International Journal of Advanced Trends in Engineering and Management
(ISSN 2583-7052)

Vol. V, No. 02, pp. 41-47, February

Explainable Vehicle CO_2 Emission Prediction Using a Hybrid TabNet and Deep Residual MLP Network

D. Lakshmi [3]

[3] Professor, Department of Electrical and Electronic Engineering
AMET University, Chennai, Tamil Nadu, India.

Corresponding Author E-mail:

ABSTRACT: The adverse effects of vehicles that emit Carbon Dioxide (CO_2) on air, causes global warming, reduced air quality and climate change which is necessary for emphasizing the need of advanced and sustainable solutions. Explainable Artificial Intelligence (XAI) plays a major role in predicting and monitoring CO_2 emission on air which is employed with deep learning network for better performance in CO_2 emission prediction. In this paper, dataset is processed initially to remove noise obtained from environment which followed by data visualization that analyse the dataset. Encoding the feature takes place to convert categorical data to numerical form in which the dataset obtained undergoes training and testing phase. Then model selection is done using structure that integrates TabNet model with deep residual Multilayer Perceptron (MLP) that paves a way for accurate measurements and provides clarity in understanding complication faced by environments due to CO_2 emission in vehicles. Features that are segmented and extracted are processed by TabNet classifier that aids in perfect decision making and attention-based feature selection. Integration of TabNet and MLP network has interpretability with deeper nonlinerar learning, which is necessary for accurate prediction of carbon emission in vehicles. SHapley Additive exPlanations (SHAP) which is categorized in XAI, addresses model interpretability and transparency, that aids in decision-making for vehicle manufacturers and policy-maker to provide sustainable transportation and predicting CO_2 emission. The proposed technique analysed using CO_2 emission dataset, achieves improved accuracy and works efficiently than other conventional approaches. This study facilitates decision making based on evidence for better development in urban transportations, which integrates feature extraction and improved predictive model.

Keywords: Greenhouse Gases, Deep Residual MLP Network, Co2 emission, TabNet, Explainable AI (XAI), Shapley additive exPlanations (SHAP).

1. Introduction

Continuous rise in Earth's average temperature and environmental problems throughout worldwide that increases intense concentration of greenhouse gases (GHGs) led to global warming. Since the sustainable energy system depends heavily on accurate CO_2 emission, that aids in minimizing climate-related risks, renewable energy network optimization and effective decarbonization strategies [1-3]. Enhanced computational efficiency, accurate predictive modelling and smart decision-making have been facilitated by rapid evolution of AI and Machine learning (ML) [4, 5]. In [6], a prediction model that employs Light Gradient-Boosting Machine (ECP-LightGBM) with XAI is proposed. Using this approach a low mean square root is obtained but the model lacks real-time adaption and complexity in hyperparameter optimization. The study proposed in [7] focuses on reducing CO_2 emission in vehicles using traffic signal based on multi-gent reinforcement learning which is employed for adaptive traffic signal control. Although its advantages provide promising outcomes, it has scalability limitations, complexity and convergence issues in training model. In [8], a hybrid model integrating complementary Ensemble Empirical Mode Decomposition with adaptive Noise (CEEMDAN) with Support Vector Machine (SVM), Artificial Neural network (ANN), Decision Tree (DT), Random Forest (RF), K-nearest Neighbors (KNN) and Naïve Bayes (NB) is proposed to predict CO_2 emission. Although it predicts carbon emission, it faces parameter sensitivity and scalability issues. Hence an innovative work is proposed in this paper that integrates TabNet and deep residual MLP network for CO_2 emission prediction [9].

2. Recent Works

Archana Sulekha Devi et al [2024] [10] have proposed a comprehensive Internet-of-Vehicle (IoV) network based on real-time evaluations for estimating and reducing CO_2 emission with Reinforcement Learning-based network (RL). Implementation is performed in on-board device and data is send to network that predicts Co2 emission of vehicles. Emission is reduced using RL algorithm, that shows reduction from 11kg/h to 150 kg/h. Since this approach focuses on speed control only, validation based on real-world is limited and faces scalability issues.

