International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Algorithmic Modeling for Predicting Carbon Emissions in an Individual Vehicles: A Machine Learning and Deep Learning Approach

[1]Rashmi B. Kale, [2]Nuzhat Faiz Shaikh,

Submitted: 05/02/2024 **Revised**: 13/03/2024 **Accepted**: 19/03/2024

Abstract: This study proposes an algorithmic model aimed to accurately predicting carbon emissions from individual vehicles by leveraging machine learning and deep learning techniques. Concerns regarding environmental sustainability and climate change have intensified the need for precise assessments of carbon footprints, particularly in the transportation sector. Traditional methods often lack the adaptability and scalability required to handle the complexity of emission prediction tasks. In contrast, machine learning and deep learning offers promising avenues for developing robust models capable of learning from vast datasets and capturing intricate patterns in vehicle emissions. The purpose of research is to address the breach by designing a deep learning algorithmic framework that integrates machine learning algorithms to analyze real time datasets with vehicle attributes, driving patterns, and fuel characteristics to predict carbon emissions. The proposed approach holds potential for enhancing our understanding of vehicle emissions dynamics and facilitating the development of targeted interventions to mitigate environmental impacts.

Keywords: *Machine Learning, Deep Learning, Air Quality Index, Green House Gases, CariQ, Mean Squared Error, Root Mean Squared Error, Mean Absolute Error, R-squared Score*

1. Introduction

The transportation sector is a significant contributor to global carbon emissions. As climate change and sustainability continue to mount, demand for effective strategies to evaluate and mitigate carbon emissions from vehicles grows. The largest problem facing all nations in the world is carbon dioxide (CO_2) emissions. The rate at which greenhouse gas emissions are rising is concerning [1]. The majority of these emissions of greenhouse gases is made up of CO_2. The emission of greenhouse gases (GHGs) is caused by a variety of factors, including forest fires, vehicle emissions, energy consumption, industrial production and waste, agricultural production and waste, marine transportation by sea, and the use of fossil fuels like coal, oil, and natural gas. The methods used to estimate emissions in the past are examined here in order to enhance and refine the new prediction models. The primary focus of this study is on carbon emissions from motor vehicles. Forecasting of vehicle carbon emissions depending on a variety of factors, including fuel type, fuel consumption, vehicle type, torque, distance traveled, emissivity, emission factor, pollution emissions, and activity level [2, 3]. These factors are incorporated into many models that forecast carbon emissions and enhance the air quality index (AQI). The difficulties in reducing carbon emissions are covered in the work. A recent survey on the calculation of carbon emission shows that a total of 36.6 billion tons of carbon dioxide is released every year approximately [4,5]. This work outlines the methodology for designing and implementing the algorithmic model, by combining the power of machine learning and deep learning with domain expertise in environmental science and transportation engineering. It aims to advance the understanding about vehicle emissions dynamics and contribute to the development of data-driven solutions for mitigating climate change.

Incorporating real-time datasets of vehicles generated through CariQ device from past 3 years as well as data collected from self-implemented and created IoT device. Data collected from both the devices over the past three years provides a valuable foundation for developing and validating the proposed algorithmic model. By leveraging this rich dataset, the system is trying to predict the carbon emission of individual vehicles.

The utilization of real-time data allows for a more accurate representation of current emission patterns, enabling the model to adapt and evolve in response to changing environmental conditions and driving behaviors.

2. Literature Survey:

In the list of pollutants, the emission of carbon dioxide is much higher. Many countries are dealing with this problem. The transportation sector's contribution to carbon emission always needs to be updated to predict the impacts on air quality [6,7,8,9]. Poor air quality index leads to different

