National College of Ireland

A Transformer-Driven Framework for Accurate CO_2 Emission Modelling and Forecasting

MSc Research Project
MSc Data Analytics

Lokesh Musunuru
Student ID: 24130753

School of Computing
National College of Ireland

Supervisor: Dr. Vikas Tomer

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Lokesh Musunuru
Student ID:	24130753
Programme:	MSc in Data Analytics **Year:** 2025
Module:	Research Practicum Part 2
Supervisor:	Dr. Vikas Tomer
Submission Due Date:	28 January 2026
Project Title:	A Transformer-Driven Framework for Accurate CO_2 Emission Modelling and Forecasting
Word Count:	8252 **Page Count** 23

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

Signature:	Lokesh Musunuru
Date:	28 January 2026

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

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

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

A Transformer-Driven Framework for Accurate CO_2 Emission Modelling and Forecasting

Lokesh Musunuru
Student ID: 24130753

Abstract

Accurately forecasting CO_2 emissions has become essential as global climate challenges intensify and transportation-related emissions continue to rise. Traditional prediction approaches struggle to model the complex, nonlinear relationships among vehicle characteristics, fuel consumption patterns, and emission behaviour, motivating the development of more robust data-driven methods. Although machine learning, deep learning, and hybrid AI techniques have shown substantial progress, many existing models face key limitations, including weak generalizability, difficulty handling high-dimensional tabular data, and limited capability in capturing long-range feature interactions. This study introduces a transformer-based framework using the FT-Transformer architecture to address these shortcomings and advance CO_2 emission forecasting. First, we construct a comprehensive modelling pipeline integrating data preprocessing, feature embedding, and attention-based learning optimized for mixed-type vehicle datasets. Second, we implement and evaluate the FT-Transformer alongside several baseline models including machine learning and deep learning to rigorously assess performance improvements in predictive accuracy, scalability, and robustness. The architecture leverages numerical and categorical embeddings, multi-head self-attention, and layer normalization to learn complex inter-feature dependencies with minimal manual feature engineering. Experimental results demonstrate that the FT-Transformer achieves the strongest performance among all models tested, delivering an MAE of 0.0219, MSE of 0.0010, RMSE of 0.0312, and an R^2 of 0.9512. These findings confirm its superior capability to capture nonlinear feature relationships and emission determinants more effectively than tree-based, linear, and neural baselines. The model's attention mechanisms also provide enhanced, enabling clearer identification of influential variables such as engine size, fuel type, and vehicle class.

1 Introduction

The fast rate of carbon dioxide (CO2) emission in the world has heightened the effects of climate change, disturbed the ecosystem, and raised the need to implement effective mitigation measures. The increasing emissions in the transportation, industry, buildings and urban sectors further endanger the stability of the environment and thus proper forecasting is a critical component of climate planning. Countries and policy-makers need to have good prediction models that can be used to predict trends in emissions and draft the right regulatory frameworks. The nonlinear, dynamic, and complex nature of the CO2 emissions caused by aspects like economic activity, fuel consumption, industrial output, and climate

influences, however, complicates the accurate forecasting of CO2 emission, which is a tasking process that requires sophisticated and data-driven analytical tools (Giannelos et al., 2024).

With the emissions of the world steadily rising, there has been a greater demand of the accurate, scale and real time prediction models. Credible forecasting helps governments assess future emission patterns, aids industries in controlling carbon footprints and also helps in the fulfilment of international climate agreements like the Paris Agreement. Traditional statistical models do not have the ability to capture nonlinear relationships as well as interacting variables. Therefore, sophisticated models of artificial intelligence (AI) and in particular those that are able to learn subtle dependencies in tabular, temporal, and multi-source data are needed to establish intelligent decision-support systems that can be used to inform the process of environmental policy and sustainable planning.

Machine learning (ML), deep learning (DL), and hybrid AI methods of CO2 prediction have already been examined in a broad scope of existing works. Gradient Boosting Regression, CatBoost, and Support Vector Machines are examples of ML models that have shown good results in modelling transportation- and socioeconomic-related emissions (Li et al., 2022; Natarajan et al., 2023; Saxena et al., 2023). Long-term and sequential emission forecasting with deep learning models, specifically LSTM, BiLSTM, GRU, and hybrid LSTM-GRU models have offered better accuracy in transportation, buildings, and industrial activity sectors (Al-Nefaie and Aldhyani, 2023; Giannelos et al., 2024; Wang et al., 2025). Hybrid artificial intelligence systems, including ARIMA-BPNN, ELM-INFO, and ensemble predictors, have solved some of the modelling shortcomings but remain unable to handle large-dimensional variables, sudden emission changes, and scalability (Hou and Chen, 2024; Feda et al., 2024). Despite the significant improvement of existing models, there are still requirements of more generalizable architectures that can effectively work with complex tabular data, mixed feature types, and real-world inconsistencies.

Research Question

How can the Feature Focused FT-Transformer architecture enhance CO_2 emission prediction accuracy compared to existing machine learning, deep learning?

Justification

This research question is consistent with the limitations found in the literature namely the failure of previous models to effectively utilize long-range dependencies, handle high-dimensional tabular data and remain stable when the data is variable and inconsistent. FT-Transformer, a tabular data-based model with self-attention mechanisms, can be an excellent solution to these issues, and it can be proposed as a potential choice to further develop the CO2 forecasting studies.

There are three significant contributions of this study. To begin with, it performs an overarching analysis of the current ML, DL, and hybrid AI models to determine particular drawbacks of CO2 prediction within industries. Second, it presents an FT-Transformer-based

framework which can model long-range interaction of the variables, mixed types of data, and enhance the robustness in incomplete or noisy datasets. Third, the leverage model is used and contrasted with the known methods to prove the positive changes in predictive accuracy, scalability. The combination of these contributions offers a great step forward in the creation of reliable and data-driven tools of CO2 emission forecasting and sustainable climate decision-making.

2 Related Work

Carbon dioxide (CO2) emission prediction has become an important tool in solving global climate problems and facilitating successful environmental policy-making. As the artificial intelligence gains momentum, machine learning and deep learning models have become more frequently utilized to predict trends in emissions in transportation, industrial, and urban sectors. This literature review discusses the current methods, their advantages and limitations and research gaps that spark the necessity of sophisticated models like the FT-Transformer.

