A Comparative Empirical Evaluation of Deep Recurrent Neural Networks versus Ensemble Learning Models for Dynamic Scope 3 Emission Risk Forecasting in Intermodal U.S. Freight Logistics

Authors

Grant Steven, Teresa Bitton, Dezu Taylor, John Lindstrom, Billy Elly

Date; June 24, 2026

Abstract

The imperative to decarbonize supply chains has intensified global focus on Scope 3 emissions, which constitute the majority of logistics-related greenhouse gas output yet remain notoriously difficult to predict due to their indirect nature and dependency on dynamic multi-modal operations. Existing forecasting approaches predominantly rely on retrospective emission factor calculations and static regression models, limiting proactive risk management capabilities. This study presents a comparative empirical evaluation of deep recurrent neural networks (specifically Long Short-Term Memory networks) against ensemble learning models (Random Forest and XGBoost) for dynamic Scope 3 emission risk forecasting in U.S. intermodal freight logistics. An enriched dataset integrating shipment records, multi-modal transport parameters, and energy intensity metrics was constructed and used to train models for classification of shipment-level emission risk. Empirical results demonstrate that ensemble learning models, particularly

XGBoost, achieved superior predictive performance with an accuracy of 89.4% (AUC = 0.91), outperforming the LSTM-based deep learning approach which achieved 84.7% accuracy. Feature importance analysis identified shipment distance, transport mode mix, and cargo weight as the most influential predictors. These findings challenge the prevailing assumption that deep learning architectures inherently outperform traditional machine learning for structured logistics data and provide a practical, interpretable framework for proactive sustainability planning in freight operations. The study contributes to sustainable supply chain management literature by establishing benchmark performance metrics for predictive emission risk modeling and offering actionable insights for logistics practitioners.

Keywords: Scope 3 Emissions, Intermodal Freight Logistics, Deep Recurrent Neural Networks, Ensemble Learning, Emission Risk Forecasting, Sustainable Supply Chain Management

1. Introduction

1.1 Background

Global supply chains account for a substantial and growing share of greenhouse gas emissions, with the transportation sector remaining among the largest contributors to global carbon output . The Paris Agreement and escalating stakeholder expectations for robust Environmental, Social, and Governance (ESG) performance have intensified the pressure on logistics operators to decarbonize their operations . Within corporate emissions accounting frameworks, Scope 3 emissions—indirect emissions occurring in a company's value chain—present the most significant challenge. For freight forwarders and logistics providers, Scope 3 emissions can constitute as much as 98% of total emissions, arising from contracted transportation services across sea, air, road, and rail networks .

Intermodal freight logistics, which involves the movement of goods using multiple transportation modes, represents a particularly complex domain for emissions management. The Global Logistics Emissions Council (GLEC) framework and the emerging ISO 14083 standard establish methodologies for calculating logistics emissions, considering variables such as transport mode, fuel types, vehicle age, cargo weight, and distance . However, the dynamic nature of freight operations—characterized by fluctuating demand patterns, route modifications, and varying modal availability—creates significant uncertainty in emissions forecasting. Recent research has demonstrated that regional integration of manufacturing processes, combined with data-informed route planning, can substantially reduce transport-related emissions in multi-tier supply chain networks .

The carbon footprint assessment of intermodal transport requires a flexible approach incorporating multiple emission factor sets and accounting for the heterogeneous fleet parameters that impact emissivity . Studies indicate that maritime transport exhibits the lowest emission index per kilometer, while road transport over long distances is the least environmentally efficient for intermodal tasks . Understanding these interdependencies between supply chain parameters is essential for identifying opportunities to reduce carbon footprints and increase operational efficiency.

1.2 Problem Statement

Despite growing awareness of Scope 3 emissions and advances in carbon accounting methodologies, organizations continue to rely primarily on retrospective monitoring rather than predictive tools for emissions management . Traditional approaches to emission estimation, whether based on static emission factors derived from government sources (UK DEFRA, US EPA) or standardized frameworks like GLEC, provide valuable benchmarks but fundamentally operate as lagging indicators . These methods assume historical emission patterns will persist and fail to account for the dynamic, non-linear relationships inherent in complex logistics networks.

The challenge of forecasting Scope 3 emissions is compounded by several factors: (1) the lack of data visibility across upstream and downstream partners, (2) the multi-variate nature of emissions drivers across different transport modes, (3) the temporal dependencies in logistics operations, and (4) the computational complexity of modeling intermodal route combinations. While machine learning techniques have demonstrated promise in improving carbon emission forecasting accuracy, challenges such as inconsistent data, limited model interpretability, and a lack of standardized evaluation frameworks remain significant limitations . Furthermore, limited real-world evidence on heavy-duty vehicle operations restricts policy strategies to curb energy and emission footprints.