Naghmeh Niroomand et al [2024] [11] have proposed a Internal Combustion (IC) engine performance model with segment-based CO_2 emission prediction using three ML techniques such as Support Vector Regression, random forest and semi-supervised deep Fuzzy C-means helps to detect inputs obtained from engine simulation software packages database. The performance metrics is calculated based on regular trips and distance travelled daily in cities. Fuzzy network provides unity approach in determination coefficient and high accuracy. Since the model predicts CO_2 emission independently, it does not combine with intelligent transport systems. The model is trained in particular region and not suitable for different climatic conditions.

Mukul Singh et al [2023] [12] have proposed a CO_2 emission prediction in vehicles using vehicle On-board Diagnostics (OBD-11) port dataset. Recurrent neural network (RNN) based on Long short-term memory (LSTM) model is used to predict CO_2 emission in vehicles in real-time. High computational cost due to use of deep learning model, interpretability and scalability issues, consideration of external factors is limited.

Liang Zheng et al [2025] [13] have proposed a dilated-convolutional block attention module, Dilated-(CBAM), which is combination of CBAM and dilated convolution in predicting carbon emission. In this study, it is noted that analysing feature is important that influence prediction of CO_2 emission mainly in urban and highways. It is about 20% total emission among all type mode of transportations. The drawbacks are limited temporal resolution, lack of generalization and real-time prediction, traffic factors such as vehicles speed, driver activities and traffic signal timing are not considered properly.

Talysson Santos et al [2023] [14] have proposed a CO_2 emission prediction technique that uses discrete Dynamic Bayesian networks without affecting its energy demand supply. It also uses an analytic threshold method for picking the data and converts into robust predicting model. Results are compared to be more accurate that ANN and Xgboost. If power sources and numbers of variable increases, the network learning becomes complex with scalability issues and uncertain energy sources.

The contribution of the study is listed below:

(a) Proposes a novel framework that integrates TabNet and a deep residual MLP for learning mechanism that aids CO_2 emission prediction.

(b) To improve stability and model convergence, residual connections are incorporated with MLP network that removes gradient problems and performs deep network training.

(c) Integrates XAI methods like SHAP, to provide details about interpretations of model prediction and observes factors that contribute in CO_2 emission in vehicles.

The methodology of proposed system is explained in following section.

3. Proposed Work Explanation

In this section, input dataset is fed into pre-processing unit in which data cleaning takes place by deleting the missing data and noise received from environment. It also examines for duplicate datasets, incorrect data type and inconsistent values. After processing the input data, data visualization takes place to analyze dataset by selecting pattern and variable relationship before training the model. Encoding unit converts the categorical data to binary values in machine-readable format. In data splitting unit, training and testing of dataset is performed and output is transferred for feature scaling process that aids in faster convergence of dataset. The proposed system block diagram is shown in Figure 1.

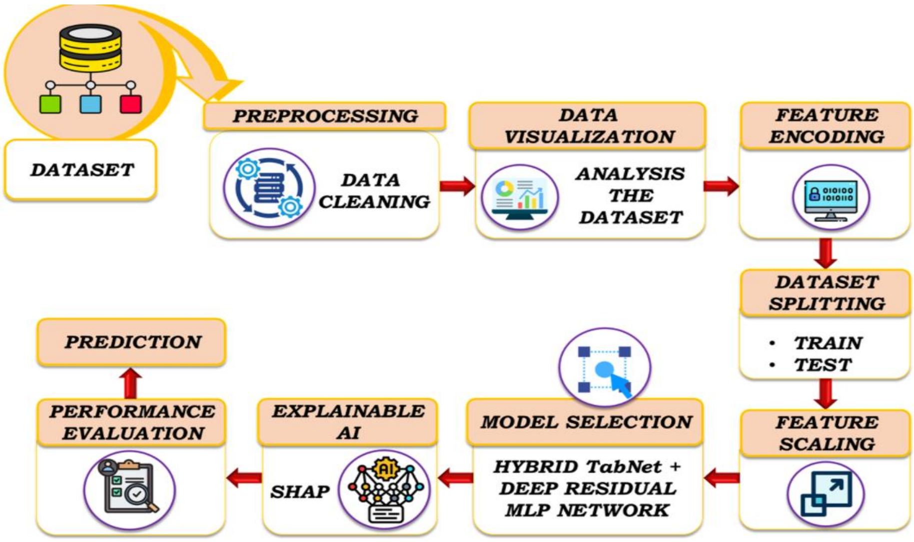

Figure 1: *Block diagram of proposed model*

For selecting the model, hybrid TabNet and Deep MLP network is employed which innovates the proposed work, that effectively predicts Co2 emission in vehicles based on fuel consumption in city and highways. SHAP which is a part of XAI is employed for model interpretability and performance is evaluated for the proposed work.

