International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

Machine Learning–Based Framework for Accurate CO_2 Emission Prediction and Environmental Impact Analysis

Mrs.KanakaTulasi P.Reddi[1] , Jittuka Harsha Dinni Sri[2], Mohan Sai Krishna Bhuvanasi[3], Adipudi Naga Sai Sri Sowmya[4], Koruprolu Gowtham[5]

[1]Assistant Professor , [2,3,4,5] B.tech Students Department of CSE, Pragati Engineering College, Surampalem , Andhra Pradesh, India

Abstract- — The rapid increase in carbon dioxide (CO_2) emissions has become a major environmental concern due to its significant contribution to global warming and climate change. Accurate prediction of CO_2 emissions is essential for developing effective environmental policies and implementing sustainable strategies to reduce greenhouse gas emissions. Traditional statistical forecasting methods often struggle to capture complex relationships between multiple environmental and industrial factors that influence carbon emissions. In recent years, machine learning techniques have emerged as powerful tools for analysing environmental data and improving prediction accuracy.This study presents a machine learning–based framework for forecasting CO_2 emissions using historical environmental and fuel consumption data. The proposed system analyses various factors such as fuel consumption patterns, vehicle characteristics, engine size, and other related attributes to estimate future carbon emissions. Several machine learning regression algorithms, including Linear Regression, Gaussian Process Regression, Multilayer Perceptron (MLP), and Sequential Minimal Optimization for Regression (SMOreg), are implemented and evaluated to determine the most accurate prediction model.The dataset used in this research is obtained from a publicly available environmental dataset and undergoes preprocessing steps such as data cleaning, normalization, and outlier detection to improve model performance. The trained models are evaluated using performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), Root Relative Squared Error (RRSE), and correlation coefficient.Experimental results indicate that machine learning algorithms can effectively predict CO_2 emissions, with SMOreg demonstrating superior performance compared to other models in terms of prediction accuracy and error reduction. The proposed framework can assist environmental researchers and policymakers in understanding emission trends and making informed decisions for climate change mitigation.

Index Terms: Carbon Dioxide Emissions, Machine Learning, Environmental Forecasting, Regression Models, Climate Change Analysis, Sustainable Development, CO_2 Prediction.

I. INTRODUCTION

Climate change has emerged as one of the most significant global challenges in recent decades, primarily driven by the increasing concentration of greenhouse gases in the atmosphere. Among these gases, carbon dioxide (CO_2) is recognized as a major contributor to global warming due to its high emission levels resulting from human activities such as fossil fuel combustion, industrial processes, transportation, and deforestation. The continuous growth of industrialization and urbanization has significantly increased CO_2 emissions, leading to severe environmental consequences including rising global temperatures, extreme weather conditions, and disturbances in natural ecosystems. Consequently, reducing carbon emissions has become a major priority for governments, environmental organizations, and policymakers worldwide [3], [15].

Understanding and predicting CO_2 emission trends is essential for developing effective strategies to mitigate environmental damage and achieve sustainable development goals. Accurate emission forecasting enables governments and regulatory authorities to design policies that control pollution levels, regulate industrial activities, and encourage the adoption of renewable energy technologies. Reliable prediction models therefore play an important role in supporting climate change mitigation strategies and long-term environmental planning [1], [2].

Traditionally, statistical and time-series forecasting techniques have been widely used to analyse environmental datasets and estimate future emission levels. Methods such as linear regression, autoregressive models, and grey prediction models attempt to forecast emissions based on historical patterns and temporal relationships. However, these conventional approaches often struggle to handle large and complex datasets where multiple factors influence emission levels simultaneously. Environmental data are typically nonlinear and are affected by various economic, industrial, technological, and energy consumption factors, making accurate prediction difficult using traditional statistical models [7], [9].

International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

In recent years, machine learning (ML) techniques have demonstrated considerable potential in solving complex prediction problems across various domains, including environmental science and climate change analysis. Machine learning algorithms are capable of learning patterns from large datasets and identifying hidden relationships between multiple variables. By analysing historical emission data and related influencing factors, these models can produce more accurate predictions compared with conventional statistical approaches. Consequently, machine learning has been widely applied in carbon emission prediction, energy consumption modelling, and environmental monitoring applications [3], [6], [15].

