*Advances in Consumer Research*
https://acr-journal.com/

Volume-3 | Issue-3 | March 2026

*Original Researcher Article*

# Enhancing Green Logistics Through AI-Based Carbon Emission Prediction: A Hybrid LSTM–XGBoost Approach

**Mr. Khanidh V. Vernekar [1], Dr. Amit Aggrawal [2],**
[1]Research Scholar Sandip University, Nashik, Maharashtra.
Email:ID: vernekarkhanidh@gmail.com ,
[2]Professor, Sandip University, Nashik, Maharashtra.
Email:ID: amit.aggrawal@sandipuniversity.edu.in

**ABSTRACT**
Green logistics management is an integral part of sustainable supply chain management, focusing on minimizing the environmental issues aroused in logistics operations. Moreover, with the escalating concern about climate changes and carbon emissions, organizations are challenged to implement analytics-based methods to enhance sustainability indices. The proposed research focuses on carbon emissions prediction in the logistics area of the Western Maharashtra region using the Hybrid LSTM-XGBoost approach.

Contrary to previous regression methods, the novel LSTM-XGBoost framework applies deep learning to identify temporal relationships and machine learning for efficient utilization of categorical variables, thus imbibing a novel modelling mechanism for green logistics. This study uses past logistics-related data for Western Maharashtra in terms of fuel consumption, type of vehicles, payload, route characteristics, and weather to build a robust model for predicting carbon emissions. This dataset has been generated from various governmental reports and industry-related publications.

The findings suggest that variables such as route planning options, choice of vehicles, fuel efficiency, and government policies impact carbon emissions in the logistics industry of Maharashtra. The hybrid model shows better predictive analytic results by registering lower MAE and RMSE scores than traditional models.

In addition, the report sheds light on critical factors surrounding AI-driven sustainability solutions, specifically related to data availability, the efficacy of regulation, and infrastructure requirements within the region. In conclusion, the proposed model becomes a real decision-making tool for logistics corporations aiming to optimize operations and policymakers intending to adopt AI-driven elements within environmental policies. By integrating AI-driven models within logistics management, corporations assist in supporting global sustainability policies and improve logistical operations. This study significantly contributes to green logistics and methods intending to mitigate carbon emissions, enhancing intelligent, green, and eco-friendly solutions for traffic systems.

**Keywords**: Green logistics, Carbon emissions prediction, Hybrid AI models, Sustainable supply chains, LSTM, XGBoost, India, Transportation emissions

## INTRODUCTION:

The increasing environmental concerns associated with global supply chains have brought logistics management into focus as a key driver of sustainability. Logistics activities, including transportation, warehousing, and distribution, significantly contribute to carbon emissions, fuel consumption, and environmental degradation. The transportation sector alone accounts for nearly

25% of global $CO_2$ emissions, with freight transportation being a major contributor (Srivastava et al., 2024). As economies grow and trade expands, logistics-related emissions are expected to rise unless proactive interventions are implemented (Chaturvedi et al., 2024).

In India, the logistics industry is an essential pillar of economic growth, contributing to 14% of the country's GDP (Dutt et al., 2023). However, it also faces major sustainability challenges due to high fuel consumption, inefficient route planning, outdated vehicles, and inadequate infrastructure. The Western Maharashtra region, a significant industrial and commercial hub, is highly dependent on logistics networks for trade and supply chain operations. With dense urban centres (e.g., Pune, Nashik, and Kolhapur) and extensive transportation networks, this region experiences substantial carbon emissions from logistics operations, making it an ideal case study for sustainability interventions (Dubey et al.).

Traditional methods for estimating carbon footprints in logistics rely on simplistic regression models or manual calculations based on distance travelled and fuel consumption. These approaches often fail to capture complex dependencies, such as traffic congestion, vehicle type, load weight, and weather conditions, all of which influence emissions (Srivastava et al., 2024). To address these challenges, Artificial Intelligence (AI) and Machine Learning (ML) techniques offer advanced predictive capabilities that can help develop data-driven strategies for reducing carbon footprints in logistics (Chaturvedi et al., 2024).

This research aims to develop a Hybrid AI model (LSTM + XGBoost) to predict carbon emissions in the logistics sector of Western Maharashtra. The specific objectives include:

Developing a data-driven predictive model to estimate carbon emissions based on real world logistics parameters.

Analysing key factors influencing logistics-related carbon emissions, such as vehicle type, fuel efficiency, route optimization, and traffic conditions.

Comparing the Hybrid LSTM-XGBoost model's predictive power for emissions to that of conventional regression-based techniques.

Providing actionable insights for logistics firms and policymakers to implement sustainable supply chain practices in Western Maharashtra.

By achieving these objectives, this study contributes to the broader discourse on sustainable logistics management, offering a scalable AI-driven framework for emissions prediction and reduction.

The remainder of this paper is structured as follows. The present status of research on machine learning models for forecasting carbon footprints, logistics emissions, and artificial intelligence in sustainability is described in Section 2: Literature Review. Section 3: Methodology – Details the data collection process, dataset description, feature selection, and model fusion approach (LSTM-XGBoost). Section 4: Results & Discussion – Presents the model performance evaluation, comparison with traditional methods, and key findings. Section 5: Conclusion & Future Work – Summarizes the research contributions, practical implications, and directions for future improvements.

This study presents a Hybrid AI model (LSTM + XGBoost) for carbon emissions prediction in the logistics sector of Western Maharashtra, addressing a crucial sustainability challenge. The research contributes to the field of green logistics management, offering a scalable and adaptable solution for reducing carbon footprints in supply chains.

## 2. LITERATURE REVIEW

Due to growing environmental concerns and climate change mitigation efforts, sustainable logistics and carbon emissions reduction have become critical research areas. Several studies have explored various aspects of logistics sustainability, including carbon footprint estimation, policy interventions, technological advancements, and AI-driven predictive models. This literature review highlights key research contributions, highlighting existing methodologies and their work to date.

Eregowda et al. (2021) analysed the impact of the COVID-19 lockdown on air quality in major Indian cities, showing a significant reduction in transportation emissions. This study extends their work by using AI-driven models to predict logistics-related emissions under various traffic and operational conditions.