2.1 Machine Learning Models for CO_2 Emission Prediction

Machine learning (ML) models have become effective and powerful in predicting CO2 emissions because they capture nonlinear relationships by the presence of environmental, transportation, and socioeconomic factors. The common types of ML have been Ordinary Least Squares (OLS), Support Vector Machines (SVM) and Gradient Boosting Regression (GBR) because they are well-predictive. As an example, in the case of modelling transportation-related emissions in the countries with high CO2 emission, ML algorithms were capable of determining the role of vehicle activity, population growth, and economic factors, and GBR was the most accurate in predicting results compared to other methods (Li et al., 2022). These results indicate the possibility of using ML to aid emission-reduction policies to provide realistic forecasts using varied data.

The vehicle emissions industry has also extensively used machine learning, where boosting and regression models have shown good results with small sets of input attributes. In particular, CatBoost has been demonstrated to be quite effective at forecasting vehicle CO2 emissions with only limited vehicle specifications, which is why it is especially appropriate in the context of real-world settings with incomplete or inconsistent data (Natarajan et al., 2023). Moreover, grey machine learning methodologies offer the effective solutions to the emission forecasting in the cases when small datasets are at hand. The authors discovered that a grey polynomial model that was optimized with Augmented Crow Search Algorithm could be used to predict energy consumption and CO2 emissions with great accuracy (Saxena et al., 2023). Such studies indicate the flexibility of ML models in different data settings.

Moreover, intelligent mobility has incorporated the use of ML to analyse traffic dynamics and predict the emission rates. Traffic congestion has been classified with models like Linear Regression, Random Forest, and XGBoost to be associated with the relevant outputs of CO2, so that standardized systems of scoring emissions of vehicles can be developed (Nagpal et al., 2024). These applications demonstrate the role of ML in sustainable urban planning in

converting real-time traffic and environmental data into useful emission data. In general, machine learning is reinforcing CO2 emission prediction systems by offering scalable data-driven solutions, which can be used to monitor the environment and develop policies.

2.2 Deep Learning Models for CO_2 Emission Prediction

Due to their capacity to learn long-term interdependences and nonlinear temporal variations in environmental and industrial data, deep learning (DL) models have become highly prominent in the forecasting of CO2 emissions. RNNs and LSTM in particular and BiLSTM are popular because of their ability to deal with sequential data. As an example, LSTM and BiLSTM models generated to predict CO2 emissions of traffic vehicles proved to be very accurate, and BiLSTM was the most successful because it could learn forward and backward data relations (Al-Nefaie and Aldhyani, 2023). In the same way, deep learning models on building-sector emissions worldwide demonstrated the efficacy of deep neural networks in regional differences and trends in long-term emissions in several countries (Giannelos et al., 2024). The findings highlight the appropriateness of the DL models to complex and multidimensional emission data.

Recently, high-level RNN structures have been used to forecast emissions at scale, and they provide better prediction accuracy than other conventional ML methods. The extensive research on long-term global CO2 prediction revealed that Stacked LSTM, BiLSTM, and Gated Recurrent Unit (GRU) networks with attention mechanisms are more successful in comparison with classical algorithms such as Random Forest and Gradient Boosted Trees (Verma et al., 2024). These DL models are especially useful when long-term forecasting is required since they are capable of learning the temporal trends, obscure patterns, and seasonality in the data of CO2 emissions. Moreover, the hybrid networks with LSTM and GRU layers have also been found to be more effective in prediction, and this is evidenced by the Omani industries annual CO2 emissions that were better predicted using the LSTM-GRU hybrid network, which produced the lowest prediction errors and the highest levels of correlation (Wang et al., 2025). The superiority of these models shows the significance of hybrid DL models in the emission trend analysis.

Deep learning methods have also been used to solve sector-specific prediction problems, including maritime operations and atmospheric CO2 variability. The DL-based models were trained with onboard ship measurements showed good predictive capability of multi-fuel propulsion emissions in maritime systems, which is useful in the intelligent voyage planning and real-time fuel optimization (Lee et al., 2024). In the meantime, atmospheric CO2 variability simulated using DLs has been able to recreate large proportions of observed variations through the combination of meteorological data, climate indicators, and vegetation indicators (Girach et al., 2022). Nevertheless, these models also demonstrate the shortcomings in the representation of the sudden spikes in emissions due to local events, indicating that it is possible to include further anthropogenic variables. Altogether, the flexibility and accuracy of DL models render them the instrumental tool to predict CO2 emissions in the short, medium, and long term with high precision in various sectors.

2.3 Hybrid AI Frameworks and Intelligent Optimization Techniques in CO_2 Forecasting

Hybrid artificial intelligence (AI) systems have become a possible solution to improve the predictability and stability of CO2 emission models. These frameworks integrate various algorithmic advantages including statistical modelling, machine learning, neural networks, and optimization to surmount the shortcomings of single-model frameworks. Examples include the combination of ARIMA to extract the linear trend and Backpropagation Neural Network (BPNN) to extract the nonlinear residual as demonstrated to enhance forecasting of urban emissions (median and long-run). The hybrid ARIMA-BPNN model was useful in identifying the most important drivers like electricity consumption and its predictions were more stable than single models (Hou and Chen, 2024). This hybridization can accommodate both structured and unstructured emission patterns, which is why it is very appropriate in the urban and regional environmental planning.

Hybrid models that are implemented using optimization also play a significant role in the refinement of the accuracy of prediction by the intelligent tuning of the model parameters. One of them is the Extreme Learning Machine (ELM) that is optimized with the use of the INFO optimization algorithm that optimizes the initial weights and results in significant error reduction and model (Feda et al., 2024). This hybrid system has seen great performance in various measures of evaluation thus showing the significance of economic growth, openness to trade and technological innovation as significant factors of emissions. Moreover, grey machine learning models that are optimized using methods like the Augmented Crow Search Algorithm have been proved to be more accurate in predicting CO2 emissions and energy consumption when limited-data conditions are used (Saxena et al., 2023). The increasing demand of such AI models with optimal computational and predictive reliability is emphasized by these optimization-based hybrid frameworks.

Another notable advancement in the hybrid AI systems is ensemble intelligence whereby two or more models are integrated to make better predictions that are more accurate and generalized. As an example, a combination of FFNN, ANFIS, and LSTM networks was trained to predict the CO2 emissions in Saudi Arabia, where the ensemble was much more effective than each individual model and had a high degree of coincidence with the real values (Nassef et al., 2023). Ensemble strategies minimize uncertainties by combining the power of rule-based system, neural networks, and time-series predictors into one forecasting system. These methods are becoming more useful in policy making scenarios where better predictions can allow governments to develop accurate plans of reducing emissions. In general, hybrid AI and optimization-based models are a strong development of CO2 forecasting, which provides increased accuracy, flexibility, and practical usefulness in a variety of environmental modelling situations.