3.1 Hybrid TabNet and Deep Residual MLP Network

In this study, TabNet classifier aids in decision making which selects and transmits features at each step, that helps in model interpretability and good performance in CO_2 emission prediction. The vector that is fused, initially projected to latent expression followed by equation (1),

$$h_0 = W_0 F_{fusion} + b_0 \quad (1)$$

The sparsemax activation function detects an attention mask M_i in every decision steps i, that enables selecting sparse subset of features, defined

$$M_i = sparsemax\big(BN(W_a h_{i-1})\big) \quad (2)$$

Where $H_i = M_i \odot h_{i-1}$ selected feature is passed into transformer consists of gated linear units (GLUs) as shown in equation (3) & (4)

$$f_i = GLU(BN(W_1 H_i + b_1)) \quad (3)$$

$$H_i = GLU(BN(W_2 f_i + b_2)) \quad (4)$$

Summation of equation (5) & (6) provides last prediction, in which a partial decision vector is produced in every decision step.

$$d_i = ReLu(W_d h_i + b_d) \quad (5)$$

$$y = softmax(\sum_{i=1}^{N_d}(d_i) \quad (6)$$

The probabilities corresponding to each CO_2 emission prediction is represented by final output y as illustrated in equation (6). Step-by-step analysis is taken using TabNet classifier that aids in achieving accurate and understandable prediction.

In the proposed methodology, TabNet model is hybrid with MLP network that has an input layer connected with 10 features followed by two hidden layer and output layer that makes use of SoftMax activation function. It also helps in reducing error value at output network thereby improves accuracy with minimized loss function.

3.2 Mathematical Expressions and symbol

In this section, mathematical expression used for evaluating the performance of classifier model to predict CO_2 emission is explained. R-squared is defined as proportion of variance in CO_2 emissions detailed in model relative to total variance in dataset as governed by equation (7),

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2)}{\sum_{i=1}^{N}(y_i - \hat{y})^2)} \quad (7)$$

Where predicted value is denoted as $\hat{y}_i$, actual emission value is denoted as y_i and mean of emission observed is denoted as $\hat{y}$. R-squared value near to unity denotes that effective prediction of Co2 emission of model. Root Mean Squared Error defined as average difference of squared value between actual and predicted emission given in equation (8),

$$MSE = \frac{1}{N}\sum_{k=0}^{n}(y_i - \hat{y}_i)^2) \quad (8)$$

The results and discussion of proposed is explained in following section where comparison graph are plotted for specified parameters.

4. Results and Discussion

Through experiments, CO_2 emission prediction in vehicles is discussed in this section. Dataset is accessed through link https://www.kaggle.com/datasets/brsahan/vehicle-co2-emissions-dataset/data. Comparison on engine size and combined fuel consumption based on feature values vs observation index is shown in Figure 2(a), CO_2 emission distribution plotted between frequency vs CO_2 emission values is shown in figure 2(b), Top 5 vehicle makes is plotted for vehicles such as Ford, Chevrolet, BMW, Mercedes-Benz and Porsche is shown in Figure 2(c), Top vehicle classes based on vehicle count is compared in figure 2(d)