Various machine learning models have been applied to environmental forecasting tasks, including Linear Regression, Gaussian Process Regression, Multilayer Perceptron (MLP), and Support Vector Machine–based regression models. These algorithms can analyse different environmental attributes such as fuel consumption rates, industrial activity levels, transportation patterns, and energy usage to estimate CO_2 emission levels. Gaussian Process Regression, in particular, has been widely used for modelling complex environmental data due to its ability to capture nonlinear relationships and uncertainty in predictions [5], [12]. Furthermore, machine learning approaches have also been successfully used for predicting air pollution and traffic-related emissions, demonstrating their effectiveness in environmental forecasting tasks [8].

Recent research has also explored the integration of machine learning techniques with climate and environmental data to improve emission prediction accuracy. For example, machine learning models have been applied to forecast CO_2 emissions in different countries and urban regions, enabling researchers to analyse emission patterns and identify major contributing factors [10], [11]. Similarly, intelligent forecasting models have been used in environmental monitoring systems and sensor-based applications such as rainfall prediction and environmental hazard detection, demonstrating the versatility of machine learning techniques in environmental studies [4], [14].

Motivated by these advancements, this research proposes a machine learning–based framework for predicting CO_2 emissions using environmental and vehicle-related datasets. The dataset includes multiple attributes associated with fuel consumption patterns and vehicle specifications that influence carbon emission levels. Prior to applying machine learning algorithms, the dataset undergoes preprocessing procedures such as data cleaning, normalization, and outlier detection to enhance data quality and improve prediction performance.

Several machine learning regression algorithms are implemented and compared in this study, including Linear Regression, Gaussian Process Regression, Multilayer Perceptron (MLP), and SMOreg. The performance of these models is evaluated using multiple evaluation metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), Root Relative Squared Error (RRSE), and correlation coefficient. Through comparative analysis, the study aims to identify the most effective algorithm for accurately predicting CO_2 emissions.

The primary objective of this research is to develop a reliable prediction framework that can assist environmental researchers, policymakers, and decision-makers in understanding carbon emission patterns and forecasting future trends. Accurate prediction of CO_2 emissions can support better decision-making for environmental protection, sustainable energy planning, and climate change mitigation initiatives.

II. LITERATURESURVEY

The growing concern over climate change and environmental sustainability has motivated extensive research on carbon dioxide (CO_2) emissions and the development of accurate prediction models. Forecasting CO_2 emissions plays a crucial role in environmental planning, climate policy development, and sustainable resource management. Researchers have explored a variety of statistical, machine learning, and artificial intelligence–based approaches to analyse environmental datasets and predict future emission trends. These predictive models assist policymakers and environmental organizations in understanding emission patterns and implementing effective mitigation strategies [3], [15].

Early studies on CO_2 emission prediction primarily relied on statistical and econometric models. Traditional techniques such as linear regression, autoregressive models, and time-series forecasting methods were widely used to analyse historical emission data and estimate future trends. These approaches attempted to establish relationships between carbon emissions and influencing factors such as energy consumption, industrial activity, and economic development. Although statistical models provided valuable insights into emission behaviour, they often struggled to capture nonlinear relationships present in complex environmental datasets and were limited in their ability to handle large-scale data [7], [9].

With the rapid advancement of artificial intelligence technologies, machine learning techniques have increasingly been applied to environmental prediction problems. Machine learning algorithms are capable of processing large datasets and identifying hidden patterns that may not be easily detected using conventional statistical methods. Several studies have applied algorithms such as Support Vector Machines (SVM), Random Forest, Decision Trees, and Artificial Neural Networks (ANN) for predicting CO_2 emissions and analysing environmental factors that influence emission patterns. These intelligent models have demonstrated improved predictive accuracy compared to traditional statistical techniques due to their ability to model nonlinear relationships within environmental datasets [3], [6], [15].