2.4 Research Gaps

Although the current research has made great advancements in predicting CO2 emission with the application of machine learning, deep learning, and hybrid AI, there are still a number of gaps and limitations present. Most of the previous research works are based on conventional ML or RNN-based models, which tend to be ineffective at capturing long-range interactions, processing high-dimensional tabular data or even being able to perform when missing or noisy inputs are present (Li et al., 2022; Girach et al., 2022). Other problems with deep learning models, including LSTM and GRU, include poor computational efficiency, slow learning, and inability to scale to large datasets, particularly data with mixed numerical and categorical variables (Verma et al., 2024). Hybrid systems are more precise, but can be difficult to train manually because of feature engineering, have complicated architectures, and cannot be generalized to other regions or emission sectors (Hou and Chen, 2024; Feda et al., 2024). In addition, a significant number of current models are not very effective in modelling sudden changes in emissions due to sudden socio-economic changes, industrial occurrences or localized anomalies. These disadvantages underscore the importance of a more scalable, flexible and data-efficient modelling approach.

In order to overcome these issues, the FT-Transformer, a novel transformer-based model tailored to tabular data, has great potential to address the weaknesses of the previous models. FT-Transformer uses self-attention to learn short- and long-range dependencies better than RNN models, enabling it to learn rich correlations between environmental, socioeconomic, transportation, and industrial variables all at the same time. It does not need sequential processing of data as LSTMs and GRUs, and thus it can be trained faster and scale better to high-dimensional data sets. The fact that it can handle both categorical and numerical data by embedding layers does not require a lot of manual feature engineering that traditional and hybrid models frequently rely on. Moreover, FT-Transformer is more resilient to missing values since its attention mechanism is capable of prioritizing the most important features, and therefore it is more resilient to real-world CO2 datasets which are often characterized by inconsistencies.

In addition, FT-Transformer is very appropriate when it comes to the modelling of sudden changes and nonlinear trends of the emissions data. The model is able to concentrate on several factors of the data at the same time through its multi-head attention system on meteorological indicators, industrial outputs, transportation intensity and economic activity which provide a more detailed view of the underlying emission dynamics. This enables FT-Transformer to be superior to the traditional ML, RNN, and hybrid systems in cases of sudden changes, extreme events, or multi-source data. It is also better through attention-weight visualisation, which allows researchers and policymakers to determine more effectively the driving forces of CO2 emissions compared to black-box RNN models. Thus, FT-Transformer has a bright future in closing the current research gaps as it represents a more precise, scalable, and generalisable approach to predicting CO2 emissions in sectors and over time.

3 Research Methodology

3.1 Methodology Workflow for CO_2 Emission Prediction

This work is based on a systematic end-to-end process to create a scalable and robust model of CO2 emissions prediction with the FT-Transformer. It begins with the research objective, which is filling gaps in previous studies on CO2 prediction by enhancing accuracy and generalization of the studies under varied conditions. The paper then analyses datasets with transportation variables, industrial activity, economic variables and environmental variables to learn feature behaviour, missing values, data quality and patterns across variables which is crucial since the emission datasets are usually noisy and heterogeneous.

The dataset is then made to suit the needs of attention-based tabular learning after comprehending the data. This involves data cleaning, normalization, categorical encoding and feature consolidation to guarantee the FT-Transformer to receive consistent numeric and categorical inputs. This model is then iteratively trained by optimizing the main settings of the model which include embedding size, number of attention heads, learning rate and depth of the layers so as to enhance the predictive performance. The last model is judged based on MAE, RMSE, and R 2 and compared with the baseline models, and the results are displayed in a readable format through the application of attention-based importance signals to highlight the most important emission drivers to be used in research and policy.

3.2 Dataset Description

The data utilized in this analysis will be the vehicle-level records that are detailed and contain both the technical specifications of the vehicle and the fuel consumption characteristics which are the main predictors of CO2 emission estimation. It comprises of vehicle make and model, vehicle class, engine size (L), number of cylinders, type of transmission, and type of fuel. Such parameters as operational features as fuel consumption in city, highway and combined driving (L/100 km), combined fuel efficiency (mpg) give the necessary evidence of the real performance. The target variable, CO2 emissions (g/km), is also available in the dataset thus facilitating supervised learning to predictive modelling. The sample that has been provided includes a wide range of vehicles and their types, which include compact cars, mid-size sedans, SUVs, sports cars and luxury cars of such manufacturers like Acura, Alfa Romeo, Aston Martin and Audi. This variety is beneficial with respect to the fact that the FT-Transformer model will be trained on a variety of engine configurations and driving conditions, which will enhance its generalization ability to predict CO2 across different types of vehicles.

3.3 Data Preprocessing

The preprocessing stage commenced with the importation of the dataset and inspection of its fundamental structure to make sure that it was properly loaded and could be used to further

modelling. The first inspection entailed getting to know the shape of the dataset, seeing the sample records, and verifying the type of the column to ensure that all the features have been identified right. The specifications of the vehicles, fuel consumption features, and the target variable which was the CO2 emissions were identified after a thorough analysis of the dataset. This move formed the basis of establishing the required transformations that need to be done prior to using the FT-Transformer model.

The intensive evaluation was performed to identify the gaps in all variables. It was discovered that the dataset was complete and there are no missing values in any of the columns. Nevertheless, a record redundancy investigation showed there are a significant number of duplicated rows, which would be a bias in the learning process and prediction accuracy distortion. These redundant records were deleted to avoid misrepresenting the models and created a more accurate and cleaner dataset. This optimization made sure that the model would be able to learn relevant patterns as opposed to reading the same records over and over again.

The dataset had a number of categorical variables such as make of the vehicle, model of the vehicle, type of the vehicle, type of transmission, and the type of fuel. As machine learning and transformer-based architectures demand numerical inputs, these categorical features were converted to numerical ones. Every one of the categories in these columns was assigned a distinct numeric value which enabled the model to handle them appropriately without losing the inherent differences between the various types and configurations of vehicles. This encoding procedure facilitated easy merging of categorical and continuous variables that is necessary in training the FT-Transformer.

After a thorough consideration, a number of fuel consumption variables were eliminated in order to simplify the dataset and avoid redundancy. These were city, highway, combined consumption in litters per 100 km, and combined fuel efficiency in miles per gallon. As these features are closely correlated with engine features and CO2 emissions, including all of them may cause multicollinearity and may fail to add value to the model. The elimination of these characteristics guaranteed a smaller set of features, yet these were the main predictors that are required to offer efficient emission forecasting.

The numerical features and target variable were normalized by a scaling technique which brings all values to a common range before the training. This is especially needed in neural models since it stabilizes training, avoids dominance of gradient, and enhances convergence. Upon scaling, the data was split into a training and a testing set, so that the model would be able to test on unseen data. The preprocessing pipeline was applied to a predetermined random state, which was necessary to guarantee reproducibility, so that the same split and transformations would be obtained in a repeat experiment.