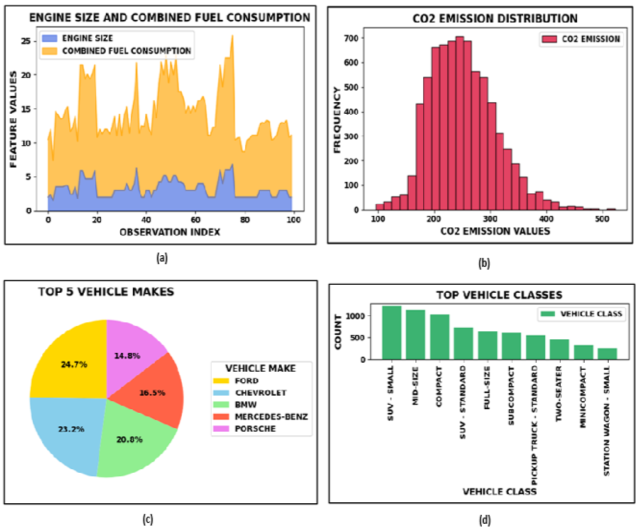

Figure 2: *Comparison on (a)engine size and combined fuel consumption (b) Co2 emission distribution (c) Top 5 vehicle make (d) Top vehicle classes.*

Figure 3 details about engine size density plotted between density vs engine size as shown in Figure 3(a), city vs highway fuel consumption that shows fuel consumption is more in highway is detailed in figure 3(b), combined fuel consumption trend is shown in figure 3(c) and cylinder variation plotted between cylinder and observation index is shown in Figure 3(d).

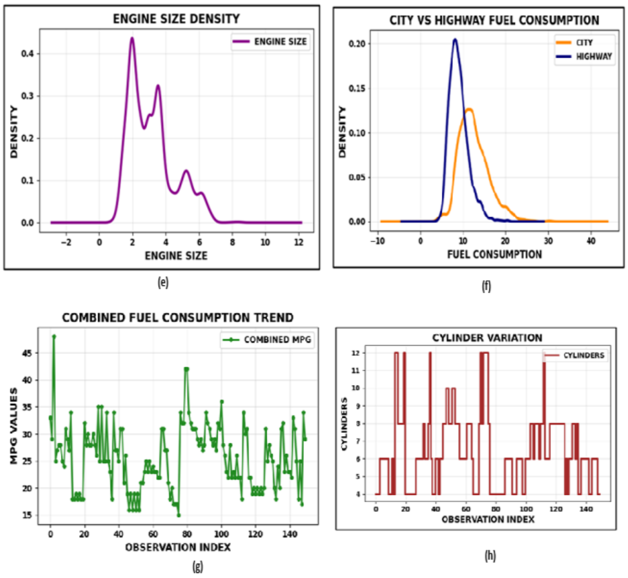

Figure 3: *Comparison on (a) energy size density (b) City vs Highway fuel consumption (c) Combined Fuel consumption trend (d) Cylinder variation*

Comparison on top transmission type plotted between count and transmission is shown in Figure 4(a), engine size of vehicle plotted between observation index and engine size is illustrated in Figure 4(b), highway fuel consumption profile is plotted between fuel consumption vs observation index is shown in Figure 4(c) and presentation of city fuel consumption plotted between fuel consumption vs observation index is shown in Figure 4(d).

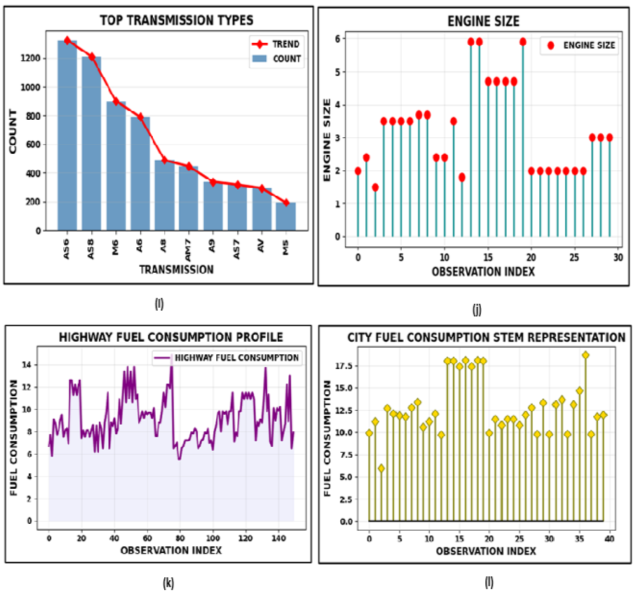

Figure 4: *Comparison on (a) Top transmission types (b) Engine size (c) Highway fuel consumption profile (d) City fuel consumption stem presentation*

Comparison on SHAP feature importance such as fuel consumption city, fuel consumption highway, fuel type, vehicle class, engine size, transmission, cylinders, model and make vs SHAP value with impact on model output is shown in Figure 5(a). Figure 5 (b) details about actual vs predicted CO_2 emissions.

