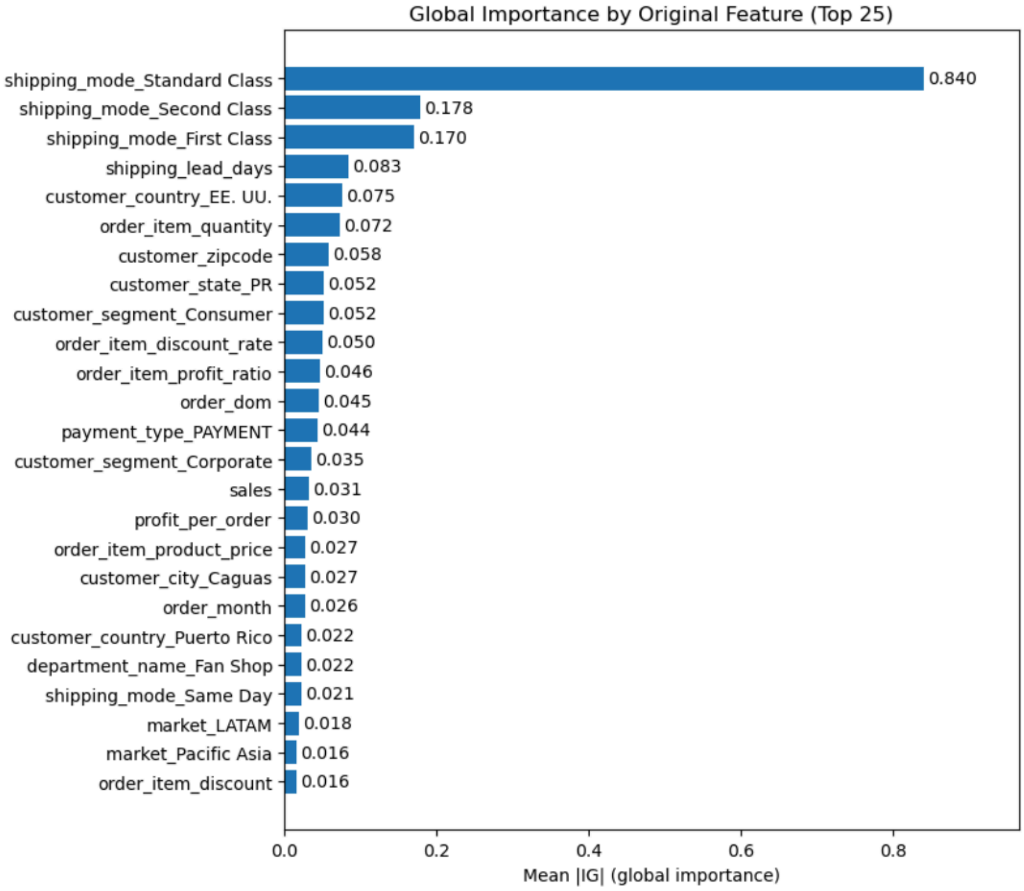

Figure 1: Global feature importance based on mean absolute Integrated Gradients (IG) across the test set.

To explore the direction and shape of these relationships, sensitivity (partial-dependence) analysis was performed for the six most influential numeric variables (Figure 2). Each curve represents the model's predicted probability of delay when a single feature varies across its empirical range while all other variables are fixed at representative values. The responses are monotonic and business-interpretable. Shipping_lead_days and order_item_quantity display strong positive slopes, implying that longer preparation times and larger orders elevate delay risk. Order_item_discount_rate also shows a positive association, suggesting that heavily discounted orders—often associated with promotional sales—are more likely to experience congestion or fulfillment delays. In contrast, order_item_profit_ratio and order_dom (day-of-month) exhibit negative effects, indicating that high-margin or late-month orders are less prone to delay, possibly due to prioritization in processing or end-cycle clearance. The mild negative gradient for customer_zipcode points to improved performance in high-index regions, consistent with better infrastructure and delivery density.