3.4 Exploratory Data Analysis (EDA)

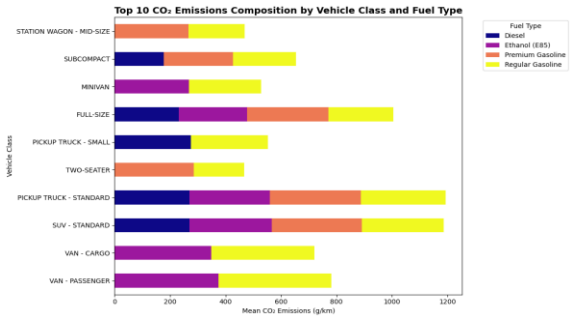

Figure 1. CO_2 Emissions by Vehicle Class

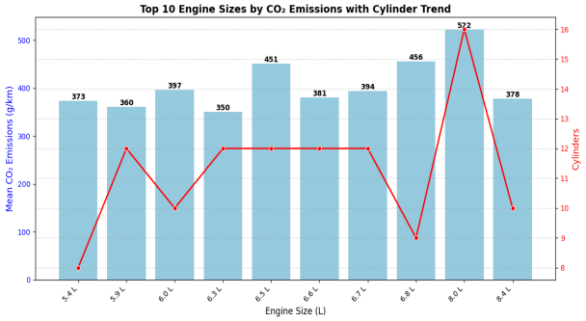

Figure 2. CO_2 Emissions by Engine Size

Figure 1 demonstrates the plot of the most CO2-emitting vehicle classes and how the various types of fuel are contributing to the total amount of CO2 emissions in each category. It emphasizes that bigger vehicles like regular SUVs, pickup trucks, and full-size cars produce a much more important amount of CO2, but regular gasoline is the largest emitter in most of the categories. Figure 2 plot illustrates that engine size has a strong relationship with the increased CO2 emissions, and the trend of cylinder count superimposed on the CO2 curve indicates that the rise of the engine capacity and the number of cylinders increases the level of emissions.

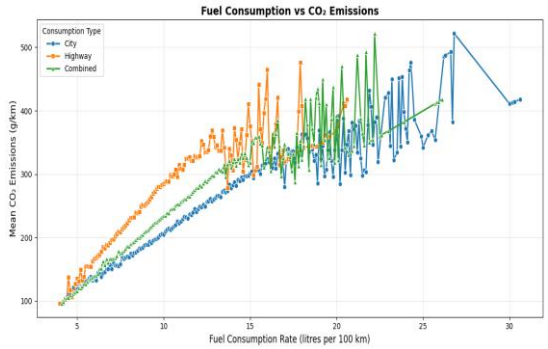

Figure 3. Fuel Consumption vs CO_2 Emissions

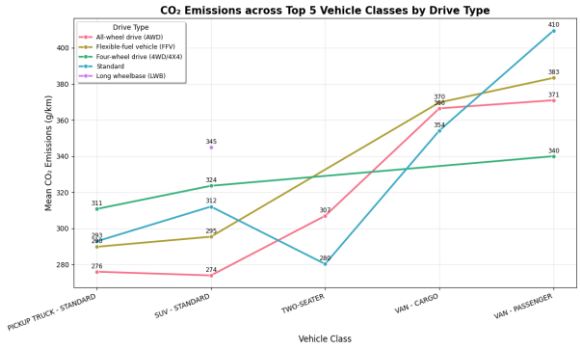

Figure 4. CO_2 Emissions by Drive Type

In figure 3, plot indicates that CO2 emissions are directly proportional to fuel consumption in the city, highway, and combined driving conditions implying that the higher the fuel consumption the higher the emissions. It further emphasizes the fact that urban driving tends to create the highest increase in CO2 emissions because it involves more frequent acceleration and changes in loads. The emissions of a vehicle by different classes and types of drives are compared in figure 4, which shows that different configurations of 4WD, long-wheelbase systems, etc., result in significantly higher CO2 emissions, particularly in larger vehicle classes like vans and trucks.

3.5 FT-Transformer-Based Framework for CO_2 Emission Prediction

In order to address the disadvantages of the existing works, this paper will apply the FT-Transformer architecture to create a more practical and scalable CO2 emission prediction model. As pointed out in the corresponding work, past machine learning, deep learning, and hybrid AI methods have shortcomings in the form of the inability to work with high-dimensional tabular data, sensitivity to missing data, manual feature engineering, and inability to model sudden changes in emissions. FT-Transformer specifically resolves these problems by using a self-attention mechanism which models both short-range and long-range associations without processing sequential data. Its embedding nature of numerical and categorical variables makes it able to provide more feature representations, which makes it avoid manual pre-processing. Moreover, the attention layers also tend to concentrate on the most informative features, which enhance resistance to noise and incomplete real-world CO2 data. The leverage model will strive to present more precise, generalizable, and explainable emission forecasts in a variety of environmental, industrial, and socioeconomic settings by adopting this architecture.

The general approach involves a number of major steps that are data acquisition, preprocessing, feature embedding, model training and evaluation. First, the datasets that are related to CO2 such as transportation, industrial activity, economic variables, and environmental variables are gathered and refined to solve the problem of inconsistency, outliers, and gaps. Finally, learned embeddings are obtained on all input features, allowing FT-Transformer to handle mixed data types in the same way. In training, the model uses multi-head self-attention to learn complicated associations among characteristics as layer normalization and elements of the feed forward make the optimization consistent. Attention head, embedding dimensions, learning rate, and batch size are hyperparameters that are optimized on validation datasets to achieve the best performance. Lastly, the model is assessed based on such metrics as RMSE, MAE, and R2, which can be compared to the baseline ML, DL, and hybrid models. By means of such systematic approach, the leverage FT-Transformer framework will address the shortcomings of the previous methods and provide a more precise and scalable prediction model.

3.6 FT-Transformer Architecture and Its Application to CO_2 Emission Prediction

The FT-Transformer model is constructed based on the transformer architecture that was initially created to handle natural language processing tasks, but specifically modified to work with high-dimensional tabular data. Contrary to the classical models of deep learning like LSTMs or GRUs, where the input is given in a sequence, FT-Transformer works with all features at once with a mixture of numeric and categorical embeddings. Categorical features are encoded as learned vectors; numerical variables are encoded using special linear encodings that maintain continuous relationships between values. These embeddings create a single feature matrix which can be processed by transformer layers. The FT-Transformer can represent the complicated interdependences between heterogeneous variables with this

structure and thus is highly applicable to CO2 emission data that incorporates vehicle characteristics, engine features, and operational features.