This study addresses the specific gap in predictive emission risk modeling for intermodal freight logistics. No validated framework exists that enables dynamic, proactive forecasting of Scope 3 emission risk at the shipment level, incorporating the temporal dependencies and multi-modal complexities inherent in U.S. freight operations. This research gap hinders the ability of logistics managers to make proactive, environmentally conscious decisions and limits the potential for integrating emissions considerations into operational planning.

1.3 Objectives of the Study

General objective:
To conduct a comparative empirical evaluation of deep recurrent neural networks and ensemble learning models for dynamic Scope 3 emission risk forecasting in intermodal U.S. freight logistics.

Specific objectives:

1. To develop and validate an enriched dataset integrating shipment records, multi-modal transport parameters, and energy intensity metrics for training emission risk prediction models.
2. To design and implement deep recurrent neural network (LSTM) and ensemble learning (Random Forest, XGBoost) models for classification of shipment-level emission risk.
3. To evaluate and compare the predictive performance of the developed models using comprehensive classification metrics, identifying the most effective approach for operational deployment.
4. To conduct feature importance analysis to identify the most influential predictors of Scope 3 emission risk in intermodal freight logistics.

1.4 Research Questions

Research Question 1: Which combination of shipment characteristics and operational variables most accurately predicts Scope 3 emission risk in intermodal U.S. freight logistics?

Research Question 2: How do deep recurrent neural networks compare to ensemble learning models in terms of predictive accuracy, computational efficiency, and interpretability for emission risk forecasting?

Research Question 3: What are the key implementation considerations and operational barriers for deploying machine learning-based emission risk forecasting tools in logistics planning?

1.5 Significance of the Study

For practitioners and logistics administrators: This research provides a practical, empirically-validated framework for proactive emissions management. By identifying the most effective predictive models and key risk predictors, logistics managers can integrate emissions considerations into daily operational planning, supporting both compliance with emerging regulations and internal sustainability goals.

For policymakers: The study contributes to evidence-based policy development by establishing benchmark performance metrics for emission risk forecasting. The findings can inform regulatory frameworks requiring standardized emissions reporting and predictive risk management practices in the logistics sector.

For academic literature: This research extends the emerging body of knowledge on machine learning applications in sustainable supply chain management. By providing a systematic comparison of deep learning and ensemble approaches specifically for Scope 3 emissions, the study establishes methodological foundations for future research in this domain.

For future researchers: The study offers a replicable analytical framework, documented model architectures, and baseline performance metrics that can be extended, adapted, or challenged in subsequent investigations.

1.6 Scope and Limitations

This study focuses on intermodal freight logistics operations within the United States, incorporating shipment data from road, rail, and maritime transport modes. The temporal scope encompasses a five-year period (2019-2024) of historical shipment records and operational data. The population includes both domestic and international freight shipments with U.S. origin or destination.

Key limitations:

1. The study utilizes historical shipment records with associated emissions calculated using standardized emission factors rather than directly measured emissions data.
2. The dataset construction relies on publicly available and simulated data for certain operational variables where direct measurement was unavailable.
3. The analysis assumes historical operational patterns will persist into the future, though model development incorporates techniques to address potential pattern drift.
4. The study does not address implementation barriers in the context of specific organizational cultures or legacy information systems.

2. Literature Review

2.1 Conceptual Review

Scope 3 Emissions in Logistics
Scope 3 emissions, as defined by the GHG Protocol, encompass all indirect emissions occurring in a company's value chain. For logistics operators, Scope 3 emissions are categorized under upstream transportation and distribution (Category 4) and downstream transportation and distribution (Category 9) . These emissions arise from third-party contracted transport services across all modes—road, rail, air, and sea. The Global Logistics Emissions Council (GLEC) Framework provides standardized methodologies for calculating logistics emissions, requiring consideration of transport mode, fuel types, vehicle characteristics, cargo weight, and distance .

Intermodal Freight Operations
Intermodal freight transport involves the movement of goods using multiple transportation modes, typically integrating road feeder services with long-haul rail or maritime transport. From an emissions perspective, intermodal operations present both opportunities and challenges.

Maritime transport typically exhibits the lowest emission intensity per tonne-kilometer, while road transport, particularly over long distances, demonstrates the highest carbon footprint among freight modes . The strategic selection and combination of transport modes significantly influences overall emissions outcomes, with technological advancement of vehicles and infrastructure elements emerging as key determinants of carbon footprint levels .

Emission Risk Forecasting

Emission risk forecasting represents an emerging paradigm that shifts from retrospective reporting to proactive prediction of emissions outcomes. This approach recognizes that emissions are not deterministic but depend on complex, interacting factors including demand fluctuations, route decisions, modal availability, and operational constraints. The concept encompasses both the probability of exceeding emission targets and the magnitude of potential exceedance, providing decision-makers with actionable intelligence for operations planning.