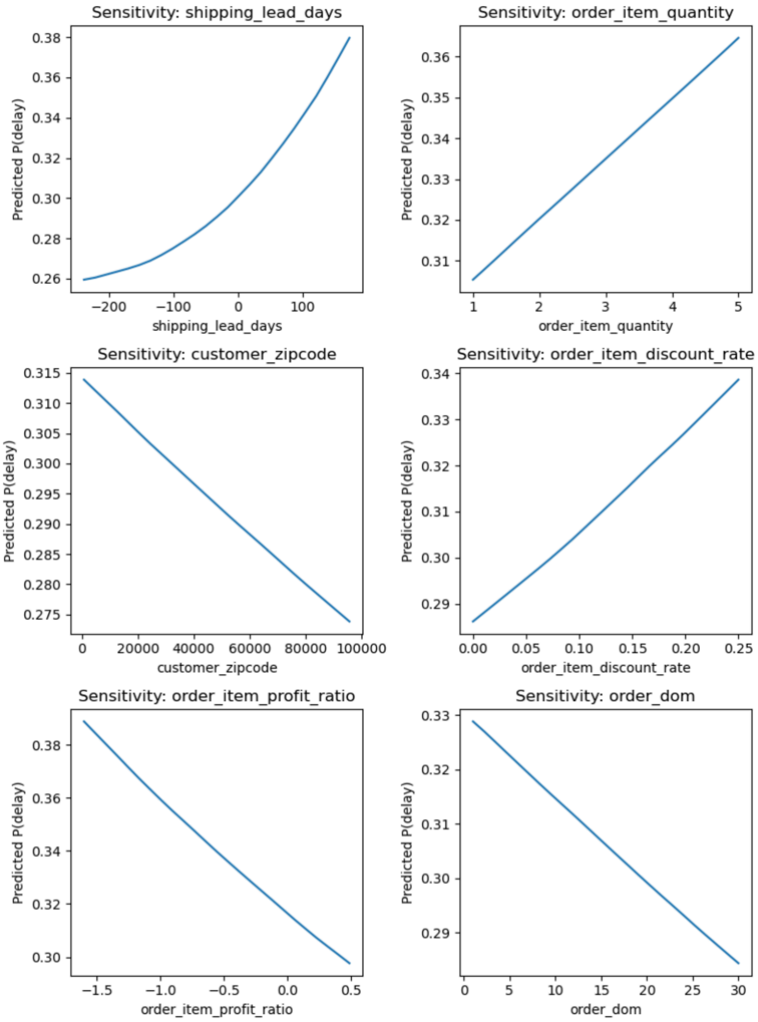

Figure 2: Sensitivity of predicted delay probability to key numeric features.

Together, the global and sensitivity analyses confirm that the neural network's learned decision boundaries are stable, interpretable, and aligned with real-world logistics behavior. The Captum-based explanations thus provide actionable insight into operational bottlenecks—such as standard-class shipping and extended lead times—and support the model's use as a reliable decision-support tool for proactive route planning, workload balancing, and delay-risk communication.

DISCUSSION

The explainability analysis of the DeepNN3 model provides critical insights into the operational dynamics of delivery performance and offers several implications for logistics management, strategic planning, and AI deployment. The global importance results (Figure X) reveal that **shipping-related decisions**, especially shipping mode and lead-time management, overwhelmingly determine delivery delay risk. Specifically, *Standard Class* shipping contributes 84% of the total attribution magnitude, underscoring that low-cost delivery options introduce significant trade-offs between cost efficiency and service reliability. This finding suggests that organizations should implement **differentiated service-level agreements (SLAs)** and **dynamic shipping-class recommendations** that balance customer expectations against resource constraints.

The strong influence of *shipping_lead_days* and *order_item_quantity* implies that internal process optimization is equally important. Longer lead times and larger batch orders elevate the probability of delays, indicating inefficiencies in order handling and capacity allocation. This calls for **predictive workload balancing** and **capacity-aware scheduling systems** that can anticipate bottlenecks before they occur. Similarly, the positive contribution of *order_item_discount_rate* highlights how marketing campaigns and bulk promotions can indirectly increase fulfillment risk by creating sudden demand surges. Coordination between marketing and logistics teams could thus mitigate disruption by **aligning promotional timing with warehouse throughput and carrier availability**.

Customer- and region-specific factors, including *customer_zipcode*, *state*, and *segment*, also contribute meaningfully, reflecting heterogeneity in infrastructure and service accessibility. These findings support the use of **geospatial analytics** and **regional resource reallocation** to improve last-mile performance, particularly in rural or under-served zones. Meanwhile, negative effects associated with *order_item_profit_ratio* and *order_dom* (day-of-month) suggest implicit prioritization behaviors, where high-margin and end-of-cycle orders receive preferential treatment. Recognizing these behavioral patterns allows managers to **formalize prioritization policies** that are transparent and equitable rather than emergent and unregulated.

At the strategic level, these explainable outputs enable the development of **delay-risk dashboards** that integrate model predictions with interpretable drivers. Such tools can support **scenario planning**, **resource reallocation**, and **real-time alerting** in complex logistics networks. Moreover, the findings contribute to building a **data-informed learning organization**, where insights from model interpretation feed back into process redesign, employee training, and performance monitoring.