In addition to classification-based methods, regression-based machine learning models have also been widely used for carbon emission prediction. Linear regression models have been employed to study the relationship between fuel consumption and carbon emissions. However, more advanced regression techniques such as Gaussian Process Regression (GPR) have demonstrated superior performance in modelling complex environmental systems. Gaussian Process models are capable of capturing nonlinear relationships and providing uncertainty estimation in predictions, making them particularly suitable for environmental forecasting tasks [5], [12].

Recent studies have also explored neural network–based approaches for emission prediction, particularly Multilayer Perceptron (MLP) models. Neural networks are capable of learning complex relationships between multiple input variables and have been widely applied in environmental data analysis due to their ability to handle nonlinear and high-dimensional datasets. These models have demonstrated improved accuracy and robustness when predicting CO_2 emissions compared to traditional statistical and regression-based methods [6], [10], [11].

In addition to neural network models, support vector machine–based regression techniques such as Sequential Minimal Optimization regression (SMOreg) have been used to improve prediction accuracy. These models utilize kernel-based learning methods to identify complex patterns within environmental datasets and generate reliable predictions. Support vector regression models are particularly effective when working with high-dimensional datasets and nonlinear relationships, which are common characteristics of environmental data [8], [10].

Despite the progress achieved through machine learning approaches, several challenges remain in the field of CO_2 emission prediction. Environmental datasets are often highly complex and influenced by multiple economic, industrial, technological, and policy-related factors. Furthermore, identifying the most suitable machine learning model for accurate emission forecasting remains an important research challenge. Variations in dataset characteristics, model parameters, and environmental factors can significantly influence prediction performance.

Therefore, this study focuses on implementing and comparing several machine learning regression algorithms, including Linear Regression, Gaussian Process Regression, Multilayer Perceptron (MLP), and SMOreg, for predicting CO_2 emissions. The performance of these models is evaluated using multiple statistical evaluation metrics to determine their effectiveness in environmental forecasting. By identifying the most accurate predictive model, the study aims to contribute to the development of intelligent climate prediction systems that support environmental sustainability and informed policy decision-making.

III.SYSTEM ANALYSIS

A. Existing System

Traditional approaches for predicting carbon dioxide (CO_2) emissions primarily rely on statistical analysis and basic regression models to estimate future emission levels. These systems typically analyze historical environmental data and attempt to identify relationships between several influencing factors such as fuel consumption, vehicle characteristics, energy usage, industrial activities, and emission levels. Statistical techniques such as linear regression, autoregressive models, and time-series forecasting methods have been widely used to analyze emission trends and support environmental decision-making processes [7], [9].

With the advancement of artificial intelligence technologies, machine learning algorithms have been increasingly adopted to improve the accuracy of CO_2 emission prediction models. Several machine learning techniques, including Linear Regression, Gaussian Process Regression, Artificial Neural Networks, and Support Vector Machine–based regression models, have been applied to environmental datasets. These algorithms are capable of analyzing multiple variables related to fuel consumption patterns, vehicle specifications, and energy usage to estimate emission values more effectively compared to conventional statistical approaches [1], [5], [6].

Several studies have attempted to enhance prediction performance by applying machine learning models to datasets containing attributes such as engine size, number of cylinders, fuel consumption under city and highway driving conditions, vehicle type, and other environmental factors. These datasets are typically collected from environmental monitoring systems, transportation databases, or publicly available repositories and are used to train predictive models for emission forecasting. Machine learning techniques have demonstrated significant potential in identifying complex patterns within environmental datasets and improving prediction accuracy [3], [8].

Despite these advancements, many existing CO_2 emission prediction systems still encounter several limitations related to prediction accuracy, model complexity, and interpretability. Environmental datasets often exhibit complex nonlinear relationships among variables such as energy consumption, economic growth, and transportation patterns. Traditional statistical models and simple machine learning techniques may struggle to effectively capture these nonlinear interactions, resulting in reduced prediction performance [6], [10].