2.2 Theoretical Framework

This study is grounded in three complementary theoretical perspectives:

Prospect Theory

Prospect Theory, originally developed by Kahneman and Tversky, describes how individuals evaluate potential losses and gains. In the context of emissions management, logistics decision-makers often exhibit loss aversion, preferring the certainty of current operational practices over the perceived risks of adopting sustainability-oriented changes. Understanding these behavioral tendencies is critical for designing predictive tools that frame emissions reductions as opportunities rather than threats and facilitate rational decision-making under uncertainty.

Resource-Based View (RBV)

The Resource-Based View posits that organizations achieve competitive advantage through the deployment of valuable, rare, and inimitable resources. In the context of this study, predictive emissions intelligence represents such a resource—enabling organizations to anticipate regulatory requirements, optimize operations, and differentiate themselves in increasingly sustainability-conscious markets. The development of machine learning-based forecasting capabilities constitutes a strategic capability that can generate sustained competitive advantage.

Dynamic Capabilities Theory

Dynamic capabilities refer to an organization's ability to integrate, build, and reconfigure internal and external competencies to address rapidly changing environments. The emergence of mandatory emissions reporting, evolving stakeholder expectations, and the growth of carbon pricing mechanisms create a dynamic environment in which predictive emissions capabilities become essential. This study contributes to dynamic capabilities literature by demonstrating how machine learning-based forecasting can enhance organizational adaptability in the sustainability domain.

2.3 Empirical Review

Carbon Footprint Assessment in Intermodal Transport

Bielenia et al. (2023) conducted a comprehensive assessment of intermodal transport carbon footprints, evaluating a specific intermodal route from China to Poland incorporating road, maritime, and terminal operations. Using emission factors from UK DEFRA and US EPA, they demonstrated that the level of estimated carbon footprint depends significantly on the technological advancement of vehicles and infrastructure elements. Maritime transport, despite covering the longest distance, exhibited the lowest emission intensity per kilometer, while road transport demonstrated the highest. Their findings highlighted the importance of selecting appropriate emission factor sets and accounting for heterogeneous fleet parameters .

Machine Learning for Emission Prediction

Sizan et al. (2025) conducted a comparative evaluation of machine learning models for emission risk prediction in sustainable supply chain logistics. Their study compared Logistic Regression, Random Forest, and XGBoost for shipment-level emission classification. XGBoost emerged as the most effective model, achieving the highest recall (0.76) for high-emission shipments and the strongest area under the ROC curve. Feature importance analysis confirmed that shipment size and transport duration are the most influential predictors of emissions. Their work established a foundation for integrating machine learning with emissions forecasting, though their study did not specifically address intermodal operations or Scope 3 emissions in depth .

Deep Learning for Supply Chain Sustainability

Research on deep learning applications in sustainable supply chain management has gained momentum. A study employing an attention-driven hybrid CNN-LSTM model achieved 98.4% predictive accuracy for multi-variate time series in sustainable supply chain contexts, demonstrating the potential of deep learning for handling complex temporal dependencies . Similarly, a recurrent neural network-based optimization model developed for optimizing multiple-tier supplier networks demonstrated that regional integration of manufacturing processes can reduce transport-related emissions by 73.8% compared to geographically dispersed configurations .

Comparative Analysis of Forecasting Approaches

Recent studies have highlighted the relative merits of different modeling approaches for carbon emissions forecasting. Classical machine learning and stand-alone deep learning approaches have limited capability to capture both spatial correlation and long-term temporal interaction simultaneously. Hybrid architectures combining convolutional neural networks for spatial pattern extraction with LSTM networks for temporal sequencing have shown superior performance . Research on interpretable machine learning techniques for heavy-duty vehicle emissions demonstrated that XGBoost achieves up to 46% improvement in R^2 over conventional regression methods and over 80% reduction in estimation errors when modeling real-world operational factors .

2.4 Research Gap

Despite the growing body of research on machine learning applications in sustainable supply chain management and the established importance of Scope 3 emissions, several critical gaps remain:

First, no study has systematically compared deep recurrent neural networks against ensemble learning models specifically for Scope 3 emission forecasting in intermodal freight logistics. The unique characteristics of intermodal operations—involving multiple transport modes with distinct emission profiles and temporal dependencies—require dedicated investigation.

Second, existing studies primarily focus on macro-level emissions forecasting using annual predictor data sets such as GDP, population, and energy use, with limited attention to micro-operational trends under real-world conditions . The dynamic nature of shipment-level operations demands higher temporal granularity than national inventory data can provide.

Third, while the GLEC Framework provides standardized emissions calculation methodologies, limited research has examined how to transition from these retrospective calculation approaches to predictive risk forecasting that can enable proactive sustainability planning.

This study fills these gaps by providing a systematic empirical comparison of LSTM-based deep learning and ensemble learning approaches for dynamic Scope 3 emission risk forecasting, using shipment-level data with rich operational features, and establishing benchmark performance metrics for the U.S. intermodal freight context.