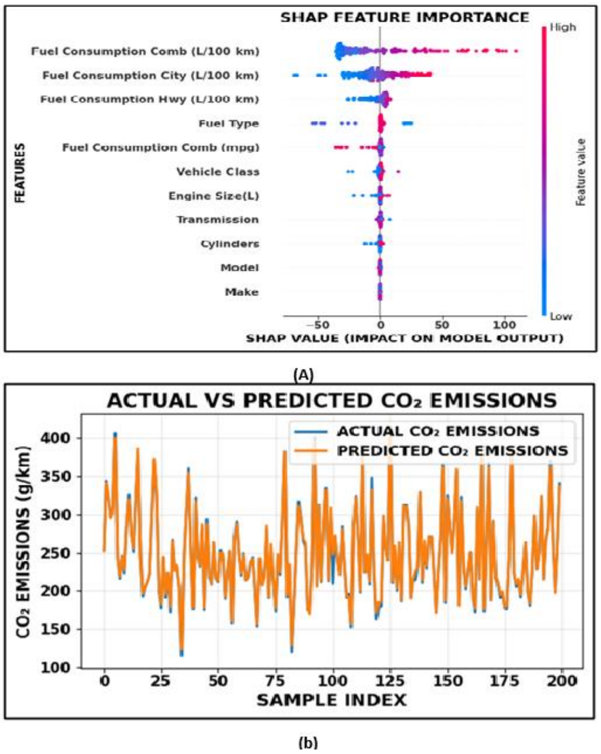

Figure 5: *Comparison on (a) SHAP feature importance (b) Actual vs predicted CO_2 emission*

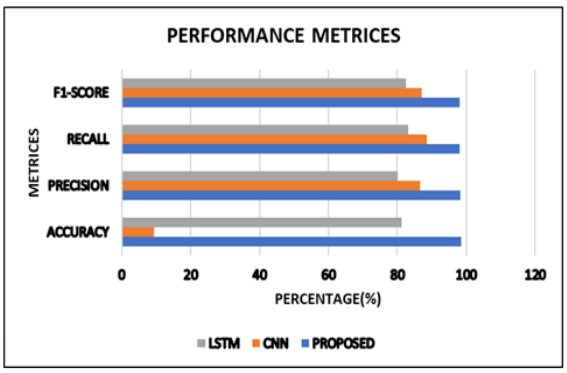

Figure 6: *Comparison graph on performance metrics*

The comparison graph for various classifiers such as Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) and proposed model is plotted in Figure 6 that details that the proposed model achieves accuracy of 98.42%, precision of 98.2%, recall value of 98.06% and F1-score is 98.13%.

5. Conclusion

In this paper, a novel structure that integrates TabNet and Deep residual MLP network is proposed to determine CO_2 emission using real-time validations. On-board device is deployed to estimate emission values of vehicle and estimated value is transmitted to network. In future research work, models integrating Mamba-Transformer network aids in better prediction with other attention mechanisms. In this work, TabNet classifier is adopted to optimize MLP network that aids in providing better performance metrics with accuracy 98.42%, precision 98.20%, recall value 98.06% and F1-measure value of 98.13%. Furthermore, analysis under XAI is conducted to provide transparent explanations with interpretable calculations. The values are estimated for fuel consumption in city vs highways and details that travelling in highway consume more fuel than city.

