# Machine Learning and Deep Learning Analysis of Vehicle Carbon Footprint

Dhyan R[1], Helen K Joy[1*], Sridevi R[1], Electa Alice Jayarani A[2], Vanusha D[3]

[1] Department of Computer Science, Christ University, Bangalore 560029, India
[2] Department of Electronics and Communication Engineering, KS Institute of Technology, Bangalore 560019, India
[3] Department of Computer Science and Engineering, SRM Institute of Technology, Chennai 603203, India

Corresponding Author Email: helenjoy88@gmail.com

**ABSTRACT**

Clearly climate change is one of the most significant hazards to mankind nowadays. And daily the situation has become worse. No other way characterises climate change except through changes in the patterns of temperature and weather. Human activity generates the primary greenhouse gas emissions. Among these activities are burning coal, oil, natural gas, as well as other fuels; agricultural techniques, industrial operations, deforestation, burning coal, oil. Mostly resulting from human activities, the average temperature of the planet has significantly increased by almost 1.1 degrees Celsius since the late 1800s. One theory holds that internal combustion engines affect roughly thirteen percent. The objective of this work is to do an analysis of a complicated dataset involving fuel consumption in urban and highway environments as well as mixed combinations since the relevance of these variables in modelling attempts dictates. Reduced $CO_2$ emissions emissions and environmental impact follow from reduced fuel use. The project used numerous machine learning and deep learning approaches to comprehend data analysis. Moreover, this work investigates the dataset to acquire knowledge and concurrently solves problems such overfitting and outliers. Control of complexity is achieved using several methods like VIF, PCA, and Cross-Validation. Models combining CNN and RNN performed really well with an accuracy of 0.99. The R-squared metrics are utilized in order to do the evaluation of the model. Apart from linear regression, support vector machines, Elastic Net with a rewardable accuracy, random forest was applied. It has rather good 0.98 accuracy. We can therefore state that our model analyzed the data properly and generated accurate output since the results we obtained during the assessment phase exactly the same ones we obtained during the training stage. Mass data cleansing is required as well as further study to increase machine learning model accuracy and performance.

## 1. INTRODUCTION

$CO_2$ is a naturally occurring greenhouse gas that has a significant impact on the Earth's atmosphere. It aids in the regulation of the planet's temperature by capturing and inhibiting the dissipation of heat into outer space. Due to use of fuels like petrol and diesel in automotive industry increasing the $CO_2$ emission, defined as carbon footprint which in turn impacts the climate change. Increased concentration of $CO_2$ enhances the greenhouse effect which results in global warming, change in weather pattern which disturbing in environment [1]. These disturbances include melting of ice or glaciers and the occurrence of dreadful weather conditions. Since 1800s due to human activities in industries like forestry, agriculture, transportation, energy generation etc., has increased the climate change. Global temperature currently is 1.1 degree Celsius greater than the temperature during the industrial era. Projections suggested that the temperature may increase up to 3 degrees Celsius by end of this period if the proper action is not taken.

As if to address the much-discussed issue of climate change it is imperative that different parties, people, governments and the automotive companies especially must come together to join efforts. A particular should be placed on discussing sustainable best practices to address the consequences of this problematic occurrence on the international level emphasis [2]. The problem requires further development of the renewable energy, rational and effective use of lands as well as conservation of the environment, reduction of emissions from vehicles and industries, and international collaboration since the problem is global in nature. Placing myself in a sustainable context, I can confirm that we have the power to save the environment for future generations and prevent the world from the more drastic consequences of climate change.

Emphasizing the Significance of Transportation and its Impact on the Environment: Transport related emissions have a decisive influence on climate change as transport sector is ranked to be contributing to over 13 to 15% of the environmental impact. Figures 1 and 2 emphasize the significant impact of transport-related emissions on climate change. In the present world with growing competition, it becomes necessary for the centre to address this particular