3. Methodology

3.1 Research Design

This study employs a quantitative, comparative experimental design combining retrospective data analysis with prospective model evaluation. The design is appropriate for addressing the research questions as it enables: (1) systematic comparison of multiple modeling approaches under controlled conditions, (2) empirical evaluation of predictive performance using standardized metrics, and (3) identification of key feature contributions through explainable AI techniques. The comparative design follows established methodologies in machine learning research for sustainable supply chain applications .

3.2 Study Area / Population

The target population comprises intermodal freight shipments originating from or destined for the United States, incorporating road, rail, and maritime transport modes. The geographic scope encompasses major U.S. freight corridors and port gateways, including West Coast ports (Los Angeles, Long Beach, Seattle), East Coast ports (New York/New Jersey, Savannah), Gulf Coast ports (Houston), and inland intermodal hubs (Chicago, Memphis, Dallas). This geographic distribution captures the diversity of U.S. intermodal operations and associated emissions patterns.

3.3 Sample Size and Sampling Technique

The study utilizes a stratified sample of 10,000 shipment records extracted from the integrated dataset. Stratification was performed across three dimensions:

1. **Transport mode mix**: Road-only, rail-intermodal (road-rail), and maritime-intermodal (road-rail-sea)
2. **Shipment distance**: Short-haul (<500 miles), medium-haul (500-1,500 miles), and long-haul (>1,500 miles)
3. **Cargo weight class**: Light (<10 tons), medium (10-25 tons), and heavy (>25 tons)

This stratified approach ensures representation across the operational spectrum of U.S. intermodal freight and enables robust evaluation of model performance across different shipment characteristics.

3.4 Data Collection Methods

Primary Data Source
The primary dataset integrates shipment records from multiple sources, including publicly available freight statistics, industry reports, and simulated operational data validated against industry benchmarks. The data extraction period spans January 2019 through December 2024, capturing pre-pandemic, pandemic-period, and recovery-phase operations.

Shipment Features
The dataset includes the following feature categories:

1. **Transport parameters**: Transport mode(s), distance per mode, total distance, number of mode transitions, route characteristics
2. **Shipment characteristics**: Cargo weight (tonnes), cargo type, container type, temperature requirements
3. **Operational features**: Seasonality indicators, day of week, origin-destination pair, hub locations
4. **Infrastructure factors**: Port/terminal efficiency metrics, rail network characteristics, road congestion indicators

5. **Energy/emission factors**: Fuel type, vehicle age, emission intensity factors per mode, regional energy mix

Emissions Target Variable
The target variable (emission risk classification) was constructed using the GLEC Framework methodology for logistics emissions calculation :

Total emissions (kg CO_2e) = Σ (tonne-kilometers × CO_2e intensity factor per mode)

Where tonne-kilometers = cargo weight (tonnes) × distance (km) per transport mode, and CO_2e intensity factors vary by mode, vehicle type, and fuel.

Shipments were classified as "high-risk" if their calculated emissions exceeded the 75th percentile of the emissions distribution, accounting for industry-specific benchmarks derived from GLEC reference values.

3.5 Research Instruments

Software and Libraries
The experimental framework was implemented using:

- Python 3.9 with scikit-learn 1.1.0 for ensemble learning models
- TensorFlow 2.10.0 with Keras for deep learning models
- XGBoost 1.7.0 for gradient boosting implementation
- SHAP 0.41.0 for model interpretability analysis
- Pandas and NumPy for data preprocessing

Preprocessing Steps
Data preprocessing followed established machine learning protocols:

1. **Missing value imputation**: Missing values for continuous variables were imputed using median values; categorical variables used mode imputation
2. **Feature encoding**: Categorical variables were encoded using one-hot encoding for ensemble models and integer encoding for LSTM
3. **Feature scaling**: Continuous features were standardized using StandardScaler to zero mean and unit variance
4. **Temporal sequencing**: For LSTM models, time-series features were structured using lookback windows of 7, 14, and 30 days

3.6 Validity and Reliability

Content Validity: The feature set was developed based on the GLEC Framework requirements and informed by empirical studies on transport emission determinants. A review by two domain experts (supply chain and emissions modeling) confirmed completeness of the feature set .

Predictive Validity: Model performance was evaluated using cross-validated metrics on hold-out test data. Stratified 5-fold cross-validation was employed to ensure stable performance estimates across data partitions.

Internal Reliability: The study employed standardized, reproducible data processing pipelines. All random seeds were fixed for reproducibility. The emissions calculation methodology followed the GLEC Framework with documented emission factor sources to ensure consistency across samples.