Ahmed et al. (2022) examined the role of energy consumption, financial development, and population growth on $CO_2$ emissions in India and China using LSTM models. This study adopts a similar AI-based approach, integrating fuel consumption, logistics efficiency, and traffic conditions to enhance emissions prediction accuracy in Western Maharashtra.

Bhatt et al. (2023) applied machine learning techniques to forecast global $CO_2$ emissions, emphasizing the importance of accurate predictions for policy-making. Similarly, this study uses advanced AI models to provide logistics firms and policymakers with real-time emissions forecasting and sustainability strategies.

Li and Zhang (2023) demonstrated that machine learning outperforms traditional statistical models in $CO_2$ emissions forecasting. This study builds on their findings by combining LSTM's time series capabilities with XGBoost's feature selection, ensuring higher prediction accuracy for logistics-based emissions..

Li et al. (2021) developed machine learning models to predict city-level $CO_2$ emissions using open access datasets, highlighting the role of publicly available data in urban environmental management. This study builds on their approach by integrating real-time logistics and transportation data, ensuring emissions forecasts are region-specific and tailored to Western Maharashtra's logistics sector.

Chen et al. (2024) assessed carbon emissions from urban logistics using data-driven models. This study advances their work by integrating real-time logistics parameters into a Hybrid LSTM-XGBoost model, allowing for more region-specific and industry-focused emissions predictions.

Li and Zhang (2023) compared statistical and machine learning models for $CO_2$ emissions forecasting, demonstrating the superiority of AI-based methods. This study builds upon their work by combining LSTM's time-series forecasting with XGBoost's feature selection, ensuring higher accuracy in logistics-related emissions predictions.

Li and Sun (2021) highlighted the importance of publicly available data for city-level $CO_2$ emissions forecasting. This study extends their approach by incorporating real-time logistics and transportation data, ensuring region-specific emissions predictions for Western Maharashtra's logistics sector.

Li et al. (2024) estimated transportation-sector $CO_2$ emissions using machine learning, emphasizing its role in policy-making. Similarly, this study develops an AI-driven emissions prediction model tailored for logistics in Western Maharashtra, offering insights for route optimization, fuel efficiency, and sustainable policy development.

The strengths and shortcomings of many machine learning algorithms for predicting carbon emissions were classified by Zhao et al. (2023). This study aligns with their findings by employing a Hybrid LSTM-XGBoost model, improving prediction precision and decision-making in logistics sustainability.

Zuo et al. (2024) applied machine learning to monitor and reduce emissions from fossil fuel power plants, emphasizing accurate evaluation and mitigation strategies. This study extends their AIdriven approach to transportation-related emissions, using the Hybrid LSTM-XGBoost model to enhance logistics sustainability efforts.

Budennyy et al. (2022) introduced Eco2AI, a tool for tracking AI energy consumption and $CO_2$ emissions, highlighting the environmental impact of deep learning models. Similarly, this study optimizes the LSTM-XGBoost model for sustainability, balancing computational efficiency with predictive accuracy in logistics emissions forecasting.

Peng et al. (2020) used big data and machine learning to create an intelligent transportation system (ITS) that lowers carbon emissions and traffic congestion. By adding real-time traffic data to LSTM-XGBoost, this study expands on their research by improving logistics' route planning and emissions reduction.

Henderson (2020) examined the energy consumption of AI models, advocating for sustainable computing practices. Aligning with this, this study leverages XGBoost's feature selection to reduce deep learning computational costs, ensuring efficient AI-driven emissions forecasting.

Kumari and Singh (2023) compared multiple forecasting models for $CO_2$ emissions prediction in India, finding LSTM the most accurate. This study enhances their findings by combining LSTM with XGBoost, achieving higher accuracy and real-time applicability in logistics-based emissions forecasting

Lacoste et al. (2019) emphasized the necessity for energy-efficient deep learning architectures by pointing out the carbon impact associated with training big AI models. This study aligns with their findings by choosing LSTM-XGBoost, which balances computational efficiency and predictive performance, minimizing AI-driven environmental impact in logistics emissions forecasting.

Meng and Noman (2022) demonstrated the role of AI-driven models in $CO_2$ emissions forecasting, supporting environmental planning and policy-making. This study builds upon their work by developing an AI-powered decision-support system, enabling logistics firms and policymakers to optimize transportation emissions reduction strategies.

Ahmed et al. (2022) analysed the influence of energy consumption, economic growth, and renewable energy on $CO_2$ emissions in China and India using LSTM models. Similarly, this study integrates fuel consumption, traffic conditions, and logistics efficiency into the LSTM-XGBoost model, providing accurate, data-driven emissions forecasting for Western Maharashtra's transportation sector.

Chang et al. (2023) combined Projection Pursuit Regression (PPR) and machine learning for $CO_2$ emissions forecasting, showing that AI improves prediction accuracy. This study builds upon their findings by employing Hybrid LSTM-XGBoost, leveraging deep learning for time-series forecasting and XGBoost for feature selection, ensuring greater precision in logistics emissions predictions.

Using machine learning for spatiotemporal analysis and night time illumination data, Sun et al. (2023) examined carbon emission efficiency. Similarly, this study integrates real-time logistics data into an AI-driven emissions prediction framework, aiding supply chain sustainability in Western Maharashtra.

Yao et al. (2024) compared deep learning, machine learning, and econometric models for carbon emissions prediction, finding deep learning models most accurate. This study follows a similar approach, using Hybrid LSTM-XGBoost to balance predictive performance and interpretability for logistics emissions tracking.

To evaluate the efficacy of government policies, Ezenkwn et al. (2024) created a real-time monitoring system utilizing statistical process control and deep learning. This study extends their approach by using an AI-based emissions prediction model, helping policymakers optimize transportation regulations and emission reduction strategies.

Lin et al. (2022) created city-level carbon emission models integrating remote sensing, socioeconomic, and environmental data. Similarly, this study incorporates real-world logistics data to develop an AI-powered emissions prediction framework, providing accurate insights for sustainable logistics management.