problem of mobility because mobility is closely related to progress [3-5]. Non-electric vehicles powered by internal combustion engines are a key factor in the production of $CO_2$ emissions and are therefore, detrimental to the atmosphere. This analysis aims to examine the complex interaction between vehicle attributes and pollution. Figure 3 is the graph showing the transmission of $CO_2$ by different vehicle class. The study aims to explore factors defining the general level of $CO_2$ emissions and the susceptibility of specific vehicles to have detrimental effects on the environment, utilizing strong data analysis, intelligent machine learning algorithms, and big data techniques. Our intention is to offer you reliable and enriching knowledge that will serve as a tool in making decisions towards making the transport sector sustainable.
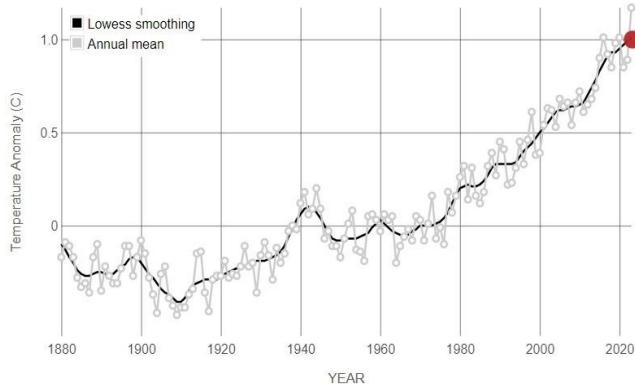


**Figure 1.** Increase in temperature since 1800
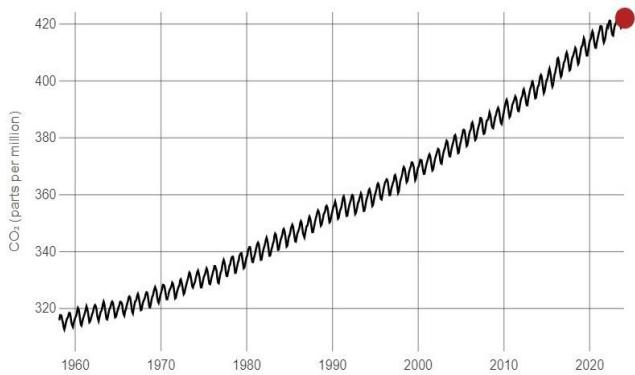


**Figure 2.** Increase in $CO_2$ since 1800

According to the given research framework, it is proposed to investigate a number of characteristics including fuel consumption in urban and highway conditions in order to forecast the level of $CO_2$ emissions. Figure 4 presents the mean $CO_2$ emissions categorized by fuel type. Therefore, it retains significance in regulating or even oftentimes lessening the emission influence in correspondence with various car models. The proposed study has identified the following machine learning and deep learning algorithms to predict $CO_2$ emissions: Linear regression, random forest, SVM, CNN, and RNN [6-8]. Out of all these algorithms, the random forest and deep learning algorithm demonstrated an extraordinarily high performance. The following analysis is useful in demonstrating how such approaches function in enhancing the management and cooperation of emissions from different models of cars in the automotive sector and as such offers factual information to automotive producers for promoting friendly environmental measures and technologies in the sustainable management.
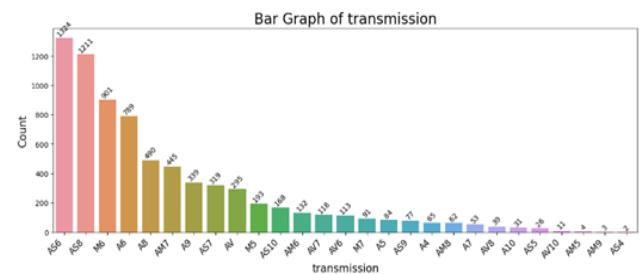


**Figure 3.** The trasmission of $CO_2$ by different engines

The $CO_2$ emissions prediction model requires certain other characteristics like the amount of gasoline utilized in city and highways. Figure 3 displays a graph illustrating the transmission of $CO_2$ emissions by different engine types. Thus, allowing for better emission control through the use of accurate projection of the major emission constituents per model of vehicle. After the application of different machine algorithms, including machine learning (linear regression, random forest, support vector machine) and deep learning (convolutional neural network, recurrent neural network), we conclude that random forest and deep learning algorithms provide the best results [9]. The outcome of this analysis shows that these approaches are also efficient in accurately quantifying the $CO_2$ emissions which may help to manage and control the emissions from a range of vehicle types. The research will be useful to auto makers since they will get to learn more thanks to this research. Ideally, it is to guide the sector toward a pursuit of a path that is sustainable for the natural environment by encouraging a culture change towards sustainable processes as well as technology acquisitions.
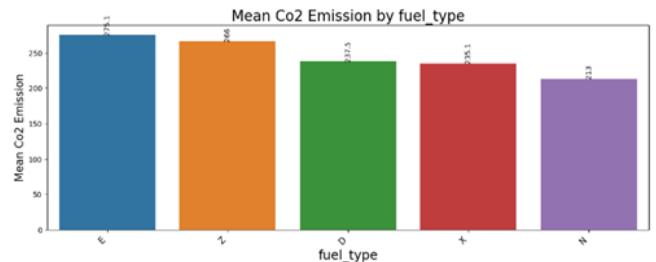