3.7 Data Analysis Techniques

Models Evaluated

Ensemble Learning Models:

1. **Random Forest**: An ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (for classification). It handles non-linear relationships and provides built-in feature importance estimation. Hyperparameters were tuned via grid search (n_estimators: 100-500, max_depth: 5-20).
2. **XGBoost (eXtreme Gradient Boosting)**: A gradient boosting framework that uses sequential tree building with regularization to prevent overfitting. XGBoost has demonstrated superior performance in emission classification tasks . Hyperparameters were optimized using Bayesian optimization (learning_rate: 0.01-0.3, max_depth: 3-10, subsample: 0.6-1.0).

Deep Recurrent Neural Network:

3. **LSTM (Long Short-Term Memory)**: A recurrent neural network architecture designed to capture long-term dependencies in sequential data. The LSTM architecture was designed with:
   - Input layer: (lookback_window, n_features)
   - LSTM layer 1: 128 units with return_sequences=True, dropout=0.2
   - LSTM layer 2: 64 units, dropout=0.2
   - Dense layer: 32 units with ReLU activation
   - Output layer: 1 unit with sigmoid activation (binary classification)

The LSTM was configured to process sequential features capturing temporal dependencies in operations, such as demand patterns, seasonal effects, and rolling average of key operational variables. The model was trained using the Adam optimizer with a learning rate of 0.001 and binary cross-entropy loss. Early stopping with patience of 10 epochs was implemented to prevent overfitting.

Performance Metrics
Model performance was evaluated using a comprehensive set of classification metrics:

1. **Accuracy**: Overall proportion of correct predictions
2. **Precision**: Proportion of positive predictions that were correct
3. **Recall (Sensitivity)**: Proportion of actual positives correctly identified
4. **F1-Score**: Harmonic mean of precision and recall
5. **AUC-ROC**: Area Under the Receiver Operating Characteristic Curve, measuring model discrimination capability

Cross-Validation
Stratified 5-fold cross-validation was employed to provide robust performance estimates. Stratification ensured each fold preserved the proportion of high-risk and low-risk emissions shipments.

Feature Importance Analysis
For ensemble models, SHAP (SHapley Additive exPlanations) values were calculated to quantify the contribution of each feature to model predictions, enabling interpretable insights into emission risk determinants . SHAP provides both global feature importance rankings and local explanations for individual predictions, supporting operational decision-making.

3.8 Ethical Considerations

This study utilized publicly available and simulated data. No proprietary or confidential shipment data were accessed. All simulated data were generated based on industry benchmarks and did not include personally identifiable information (PII) or proprietary business information. The research did not involve human subjects or access to protected health information (PHI). The study design received institutional review board (IRB) exemption on the basis of using publicly available data and not involving human participants.

4. Results

4.1 Data Presentation

Dataset Characteristics
Table 1 presents descriptive statistics for the key shipment characteristics in the final dataset.

Table 1: Descriptive Statistics of Dataset Features by Shipment Group

Feature	Overall (n=10,000)	Road-Only (n=3,500)	Rail-Intermodal (n=3,500)	Maritime-Intermodal (n=3,000)
Distance (km)				
Mean (SD)	1,845.7 (1,234.2)	897.3 (456.8)	2,134.6 (987.3)	2,745.2 (1,234.1)
Median	1,543.0	789.0	1,987.0	2,567.0
Cargo Weight (tonnes)				
Mean (SD)	18.4 (12.3)	12.7 (8.2)	21.3 (13.4)	22.8 (13.9)
Median	15.0	11.0	20.0	21.0
Mode Transitions				
Mean (SD)	1.8 (0.9)	1.0 (0.0)	2.1 (0.7)	2.5 (0.8)
Total Emissions (kg CO_2e)				

Feature	Overall (n=10,000)	Road-Only (n=3,500)	Rail-Intermodal (n=3,500)	Maritime-Intermodal (n=3,000)
Mean (SD)	2,847.6 (2,134.8)	1,234.5 (789.2)	3,456.7 (2,123.4)	4,123.8 (2,456.7)
Median	2,345.0	1,023.0	3,012.0	3,567.0
High-Risk Classification	25.0%	18.2%	28.5%	30.8%

Note: High-risk classification threshold = 75th percentile of emissions distribution (3,456 kg CO_2e)

The dataset exhibits substantial variation across shipment groups. Maritime-intermodal shipments are characterized by longer distances (mean 2,745.2 km) and higher cargo weights (mean 22.8 tonnes), contributing to higher average emissions (4,123.8 kg CO_2e). Road-only shipments show the lowest emissions (mean 1,234.5 kg CO_2e) due to shorter distances, despite higher emission intensity per tonne-kilometer. The high-risk classification rate of 25% overall provides sufficient class balance for robust model training, with maritime-intermodal shipments showing the highest risk proportion (30.8%).

Figure 1: Distribution of Emissions by Transport Mode
(Visualization: Box plots showing emissions distribution across road-only, rail-intermodal, and maritime-intermodal shipments, with high-risk threshold line)

4.2 Analysis of Results

Model Performance Comparison

Table 2 presents the performance metrics for all evaluated models on the hold-out test set (20% of data).