[1]*Department of Computer Engineering Smt. Kashibai Navale College of Engineering (SPPU), Pune, India*
rashmi.kale2705@gmail.com
ORCID:0009-0004-8997-019X
[2]*Department of Computer Engineering, Wadia College of Engineering, Pune, India*
Email:nfshaikh76@gmail.com

health issues and effects on living beings. The number of vehicles increases every year and so is the increase in the emission of carbon gas. In urban areas or in metropolitan cities vehicular emission is significantly higher than the rural areas. Carbon emission is the major factor that contributes to global warming across the world. Transit buses, heavy motor vehicles, and personal transport or basically automobiles are the primary causes of carbon emissions from vehicles including the burning of fossil fuels. The need to use ML and DL for the prediction or classification of carbon emission is of very importance now. The newly designed neural network algorithms play an important role in it. There are many applications of ML and DL in prediction. Numerous examples from the present literature, spanning from automated quality inspections to inline fault forecasts, show that ML or DL-based predictive quality is feasible [11]. In the field of biomedicine [12,13] also, the prediction is carried out by using ML or deep learning only. Similarly, the prediction of carbon emission is carried out by using different existing ML algorithms. But it provides a higher mean squared error (MSE). Therefore, in the next session the proposed methodology is introduced which gives very less MSE.

Here, the problem is that people don't know how much pollution they produce every year or month through their vehicles. This work proposes an unconventional approach for predicting carbon emissions in an individual vehicle using machine learning and Deep learning techniques. It offers a powerful framework for analyzing complex datasets and extracting meaningful insights, making it well-suited for tackling the multifaceted nature of vehicle emissions prediction [14,15,16,17]. By leveraging learning algorithms, the study aims to develop a robust predictive model capable of accurately estimating carbon emissions based on various factors such as vehicle characteristics, driving behavior, and fuel type.

3. Methodology

The integration of real-time individual carbon emission data from the past three years through IoT device serves as a cornerstone for the development of a robust and effective algorithmic model for predicting vehicle emissions. By harnessing the power of data-driven insights, system seeks to improve environment friendly transportation options to slow down global warming.

.

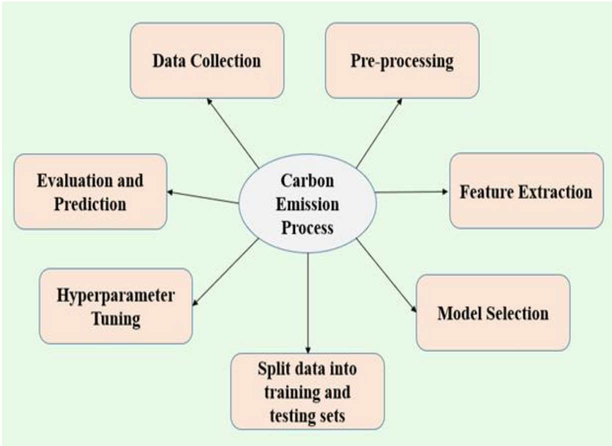

Fig.1: Dataset Processing

The dataset collected on vehicle characteristics like Vehicle Type, Company, Model, Engine Size, Cylinders, Vehicle No., Date of Purchase, Fuel Type, Amount Spend on Fuel, Fuel Consumption, Km Travelled, CO_2 Emission, Fuel Price, Emission Factor. Extracted relevant features from the dataset. Normalized the features to ensure it has similar scale. Chosen a Neural Networks Deep learning model for prediction. Divided the gathered information into testing and training sets. System used characteristics as input and carbon emissions as the goal variable to train the selected model on the training set. Hyperparameters were adjusted to maximize model performance. Metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and R-squared Score can be used to evaluate the accuracy of the trained model by comparing its performance to the testing dataset[18]. The model can be used to forecast carbon emissions for fresh or unseen data once it has been trained and assessed.

A. Basic Algorithmic Steps:

Basic Steps to design the algorithm shown in fig. 2. Collected datasets, undergoes smoothing using a rotating window technique. This smoothing helps to reduce noise and fluctuations in the data. The rotating window refers to a moving window that slides over the sensor data. The smoothed data reflects a more consistent and averaged representation of the sensor readings. It is useful for sensor data that may contain variations due to environmental factors, measurement errors and noise. Rotating window size of 2 that slides over the sensor data for averaging each data point with its adjacent one. So, each data point smoothed with its immediate neighbor.