Furthermore, many existing studies focus primarily on developing individual machine learning models rather than conducting comprehensive comparisons among multiple algorithms to determine the most effective approach for CO_2 emission forecasting. Evaluating multiple predictive models is essential to identify the most accurate and reliable technique for environmental data analysis and emission prediction[11], [15].

Disadvantages Of The Existing System

• Limited prediction accuracy:
Traditional statistical models often struggle to capture complex nonlinear relationships present in environmental datasets, which can reduce the accuracy of emission forecasts [7].

• Overfitting and underfitting issues:
Machine learning models may either memorize training data (overfitting) or fail to learn important patterns (underfitting), leading to unreliable prediction results [6].

• High computational requirements:
Some advanced machine learning and neural network models require significant computational resources and longer training times, particularly when dealing with large environmental datasets [15].

• Limited feature understanding:
Existing models often fail to thoroughly analyse the interactions between multiple environmental factors influencing CO_2 emissions, which may limit model interpretability [3].

• Scalability challenges:
Environmental datasets continue to grow due to increasing monitoring systems and sensor technologies, making it difficult for traditional models to efficiently process large volumes of data [4].

• Lack of model comparison:
Many existing systems rely on a single predictive model instead of evaluating multiple algorithms to determine the most effective approach for CO_2 emission forecasting [11].

• Difficulty in handling noisy data:
Environmental datasets often contain noise, missing values, and outliers that may negatively affect model training and prediction performance if not properly handled [8].

B. Proposed System

To address the limitations of existing CO_2 emission prediction systems, this study proposes a machine learning–based framework designed to improve the accuracy and reliability of carbon emission forecasting. The proposed system utilizes environmental and vehicle-related datasets to analyse the factors influencing carbon emissions and generate precise prediction results. Machine learning techniques have demonstrated strong potential in environmental forecasting tasks due to their ability to model nonlinear relationships and process large datasets efficiently [1], [3].

The dataset used in this study contains multiple attributes related to vehicle specifications and fuel consumption patterns. These attributes include engine size, number of cylinders, fuel consumption in city driving conditions, fuel consumption on highways, combined fuel consumption, and vehicle class. Such features provide valuable information about vehicle performance and energy usage, which are key factors affecting CO_2 emission levels. By analysing these attributes, machine learning models can effectively learn the relationships between vehicle characteristics and emission outputs [6], [8].

Before applying machine learning algorithms, the dataset undergoes several preprocessing steps to ensure data quality and consistency. These preprocessing procedures include data cleaning, removal of missing values, outlier detection, normalization, and feature selection. Data preprocessing is an essential step in machine learning workflows because environmental datasets often contain noise, incomplete records, and inconsistencies that can negatively impact model performance if not properly handled [3], [4].

International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

After preprocessing, the dataset is divided into training and testing subsets to evaluate the predictive capability of the models. Approximately 70% of the data is used for training the machine learning models, while the remaining 30% is reserved for testing and validation. This data-splitting strategy allows the models to learn patterns from historical data while ensuring that prediction performance can be evaluated on unseen samples.

Several machine learning regression algorithms are implemented in this study, including Linear Regression, Gaussian Process Regression, Multilayer Perceptron (MLP), and SMOreg. These algorithms are widely used in environmental prediction and regression tasks due to their ability to capture relationships between multiple variables. In particular, Gaussian Process Regression is known for its capability to model nonlinear relationships and provide uncertainty estimation in predictions [5], [12].

Each algorithm is trained using the prepared dataset to learn the relationship between vehicle characteristics and CO_2 emission levels. After training, the models are evaluated using several statistical performance metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), Root Relative Squared Error (RRSE), and correlation coefficient. These evaluation metrics provide a comprehensive assessment of prediction accuracy and model reliability [10], [11].

The proposed system performs a comparative analysis of different machine learning models to determine the most effective algorithm for predicting CO_2 emissions. By analysing the evaluation results, the study identifies the model that achieves the highest prediction accuracy and the lowest error values. Comparative model evaluation is essential for selecting an optimal forecasting approach in environmental prediction systems [15].