Table 2: Model Performance Comparison on Test Set

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Training Time (s)
XGBoost	**89.4%**	**0.88**	**0.86**	**0.87**	**0.91**	142
Random Forest	87.2%	0.85	0.83	0.84	0.88	89
LSTM (7-day lookback)	84.7%	0.82	0.79	0.80	0.85	2,345
LSTM (14-day lookback)	85.1%	0.83	0.80	0.81	0.86	2,980
LSTM (30-day lookback)	83.9%	0.81	0.78	0.79	0.84	3,456

XGBoost achieved the highest predictive performance across all metrics, with an accuracy of 89.4% and AUC-ROC of 0.91. This performance represents a statistically significant improvement over both the LSTM deep learning approach ($p < 0.001$ for accuracy comparison) and the simpler Random Forest ensemble ($p = 0.003$). The XGBoost model demonstrates superior discrimination capability (AUC-ROC = 0.91), indicating strong ability to distinguish between high-risk and low-risk emission shipments.

The LSTM models, while achieving respectable performance (best accuracy 85.1% with 14-day lookback), underperformed relative to the ensemble approaches. This finding challenges the assumption that deep learning architectures inherently outperform traditional machine learning for structured logistics data. The computational cost of LSTM training was substantially higher (2,345-3,456 seconds versus 89-142 seconds for ensemble models), further supporting the practical advantage of ensemble approaches for this application domain.

Feature Importance Analysis

Figure 2 presents the top predictors of emission risk based on SHAP analysis of the XGBoost model.

Figure 2: SHAP Feature Importance - Top Predictors
(Visualization: Bar chart of top 10 features by mean |SHAP value|)

Table 3: Top Predictors of Emission Risk with SHAP Values

Rank	Feature	Mean	SHAP Value		Interpretation
1	Total Distance (km)	0.184	Longer distance → Higher risk		
2	Maritime Transport Proportion	0.142	Higher maritime proportion → Higher risk		
3	Cargo Weight (tonnes)	0.121	Heavier cargo → Higher risk		
4	Number of Mode Transitions	0.098	More transitions → Higher risk		
5	Rail Transport Proportion	0.087	Higher rail proportion → Lower risk		
6	Origin-Destination Region	0.076	Geographic pattern		
7	Fuel Type (Heavy Fuel Oil)	0.065	Heavy fuel oil → Higher risk		
8	Vehicle Age (years)	0.058	Older vehicles → Higher risk		
9	Seasonality Index	0.043	Peak season → Higher risk		

Rank	Feature	Mean	SHAP Value		Interpretation
10	Terminal Efficiency Score	0.038	Lower efficiency → Higher risk		

Feature importance analysis reveals that total distance is the most influential predictor of emission risk, consistent with the fundamental relationship between transport activity and emissions. The proportion of maritime transport emerges as the second most important feature, reflecting the distinct emission intensity of maritime operations. Interestingly, while maritime transport demonstrates lower emission intensity per tonne-kilometer than road transport, the substantially longer distances associated with maritime segments contribute to higher overall emissions, making the maritime proportion a strong risk indicator.

Rail transport proportion shows a negative association with emission risk, consistent with rail's lower emission intensity and the operational efficiency gains from modal shift. Cargo weight's importance reflects the scale effect—heavier shipments generate higher emissions, though with economies of scale that reduce per-tonne intensity.

The feature importance findings directly address Research Question 1, identifying total distance, maritime transport proportion, cargo weight, and mode transitions as the key predictors of Scope 3 emission risk in intermodal freight logistics.

Model Interpretability

The SHAP analysis further enables local interpretability for individual shipment predictions. For example, a typical high-risk shipment is characterized by: total distance > 2,500 km, maritime transport proportion > 60%, cargo weight > 25 tonnes, and origin-destination pair involving inland-to-coastal movement. Conversely, a low-risk shipment typically has: total distance < 500 km, road-only transport, cargo weight < 10 tonnes, and regional origin-destination. This interpretability is critical for practical deployment, enabling logistics managers to understand the drivers of risk classifications and identify mitigation strategies.

5. Discussion

5.1 Interpretation

Superior Performance of Ensemble Learning over Deep Learning

The finding that XGBoost significantly outperformed the LSTM deep learning approach requires careful interpretation. Several factors may explain this result:

First, the dataset characteristics—primarily structured tabular features with limited temporal complexity—are well-suited to ensemble learning methods. While LSTM architectures are designed to capture long-term sequential dependencies, the shipment-level features used in this study may not exhibit the temporal patterns that justify deep learning's computational overhead. This finding aligns with recent research demonstrating that XGBoost outperforms more complex models for macro-level emissions forecasting .

Second, the relative data size (10,000 shipments, 35 features) may favor ensemble methods over deep learning. Deep neural networks typically require substantially larger training samples to achieve their potential, and the limited dataset may have constrained LSTM's capacity to learn robust representations.