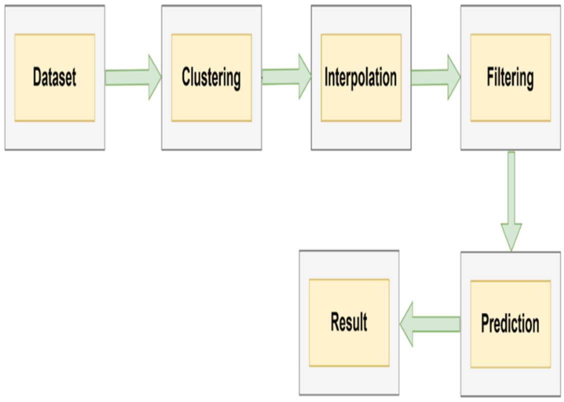

Fig.2: Implementation Flow of the Algorithm

Basic Algorithm:

1. CO_2 emission data collection from IoT device connected to various cars run across Maharashtra
2. Pre-Processing of the Dataset (cleaning, removing duplicates etc.)
3. Statistical analysis of the dataset
4. Divide the datasets into Clusters as per Vehicle Number
5. Detect the unequal intervals through Interpolation
6. Apply Filtering to the dataset for avoiding ambiguities and smoothening of the data.
7. Feature extraction and scaling
8. Divide the dataset into training and testing
9. Send the data to deep learning model through lookback
10. Predict the carbon emission through Predictive Model

Algorithm1: Preprocessing of the Dataset

1. Start
2. Set window_size to 2.
3. Perform moving average filtering on per_hundred_KM_CO_2_emission using the window_size.

   per_hundred_KM_CO_2_emission= moving_average (per_hundred_KM_CO_2_emission, window_size)
4. Print the filtered per_hundred_KM_CO_2_emission.
5. Initialize XX as an empty list
6. Initialize predict_XX as an empty list.
7. Iterate i over the length of per_hundred_KM_CO_2_emission
8. Append the i to XX
9. Iterate i over twice the length of per_hundred_KM_CO_2_emission: Append (i + length of per_hundred_KM_CO_2_emission) to predict_XX.
10. Initialize pp as an empty list.
11. Iterate kk over per_hundred_KM_CO_2_emission:
12. Append [kk] to pp
13. Set look_back to 1
14. Create datasets trainX and trainY using create_dataset function with (np.array(pp), look_back)
15. Print trainX.

Algorithm2: Novel LSTM model to predict CO_2 Emission

1. Start
2. Take first input trainX and reshape it to have dimensions (samples, 1, features)
3. Take second input trainX and reshape it to have dimensions (samples, 1, features)
4. Define and compile an LSTM model:

   - Create a Sequential model.
   - Add an LSTM layer with 4 units and input shape (1, look_back).
   - Add a Dense layer with 1 unit.
   - Compile the model
5. Fit the LSTM model on trainX and trainY with 100 epochs and batch size 1.
6. Predict using the trained model on trainX:
7. trainPredict = model.predict(trainX)
8. Reshape XX and per_hundred_KM_CO_2_emission for regression:

   X11 = np.array(XX).reshape(-1, 1)

   Y11 = pp.array(per_hundred_KM_CO_2_emission) .reshape(-1, 1)
9. Prepare data for prediction:

   to_predict_x = np.array(predict_XX).reshape(-1, 1)
10. Make predictions using the regression model (regsr):

   predicted_y1 = regsr.predict(to_predict_x)
11. Calculate mean squared error (mse) between y_true and predicted_y1:

   mse = mean_squared_error(y_true, predicted_y1)

12. Print the mse.

13. Stop

Collected datasets, undergoes smoothing using a rotating window technique. This smoothing helps to reduce noise and fluctuations in the data. The rotating window refers to a moving window that slides over the sensor data. The smoothed data reflects a more consistent and averaged representation of the sensor readings. It is useful for sensor data that may contain variations due to environmental factors, measurement errors and noise. Rotating window size of 2 that slides over the sensor data for averaging each data point with its adjacent one. So, each data point smoothed with its immediate neighbor.