**Figure 4.** Mean $CO_2$ emission by fuel type
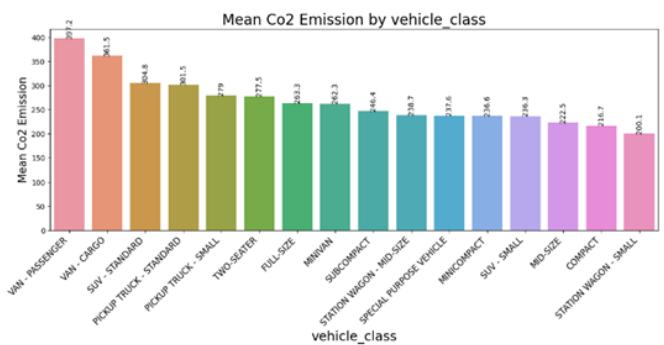


**Figure 5.** Transmission of $CO_2$ by different vehicle class

## 2. LITERATURE REVIEW

Offering convenience and access that constitute the modern living, the automotive sector is at the core of the distinguishing traits of the modern society. However, within the broader

framework of environmental responsibility, $CO_2$ and vehicle emissions have become increasingly significant issues [10] (Figure 5). As a natural component of the Earth's atmosphere, $CO_2$ is a known gas behind the greenhouse effect. Through insulating properties, it plays a major role in controlling planetary temperature. Still, human activities—especially in the automotive sector—have been discovered to spew enormous volumes of $CO_2$ by burning fossil fuels. All of these rising $CO_2$ levels significantly help to explain climate change. This results in the enhanced greenhouse effect, leading to global warming, rising temperatures, intense storms, and disruptions in the habitats of various species [11]. Melting of the polar ice caps, rise in frequency and intensity of natural disasters, and changes in the geographical range of some animal species are among the obvious physical changes noted from this gradual environmental change [12]. Nowadays, the automotive sector as well as customers and legislators should give reducing $CO_2$ emissions from vehicles first priority. Furthermore, improvements in hybrid systems and higher benchmarks for real ignition motors meant to lower transportation sector overall greenhouse gas emissions. In an effort to lower the $CO_2$ emissions in the automotive sector [13], global regulatory authorities are pushing non-conventional fuel use, mass transportation systems that are ecologically friendly, and better emission standards in automobile manufacturers. Every sector is calling for change. One example of how businesses aim for improved practices and pursue the integration of green materials [14], enhancing fuel efficiency, and adopting new technologies to contribute to a better planet, is manufacturing. This work is concentrated on the thorough and meticulous analysis of the surroundings connected to several car engines and personal driving behavior. We presented contingency strategies for engine efficiency improvement and foot-print minimization by separating the internalization process into several distinct categories. We want to apply strategic measures for optimizing engine efficiency and lowering footprints by breaking apart the subtle ways in which people affect the environment. Incorporating important criteria including make, gearbox, fuel type, and several fuel economy measures, the thorough study spans a varied dataset [15]. The aim is to measure the combined influence of these factors on $CO_2$ emissions by means of correlations between them. Furthermore, included is the field of artificial intelligence (AI) integration to improve the accuracy and efficiency of the evaluation. Deeper understanding of the complex dynamics of environmental effect inside the automobile sector is promised by the possible integration of artificial intelligence technologies [16]. By means of data-driven models and machine learning techniques, we want to expose trends and patterns that would guide focused interventions and creative ideas [17]. Research by this kind will help automotive producers to get understanding and hence advantage. Eventually, we intend to encourage a paradigm change towards more environmentally friendly technology and practices, therefore guiding the sector down a road of higher environmental sustainability.