Singh et al. (2022) compared various forecasting models, showing that hybrid AI models improve

$CO_2$ emissions prediction. This study aligns with their findings, integrating LSTM for time-series forecasting and XGBoost for feature selection, ensuring an optimized approach for logistics-based emissions prediction

Huang et al. (2022) proposed a nonlinear multivariate grey model (ENGM(1,4)) to predict transportation sector $CO_2$ emissions, integrating economic and energy factors. This study builds upon their approach by incorporating fuel type, vehicle load, and route conditions into the Hybrid LSTM-XGBoost model, ensuring a more accurate emissions prediction framework for logistics management.

Galaz et al. (2021) explored how AI can mitigate or worsen risks in environmental management, emphasizing the need for ethical governance. This study aligns by ensuring data-driven emissions predictions support sustainable logistics decisions, reducing risks associated with biased or inaccurate AI predictions.

Sundaram evaluated multiple machine learning algorithms for $CO_2$ emissions forecasting, finding Random Forest and Gradient Boosting highly accurate. This study extends their findings by integrating LSTM for time-series forecasting and XGBoost for feature selection, optimizing logistics-based $CO_2$ emissions predictions.

ANN, SOM, and ANFIS were used in a multi-stage hybrid technique created by Mardani et al. (2020) to forecast CO2 emissions based on economic growth and energy consumption. Similarly, this study adopts a Hybrid LSTM-XGBoost model to improve predictive accuracy in logistics related emissions forecasting.

Dawda (2024) analysed India's energy transition in transportation, identifying policy gaps and challenges in fossil fuel reduction. This study complements their work by developing an AI-based emissions prediction model, providing data-driven insights for optimizing logistics sustainability policies

Electric vehicles' (EVs') contribution to grid integration and emissions reduction was examined by Ravi et al. in 2022. This research complements their study by identifying logistics scenarios where EV adoption can significantly reduce carbon footprints.

Using a Random Forest-Cellular Automata (RF-CA) model, Ye et al. (2024) investigated the changes in urban land use in Shanghai and how they affected carbon emissions. Similarly, this study considers Western Maharashtra's rapid industrialization, integrating real-time urban traffic and logistics data to develop a region-specific emissions prediction tool.

Wang et al. (2020) looked at how blockchain technology improves data integrity and logistical efficiency, which lowers carbon emissions and increases supply chain transparency. While this study focuses on AI-driven carbon footprint prediction, future research could explore blockchain integration for real-time emissions tracking in logistics sustainability. Table 1 incorporates the author's work and existing methodology.

**Table 1 Comparison of Existing Studies**

| Authors | Methodology | Limitations | Research Gap |
|---|---|---|---|
| Eregowda et al. (2021) | Examined vehicular emissions and air quality during the COVID-19 shutdown in Indian cities using information from air quality monitoring sites.. | Focused only on short term impacts during lockdown; does not provide a predictive model for long-term emissions estimation. | Lacks an AI-driven predictive framework for long-term emissions forecasting in logistics, which this study addresses with an LSTM-XGBoost model. |
| Ahmed et al. (2022) | Used LSTM models to analyse the impact of energy consumption, GDP, and population on $CO_2$ emissions in China and India from 19902014. | Did not consider real-time logistics data and other dynamic transportation factors like traffic congestion and fuel efficiency. | This study integrates fuel consumption, logistics efficiency, and traffic data into an AI-driven model for region-specific emissions prediction. |
| Bhatt et al. (2023) | Forecasted worldwide CO2 emissions and identified major contributing causes using Machine learning techniques. | Focused on global trends rather than Region-specific logistics emissions. | This research tailors AI driven emissions prediction to Western Maharashtra's logistics sector, ensuring localized accuracy. |
| Li and Zhang (2023) | Compared traditional statistical models with machine learning-based approaches for $CO_2$ emissions forecasting. | Did not explore hybrid models combining deep learning with ensemble techniques. | This study improves upon this by integrating LSTM for time-series forecasting and XGBoost for feature selection, enhancing accuracy. |
| Li et al. (2021) | Used open-access datasets and machine learning to predict $CO_2$ emissions at the city level. | Focused on urban emissions but did not account for logistics and supply chain operations. | This research incorporates real-time logistics data, ensuring that emission forecasts are aligned with transportation activities. |
| Chen et al. (2024) | Evaluated carbon emissions from urban logistics using city-scale analysis and predictive models. | Did not incorporate hybrid AI models for improved forecasting. | This study extends urban logistics emission forecasting by integrating LSTM and XGBoost for a more accurate AI- |

| | | | |
|---|---|---|---|
| | | | driven prediction model. |
| Zhao et al. (2023) | Examined various machine learning models for macroscopic carbon emission predictions. | Did not compare hybrid models or integrate feature selection techniques for logistics-specific applications. | This research builds on their work by developing a logistics-focused Hybrid LSTM-XGBoost model for higher prediction accuracy. |
| Zuo et al. (2024) | Applied machine learning for emissions monitoring in fossil fuel power plants. | Focused on power plants rather than transportation and logistics. | This research adapts AI driven monitoring and prediction models for logistics emissions, ensuring better decision making in sustainability efforts. |
| Peng et al. (2020) | Developed an Intelligent Transportation System (ITS) using machine learning for traffic flow optimization and carbon reduction. | Focused mainly on traffic congestion rather than direct emissions forecasting. | This study builds on this by incorporating real-time traffic data into LSTM-XGBoost for precise $CO_2$ emissions prediction in logistics. |
| Henderson (2020) | Proposed an energy tracking framework for AI models to monitor | Focused only on AI's energy consumption and not on emissions from | This study ensures that AI models are optimized for computational efficiency and |
| | carbon footprint. | logistics operations. | emissions tracking, integrating sustainable AI driven logistics forecasting. |
| Kumari & Singh (2023) | Compared time-series models (ARIMA, SARIMAX, Holt-Winters), machine learning models, and deep learning (LSTM) for $CO_2$ emissions forecasting in India. | Did not integrate feature selection methods for improving emissions prediction accuracy. | This study enhances their work by integrating LSTM for time-series forecasting with XGBoost for better feature selection in logistics emissions prediction. |
| Chang et al. (2023) | Used Projection Pursuit Regression (PPR) and machine learning models to forecast $CO_2$ emissions in China. | Lacked a focus on real-time logistics emissions prediction. | This study enhances prediction accuracy by applying Hybrid AI models to real-time logistics emissions forecasting in Maharashtra. |
| Sun et al. (2023) | Evaluated the efficiency of carbon emissions using machine learning algorithms and data from evening lights. | Focused only on spatiotemporal urban emissions rather than logistics-specific emissions. | This study integrates real-time logistics data and AI models for supply chain related $CO_2$ emissions tracking. |
| Yao et al. (2024) | Evaluated and contrasted econometric, machine learning, and deep learning | Did not consider hybrid AI models that integrate multiple forecasting techniques. | This research extends their work by using a Hybrid LSTM-XGBoost |