Here's how the process works:

1. Initial Position: Start with the first two data points.

2. Calculate Average: Take the average of these two points.

3. Move Window: Slide the window by one data point.

4. Repeat: Repeat steps 2-3 until the end of the dataset is reached.

Filtering of data typically involves processing data to remove noise or unwanted information while retaining the essential features. Filtering improves data quality and makes it more suitable for applications. The selection of a filtering method is contingent upon the characteristics of the sensor data, the nature of the noise, interference present, and the specific requirements of the application. Experimentation and evaluation are often necessary to determine the most appropriate filtering approach for a given dataset.

Here, in this predictive modeling Moving average filter is used. This technique involves replacing each data point with the average value of neighboring points within a fixed-size window. Moving average filtering can smooth out fluctuations in the data and provide a clearer representation of the underlying trend.

In the conventional approach, machine learning algorithms are used to predict carbon emissions based on sensor readings. However, to enhance accuracy and reduce data discrepancies, a neural network model is employed. The validation of the system hinges on the input provided and the method used to extract output from the model. This validation process determines the system's overall accuracy. To achieve this, the system utilizes neural networks for analysis.

The sequential model incorporates additional components, including a neural network with two layers. The first layer comprises LSTM units with four neurons, while the second layer is dense. A lookback is configured for the post value in the dense layer. Memory allocation within the neural network is strategically designed to generate subsequent values based on previous ones, facilitated by a designated function. Following the application of filtering and interpolation, curve fitting becomes more manageable. A proper curve fit indicates the neural network's efficacy, resulting in accurate and desirable outcomes.

B. Equations

MSE is computed as follows if the dataset contains *n* data points, predicted values $\hat{y}_i$ and actual values y_i for *i=1,2,...,n*, then the MSE is calculated as:

$$\text{MSE}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \quad (1)$$

The RMSE is simply the square root of the MSE. It is a measurement of the errors' mean magnitude. RMSE is calculated as:

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

The average of the absolute Errors is measured by the MAE. In comparison to MSE, it is less susceptible to outliers. MAE is calculated as:

$$\text{MAE}=\frac{1}{n}\sum_{i=1}^{n}\left(\left|y_i-\hat{y}_i\right|\right)^2 \quad (3)$$

The percentage of the dependent variable's variance that can be predicted from the independent variables is expressed as the R-squared score, sometimes referred to as the coefficient of determination. On a scale of 0 to 1, 1 denotes an ideal fit. R^2 is calculated as:

$$R^2=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \quad (4)$$

4. Dataset

The dataset is formed using the real time CO_2 emission readings collected using IOT device [add reference of your paper] connected to various cars. These cars are run on various kinds of roads across Maharashtra, India. The data collection provides an accurate representation of the range of potential CO2 emissions for a vehicle based on several parameters. It has fourteen features. Table 1 displays the dataset's features, with int64 representing the actual number and fload64 representing a combination of positive and negative values separated by a decimal point. Some sample values showed in Table 2.

TABLE I. FEATURES OF THE DATASET

Feature Variable	***Data Types***
#SR. NO.	int64
#VEHICLE TYPE	int64
#COMPANY	object

Feature Variable	Data Types
#MODEL	object
#ENGINE SIZE	float64
#CYLINDERS	int64
#VEHICLE NO.	object
#DATE OF PURCHASE	object
#FUEL TYPE	object
#AMOUNT SPEND ON FUEL	int64
#FUEL CONSUMPTION (Lit.)	float64
#KM TRAVELLED	float64
#CO2 EMISSION	float64
#FUEL PRICE	float64
#EMISSION FACTOR	float64

TABLE II. SAMPLE VALUES

Feature Variable	*Sample #1*	*Sample #2*	*Sample #3*
SR. NO.	236	169	295
VEHICLE TYPE	CAR	CAR	CAR
COMPANY	HYUND AI	MARUT I	HYUN DAI
MODEL	I20	Swift Dezire	GRAN D i10
ENGINE SIZE	1.2	1.2	1.2
CYLINDERS	4	4	4
VEHICLE NO.	MH 05 CM-5993	MH 28 V-3753	MH 12 MB-0357
DATE OF PURCHASE	12/12/20 15	15/03/20 12	15/5/20 15
FUEL TYPE	Petrol	Diesel	Petrol
AMOUNT SPEND ON FUEL (Rs)	435	41	122
FUEL CONSUMPTION (Lit.)	4.07	0.4	1.16
KM TRAVELLED (Km)	54	6	16.5
CO2 EMISSION	9.43	1.2	2.67
FUEL PRICE	103	92.03	105
EMISSION FACTOR	2.29	2.68	2.29

5. Results and Discussion

This section illustrates statistical analysis and predictive analysis.