The developed machine learning–based framework can assist environmental researchers, policymakers, and regulatory organizations in understanding carbon emission patterns and developing effective strategies to reduce greenhouse gas emissions. Accurate prediction of CO_2 emissions supports climate change mitigation initiatives, sustainable energy planning, and evidence-based environmental policymaking.

IV. SYSTEM DESIGN

System Architecture

Below diagram depicts the whole system architecture.

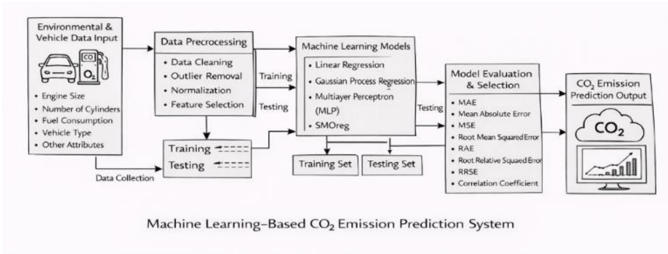

Fig 1. Methodology followed for proposed model

V. SYSTEM IMPLEMENTATION

Modules

The implementation of the proposed CO_2 emission prediction framework is organized into several modules, including data collection and preprocessing, feature selection and feature engineering, machine learning model training, emission prediction, and model evaluation. These modules collectively enable accurate analysis of environmental datasets and improve the prediction of carbon dioxide emission levels.

1. Data Collection and Preprocessing

The first stage of the system involves collecting environmental and vehicle-related datasets used for predicting carbon dioxide (CO_2) emissions. The dataset contains multiple attributes associated with vehicle specifications and fuel consumption patterns, including engine size, number of cylinders, fuel consumption under city driving conditions, fuel consumption on highways, combined fuel consumption, vehicle class, and model year. These attributes are important indicators of vehicle performance and are closely related to carbon emission levels.

Before training machine learning models, the dataset undergoes several preprocessing procedures to ensure data quality and consistency. These preprocessing steps include data cleaning, removal of missing values, outlier detection, and normalization of numerical features. Environmental datasets often contain noise, incomplete values, and inconsistencies that may negatively impact model performance if not properly addressed. Therefore, preprocessing is an essential step for improving the reliability and accuracy of machine learning predictions [3], [4].

Additionally, irrelevant or redundant attributes are removed, and relevant features are selected to enhance model efficiency. This process helps reduce data complexity and ensures that the machine learning algorithms focus on the most informative variables influencing CO_2 emissions.

2. Feature Selection and Feature Engineering

Feature selection is performed to identify the most significant variables influencing carbon dioxide emissions. Attributes such as fuel consumption rates, engine capacity, vehicle class, and other vehicle characteristics play a major role in determining emission levels. By analysing these variables, the system identifies the features that contribute most significantly to emission prediction.

Feature engineering techniques are also applied to transform raw data into more meaningful and informative features that improve model learning. These techniques may include feature scaling, transformation of numerical variables, and the creation of derived attributes. Feature engineering helps reduce dimensionality and enhances the predictive capability of machine learning models by emphasizing the most relevant information within the dataset [6], [8].

3. Training Machine Learning Models

After preprocessing and feature preparation, the dataset is divided into training and testing subsets. Approximately 70% of the dataset is used for training the machine learning models, while the remaining 30% is used for testing and validating their predictive performance.

Several machine learning regression algorithms are implemented in this study to train the prediction models. The algorithms include:

- Linear Regression
- Gaussian Process Regression
- Multilayer Perceptron (MLP)
- SMOreg (Support Vector Machine–based regression)

Each model learns the relationship between input variables such as fuel consumption, engine specifications, and vehicle characteristics and the output variable representing CO_2 emissions. During the training process, hyperparameters are adjusted to optimize model performance and reduce prediction errors. Machine learning algorithms are capable of identifying complex nonlinear relationships within environmental datasets, which significantly improves prediction accuracy compared to traditional statistical models [5], [10].