The multi-head self-attention mechanism is at the heart of the FT-Transformer, as it enables the model to understand the features of inputs that are the most important in predicting CO2 emissions. Self-attention calculates pairwise associations among all features such that the model can dynamically assign them a certain weight instead of following a sequence of assumptions or a predetermined hierarchy. As an example, attention layers can discover the interaction between engine size and vehicle type, how fuel type will alter fuel consumption behaviour, or how transmission type will affect real-world emission behaviour. The attention heads learn different patterns or feature relationships and their results are a complete learned representation of the emission drivers. This ability to learn the nonlinear and non-additive interactions on tabular data makes the FT-Transformer have a great edge compared to tree-based models and the standard neural networks.

The FT-Transformer also has a number of architectural improvements that enhance the robustness, scaling and comparative to the old methods. Feed-forward networks and layer normalization stabilize training and enable the model to work with large datasets without the vanishing or exploding gradient issues that RNNs are prone to. Its embedding nature feature processing lessens the manual feature engineering, and its attention scores give an inbuilt one by showing which variables affect CO2 predictions the most. This is especially useful in environmental modelling, where policymakers need to have clear understanding of the most important determinants of emissions. In general, the FT-Transformer is a versatile and strong design that can handle mixed data types, learn complicated feature interactions and provide high-accuracy CO2 emission predictions in a wide range of vehicle types and operating situations.

3.7 Evaluation Metrics and Validation Strategy

This paper will assess the FT-Transformer model based on three main regression measures, namely Mean Absolute Error (MAE), Root Mean Squared error (RMSE) and the coefficient of determination (R2). The choice of these metrics is based on the fact that they are effective in the measurement of prediction error, as well as the model accuracy of continuous emission values. The performance of the model is checked on unseen data by using a standard train-test split. The validation strategy is carried out in all experiments in a similar manner. This would make sure that the model is tested on the basis of robust and comparable performance measures.

4 Design Specification

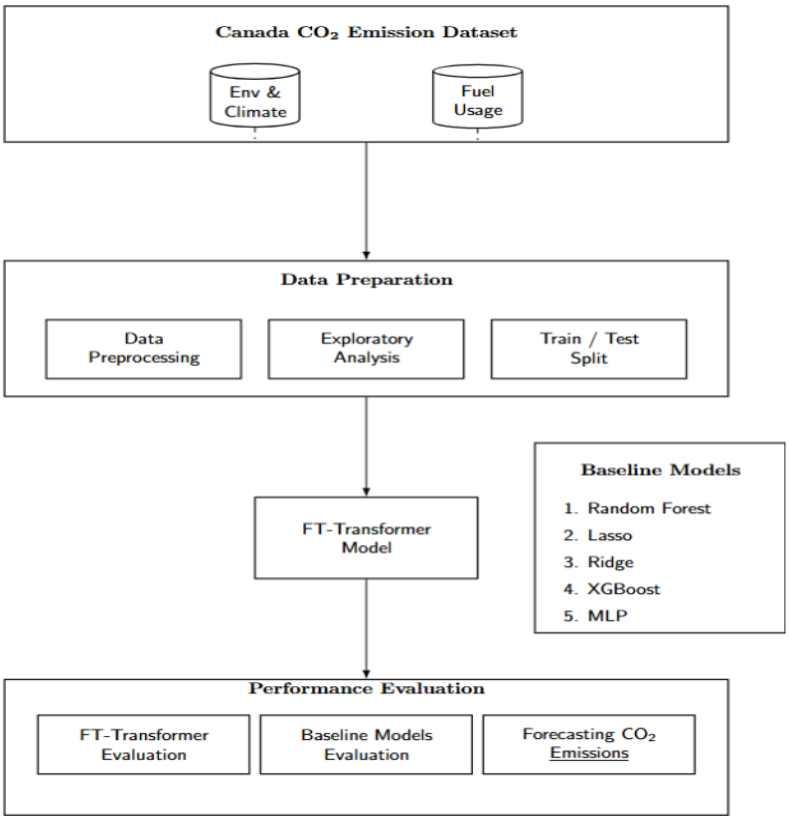

Figure 5. Architecture Overview

4.1 System Architecture Overview

The system architecture is structured into a lean as illustrated in figure 5, end-to-end pipeline that starts with the gathering of Canada CO2 emission dataset which comprises environmental, climatic as well as fuel-usage characteristics vital in analysing the emissions. This data is fed into a structured data preparation phase where preprocessing, exploratory data analysis and train-test splitting is performed to guarantee that the data is quality and ready to be used in modelling. The FT-Transformer model is at the centre of the architecture and it processes the cleaned and encoded tabular data by its attention-based mechanism. To make the performance benchmarking fare, some baseline models such as Random Forest, Lasso, Ridge, XGBoost and MLP are also trained using the same dataset. Lastly, all the outputs of the models are transferred to the performance evaluation phase where the accuracy of predictions is evaluated and the most successful model to utilize in CO2 prediction is determined. It is a high-dimensional tabular emissions architecture that is scaled and modular.

4.2 Workflow and Pipeline Implementation

The implementation of the workflow and pipeline has a systematic flow starting with the ingestion of the data and its validation where each feature of the dataset is properly formatted and consistent. Once loaded with the data, the pipeline runs some preprocessing steps which

include removing duplicates, categorical encoding, normalization, and feature selection and then runs exploratory analysis to gain insight into patterns, correlations, and trends of emissions. The processed dataset is then divided into training and testing subsets and then the FT-Transformer model is trained on embedded feature representations and attention mechanisms with the baseline models being trained simultaneously. After training the models, predictions are made and forwarded to an evaluation module which calculates measurements such as RMSE, MAE and R2. This workflow is used in handling data efficiently, making sure that it is reproducible and reliable in performance assessment of raw data up to final emission forecasting.

5 Implementation

5.1 Machine Learning Models Implementation

In the case of the Random Forest Regressor, hyperparameter optimization was performed by means of the GridSearchCV to identify the most suitable value of decision trees in the ensemble. Only number of estimators of 50, 80, 100, 150, 200 were used in the parameter grid and other model settings were held constant to allow controlled experimentation. Upon cross-validation over all the configurations, the model with number of estimators gave the highest level of performance which was 100 meaning that this number of trees gave the best trade-off between learning depth and ensemble stability as depicted in figure 6.