## 3. DATASET

This dataset contains all the used data regarding the amount of $CO_2$ emissions for the specified vehicle for the last seven years. It consists of 7,385 rows and 12 columns. It also defines some car features, mainly the model, using special codes such

as 4WD/4X4 for four-wheel drive, AWD for all-wheel drive, FFV for flex fuel vehicle, SWB for short wheelbase [18], LWB for long wheelbase, and EWB for extended wheelbase. The transmission types are divided as follows: A symbolizes automatic transmission; AM stands for automated manual; AS for automated select shift; AV for continuum variable transmission; M represents manual transmission with 3-10 gears [19]. The fuel type is indicated by letters: X for regular grade with 87 octane, Z for premium grade with 93 octane, D for diesel, E for ethanol with E85, N for natural gas. Additionally, fuel efficiency is assessed in L/100km for city [20], highway, and combination (55% city and 45% highway), as well as in miles per gallon (mpg). Therefore, the main variable of interest is carbon emissions stated in grammes $CO_2$ per kilometre for both city and highway cycles. This information is gathered from the official open data [21] source maintained by the Canada Government and applied for environmental and automotive research aiming at establishing the link between vehicle features and emissions.

## 4. METHODOLOGY USED

Dataset: A dataset containing information about various vehicle makes, models, including their engine size, number of cylinders, fuel type and fuel consumption rates in both city and highway driving conditions is gathered by combining various existing data.

Data Preprocessing: The data cleaning of the dataset is done to remove any inconsistencies, missing values, and duplicated values ensuring the accuracy and reliability of the data for analysis.

Features Selection: Key features were identified using random forest feature importance that have more impact on the $CO_2$ emission, and we focus more on fuel type and on various fuel consumption in city and highway. Figure 6 shows the feature importance determined by a random forest model. Using a Random Forest Regressor, the produced diagram from the given code shows the feature importance in estimating CO2 emissions (g/km). X_train was the predictor variable for the model; y_train was the target variable, $CO_2$ emissions. Computed and plotted in descending order, feature significance values—which show the relative contribution of every feature in producing accurate predictions—were the bar chart shows various importances; the leftmost most important aspect is shown here. The y-axis shows the significance score; the x-axis names every feature. This visualisation guides judgements on feature selection and offers understanding of the elements most significantly linked with $CO_2$ emissions, therefore helping to find the most important aspects in the dataset.

Figure 7 displays a correlation matrix illustrating the relationships between various parameters. The pandas tool df_labeled.corr was used to compute the correlation matrix. A heatmap created with the seaborn library clearly shows the correlation matrix. The heatmap was produced using the seaborn function sns heatmap under the following main guidelines: correlation matrix: The input matrix with the correlation coefficients. annotation=True: This value allowed every cell to have its matching correlation coefficient annotated, therefore enabling exact interpretation. Cmap = 'coolwarm': With warm colours denoting positive values and cool colours denoting negative values, a diverging colormap was used to adequately separate positive from negative

correlations. Plot title "Correlation Matrix Heatmap" using plt.title to set the scene for the visualization. The generated heat map offers a whole visual overview of the dataset's variable interactions. For data scientists and academics, this visualization is a great tool for finding and comprehending underlying trends in their data, so directing next analytical and research projects.
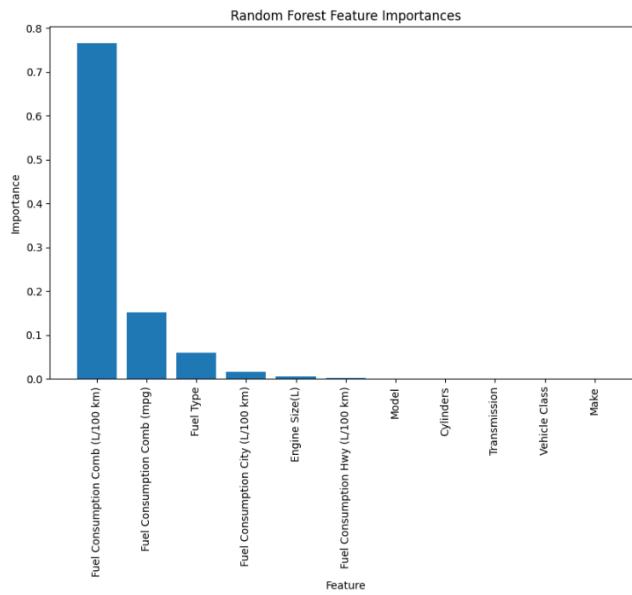