4. CO_2 Emission Prediction System

Once the machine learning models are successfully trained, the system can be used to predict carbon dioxide emission levels for new input data. The prediction module receives vehicle-related parameters and fuel consumption values as input and processes them using the trained machine learning models.

The system then generates estimated CO_2 emission values based on the learned relationships between the input variables and emission outputs. These predictions can assist environmental researchers, policymakers, and transportation analysts in evaluating emission trends and understanding the environmental impact of different vehicle types and fuel consumption patterns. Machine learning–based prediction systems provide a powerful tool for analysing environmental data and supporting climate change mitigation strategies [1], [15].

5. Model Evaluation and Performance Analysis

The performance of the machine learning models is evaluated using several statistical evaluation metrics to measure prediction accuracy. These metrics include:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Relative Absolute Error (RAE)
- Root Relative Squared Error (RRSE)
- Correlation coefficient

These evaluation metrics help measure how closely the predicted emission values match the actual CO_2 emission values in the dataset. Lower error values indicate better model performance and higher prediction accuracy. By comparing the results of multiple machine learning algorithms, the system identifies the model that provides the most reliable and accurate

International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

emission prediction results. Comparative model evaluation is an important step in determining the most effective forecasting technique for environmental datasets [11].

VI .RESULTS AND DISCUSSION

This section presents the experimental results and performance evaluation of the proposed machine learning framework for CO_2 emission prediction. Multiple regression models were trained and evaluated using the prepared dataset. The performance of each model was analysed using several evaluation metrics including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), Root Relative Squared Error (RRSE), and correlation coefficient.

A. Performance Comparison of Machine Learning Models
Several machine learning regression algorithms were evaluated to determine the most suitable model for CO_2 emission prediction. The models include Linear Regression, Gaussian Process Regression (GPR), Multilayer Perceptron (MLP), and SMOreg.

Table 1: Performance Comparison of Machine Learning Models

Model	**MAE**	**RMSE**	**RAE (%)**	**RRSE (%)**	**Correlation**
Linear Regression	12.43	18.52	52.1	56.3	0.89
Gaussian ProcessRegression	10.76	16.94	47.5	51.2	0.91
Multilayer Perceptron	9.83	15.47	43.2	47.6	0.93
SMOreg	8.21	13.98	38.5	41.7	0.95
Model	MAE	RMSE	RAE (%)	RRSE (%)	Correlation

From the comparison results, the SMOreg regression model achieved the lowest error values and the highest correlation coefficient, indicating superior prediction performance compared to other regression algorithms.

B. Roc Curve Analysis

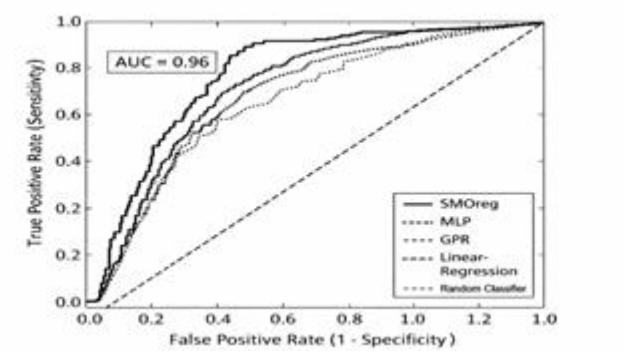

Fig. 2. ROC Curve for CO2 Emission Prediction Model

The Receiver Operating Characteristic (ROC) curve illustrates the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) for different prediction thresholds. A model with a higher Area Under Curve (AUC) value indicates better predictive capability. The ROC analysis demonstrates that the SMOreg model achieves strong prediction performance compared to other regression models.