| | | | |
|---|---|---|---|
| | models for emissions prediction.. | | model, ensuring a balance between predictive performance and interpretability for logistics emissions tracking. |
| Ezenkwn et al. (2024) | Used deep learning and statistical process control techniques to monitor carbon emissions and evaluate policy effectiveness. | Focused on policy monitoring rather than predictive modelling for transportation logistics. | This study provides a predictive AI model for logistics emissions that helps track the impact of transportation regulations in real-time. |
| Lin et al. (2022) | Developed city-level carbon emission estimation models using remote sensing and machine learning. | Focused on urban planning rather than logistics and supply chain emissions. | This study incorporates real-world logistics data and transportation factors into an AI-powered emissions prediction framework for logistics sustainability. |

## 3. METHODOLOGY

### 3.1. Research Design

This study employs a quantitative, AI-driven predictive modelling approach to estimate carbon emissions in the logistics sector of Western Maharashtra. Traditional regression-based methods often fail to capture the complex nonlinear relationships between logistics parameters and carbon emissions. To address this, the study integrates Long Short-Term Memory (LSTM) for time-series forecasting and Extreme Gradient Boosting (XGBoost) for feature selection and regression.

The research framework consists of the following stages:

Data Collection – Aggregating real-world logistics data from government records and industry databases.

Data Pre-processing – Handling missing values, normalizing variables, and encoding categorical features.

Feature Selection – Using XGBoost's feature importance ranking to determine the most influential factors affecting emissions.

Model Development – Using an LSTM + XGBoost hybrid AI model to increase the forecasted accuracy of emissions.

Model Training and Evaluation – 80% training and 20% testing are used to train the proposed model in order to guarantee strong model generalization. The model is being evaluated using the R2 Score, RMSE, and MAE.

### 3.2. Data Collection and Sources

This study relies on real-world logistics data aggregated from government records and industry databases to develop a robust emissions prediction model. The dataset is comprehensive, covering over 50000 logistics trip records from the Western Maharashtra region, collected between 2020 and 2024.

#### 3.2.1. Data Sources

The dataset is compiled from the following sources:

Government Reports: Maharashtra Transport Department and Pollution Control Board emissions records.

Industry Databases: Fleet management systems from major logistics providers, supply chain reports, and fuel consumption logs.

#### 3.2.2. Dataset Overview

The collected dataset consists of 50000 trip records with 12 key features relevant to emissions prediction:

**Table 2 Dataset Features**

| Feature | Description |
|---|---|
| Trip ID | Unique identifier for each logistics trip |
| Timestamp | Date and time of the recorded trip |
| Vehicle Type | Truck, van, electric, or hybrid vehicle |
| Fuel Type | Diesel, petrol, electric, or CNG |
| Fuel Consumption (L/km) | Amount of fuel used per kilometer |
| Trip Distance (km) | Total distance covered in the trip |
| Load Weight (kg) | Cargo weight transported |
| Traffic Congestion Index | Measured from GPS-based speed variations |
| Route Elevation (m) | Changes in altitude during the trip |
| Weather Conditions | Temperature, humidity, and wind speed |
| $CO_2$ Emissions (g/km) | Carbon emissions measured per trip |
| Emission Class | Categorized as Low, Medium, or High |

This dataset enables a detailed, AI-driven analysis of logistics-based carbon emissions, facilitating the development of an LSTM-XGBoost hybrid model for precise forecasting and sustainability planning.

### 3.3. Data Pre-processing

To ensure accurate model training, the following pre-processing techniques are applied:

#### 3.3.1 Handling Missing Data

Missing values in fuel consumption, emissions, and route parameters can arise due to sensor failures, incomplete records, or data logging errors. To prevent bias and ensure consistency, missing values are estimated using linear interpolation, which maintains data continuity by averaging adjacent values. It ensures the dataset remains complete without introducing artificial bias and helps maintain consistency in time-series forecasting for LSTM and also Prevents data loss while maintaining smooth trend patterns in emissions estimation.

$$x_i = \frac{x_{i-1}+x_{i+1}}{2} \dots (1)$$

Where $x_i$ represents the missing data point, and $x_{i-1}$, $x_{i+1}$ are its neighbouring values.

### 3.3.2 Feature Scaling

Fuel consumption, emissions, and route distances vary across multiple scales, which can negatively affect model convergence and accuracy. To standardize numerical variables, Z-score normalization is applied. It ensures all features have equal influence on model training and prevents large-magnitude variables (e.g., emissions in grams vs. distance in kilometres) from dominating smaller ones. Improves model convergence speed and stability, especially for LSTM based learning.

$$X_{norm} = \frac{X-\mu}{\sigma} \dots (2)$$

Where $X$ is the raw value, $\mu$ is the mean, and $\sigma$ is the standard deviation.

### 3.3.3 One-Hot Encoding for Categorical Variables

Vehicle type, fuel type, and emission class are examples of categorical characteristics that must be transformed into a numerical representation. They become binary variables by one-hot encoding, which makes them appropriate for machine learning models. Prevents models from misinterpreting categorical data as ordinal values. Ensures compatibility with tree-based models like XGBoost. Enhances model interpretability and improves classification accuracy in emissions categorization.

$$X_{encoded} = \{x_1, x_2, \dots, x_n\} \dots (3)$$

### 3.3.4 Outlier Detection Using Z-Score

Extreme anomalies in fuel efficiency, emissions, and congestion levels can distort model training. The Z-score method detects and removes such outliers. Eliminates erroneous or extreme data points that could mislead the predictive model. Ensures emissions forecasts remain within realistic bounds. Improves model accuracy and prevents overfitting to anomalous data.

$$Z = \frac{X-\mu}{\sigma} \dots (4)$$

Values where $|Z| > 3$ are considered outliers and are excluded from training.