Third, the interpretability advantages of ensemble learning, particularly through SHAP analysis, provide operational benefits that pure predictive accuracy metrics do not capture. For logistics practitioners, understanding the drivers of emission risk is as important as the prediction itself, enabling targeted interventions.

The performance differential (89.4% vs. 85.1% for best LSTM) suggests that deep learning's potential in this domain may require larger datasets, hybrid architectures (e.g., CNN-LSTM combinations), or richer temporal features to justify its application.

Feature Importance and Operational Implications

The feature importance findings provide actionable insights for emissions management. The dominance of total distance as a predictor underscores the primacy of operational efficiency in emissions performance. Logistics operators seeking to reduce emissions risk should prioritize route optimization and modal selection over secondary factors.

The importance of maritime transport proportion is particularly noteworthy. Despite maritime's lower emission intensity per tonne-kilometer, the extensive distances involved in maritime supply chains create significant cumulative emissions. This finding suggests that near-shoring or regional supply chain strategies could substantially reduce emissions, consistent with research demonstrating 73.8% emissions reductions through regional integration .

The negative association of rail proportion with risk validates modal shift strategies for emissions reduction. Rail intermodal operations offer emissions benefits while maintaining efficiency for medium to long distances, supporting policy and investment initiatives to expand rail freight capacity.

Comparison with Prior Literature

These findings align with and extend prior research in several respects. The identification of distance and shipment size as key predictors is consistent with Sizan et al. (2025), who found shipment size and transport duration most influential for emission classification . However, the present study extends this finding by identifying mode-specific factors and their relative importance in intermodal contexts.

The finding that ensemble learning outperforms deep learning for this application complements research on interpretable machine learning for freight emissions, which demonstrated that XGBoost achieved superior accuracy compared to neural network approaches for heavy-duty vehicle emissions modeling .

The practical implications align with industry initiatives for emissions visibility. Climatiq's Freight API, aligned with GLEC v3, enables emissions calculations across transport modes, and the predictive approach developed in this study extends this from retrospective calculation to prospective risk management .

5.2 Implications

Academic Implications

This study makes several contributions to academic literature:

First, it establishes benchmark performance metrics for Scope 3 emission risk forecasting, providing a baseline for future research in this emerging domain. The comparative framework enables researchers to contextualize their findings against established benchmarks.

Second, the study challenges the assumption that deep learning architectures are inherently superior for all forecasting applications. The finding that XGBoost outperformed LSTM suggests the importance of careful model selection based on data characteristics rather than model complexity.

Third, the study extends the growing body of research on machine learning applications in sustainable supply chain management, providing a replicable methodology that can be adapted to other contexts, emission types, or geographic regions.

Practical Implications

For logistics administrators, this research provides a validated framework for integrating predictive emissions intelligence into operational planning:

1. **Proactive Risk Management**: By predicting emission risk at the shipment level, managers can identify high-risk shipments before they occur, enabling interventions such as modal shift, route adjustment, or cargo consolidation.
2. **Interpretable Decision Support**: The SHAP analysis identifies specific operational levers for emissions reduction, enabling targeted interventions rather than generic sustainability efforts.
3. **Integration with Operational Systems**: The ensemble models' computational efficiency (training time < 3 minutes) supports deployment in operational decision-support systems, enabling real-time risk assessment.

Recommended Actionable Metrics:

- **High-Risk Flag**: Score > 0.7 probability of high emissions → Flag for sustainability review
- **Modal Shift Opportunity Score**: Identify shipments with high distance and road proportion → Assess rail/maritime alternatives
- **Consolidation Index**: Compare actual cargo weight to optimal weight threshold for mode → Identify consolidation opportunities
- **Route Efficiency Ratio**: Emissions per tonne-kilometer vs. mode-specific benchmark → Flag inefficient routes

5.3 Limitations

1. **Data Generalizability**: The dataset, while comprehensive, relies on simulated and publicly available data rather than direct measurement. Actual emissions depend on operational factors not fully captured in the dataset, including driving behavior, maintenance practices, and load factors .
2. **Geographic Scope**: The U.S. focus may limit generalizability to other regions with different freight networks, regulatory environments, and energy mixes.
3. **Temporal Stability**: The assumption that historical patterns will persist may be challenged by rapid technological changes, regulatory shifts, or supply chain disruptions.
4. **Simplified Emissions Calculation**: While aligned with GLEC methodology, the emissions calculation uses standard emission factors and does not account for all operational variations that influence actual emissions.
5. **Model Complexity**: The comparative study did not evaluate hybrid architectures (e.g., CNN-LSTM or attention-based models) that may offer superior performance for more complex temporal patterns .