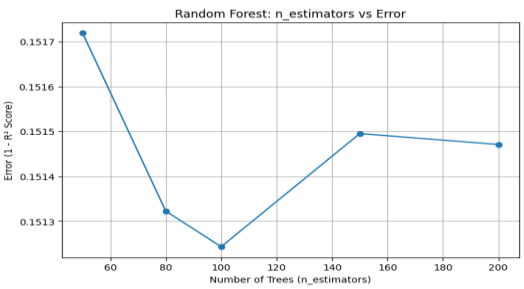

Figure 6. RF Error vs. Estimators

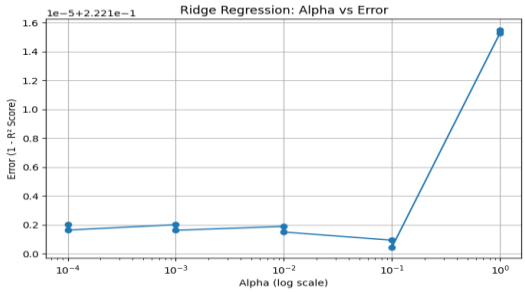

Figure 7. Ridge: Alpha vs. Error

In the case of Ridge Regression model presented in figure 7, tuning procedure was directed to finding the optimal combination of regularization strength and solver method. The parameter grid that has been evaluated on various alpha values is 0.0001, 0.001, 0.1, 1.0 and two solver options (auto and saga). The cross-validation search revealed the best configuration to be alpha is 0.1 using the saga solver, which is the most appropriate regularization and optimization strategy to use on the dataset at hand.

The Lasso Regression model was hyperparameter tuned to ensure that it has a better regularization control and convergence. The grid search that has been investigated on alpha values is 0.0001, 0.001, 0.1 with various mixite settings to make the fitting stable. It was found that the alpha is 0.0001 with mixite to be the best performing configuration and this

implies that the light regularization and moderate iteration limit gave the optimum model setting.

In the case of the XGBoost Regressor, the hyperparameter optimization was performed on the important boosting parameters that affect the learning rate and the complexity of the model. The parameter grid that comprised a set of values of number of estimators was 30, 50, 80, 100, 150, 200 and two learning rates, which allowed exploration of boosting iterations and update magnitudes systematically. After analysis, it was determined that the best configuration is number of estimators is 200 with a learning rate as the most successful combination to represent the nonlinear feature relationships in the data set.

5.2 Deep Learning Models Implementation

The deep learning model in this analysis is a high-accuracy Multilayer Perceptron (MLP) that was trained in PyTorch to acquire a nonlinear relationship between the CO2 emission dataset. PyTorch tensors representing the input features and target values were then created and a deep fully connected, architecture was implemented with sequential linear layers of hidden dimensions of 256, 128, 64, and 32 with each layer being activated using LeakyReLU to enhance gradient flow and model stability. The final linear layer produced continuous emission predictions. Model training was conducted using the Adam optimizer with a learning rate of 0.0015 and a weight decay of 1e-6, while Mean Squared Error (MSE) was employed as the loss function. A StepLR scheduler was used to reduce the learning rate every 80 epochs, and an early-stopping mechanism with a patience of 40 epochs was applied to prevent overfitting. Training was performed for a maximum of 600 epochs, and the best-performing model state was saved based on the lowest training loss. Following training, the final model was restored for inference, and inverse scaling was applied to both predicted and true CO_2 values to return them to original units. A training loss curve was also generated to visualize the model's convergence behaviour throughout the training process.

5.3 FT-Transformer Implementation

The implementation of the FT-Transformer was based on the PyTorch to take advantage of its transformer-based model to predict tabular CO2 emissions. The preprocessing phase involved standardization of input features and target values with the use of two different StandardScaler objects in order to achieve consistent optimization. A custom dataset class and Data Loader objects were made so that they can facilitate efficient batch processing in the training. The model architecture was a feature tokenizer which transformed each numerical feature into a learnable token embedding, and then a learnable CLS token which represented the globe was added. The main Transformer encoder was a series of stacked layers with the multi-head self-attention, feed-forward networks, GELU activation and dropout regularization so that the model could learn intricate inter-feature interactions. The CLS embedding was converted into the final regression output by a prediction head that used layer normalization, linear projection and GELU activation. The AdamW optimizer was used with the learning rate of 0.001 and a ReduceLROnPlateau scheduler, which varies the

learning rates according to the validation performance. A patience of 15 epochs was used to provide an early-stopping mechanism that prevented overfitting by monitoring the validation loss and keeping the best-performing model weights. The training loop was continued until the completion of 200 mini-batch updates, and the best state of the model stored was then loaded to be used during inference. The predictions were generated in evaluation mode and inverted to original units and compared to training and validation loss curves to see the learning mechanism and stability of the FT-Transformer model.

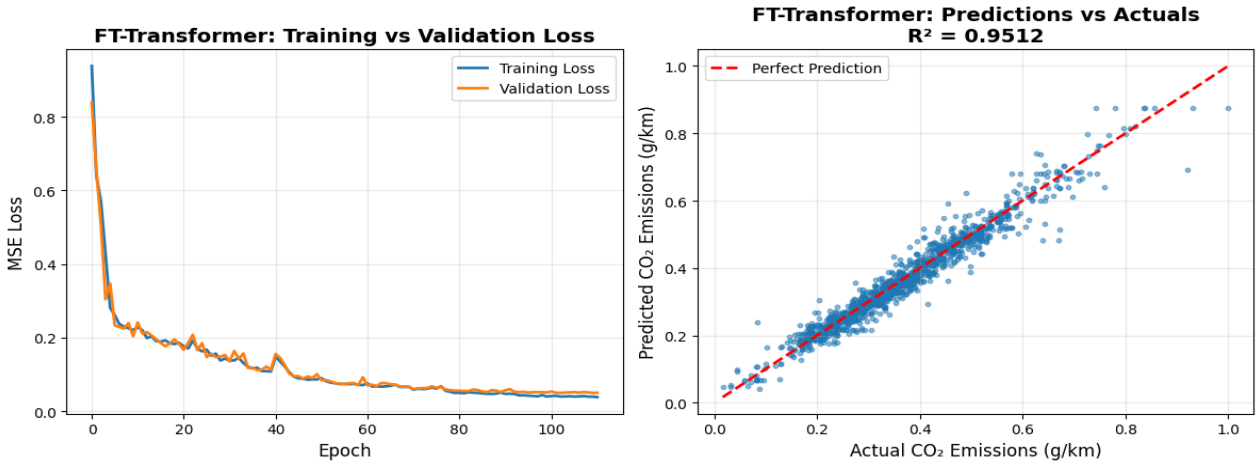

Figure 8. FT-Transformer Performance Overview

As can be seen in figure 8, the left plot indicates that training and validation loss reduces consistently with the number of epochs, which implies good learning with avoiding overfitting. The right plot pits the predicted and the actual CO2 emissions and indicates points that are very close to the diagonal line of perfect prediction. R2 is high (0.9512) which reveals that the model has a high level of predictability of CO2 emissions by the FT-Transformer model.