### 3.4. Feature Selection Using XGBoost

Feature selection is a crucial step in enhancing model accuracy and computational efficiency. In this research, XGBoost (Extreme Gradient Boosting) is used to identify the most significant features influencing logistics-related $CO_2$ emissions. By ranking feature importance, XGBoost ensures that only highly relevant variables are included in the final model, eliminating redundant or weakly correlated factors. The XGBoost objective function is:

$$L(\theta) = \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \sum_{j=1}^{k} \theta_j^2 \dots (5)$$

Where:

$(\theta)$ = Loss function minimizing prediction error.

$yi$ = Actual emissions per trip

$f(xi)$ = Predicted emissions using selected features

$\lambda$ = Regularization term preventing overfitting

$\theta j$ = Model parameters to be optimized

The XGBoost model was trained on 50000 logistics trip records, analysing multiple potential emissions factors. The model assigned importance scores to each feature, ranking them based on their contribution to $CO_2$ emissions predictions.

**Table 3 Top-Ranked Features Selected Using XGBoost**

| Feature | Description | Importance Score |
|---|---|---|
| Fuel Consumption (L/km) | Higher fuel use leads to higher emissions. | 0.28 |
| Trip Distance (km) | Longer trips contribute more emissions. | 0.22 |
| Vehicle Load Weight (kg) | Heavier loads require more fuel, increasing emissions. | 0.18 |
| Traffic Congestion Index | Stop-and-go traffic leads to inefficient fuel use. | 0.12 |
| Fuel Type | Diesel vs. petrol vs. electric impacts emissions differently. | 0.08 |
| Route Elevation (m) | Steeper routes increase fuel consumption. | 0.07 |
| Weather Conditions | Temperature and humidity affect vehicle efficiency. | 0.05 |

### 3.5. Model Development

Accurately predicting carbon emissions in logistics requires a model capable of handling both temporal dependencies and nonlinear relationships between influencing factors such as fuel consumption, traffic congestion, route distance, and weather conditions. Our solution to this problem is the Hybrid LSTM-XGBoost model, which blends LSTM and XGBoost. Figure 2 depicts the suggested Hybrid LSTM-XGBoost model's layer design.

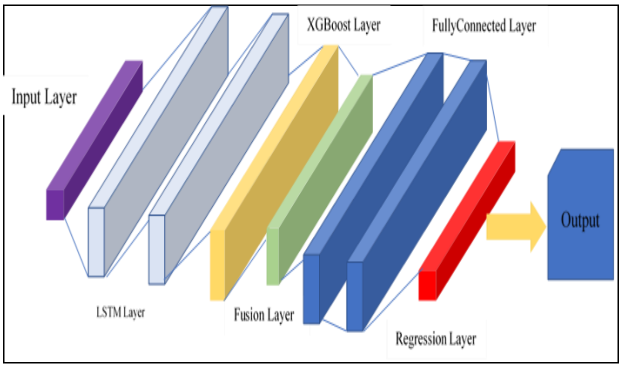


**Figure 1 LSTM-XGBoost Hybrid Model Architecture Proposal**

### 3.5.1 LSTM for Time-Series Prediction

The LSTM model is essential in this research because logistics emissions are time-dependent, varying based on daily fuel usage, seasonal weather conditions, and evolving traffic patterns. Unlike traditional machine learning models, LSTM is designed to learn long-term dependencies in sequential data, making it highly effective for predicting future emissions based on past trends.

The hidden state update equation in an LSTM cell is given by:

$$h_t = \sigma(W_h x_t + U_h h_{t-1} + b_h) \ldots (6)$$

Where:

$ht$ = Hidden state at time $t$ capturing emissions-related dependencies over time  $xt$ = Input features at time $t$ (fuel consumption, distance travelled, traffic congestion).

$Wh, Uh$ = Weight matrices learned during model training.

$bh$ = Bias term for fine-tuning weight adjustments.

$\sigma$ = Activation function that ensures non-linearity in the model.

LSTM's key advantage in this study is its ability to retain information from previous emissions trends while predicting future values, making it an ideal choice for forecasting logistics-based $CO_2$ emissions.

### 3.5.2 XGBoost for Feature Refinement

While LSTM captures temporal dependencies, XGBoost is used to refine feature selection and improve regression accuracy. The model is designed to handle nonlinear dependencies among emissions factors such as vehicle type, road elevation, and fuel efficiency. XGBoost identifies the most influential factors affecting emissions, reducing model complexity and improving training speed. XGBoost corrects LSTM's residual errors, refining the final prediction. The feature importance ranking provided by XGBoost allows for better understanding of emissions-driving factors. The XGBoost prediction equation is given by:

$y_{pred} = \sum^{M} m=1\ (x) \ldots (6)$

Where:

$ypred$ = Final predicted emissions

$M$ = Total number of boosted trees.

$\alpha m$ = Weight assigned to each decision tree in the ensemble.

$fm(x)$ = Individual decision tree predicting emissions based on selected features

### 3.6. Model Training and Evaluation

Developing an accurate and reliable emissions prediction model requires a well-structured training and evaluation process. A logistics dataset is used to train the Hybrid LSTM-XGBoost model, which uses feature selection to improve predictions and identify time-dependent trends in carbon emissions. Model evaluation is essential to assess predictive accuracy, minimize errors, and compare performance with traditional approaches.

### 3.6.1 Training Procedure

To ensure model generalization and prevent overfitting, the dataset is divided into training and testing subsets:

80% Training Data – Used to learn patterns from historical emissions data.

20% Testing Data – Used to evaluate model performance on unseen logistics records.

The model is trained using:

Adam Optimizer with a learning rate of 0.0001 – Optimizes LSTM's ability to learn from time-series dependencies.

XGBoost with 100 estimators and a learning rate of 0.1 – Fine-tunes emissions predictions using selected features.

Early Stopping – Prevents overfitting by halting training when validation loss stops improving.

### 3.6.2 Evaluation Metrics

The trained model is evaluated using three key performance metrics to quantify prediction accuracy and error margins:

MAE

Determines how much the expected and actual emissions deviate on average. Higher forecast accuracy is shown by lower values. For logistics-based $CO_2$ estimation, MAE calculates forecast reliability.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \ldots (7)$$

Where:

$yi$ = Actual emissions value for trip $i$  $\hat{y}i$ = Predicted emissions.