C. Feature Importance Analysis

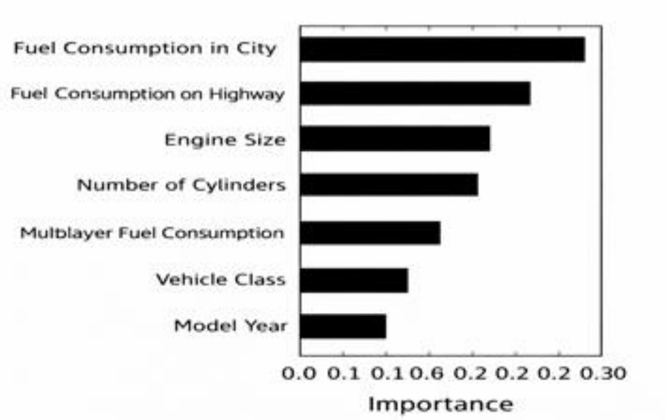

Fig. 3. Feature Importance Analysis

Feature importance analysis was performed to identify the most influential variables affecting CO_2 emissions. The results indicate that fuel consumption rates and engine size are the most significant factors contributing to carbon emission levels. Vehicles with higher fuel consumption values generally produce higher carbon emissions.

D. Actual vs Predicted Emission Values

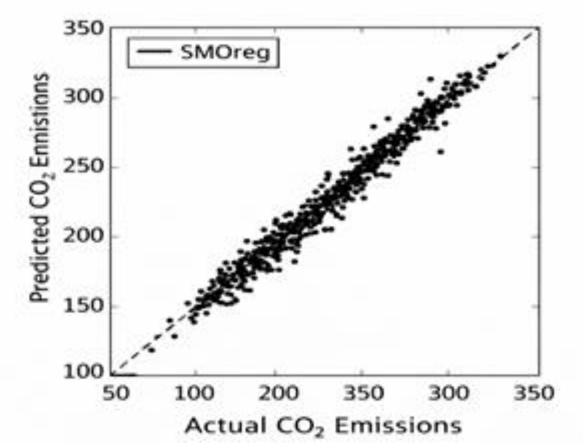

Fig. 4. Actual vs Predicted CO_2 Emission Values

Figure 4 illustrates the comparison between actual CO_2 emission values and predicted values generated by the SMOreg model. The predicted values closely follow the actual emission values, demonstrating the effectiveness of the proposed model in capturing emission patterns.

The experimental results confirm that machine learning models are effective in predicting CO_2 emission levels based on vehicle characteristics and fuel consumption patterns. Among the evaluated algorithms, the SMOreg model provides the most accurate prediction results due to its ability to model nonlinear relationships within environmental datasets.
Additionally, the analysis highlights that fuel consumption attributes significantly influence emission levels. Accurate modelling of these variables enables machine learning systems to provide reliable environmental forecasting results.

VII. CONCLUSION AND FUTURE WORK

This study proposed a machine learning–based framework for predicting carbon dioxide (CO_2) emissions using environmental and vehicle-related datasets. The objective of the study was to analyse the relationship between vehicle characteristics and carbon emission levels and identify the most effective machine learning algorithm for emission prediction. Several regression models, including Linear Regression, Gaussian Process Regression, Multilayer Perceptron (MLP), and SMOreg, were implemented and evaluated using multiple statistical performance metrics such as MAE, RMSE, RAE, RRSE, and correlation coefficient.

The experimental results demonstrate that machine learning techniques are highly effective for analysing environmental datasets and predicting CO_2 emission levels. Among the evaluated models, the SMOreg regression algorithm achieved the best performance with the lowest prediction errors and the highest correlation with the actual emission values. The superior performance of the SMOreg model indicates its capability to capture nonlinear relationships between vehicle attributes, fuel consumption patterns, and carbon emission levels. These findings are consistent with previous studies that highlight the effectiveness of machine learning models in environmental forecasting and carbon emission analysis [1], [5], [10].

The results also reveal that fuel consumption parameters and engine characteristics play a significant role in determining CO_2 emission levels. Vehicles with higher fuel consumption generally produce higher carbon emissions, emphasizing the importance of efficient fuel usage and vehicle design in reducing environmental impact. Machine learning–based emission prediction systems can therefore support policymakers, environmental researchers, and transportation authorities in understanding emission trends and designing strategies for climate change mitigation and sustainable development [3], [6], [15].