Statistical analysis:

The gathering, handling, analyzing, and presenting of data is the main focus of statistical analysis. A statistical analysis of the gathered dataset is shown in Table 3. This statistical analysis aims to monitor fuel consumption and emissions of carbon dioxide from various vehicle models, brands, classes, cylinder counts, engine sizes, transmissions, and fuel types.

TABLE III. STATISTICAL ANALYSIS RESULTS

Variable	*Me an*	*Std Dev*	*Mi n*	*25 %*	*50 %*	*75 %*	*M ax*
Engine Size	1.34	0.35	1.1	1.1	1.2	1.2	2
Cylinders	4	0	4	4	4	4	4
Amount Spend On Fuel (Rs)	225. 2	386. 24	11	45	78	18 2.5	26 62
Fuel Consump tion (Lit.)	2.29	4.02	0.1	0.5	0.8	1.9	28. 9
Km Travelled	34.2 6	61.0 9	1	7	12. 2	29	44 8
Co2 Emission	5.85	10.4 9	0.2 3	1.18	2	5.0 5	77. 5
Fuel Price	98.9 6	4.92	92. 03	95.3 7	96. 35	10 5	10 5
Emission Factor	2.52	0.19	2.3	2.3	2.6 8	2.6 8	2.6 8

Fuel consumption liters per 100 Km is given in Table 4.

TABLE IV. FUEL CONSUMPTION OF VARIOUS MAKERS PER 100 KM

Car Company	*Fuel consumption Liters/100 Km*
HYUNDAI	13.59
MARUTI	15.26

Car Company	*Fuel consumption Liters/100 Km*
TATA	16.08
VOLKSWAGEN	15.86

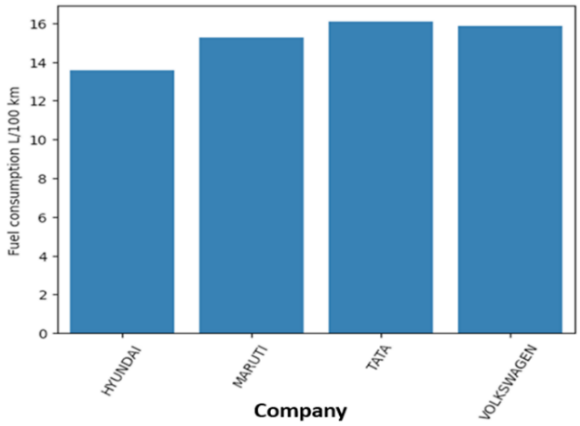

Fig.3: Fuel Consumption of per hundred kilometers by Companies

To specify the degree of strength of two characteristics in the dataset and to analyze how the brand, model, vehicle class, cylinder, engine size, transmission type, and fuel type correlate