**Figure 6.** Random forest feature importance
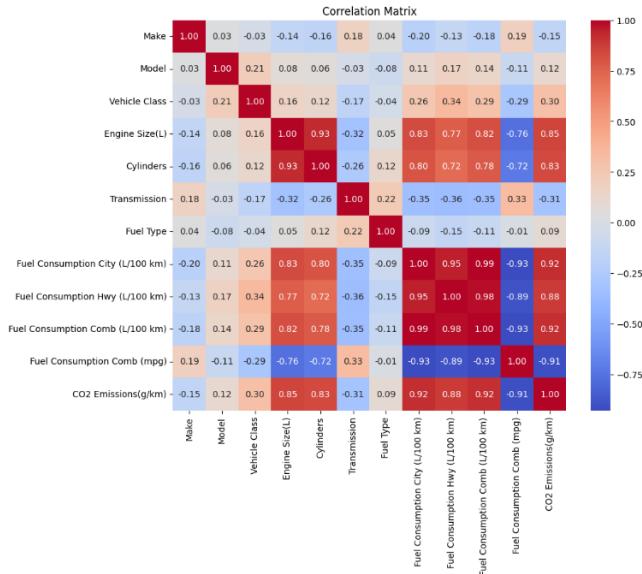


**Figure 7.** The relationships between various parameters

Model Training: In order to develop predictive models of $CO_2$ emission based on the selected feature we have used machine learning and deep learning algorithms like linear regression, random forest, support vector machine, convolutional neural networks, recurrent neural networks.

Model Evaluation: Each model is evaluated using the metrics like R-squared value which indicates how good the model fits the data for its performance, we also evaluated the computational costs associated with each model.

Model Selection: Based on the Evaluation result, we have selected the random forest and deep learning model as it

resulted in good performance with the high R-squared value of 0.933 and efficient computational costs.

Final Model Training: To make the model even more precise, it was trained on the whole data and made it even finer.

Model testing: The final model of CNN was then used to make the correct prediction of $CO_2$ emission of different vehicle types

Analysis and Interpretation: Analysis is done for findings' part to determine the key influential factors on $CO_2$ emissions in cars. This gave us a better feel of which vehicle characteristics more or less impact the environment.

Implications and Recommendations: Our discussion of implications with respect to automobile manufactures, policy makers and consumers was also presented. Concerning the fourth threat, we highlighted how necessary fuel-efficient technologies and environmentally friendly methods are for the reduction of $CO_2$ emissions and effective combating of climate change.

This system of arterialization earns an R-squared score of 0. 933. When compared with all the other models, the Random Forest model outperformed all of them. It was, therefore, found to very naturally fit the facts. While the Random Forest model generates a strong connection between features and incidents, Random Forests can understand several relations with a number of factors. Further, instead of producing very specific predictions, which can be risky, they aggregate several decision trees. Similar to numeric variables, Random Forests also help to manage a lot of features and at the same time do not lose in performance. It performed well on the tests as well as during the course of training. Besides, the CNN and RNN are the suitable deep-learning models for resolving the regression problems. It will enable them to identify trends into time and have an understanding of different datasets in entirety. Most beneficial when the data is high-dimensional and intricate, mainly for regression problems, CNNs and RNNs have the excellent capability of finding out valuable features on its own. During testing, the R-squared values are about 0. 87, linear regression and SVM performed equally well comprehensively and the result can be seen below.