References

1. Neng Ayu Herawati; Asyraf Atthariq Putra Gary; Erna Hikmawati; Kridanto Surendro, Year: 2024, “A Hybrid Predictive Model as an Emission Reduction Strategy Based on Power Plants’ Fuel Consumption Activity”, in IEEE Access, Vol: 12, pp. 47119-47133.
2. Omar Farrag; Ahmad Mansour; Baraa Abed; Amr Abu Alhaj; Taha Landolsi; Abdul-Rahman Al-Ali, Year: 2025, “IVEMPS: IoT-Based Vehicle Emission Monitoring and Prediction System”, in IEEE Access, Vol: 13, pp. 95628-95646.
3. Pankaj Verma; Krishna Gandhi; Salma Idris; Faten S. Alamri; Muhammad I. Khan, Year: 2026, “STDSE-Net: A GIS-Driven AI Framework for Dynamic Ensemble-Based Energy Emission Prediction”, in IEEE Access, Vol: 14, pp. 4305-4326.
4. Pritam Singh Balai; Asaruddin Sheikh; Garima Rabha; Samiran Das; Bhaba Krishna Kuli; Mohit Raj, Year: 2025, “Revolutionizing agricultural machinery: The role of AI, IoT, and renewable energy in enhancing efficiency and sustainability”, International Journal of Scientific Research in Science and Technology, Vol: 12, No: 2, pp. 813–830.
5. Prema Nedungadi; Simi Surendran; Kai-Yu Tang; Raghu Raman, Year: 2024, “Big data and AI algorithms for sustainable development goals: A topic modeling analysis”, IEEE Access, pp. 1.
6. Sundus Munir; Manas Ranjan Pradhan; Sagheer Abbas; Muhammad Adnan Khan, Year: 2024, “Energy Consumption Prediction Based on LightGBM Empowered With eXplainable Artificial Intelligence”, in IEEE Access, Vol: 12, pp. 91263-91271.
7. Hyemin Kim; Jinhyuk Park; Dongbeom Kim; Chulmin Jun, Year: 2025, “Cooperative Control of Intersection Traffic Signals Based on Multi-Agent Reinforcement Learning for Carbon Dioxide Emission Reduction”, in IEEE Access, Vol: 13, pp. 33485-33495.
8. Muhammad Ali; Dost Muhammad Khan; Huda M. Alshanbari; Omalsad H. Odhah, Year: 2025, “Forecasting Carbon Dioxide Emission Using Hybrid Machine Learning and Nonlinear Data Decomposition Methods”, in IEEE Access, Vol: 13, pp. 212942-212958.
9. Rohit Ravi; P. Madhavan, Year: 2026, “MCAF_TabNet: A Multiscale Convolutional–Attention Fusion Network With TabNet Classifier for Interpretable Heart Sound Analysis”, in IEEE Access, Vol: 14, pp. 34137-34150.
10. Archana Sulekha Devi; Milagres Mary John Britto; Zian Fang; Renjith Gopan; Pawan Singh Jassal;

Mohammed MH Qazzaz; Sujan Rajbhandari; Farah Mahdi Al-Sallami, Year: 2024, “Internet-of-Vehicles Network for CO_2 Emission Estimation and Reinforcement Learning-Based Emission Reduction”, in IEEE Access, Vol: 12, pp. 110681-110690.

11. Naghmeh Niroomand; Christian Bach, Year: 2024, “Integrating Machine Learning for Predicting Internal Combustion Engine Performance and Segment-Based CO2 Emissions Across Urban and Rural Settings”, in IEEE Access, Vol: 12, pp. 66223-66236.
12. Mukul Singh; Rahul Kumar Dubey, Year: 2023, “Deep Learning Model Based CO2 Emissions Prediction Using Vehicle Telematics Sensors Data”, in IEEE Transactions on Intelligent Vehicles, Vol: 8, No: 1, pp. 768-777.
13. Liang Zheng; Shenshen Li; Xuefei Hu; Kun Cai; Yang Liu, Year: 2025, “The Downscaling Prediction Algorithm of Traffic Source Carbon Emissions Based on Multisource Remote Sensing Data and Deep Learning”, in IEEE Transactions on Geoscience and Remote Sensing, Vol: 63, pp. 1-18, Art no. 4112018.
14. Talysson Santos; Michel Bessani; Ivan Da Silva, Year: 2023, “Evolving Dynamic Bayesian Networks for CO2 Emissions Forecasting in Multi-Source Power Generation Systems”, in IEEE Latin America Transactions, Vol: 21, No. 9, pp. 1022-1031.

IJATEM-V05I02P5