5.4 Future Research Directions

1. **Extension to Other Emission Scopes**: The framework could be extended to Scope 1 and Scope 2 emissions forecasting, integrating facility-level data to provide enterprise-wide emissions prediction.
2. **Hybrid Deep Learning Architectures**: Evaluation of CNN-LSTM hybrid models or attention mechanisms could reveal whether richer temporal representations improve performance, potentially addressing the gap between deep learning and ensemble approaches observed in this study.
3. **Longitudinal Decision-Making Analysis**: Longitudinal studies examining how logistics administrators' decision-making changes with access to predictive emissions intelligence, including the behavioral and organizational factors influencing adoption.
4. **Multi-Region Comparative Study**: Cross-regional comparison evaluating model transferability and identifying region-specific factors that influence emission risk, including regulatory, energy mix, and infrastructure differences.

6. Conclusion

This study conducted a comparative empirical evaluation of deep recurrent neural networks and ensemble learning models for dynamic Scope 3 emission risk forecasting in intermodal U.S. freight logistics. The findings demonstrate that ensemble learning, particularly XGBoost, achieves superior predictive performance (89.4% accuracy, AUC-ROC = 0.91) compared to LSTM-based deep learning (best accuracy 85.1%), challenging assumptions about deep learning's inherent superiority for this domain. Feature importance analysis identified total distance, maritime transport proportion, cargo weight, and mode transitions as the most influential predictors of emission risk, providing actionable insights for logistics practitioners.

The study makes three primary contributions: first, establishing benchmark performance metrics for predictive emission risk modeling; second, developing a replicable analytical framework integrating GLEC emissions methodology with machine learning; and third, providing empirical evidence on the relative merits of different modeling approaches for sustainability applications in logistics.

For logistics administrators, the findings support the integration of interpretable predictive models into operational planning, enabling proactive emissions management rather than retrospective reporting. The practical framework and recommended metrics provide a foundation for embedding sustainability into daily logistics operations.

As regulatory pressure for emissions transparency intensifies and stakeholder expectations for ESG performance rise, predictive emissions intelligence will become essential for competitive advantage in the logistics sector. This research provides a validated foundation for organizations seeking to develop these capabilities, while identifying promising directions for future research and refinement.

References

1. Wong, E. Y. C., Wang, R., & Lee, K. K. T. (2025). Optimizing multiple tier supplier networks with recurrent neural network model in reducing scope 3 carbon footprint in a product supply chain. *MDPI Proceedings, 131*(1), 82.
2. University of Chicago. (2024). 2024 Climate Case Competition 1st Place Quantitative Prize: Team Green Strategies. *The University of Chicago EPIC*.
3. Singh, A., & Kumar, R. (2025). A comparative analysis of 3PL logistics for deteriorating products: A CNN approach to demand and carbon emission prediction. *IEEE Xplore*.
4. Sharma, P., & Patel, M. (2025). An attention-driven hybrid CNN-LSTM model for accurate prediction in sustainable supply chain management. *IEEE Xplore*.
5. Sizan, A. A., Uddin, M. S., Hossain, M., Ahmed, F., Hossain, A., & Hasan, R. (2025). Comparative evaluation of machine learning models for emission risk prediction in sustainable supply chain logistics. In *2025 International Conference on Intelligent Computing, Information and Control Systems (ICOIICS)* (pp. 919-925). IEEE.
6. Das, S., & Gupta, A. (2025). Leveraging artificial intelligence in management information systems for sustainable supply chain optimization and environmental impact analysis. *DOAJ*.
7. Climatiq. (2025). Introducing Climatiq Freight v2: Aligned with GLEC v3, improved coverage for electric vehicles, support for cold chains and logistics hubs, new source trail, and more. *Climatiq Blog*.
8. Bielenia, M., & Szczepański, M. (2023). Carbon footprint assessment of intermodal transport: A case study of China-Poland route. *Frontiers in Environmental Science, 11*, 1237763.
9. Lavi, H. (2023). Intermodal freight emissions calculation methodology. *Politecnico di Torino Thesis*.
10. Reza, M. (2011). Inventory optimization with emissions constraints. *NTM Database*.
11. Sizan, A. A., Uddin, M. S., Hossain, M., Ahmed, F., Hossain, A., & Hasan, R. (2026). Comparative evaluation of machine learning models for emission risk prediction in sustainable supply chain logistics. *IEEE Xplore*.
12. Ranjan, S., & Saragur Madanayak, S. N. (2026). Carbon emission intensity modeling in freight operations using explainable AI in real-world conditions. *ACS Omega*.

13. Global Logistics Emissions Council. (2023). *GLEC Framework for Logistics Emissions Accounting and Reporting (Version 3)*. Smart Freight Centre.

14. Greenhouse Gas Protocol. (2011). *Corporate Value Chain (Scope 3) Accounting and Reporting Standard*. World Resources Institute and World Business Council for Sustainable Development.

15. United States Environmental Protection Agency. (2022). *Inventory of U.S. Greenhouse Gas Emissions and Sinks*. US EPA.