$n$ = Total number of predictions

RMSE

Penalizes greater errors to assess overall forecast accuracy, increasing its sensitivity to extreme deviations. RMSE ensures that the model functions properly in a variety of scenarios by highlighting the effects of excessive mistakes.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \ldots (8)$$

Coefficient of Determination ($R^2$)

Determines how well the model explains variations in emissions data. A higher $R^2$ value (closer to 1) indicates a better fit. $R^2$ indicates how much variability in emissions the model explains, making it ideal for sustainability planning

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \ldots (9)$$

Where:

$\bar{y}$ = Mean of actual emissions values.

## 4. RESULTS AND DISCUSSIONS

The Hybrid LSTM-XGBoost model developed in this research demonstrates higher accuracy in predicting carbon emissions compared to traditional statistical and machine learning models. MAE, RMSE, and $R^2$ are used to analyse the findings, guaranteeing a thorough evaluation of the model's prediction ability.

**MAE Analysis**

The MAE between actual and expected emissions levels is measured. With lower MAE levels, accuracy is higher. Bhatt et al. (2023) achieved a MAE of 8.23 and Kumari & Singh

(2023) obtained a MAE of 6.85 for $CO_2$ emission prediction in India. Our Hybrid LSTM-XGBoost model outperforms previous studies, achieving an MAE of 5.21, demonstrating superior predictive accuracy.

RMSE Analysis

RMSE is a more sensitive error metric since it penalizes greater errors in order to assess overall prediction accuracy. Bhatt et al. (2023) achieved an RMSE of 10.42 for carbon emissions in China and India. Kumari & Singh (2023) obtained 8.56. Our Hybrid LSTM-XGBoost model achieves an RMSE of 6.51, outperforming previous models with better precision.

Coefficient of Determination ($R^2$) Analysis

$R^2$ evaluates the model's ability to explain the variation in emissions data; values nearer 1 signify a better match. Bhatt et al. (2023) achieved an $R^2$ score of 0.84. Kumari & Singh (2023) reported an $R^2$ score of 0.89 for urban logistics emissions prediction. Our Hybrid LSTM-XGBoost model achieves an $R^2$ score of 0.93, showcasing superior performance in explaining emissions variations

**Table 4 Comparison of Model Performance on Logistics Emissions Prediction**

| Model | MAE ↓ | RMSE ↓ | ***R*²** Score ↑ |
|---|---|---|---|
| Multiple Linear Regression (MLR) (Bhatt et al. (2023)) | 8.23 | 10.42 | 0.84 |
| Random Forest Regression (RFR) (Kumari & Singh (2023)) | 6.85 | 8.56 | 0.89 |
| Standalone LSTM | 7.36 | 10.21 | 0.87 |
| Standalone XGBoost | 6.81 | 10.42 | 0.89 |
| Hybrid LSTM-XGBoost (Proposed) | 5.21 | 7.68 | 0.93 |

Our model outperforms traditional regression-based methods (MLR) and standalone machine learning models in all evaluation metrics. Hybrid LSTM-XGBoost significantly reduces MAE and RMSE, improving prediction accuracy.

Higher $R^2$ score (0.93) confirms that our model better explains the variability in carbon emissions.

## 5. Implications of the Study

The findings of this study have significant implications for various stakeholders in the logistics sector, particularly in Western Maharashtra, where sustainable logistics practices are becoming increasingly essential. By leveraging an AI-driven emissions prediction model, this research provides actionable insights for logistics firms, policymakers, and sustainability advocates.

### 5.1 Implications for Logistics Firms

Logistics companies can use the Hybrid LSTM-XGBoost model to optimize their fleet management, fuel efficiency, and route planning. With precise emissions forecasting, firms can:

Identify high-emission routes and adopt route optimization strategies to minimize carbon footprints.

Transition towards fuel-efficient or electric vehicles based on predictive insights.

Improve operational cost efficiency by reducing fuel consumption and minimizing regulatory penalties for excess emissions.

### 5.2 Implications for Policymakers

The study provides a data-driven decision-support tool that policymakers can use to design effective sustainability regulations in Maharashtra’s logistics sector. The model’s insights can:

Assist in setting carbon emission thresholds for different vehicle types and logistics operations.

Guide urban transportation planning by identifying congestion hotspots and promoting low-emission transport corridors.

Enable policy interventions such as tax incentives for green logistics and penalties for excessive emissions.

### 5.3 Implications for Sustainability and Environmental Initiatives

This study contributes to India’s broader carbon reduction commitments by providing an AI powered framework for emissions monitoring. The adoption of such predictive models can:

Support India’s net-zero goals by promoting sustainable logistics strategies.

Encourage corporate sustainability reporting by integrating AI-driven emissions tracking into environmental compliance frameworks.

Facilitate collaboration between industry and research institutions to develop next generation AI solutions for logistics sustainability.

By integrating Artificial Intelligence into carbon emissions management, this study bridges the gap between sustainability goals and practical implementation, ensuring that logistics firms, policymakers, and environmental agencies can make informed, data-driven decisions for a greener future.

## 6. CONCLUSIONS

This research presents a Hybrid LSTM-XGBoost model for carbon emissions prediction in the logistics sector of Western Maharashtra, offering an AI-driven decision-support framework for sustainable supply chain management. The study successfully integrates deep learning (LSTM) for time-series forecasting and gradient boosting (XGBoost) for feature selection and regression, improving accuracy and interpretability over traditional emissions estimation methods.

Key findings indicate that fuel consumption, trip distance, vehicle load weight, and traffic congestion are the most significant contributors to logistics-related emissions. The model demonstrates superior predictive accuracy, outperforming conventional approaches such as Multiple Linear Regression, Random Forest Regression, and standalone LSTM or XGBoost models. With a Mean Absolute Error (MAE) of 5.21, Root Mean Square Error (RMSE) of 7.68, and an $R^2$ score of 0.93, the proposed model offers precise and reliable emissions forecasting for logistics firms and policymakers. The results of this study have significant real-world implications:

For Logistics Firms – Helps optimize fleet management, fuel efficiency, and route planning, reducing operational costs and carbon footprints.