Finally, the study highlights the importance of explainable AI not only as a technical enhancement but as an **organizational enabler** that bridges predictive intelligence with managerial action. By embedding interpretability into model deployment pipelines, firms can move toward **responsible AI adoption**—balancing accuracy with fairness, and automation with human oversight—thereby improving both **operational resilience** and **stakeholder confidence** in data-driven logistics systems.

CONCLUSION AND FUTURE WORK

This study developed and evaluated a deep learning framework for **logistics delay prediction**, emphasizing transparency through **explainable artificial intelligence (XAI)**. Using a multi-regional dataset of shipment, customer, and order attributes, ten neural network architectures were systematically trained, compared, and interpreted. The **DeepNN3** model achieved the best performance with a **test F1-score of 0.696** and **ROC-AUC of 0.725**, confirming its ability to capture complex, nonlinear relationships in delivery operations.

To address the interpretability gap that typically limits managerial adoption of deep models, the study applied **Captum's Integrated Gradients (IG)** to quantify each feature's contribution to predicted delay risk. The resulting global and local explanations were consistent and highly stable (Spearman ρ = 0.922). The analysis revealed that **shipping-related attributes**—notably *shipping mode* and *lead time*—dominate model reasoning, while *order quantity*, *discount rate*, and *profit ratio* exert secondary but interpretable effects. These findings align closely with established logistics principles, reinforcing confidence that the model has internalized realistic operational dependencies.

From a managerial standpoint, the explainable framework provides actionable insights for **risk mitigation and process optimization**. Firms can use these results to redesign service-level policies, synchronize marketing promotions with capacity planning, and deploy real-time dashboards that highlight high-risk shipments. Moreover, the study demonstrates that integrating explainability into model development strengthens **AI governance, accountability, and regulatory compliance**, ensuring that predictive analytics contribute to transparent and responsible decision-making.

Future research can extend this work in several directions. First, incorporating **temporal and sequential models** (e.g., LSTM, Transformer architectures) would enable dynamic prediction of delay propagation across supply-chain stages. Second, combining structured data with **multimodal information**—such as weather, traffic, or IoT sensor feeds—could enhance context awareness and robustness. Third, comparative studies using multiple XAI frameworks (e.g., SHAP, LIME, DeepLIFT) could assess consistency across interpretability methods and establish best-practice standards for logistics analytics.

Finally, integrating the proposed model into **real-time decision-support platforms** would allow continuous learning from operational feedback, paving the way toward **adaptive, human-centered, and trustworthy AI systems** for supply-chain management.

In conclusion, the research contributes both a **methodological blueprint** and **empirical evidence** that deep neural networks, when coupled with robust interpretability mechanisms, can provide reliable, transparent, and actionable insights for modern logistics operations.

References

[1] R. R. J. - and R. S. -, “A Study on Challenges and Optimization of Last Mile Delivery and its Impact on Customer Satisfaction.,” *Int. J. Multidiscip. Res.*, vol. 6, no. 4, p. 24486, July 2024, doi: 10.36948/ijfmr.2024.v06i04.24486.

[2] H. Uvet, “Importance of Logistics Service Quality in Customer Satisfaction: An Empirical Study,” *Oper. Supply Chain Manag. Int. J.*, pp. 1–10, Feb. 2020, doi: 10.31387/oscm0400248.

[3] Y. Issaoui, A. Khiat, K. Haricha, A. Bahnasse, and H. Ouajji, “An Advanced System to Enhance and Optimize Delivery Operations in a Smart Logistics Environment,” *IEEE Access*, vol. 10, pp. 6175–6193, 2022, doi: 10.1109/ACCESS.2022.3141311.

[4] E. Kalpana and S. V. Raju, “LEVERAGING MACHINE LEARNING FOR BUSINESS SUCCESS: A CASE STUDY OF SUPPLY CHAIN OPTIMIZATION IN A LOGISTICS COMPANY,” *Int. J. Adv. Bus. Manag. Res.*, vol. 01, no. 01, pp. 01–09, 2023, doi: 10.62674/ijabmr.2023.v1i01.001.

[5] USA and A. Chandramouli, “Optimizing Last-Mile Delivery Operations: Leveraging Predictive Analytics, Technology Integration, and Sustainable Practices,” *J. Math. Comput. Appl.*, vol. 2, no. 3, pp. 1–7, Sept. 2023, doi: 10.47363/JMCA/2023(2)145.