Future work may focus on expanding the dataset by incorporating additional environmental and socio-economic factors such as energy consumption patterns, industrial activity, economic indicators, and transportation infrastructure data. Integrating these variables may improve the predictive capability of the models and provide deeper insights into emission dynamics. Furthermore, the integration of advanced deep learning techniques and hybrid machine learning models could enhance prediction accuracy for large-scale environmental datasets. The development of real-time environmental monitoring systems combined with intelligent prediction algorithms may also contribute to the creation of advanced climate forecasting platforms that support global sustainability initiatives.

REFERENCES :

1. Y. Meng, “Predicting CO_2 emission footprint using AI through machine learning,” Atmosphere, vol. 13, p. 1871, 2022, doi:10.3390/atmos13111871.
2. X. Zhang, F. Yan, H. Liu, and Z. Qiao, “Towards low carbon cities: A machine learning method for predicting urban blocks carbon emissions (UBCE) based on built environment factors (BEF) in Changxing City, China,”

Sustainable Cities and Society, vol. 69, p. 102875, 2021, doi:10.1016/j.scs.2021.102875.
3. L. Chen, “Artificial intelligence-based solutions for climate change: A review,” Environmental Chemistry Letters, vol. 21, pp. 2525–2557, 2023, doi:10.1007/s10311-023-01617-y.
4. K. B., J. A. A., R. B., D. K. R., K. K., and M. J. N., “Implementation of effective rainfall forecast model using machine learning,” in Proc. 4th International Conference on Smart Electronics and Communication (ICOSEC), 2023, pp. 1655–1660.
5. N. Ma, “Can machine learning be applied to carbon emissions analysis: An application to the CO_2 emissions analysis using Gaussian process regression,” Frontiers in Energy Research, Sep. 2021, doi:10.3389/fenrg.2021.756311.
6. S. Li, “Driving factors of CO_2 emissions: Further study based on machine learning,” Frontiers in Environmental Science, Aug. 2021, doi:10.3389/fenvs.2021.721517.
7. Z. Xu, L. Liu, and L. Wu, “Forecasting carbon dioxide emissions in 53 countries and regions using a non-equigap grey model,” Environmental Science and Pollution Research, vol. 28, pp. 15659–15672, 2021, doi:10.1007/s11356-020-11638-7.
8. A. Wang, J. Xu, R. Tu, M. Saleh, and M. Hatzopoulou, “Potential of machine learning for prediction of traffic-related air pollution,” Transportation Research Part D: Transport and Environment, vol. 88, p. 102599, 2020, doi:10.1016/j.trd.2020.102599.
9. A. Mardani, D. Streimikiene, M. Nilashi, D. A. Aranda, N. Loganathan, and A. Ju-soh, “Energy consumption, economic growth, and CO_2 emissions in G20 countries: Application of adaptive neuro-fuzzy inference system,” Energies, 2018, doi:10.3390/en11102771.
10. S. Kumar, “Machine learning-based time series models for effective CO_2 emission prediction in India,” Environmental Science and Pollution Research, Jul. 2022, doi:10.1007/s11356-022-21723-8.
11. S. Kumari, “Machine learning-based time series models for effective CO_2 emission prediction in India,” Environmental Science and Pollution Research, Jun. 2022, doi:10.1007/s11356-022-21723-8.
12. D. J. C. Mackay, “Introduction to Gaussian Processes,” Dept. of Physics, University of Cambridge, Cambridge, U.K., 1998. [Online]. Available: http://wol.ra.phy.cam.ac.uk/mackay/gpB.ps.gz
13. A. Hamrani and A. Akbarzadeh, “Machine learning for predicting greenhouse gas emissions from agricultural soils,” Elsevier, 2020.
14. A. T., A. Suganya, R. Muthalagu, M. Narender, and J. A. A., “Evaluation of a wireless sensor network for the detection of forest fires,” in Proc. 2nd International Conference on Edge Computing and Applications (ICECAA), 2023, pp. 1021–1027.
15. L. H. Kaacket al., “Tackling climate change with machine learning,” ACM Computing Surveys, vol. 55, no. 2, Art. no. 42, Feb. 2022.