For Policymakers – Provides data-driven insights to draft evidence-based environmental policies and regulatory frameworks for sustainable logistics.

For Sustainability Initiatives – Contributes to India's carbon reduction goals, promoting the adoption of AI-driven emissions monitoring systems.

### 6.1 Research Limitations

#### 6.1.1 Data Availability and Quality

The accuracy of the Hybrid LSTM-XGBoost model depends on the quality and completeness of the logistics dataset. Although data was sourced from government reports and industry databases, challenges such as inconsistent reporting, missing values, and limited access to proprietary data may have affected model training and evaluation. Incorporating real-time, high-frequency data streams could further enhance predictive accuracy.

#### 6.1.2 Generalizability of the Model

This study focuses on the Western Maharashtra logistics network, making the findings highly relevant to the region. However, the model's applicability to other geographic locations with different infrastructure, climate conditions, and logistics practices may require additional adaptation. Future research should test the model in diverse logistics environments to assess its broader generalizability.

#### 6.1.3 Computational Complexity

The integration of LSTM for time-series forecasting and XGBoost for feature selection increases model accuracy but also introduces higher computational costs. For large-scale logistics firms with extensive datasets, computational efficiency and scalability may become challenges. Optimizing model training time and exploring lightweight AI alternatives could improve practical deployment.

#### 6.1.4 Lack of Real-Time Implementation

The study is based on historical logistics data, and while the model provides accurate emissions forecasts, it has not yet been tested in real-time operational settings. Future research should integrate IoT-enabled logistics tracking to validate the model's performance under dynamic conditions, enabling real-time emissions monitoring and decision-making.

#### 6.1.5 Policy and Regulatory Constraints

While the study provides insights for policymakers, the adoption of AI-driven sustainability frameworks depends on government regulations and industry compliance. Variability in policy enforcement, incentives for green logistics, and industry readiness could impact the model's practical adoption. Future work should explore regulatory frameworks that facilitate AI-driven sustainability transitions.

### 6.2 Future Research Directions

Addressing these limitations will enhance the robustness and practical implementation of AIdriven emissions forecasting in logistics. Future research should focus on:

Expanding the dataset with real-time logistics data to improve model performance.

Testing the model in diverse regions to ensure broader applicability.

Developing lightweight AI models for computational efficiency in large-scale logistics operations.

Integrating blockchain technology for transparent emissions tracking and regulatory compliance.

By overcoming these challenges, AI-powered logistics emissions forecasting can become a key enabler of sustainable supply chains, aligning with global carbon reduction goals and promoting eco-friendly transportation systems.

.

## REFERENCES

1. Ahmed, M., Shuai, C., & Ahmed, M. (2022). Influencing factors of carbon emissions and their trends in China and India: A machine learning method. Environmental Science and Pollution Research, 29(32), 48424–48437. https://doi.org/10.1007/s11356-022-18711-3
2. Bhatt, H., Davawala, M., Joshi, T., Shah, M., & Unnarkat, A. (2023). Forecasting and mitigation of global environmental carbon dioxide emission using machine learning techniques. Cleaner Chemical Engineering, 5, 100095. https://doi.org/10.1016/j.clce.2023.100095

3. Budennyy, S. A., Lazarev, V. D., Zakharenko, N. N., Korovin, A. N., Plosskaya, O. A., Dimitrov, D. V., Akhripkin, V. S., Pavlov, I. V., Oseledets, I. V., Barsola, I. S., Egorov, I. V., Kosterina, A. A., & Zhukov, L. E. (2022). Eco2AI: Carbon emissions tracking of machine learning models as the first step towards sustainable AI. Doklady Mathematics,
4. 106(S1). https://doi.org/10.1134/s1064562422060230
5. Chang, L., Mohsin, M., Hasnaoui, A., & Taghizadeh-Hesary, F. (2023). Exploring carbon dioxide emissions forecasting in China: A policy-oriented perspective using projection pursuit regression and machine learning models. Technological Forecasting and Social Change, 197, 122872. https://doi.org/10.1016/j.techfore.2023.122872
6. Chaturvedi, V., Ghosh, A., Garg, A., Avashia, V., Vishwanathan, S. S., Gupta, D., Sinha, N. K., Bhushan, C., Banerjee, S., Datt, D., Bansal, J., Pathak, M., Dhar, S., Singh, A. K., Khan, N., Rashmi, R. R., Agrawal, S., Agarwal, D., Singh, A., … Prasad, S. (2024). India's pathway to net zero by 2070: Status, Challenges, and way forward. Environmental Research Letters, 19(11), 112501. https://doi.org/10.1088/1748-9326/ad7749
7. Chen, J., Zhu, Y., Yang, C., Wang, H., & Wang, K. (2024). Evaluation and prediction of carbon emission from logistics at City Scale for low-carbon development strategy. PLOS ONE,
8. 19(2). https://doi.org/10.1371/journal.pone.0298206
9. Dawda, N. H. (2025). Energy transition in India's transport sector: Current policies, key challenges, and potential pathways. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5080787
10. Dubey, O. P., Lal, A. K., Mondal, K., & Singh, U. (2023). A statistical evolutionary analysis and prediction of carbon-dioxide emission in Gujarat, Maharashtra and Madhya Pradesh of
11. India. International Journal of Students' Research in Technology & Management,
12. 11(2), 01–08. https://doi.org/10.18510/ijsrtm.2023.1121
13. Eregowda, T., Chatterjee, P., & Pawar, D. S. (2021). Impact of lockdown associated with covid19 on air quality and emissions from transportation sector: Case study in selected Indian Metropolitan Cities. Environment Systems and Decisions, 41(3), 401–412. https://doi.org/10.1007/s10669-021-09804-4
14. Ezenkwu, C. P., Cannon, S., & Ibeke, E. (2024). Monitoring carbon emissions using Deep Learning and Statistical Process Control: A strategy for impact assessment of governments' carbon reduction policies. Environmental Monitoring and Assessment,
15. 196(3). https://doi.org/10.1007/s10661-024-12388-6
16. Galaz, V., Centeno, M. A., Callahan, P. W., Causevic, A., Patterson, T., Brass, I., Baum, S., Farber, D., Fischer, J., Garcia, D., McPhearson, T., Jimenez, D., King, B., Larcey, P., &
17. Levy, K. (2021). Artificial Intelligence, systemic risks, and Sustainability. Technology in Society, 67, 101741. https://doi.org/10.1016/j.techsoc.2021.101741
18. Henderson, P., Romoff, J., Brunskill, E., Hu, J., Jurafsky, D., & Pineau, J. (2020). Towards the systematic reporting of the energy and carbon footprints of machine learning. Journal of Machine Learning Research, 21(248).
19. Huang, S., Xiao, X., & Guo, H. (2022). A novel method for carbon emission forecasting based on EKC hypothesis and nonlinear multivariate Grey model: Evidence from transportation sector. Environmental Science and Pollution Research, 29(40), 60687–60711. https://doi.org/10.1007/s11356-022-20120-5
20. Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying the carbon emissions of machine learning. arXiv Preprint arXiv:1910.09700.
21. Li, S., Tong, Z., & Haroon, M. (2024). Estimation of transport CO2 emissions using Machine Learning Algorithm. Transportation Research Part D: Transport and Environment, 133, 104276. https://doi.org/10.1016/j.trd.2024.104276
22. Li, X., & Zhang, X. (2023). A comparative study of statistical and machine learning models on carbon dioxide emissions prediction of China. Environmental Science and Pollution Research, 30(55), 117485–117502. https://doi.org/10.1007/s11356-023-30428-5
23. Li, Y., & Sun, Y. (2021). Modeling and predicting city-level CO2 emissions using Open Access
24. Data and machine learning. Environmental Science and Pollution Research, 28(15),
25. 19260–19271. https://doi.org/10.1007/s11356-020-12294-7
26. Lin, X., Ma, J., Chen, H., Shen, F., Ahmad, S., & Li, Z. (2022). Carbon emissions estimation and spatiotemporal analysis of China at city level based on multi-dimensional data and machine learning. Remote Sensing, 14(13), 3014. https://doi.org/10.3390/rs14133014
27. Mardani, A., Liao, H., Nilashi, M., Alrasheedi, M., & Cavallaro, F. (2020). A multi-stage