5.4 Libraries Used

The research used some of the most important Python libraries in order to assist in data processing, model training and evaluation. Preprocessing, scaling, train-test splitting, hyperparameter tuning and evaluation metrics were done using scikit-learn. PyTorch was used as the fundamental deep learning platform to run and train the MLP and FT-Transformer models and it offered neural network layers, optimizers, schedules and acceleration on a GPU. The gradient-boosting baseline model was implemented using XGBoost, and the NumPy was used to perform numerical computations and array operations. The use of Matplotlib was in creating visualizations like training-loss curves and prediction plots. These libraries combined to offer a consistent and effective platform to construct and compare the machine learning and deep learning models that were utilized in this study.

6 Evaluation

In this section, all the machine learning and deep learning methods are compared and contrasted before analyzing the individual models, which are applied in making predictions regarding CO2 emission. Standard regression measures such as MAE, MSE, RMSE, and R2 score were used to evaluate the models and a comparable and significant comparison of their predictive ability was made. The results of each type of the models are given in the following subsections, with their strengths, weaknesses, and general effectiveness in the research as indicated in table 1.

Table 1. Model Performance Comparison Across Regression Metrics

Model	MAE	MSE	RMSE	R2 Score
Random Forest	0.0401	0.0029	0.0538	0.8543
Lasso	0.0495	0.0042	0.0649	0.7883
Ridge	0.0495	0.0042	0.0649	0.7885
XGBoost	0.0388	0.0028	0.0525	0.8614
MLP	18.2487	563.4910	23.738	0.8439
FT-Transformer	0.0219	0.0010	0.0312	0.9512

6.1 Machine Learning Model Results

The tuned Random Forest model showed a high level of predictive ability with an MAE of 0.0401, MSE of 0.0029, RMSE of 0.0538 and a R2 value of 0.8543. These findings suggest that the ensemble method was able to obtain the nonlinear relationships in the data, and it was able to generate precise predictions with minimal error. The low MAE and RMSE values indicate that the model has always forecasted the CO2 emissions near to the actual values, whereas the large R2 value indicates that the model can explain a significant amount of variation in the emission values. This experimental outcome shows the strength of Random Forest as a baseline predictor of CO2, especially because of its ability to deal with feature interactions and overfitting with averaging over several trees.

The Ridge Regression model produced an MAE of 0.0495, an MSE of 0.0042, RMSE of 0.0649, and a score of 0.7885 in R 2, which is moderate in terms of predictive power when

contrasted with ensemble techniques. Being a regularized linear model, Ridge Regression is useful in controlling multicollinearity and stabilizing coefficient estimates, but is restricted by its linearity in controlling complex nonlinear patterns in the data. The reduced prediction errors and decreased R2 value compared to Random Forest indicate that the correlation between the characteristics of the vehicle and emissions of CO2 is not linear, and thus, cannot be completely represented by the use of linear regularization. Nevertheless, Ridge Regression offers a stable and benchmark model.

In the Lasso Regression model, the performance metrics were very close to the Ridge Regression with an MAE of 0.0495, MSE of 0.0042, RMSE of 0.0649 and R2 of 0.7883. The sparsity-inducing regularization used by Lasso can be used to determine the most significant features by shrinking less significant coefficients to zero, but it has the same weakness as Ridge Regression when it comes to modelling nonlinear interactions. The fact that the two models exhibit almost identical performance is an indication that linear predictors are not good enough to capture the complete complexity of the patterns of CO2 emissions, which supports the use of more complex or nonlinear machine learning methods. Nevertheless, Lasso will provide the option of selecting the features and help to comprehend which variables have a significant impact on emissions.

The XGBoost model with the best overall performance of the machine learning methods had an MAE of 0.0388, MSE of 0.0028, RMSE of 0.0525 and R2 of 0.8614. These findings indicate that XGBoost has a high capacity to acquire complex, nonlinear relationships and interaction by use of gradient-boosted decision trees. Its R2 and error values are slightly lower than that of the random forest implying a greater accuracy in prediction, probably because of its boosting structure that corrects the mistakes made in the previous stages. The high level of performance of XGBoost justifies the use of advanced boosting methods in the model of CO2 emissions and its applicability in the high-performing baseline model before switching to deep learning and transformer-based models.

6.2 Deep Learning Model Results

The Multilayer Perceptron (MLP) model (tuned) exhibited a good predictive ability and the resulting MAE of 18.2487, MSE of 563.4910, RMSE of 23.738, and R2 value of 0.8439 were attained when tested on the original unit of CO2 emissions. These findings suggest that the deep neural network could learn meaningful nonlinear correlations between vehicle characteristics and emissions, and it is able to perform similarly to the most successful machine learning models. The moderate MAE and RMSE values indicate that the model predictions remained within a relative proximity to the real emission productions within a large variety of different vehicles. At the same time, the large R2 value indicates that the MLP was able to explain a large amount of the variance in the data, which shows that the network architecture with its multi-layered hidden layers and nonlinear activations was able to capture the intricate interaction of features.

6.3 FT-Transformer Results

The FT-Transformer model has shown the best results in this study with an MAE of 0.0219, an MSE of 0.0010, an RMSE of 0.0312 and a very high R2 value of 0.9512. The findings of these results are that the model can be used to model intricate interactions between features with a high degree of accuracy, in part because of its attention-based architecture, which allows the model to concentrate on the most informative features in the tabular data. The error values were very low indicating that the predictions made by the FT-Transformer were very close to the real values of CO2 emissions and it was found to be better than both the traditional machine learning models and the deep learning MLP. The high R 2 value is also an indication that the model is in a position to explain almost all the variability in the emission data, which is an indication that the model is superior and effective in dealing with high-dimensional, mixed-type features. In general, the FT-Transformer is a very precise and strong model to predict CO2 emission, which confirms that this is an appropriate model to use to model complex datasets on the environment.

6.4 Findings of the Study and Superiority of the FT-Transformer

The results of this paper clearly indicate a performance trend through the various modelling techniques, whereby the traditional machine learning models offer fairly good accuracy, the deep learning MLP offers even better results, and FT-Transformer demonstrates an unquestionable better performance than the other models used in the research. Although the nonlinear relationships were effectively captured by models like the Random Forest and XGBoost, Ridge and Lasso were able to provide stability and none of them could possibly match the extremely high accuracy of the FT-Transformer. Its very high R2 value and very low values of error show that it has learned the underlying structure of the data on CO2 emissions much more comprehensively and evidences the better ability of the model to deal with high-dimensional, mixed-type tabular data. This advantage becomes even clearer when the performance of this model is compared to that of the MLP, since the FT-Transformer had much lower error rates even though both models have nonlinear architectures.