[6] Z. C. Lipton, “The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery.,” *Queue*, vol. 16, no. 3, pp. 31–57, June 2018, doi: 10.1145/3236386.3241340.

[7] W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K.-R. Muller, “Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications,” *Proc. IEEE*, vol. 109, no. 3, pp. 247–278, Mar. 2021, doi: 10.1109/JPROC.2021.3060483.

[8] F. Doshi-Velez and B. Kim, “Towards A Rigorous Science of Interpretable Machine Learning,” 2017, *arXiv*. doi: 10.48550/ARXIV.1702.08608.

[9] N. Kokhlikyan *et al.*, “Captum: A unified and generic model interpretability library for PyTorch,” 2020, *arXiv*. doi: 10.48550/ARXIV.2009.07896.

[10] H. Mirzaei, A. Daneshvar, and B. Nahavandi, “Predicting the duration of goods transportation delays based on machine learning methods,” *Mod. Supply Chain Res. Appl.*, pp. 1–20, Sept. 2025, doi: 10.1108/MSCRA-02-2025-0011.

[11] Q. Qiu, “Based on Bayesian Regression and Logistic Regression Model Forecast Logistics Transportation Delay,” *Appl. Comput. Eng.*, vol. 135, no. 1, pp. 175–183, Feb. 2025, doi: 10.54254/2755-2721/2025.21217.

[12] R. Sahoo, A. K. Pasayat, B. Bhowmick, K. Fernandes, and M. K. Tiwari, “A hybrid ensemble learning-based prediction model to minimise delay in air cargo transport using bagging and stacking,” *Int. J. Prod. Res.*, vol. 60, no. 2, pp. 644–660, Jan. 2022, doi: 10.1080/00207543.2021.2013563.

[13] G. D. Mendonça, S. R. D. M. Oliveira, O. F. Lima Jr, and P. T. V. D. Resende, “Intelligent algorithms applied to the prediction of air freight transportation delays,” *Int. J. Phys. Distrib. Logist. Manag.*, vol. 54, no. 1, pp. 61–91, Jan. 2024, doi: 10.1108/IJPDLM-10-2022-0328.

[14] Sayed Mahammad Umez and K.Muddu Swamy, “Short-Term Arrival Delay Time Prediction inFreight Rail Operations Using Data-DrivenModels,” *Int. J. Eng. Technol. Manag. Sci.*, vol. 9, no. 2, pp. 613–619, 2025, doi: 10.46647/ijetms.2025.v09i02.078.

[15] J. Wang and X. Li, “Dynamic Optimization of Transportation Networks in Logistics Using Long Short-Term Memory Neural Networks,” Dec. 27, 2024, *Engineering*. doi: 10.20944/preprints202412.2340.v1.

[16] Venkata Manikesh Iruku, “Multi-dimensional XAI Framework Revealing Critical Supply Chain Vulnerability Drivers,” *World J. Adv. Eng. Technol. Sci.*, vol. 15, no. 3, pp. 2141–2152, June 2025, doi: 10.30574/wjaets.2025.15.3.1154.

[17] Enoch Oluwademilade Sodiya *et al.*, “Reviewing the role of AI and machine learning in supply chain analytics,” *GSC Adv. Res. Rev.*, vol. 18, no. 2, pp. 312–320, Feb. 2024, doi: 10.30574/gscarr.2024.18.2.0069.

[18] F. Olan, K. Spanaki, W. Ahmed, and G. Zhao, “Enabling explainable artificial intelligence capabilities in supply chain decision support making,” *Prod. Plan. Control*, vol. 36, no. 6, pp. 808–819, Apr. 2025, doi: 10.1080/09537287.2024.2313514.

[19] E. E. Kosasih, E. Papadakis, G. Baryannis, and A. Brintrup, “A review of explainable artificial intelligence in supply chain management using neurosymbolic approaches,” *Int. J. Prod. Res.*, vol. 62, no. 4, pp. 1510–1540, Feb. 2024, doi: 10.1080/00207543.2023.2281663.

[20] A. El Jaouhari and S. Driouiche, “Towards an Explainable Artificial Intelligence in the Realm of Supply Chain Management: A Systematic Literature Review,” in *Advances in Computational Intelligence and Robotics*, J. Arif and F. Jawab, Eds., IGI Global Scientific Publishing, 2025, pp. 161–184. doi: 10.4018/979-8-3373-0923-1.ch007.

[21] [Online]. Available: https://www.incom2024.org/data-challenge/