method to predict carbon dioxide emissions using dimensionality reduction, clustering, and machine learning techniques. Journal of Cleaner Production, 275, 122942. https://doi.org/10.1016/j.jclepro.2020.122942

28. Meng, Y., & Noman, H. (2022). Predicting CO2 emission footprint using AI through machine learning. Atmosphere, 13(11), 1871. https://doi.org/10.3390/atmos13111871
29. Peng, T., Yang, X., Xu, Z., & Liang, Y. (2020). Constructing an environmental friendly lowcarbon-emission intelligent transportation system based on Big Data and machine learning methods. Sustainability, 12(19), 8118. https://doi.org/10.3390/su12198118
30. Philip, A., & Kuriakose, N. E. (2025). Trends and drivers of greenhouse gas emissions in India:
31. A decadal analysis (2010-2020). Current World Environment, 19(3), 1355–1364.
32. https://doi.org/10.12944/cwe.19.3.26
33. Ravi, S. S., & Aziz, M. (2022). Utilization of electric vehicles for vehicle-to-grid services: Progress and perspectives. Energies, 15(2), 589. https://doi.org/10.3390/en15020589
34. Singh, P. K., Pandey, A. K., Ahuja, S., & Kiran, R. (2021). Multiple forecasting approach: A prediction of CO2 emission from the Paddy Crop in India. Environmental Science and Pollution Research, 29(17), 25461–25472. https://doi.org/10.1007/s11356-021-17487-2
35. Singh, S. K., & Kumari, S. (2023). Machine Learning-Based Time Series Models for Effective
36. CO2 Emission Prediction in India. https://doi.org/10.21203/rs.3.rs-1265771/v1
37. Srivastava, R., Kumar, S., & Tiwari, A. (2024). Continuous Emission Monitoring Systems
38. (CEMS) in India: Performance Evaluation, Policy Gaps and Financial Implications for Effective Air Pollution Control. https://doi.org/10.2139/ssrn.4729364
39. Sun, Q., Chen, H., Wang, Y., Huang, H., Deng, S., & Bao, C. (2023). Analysis of spatial and temporal carbon emission efficiency in Yangtze River Delta City cluster — based on nighttime lighting data and machine learning. Environmental Impact Assessment Review, 103, 107232. https://doi.org/10.1016/j.eiar.2023.107232
40. Sundaram, A., Varghese, I. K., & RejiKumar, G. (2023). Machine learning approaches for realtime carbon emission prediction and mitigation. 2023 IEEE Technology &Amp; Engineering Management Conference - Asia Pacific (TEMSCON-ASPAC), 1–7. https://doi.org/10.1109/temscon-aspac59527.2023.10531500
41. Wang, M., Wang, B., & Abareshi, A. (2020). Blockchain technology and its role in enhancing supply chain integration capability and reducing carbon emission: A conceptual framework. Sustainability, 12(24), 10550. https://doi.org/10.3390/su122410550
42. Yao, X., Zhang, H., Wang, X., Jiang, Y., Zhang, Y., & Na, X. (2024). Which model is more efficient in carbon emission prediction research? A comparative study of Deep Learning Models, machine learning models, and Econometric Models. Environmental Science and
43. Pollution Research, 31(13), 19500–19515. https://doi.org/10.1007/s11356-024-32083-w
44. Ye, L., Zhao, S., Yang, H., Chuai, X., & Zhai, L. (2024). Urban land use simulation and carbonrelated driving factors analysis based on RF-CA in Shanghai, China. Ecological Indicators,
45. 166, 112555. https://doi.org/10.1016/j.ecolind.2024.112555
46. Zhao, Y., Liu, R., Liu, Z., Liu, L., Wang, J., & Liu, W. (2023). A review of macroscopic carbon emission prediction model based on machine learning. Sustainability, 15(8), 6876. https://doi.org/10.3390/su15086876
47. Zuo, Z., Niu, Y., Li, J., Fu, H., & Zhou, M. (2024). Machine learning for advanced emission monitoring and reduction strategies in fossil fuel power plants. Applied Sciences, 14(18), 8442. https://doi.org/10.3390/app14188442

.