## 5. EVALUATION OF THE PROPOSED MODEL

The model showed exceptional performance during the evaluation stage of our project, obtaining a really high R-squared value of almost 0.98. The CNN model, in particular, demonstrated amazing accuracy, emphasizing the efficacy of our approach in predicting $CO_2$ emissions among various car models. One must consider the computational costs connected to several models applied in our study. Though they produce amazing results, deep learning models including CNNs and RNNs usually demand more computer resources and training time than simpler models as linear regression or Random Forests. Given the performance improvements these deep learning models offer, it is important to carefully evaluate their computational complexity. Maximizing the training process and ensuring practicality in real-world applications depend on striking the right balance between accuracy and computational efficiency. Figure 8 presents a graph comparing the R² values for different models. With a comprehensive knowledge and careful handling of various computational factors, we may make educated choices about resource allocation and model choice. This will eventually enable us to maximize the predictive analysis efficiency and efficacy.
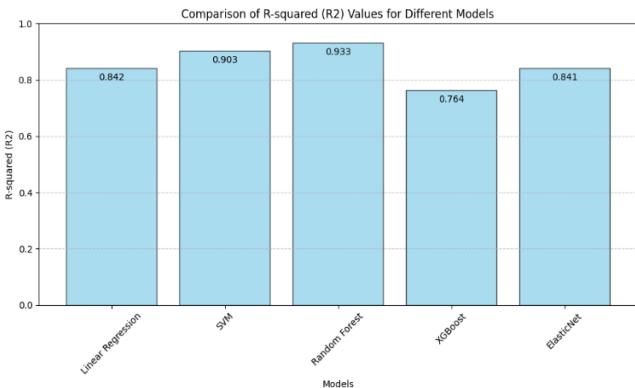
**Figure 8.** The comparison of $R^2$ values for different models

## 6. CONCLUSIONS

This study highlights the vital importance of the automotive industry in addressing climate change-related problems. By means of modern machine learning and deep learning techniques, we have achieved remarkable accuracy in predicting the $CO_2$ emissions across a spectrum of several vehicle types and models. The information acquired from this study is rather useful to create particular plans meant to lower emissions. It also provides direction for legislators developing regulations grounded in strong data and encouraging the application of greener transport options. The findings of this research enable automakers to design cars with better carbon emissions and fuel economy. Transposing the results of research into simply comprehensible knowledge will enable us to assist individuals in making wise decisions about vehicle purchase and so raise environmental consciousness in society. Making great progress towards a more sustainable future depends on adopting a complete strategy that takes consumer behaviour as well as industry activity into account. Further research into deep learning models holds the potential for even more significant advancements in $CO_2$ emission prediction. Improving and refining ways to lower emissions will depend much on the knowledge acquired from this study, thereby enabling a significant contribution to the worldwide endeavour to mitigate the detrimental consequences of climate change. We have an obligation to acknowledge the effects of our activities on the surroundings and strive for a more bright future by means of sustainable and responsible behaviour. By means of thorough investigation and group effort, we can steer the car industry towards a time when environmental responsibility takes front stage. For next generations, this will produce an environmentally friendly planet.

## REFERENCES

[1] Akpan, U.F., Chuku, A. (2011). Economic growth and environmental degradation in Nigeria: Beyond the environmental kuznets curve. https://mpra.ub.uni-muenchen.de/31241/.

[2] Li, S., Siu, Y.W., Zhao, G. (2021). Driving factors of $CO_2$ emissions: Further study based on machine learning. Frontiers in Environmental Science, 9: 721517. https://doi.org/10.3389/fenvs.2021.721517

[3] Chong, H.S., Kwon, S., Lim, Y., Lee, J. (2020). Real-world fuel consumption, gaseous pollutants, and $CO_2$ emission of light-duty diesel vehicles. Sustainable Cities and Society, 53: 101925. https://doi.org/10.1016/j.scs.2019.101925

[4] Chong, H.S., Park, Y., Kwon, S., Hong, Y. (2018). Analysis of real driving gaseous emissions from light-duty diesel vehicles. Transportation Research Part D: Transport and Environment, 65: 485-499. https://doi.org/10.1016/j.trd.2018.09.015

[5] Jaikumar, R., Nagendra, S.S., Sivanandan, R. (2017). Modal analysis of real-time, real world vehicular exhaust emissions under heterogeneous traffic conditions. Transportation Research Part D: Transport and Environment, 54: 397-409. https://doi.org/10.1016/j.trd.2017.06.015

[6] Goyal, Y., Meena, S., Singh, S.K., Kulshrestha, M. (2023). Real-time emissions of gaseous pollutants from vehicles under heterogeneous traffic conditions. Zeszyty Naukowe. Transport/Politechnika Śląska, 118: 55-75. http://dx.doi.org/10.20858/sjsutst.2023.118.5

[7] Joy, H.K., Kounte, M.R. (2024). Deep CNN based interpolation filter for high efficiency video coding. In 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, pp. 519-524.