The excellence of FT-Transformer can be greatly explained by the fact that it has a developed system of self-attention, which provides the model to consider the relationships between all pairs of features at the same time. This enables it to model fine, intricate dependencies, including the interplay between fuel consumption behaviour, engine behaviour and vehicle type, that simple models are either too simplistic to capture or do not observe. Its feature tokenization process also improves performance by encoding each input variable into a learned embedding, which enables more informative and richer representations than simple numerical encoding. The multi-head attention structure also allows the model to simultaneously learn many different kinds of interactions, providing the model with a more holistic and detailed view of the data. Combined, these architectural developments justify why the FT-Transformer was the most accurate and showed evident superiority to machine learning and deep learning baselines when it came to predicting CO2 emissions.

7 Conclusion and Future Work

7.1 Conclusion

This paper has managed to outline a complete CO2 emission forecasting model that evaluated the performance of conventional machine learning models, a deep learning model of multilayer perceiving, and a more sophisticated model of the FT-Transformer, and finally confirmed that the FT-Transformer outperforms the other two models in terms of predictive accuracy of high-dimensional tabular data. The systematic preprocessing, model tuning, and evaluation revealed the findings of the research where classical linear models and tree-based algorithm identify significant patterns but are restricted when it comes to modelling the complex interactions. By contrast, the FT-Transformer made use of feature tokenization and self-attention to learn rich relationships between vehicle specifications and operational properties and this model produced the highest accuracy of all models used. The main contribution of the study is that it shows that the transformer-based architectures are effective in environmental prediction tasks and can better perform than machine learning as well as traditional deep learning methods. This work offers a solid basis on the future development of CO2 emission modelling through the creation of a robust and scalable methodology and contributes to the overall application of attention-driven architectures to sustainability analytics.

7.2 Future Work

The findings of this study can be used in future work to develop the study by examining various directions to optimally improve the prediction of CO2 emissions. The first opportunity is to combine bigger and more heterogeneous data sets, which would include more environmental, socioeconomic, and behavioural factors to enhance the model generalization in regions and vehicle types. The inclusion of time-dependent data would also make it possible to create hybrid transformer designs that are able to capture longitudinal trends and cross-feature interactions. The use of explainable AI (XAI) methods, like integrated gradients or attention-based interpretation tools, is another possible direction that would give more insight into the impacts of certain features on emission patterns. Moreover, the practical usage of FT-Transformer in real-life systems like smart transport networks or emission-monitoring platforms might confirm its usefulness in practice and aid in the real-time decision-making. Lastly, further studies can be done on model optimization techniques such as pruning, quantization, and edge deployment to minimize the cost of computation without compromising predictive accuracy.

References

Al-Nefaie, A.H. and Aldhyani, T.H., 2023. Predicting CO2 emissions from traffic vehicles for sustainable and smart environment using a deep learning model. Sustainability, 15(9), p.7615.

Bhatt, H., Davawala, M., Joshi, T., Shah, M. and Unnarkat, A., 2023. Forecasting and mitigation of global environmental carbon dioxide emission using machine learning techniques. Cleaner Chemical Engineering, 5, p.100095.

Feda, A.K., Adegboye, O.R., Agyekum, E.B., Hassan, A.S. and Kamel, S., 2024. Carbon emission prediction through the harmonization of extreme learning machine and INFO algorithm. IEEE Access, 12, pp.60310-60328.

Giannelos, S., Bellizio, F., Strbac, G. and Zhang, T., 2024. Machine learning approaches for predictions of CO2 emissions in the building sector. Electric Power Systems Research, 235, p.110735.

Girach, I.A., Ponmalar, M., Murugan, S., Rahman, P.A., Babu, S.S. and Ramachandran, R., 2022. Applicability of Machine Learning Model to Simulate Atmospheric CO_2 Variability. IEEE Transactions on Geoscience and Remote Sensing, 60, pp.1-6.

Hou, L. and Chen, H., 2024. The prediction of medium-and long-term trends in urban carbon emissions based on an ARIMA-BPNN combination model. Energies, 17(8), p.1856.

Lee, J., Eom, J., Park, J., Jo, J. and Kim, S., 2024. The development of a machine learning-based carbon emission prediction method for a multi-fuel-propelled smart ship by using onboard measurement data. Sustainability, 16(6), p.2381.

Li, X., Ren, A. and Li, Q., 2022. Exploring patterns of transportation-related CO2 emissions using machine learning methods. Sustainability, 14(8), p.4588.

Meng, Y. and Noman, H., 2022. Predicting CO2 emission footprint using AI through machine learning. Atmosphere, 13(11), p.1871.

Nagpal, A., Nanda, S., Saini, P., Varshney, S., Sachan, S. and Rekh, S., 2024, December. Quantifying Carbon Emissions: Integrating Machine Learning with IoT for Traffic-Aware Emission Scores. In 2024 1st International Conference on Sustainability and Technological Advancements in Engineering Domain (SUSTAINED) (pp. 601-607). IEEE.

Nassef, A.M., Olabi, A.G., Rezk, H. and Abdelkareem, M.A., 2023. Application of artificial intelligence to predict CO2 emissions: Critical step towards sustainable environment. Sustainability, 15(9), p.7648.

Natarajan, Y., Wadhwa, G., Sri Preethaa, K.R. and Paul, A., 2023. Forecasting carbon dioxide emissions of light-duty vehicles with different machine learning algorithms. Electronics, 12(10), p.2288.

Saxena, A., Zeineldin, R.A. and Mohamed, A.W., 2023. Development of grey machine learning models for forecasting of energy consumption, carbon emission and energy generation for the sustainable development of society. Mathematics, 11(6), p.1505.

Verma, A., Routhu, L.C., Chigrupaatii, R., Choubey, A., Misra, R. and Singh, T.N., 2024, December. Big Data-Driven Long Term Accurate CO_2 Emission Forecasting Using Advanced Recurrent Neural Networks. In 2024 IEEE International Conference on Big Data (BigData) (pp. 4438-4447). IEEE.

Wang, C., Zhang, X., Nie, Z. and Gajbhiye Meshram, S., 2025. Prediction of Annual Carbon Emissions Based on Carbon Footprints in Various Omani Industries to Draw Reduction Paths with LSTM-GRU Hybrid Model. Sustainability, 17(11), p.4940.