[8] Aliramezani, M., Koch, C.R., Shahbakhti, M. (2022). Modeling, diagnostics, optimization, and control of internal combustion engines via modern machine learning techniques: A review and future directions. Progress in Energy and Combustion Science, 88: 100967. https://doi.org/10.1016/j.pecs.2021.100967

[9] Norouzi, A., Heidarifar, H., Shahbakhti, M., Koch, C.R., Borhan, H. (2021). Model predictive control of internal combustion engines: A review and future directions. Energies, 14(19): 6251. https://doi.org/10.3390/en14196251

[10] Magazzino, C., Mele, M. (2022). A new machine learning algorithm to explore the $CO_2$ emissions-energy use-economic growth trilemma. Annals of Operations Research, 1-19. https://doi.org/10.1007/s10479-022-04787-0

[11] Mądziel, M., Jaworski, A., Kuszewski, H., Woś, P., Campisi, T., Lew, K. (2021). The development of $CO_2$ instantaneous emission model of full hybrid vehicle with the use of machine learning techniques. Energies, 15(1): 142. https://doi.org/10.3390/en15010142

[12] Joy, H.K., Kounte, M.R. (2019). An overview of traditional and recent trends in video processing. In 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT). Tirunelveli, India. pp. 848-851. https://doi.org/10.1109/ICSSIT46314.2019.8987896

[13] Nishio, Y., Shen, T. (2021). Model predictive control with traffic information-based driver's torque demand prediction for diesel engines. International Journal of

Engine Research, 22(2): 674-684. https://doi.org/10.1177/1468087419851678

[14] Xing, J., Zheng, S.X., Ding, D., Kelly, J.T., Wang, S.X., Li, S.W., Qin, T., Ma, M., Dong, Z.X., Jang, C., Zhu, Y., Zheng, H.T., Ren, L., Liu, T.Y., Hao, J.M. (2020). Deep learning for prediction of the air quality response to emission changes. Environmental Science & Technology, 54(14): 8589-8600. https://doi.org/10.1021/acs.est.0c02923

[15] Farahzadi, L., Kioumarsi, M. (2023). Application of machine learning initiatives and intelligent perspectives for $CO_2$ emissions reduction in construction. Journal of Cleaner Production, 384: 135504. https://doi.org/10.1016/j.jclepro.2022.135504

[16] Candau, F., Dienesch, E. (2017). Pollution haven and corruption paradise. Journal of environmental economics and management, 85: 171-192. https://doi.org/10.1016/j.jeem.2017.05.005

[17] Subramani, C., Vijayakumar, K., Dakyo, B., Dash, S.S. (2022). Proceedings of International Conference on Power Electronics and Renewable Energy Systems: ICPERES 2021. Singapore, Springer.

[18] Khan, Z., Murshed, M., Dong, K., Yang, S. (2021). The roles of export diversification and composite country risks in carbon emissions abatement: Evidence from the Signatories of the Regional Comprehensive Economic Partnership Agreement. Applied Economics, 53(41): 4769-4787. https://doi.org/10.1080/00036846.2021.1907289

[19] Liu, J.Y., Murshed, M., Chen, F.Z., Shahbaz, M., Kirikkaleli, D., Khan, Z. (2021). An empirical analysis of the household consumption-induced carbon emissions in China. Sustainable Production and Consumption, 26: 943-957. https://doi.org/10.1016/j.spc.2021.01.006

[20] Murshed, M., Ali, S.R., Banerjee, S. (2020). Consumption of liquefied petroleum gas and the EKC hypothesis in South Asia: Evidence from cross-sectionally dependent heterogeneous panel data with structural breaks. Energy, Ecology and Environment, 6: 353-377. https://doi.org/10.1007/s40974-020-00185-z

[21] Rahman, M.M., Saidi, K., Mbarek, M.B. (2020). Economic growth in South Asia: The role of $CO_2$ emissions, population density and trade openness. Heliyon, 6(5): e03903. https://doi.org/10.1016/j.heliyon.2020.e03903