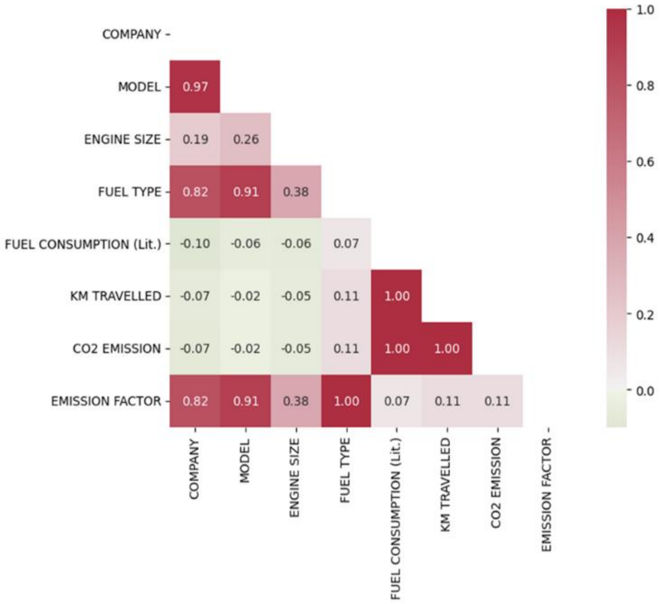

Fig.4: Correlation of Dataset Attributes through Heat map

in order to produce correlation coefficients, a correlation method has been created for the emissions and consumption of different cars. This statistic's goal in this study is to identify the parameter that most closely correlates with CO2 emissions. In order to do this, a correlation heat map, depicted in Figure 4, was created by applying and computing Pearson's correlation coefficients between all features across all cars. It is seen that CO2 emission has highest correlation with Fuel consumption and distance travelled in kilometers.

Various machine learning algorithms such as Decision Tree Regressor, Random Forest Regressor Gradient Boost Regressor and our proposed novel LSTM are applied to this regression problem. Table 4 shows the comparative results of various algorithms with our proposed LSTM algorithm.

TABLE V. SAMPLE VALUES

Model	*MSE*	*MAE*	*RMSE*	*R2 Score %*
Decision Tree Regressor	2.205	0.439	1.485	97.4
Random Forest Regressor	3.174	0.504	1.781	96.3
Gradient Boost Regressor	1.799	0.372	1.341	97.9
Proposed LSTM Regressor	0.99	0.12	0.76	98.1

From the results it is observed that the proposed LSTM model has low MSE, MAE, RMSE and high accuracy.

References

[1] Sannigrahi, S., Pilla, F., Basu, B., Basu, A. S., Sarkar, K., Chakraborti, S Roy, P. S. (2020). Examining the effects of forest fire on terrestrial carbon emission and ecosystem production in India using remote sensing approaches. Science of The Total Environment, 138331. doi:10.1016/j.scitotenv.2020.1383

[2] Rebolledo-Leiva, R., Angulo-Meza, L., Iriarte, A., & González-Araya, M. C. (2017). Joint carbon footprint assessment and data envelopment analysis for the reduction of greenhouse gas emissions in agriculture production. Science of The Total Environment, 593-594, 36–46. doi:10.1016/j.scitotenv.2017.03.1

[3] Lin, X., Zhu, X., Han, Y., Geng, Z., & Liu, L. (2020). Economy and carbon dioxide emissions effects of energy structures in the world: Evidence based on SBM-DEA model. Science of The Total Environment, 138947. doi:10.1016/j.scitotenv.2020.1389

[4] Suzanne Greene a, Haiying Jia b, Gabriela Rubio-Domingo c. (2020). Well-to-tank carbon emissions from crude oil maritime transportation. Transportation Research, 102587. sci-hub.se/10.1016/j.trd.2020.102587

[5] Xiaojun Ma, Changxin Wang, Biying Dong, Guocui Gu, Ruimin Chen, Yifan Li, Hongfei Zou, Wenfeng Zhang, Qiunan Li (2019). Carbon emissions from energy consumption in China: Its measurement and

driving factors. Science of The Total Environment, Volume 648, 15 January 2019, Pages 1411-1420

[6] Sun, S., Li, L., Wu, Z., Gautam, A., Li, J., & Zhao, W. (2020). Variation of industrial air pollution emissions based on VIIRS thermal anomaly data. Atmospheric Research, 105021. doi:10.1016/j.atmosres.2020.10502

[7] Luo, L., & Chen, Y. (2020). Carbon Emission Energy Management Analysis of LCA-Based Fabricated Building Construction. Sustainable Computing: Informatics and Systems, 100405. doi:10.1016/j.suscom.2020.100405

[8] Ali Newaz Bahar1, Nazrul Islam, Shougat Hossain, Ruhul Amin Sujon (2015). A New Automation Approach for Fuel Station Management System, Nevşehir Bilim ve Teknoloji Dergisi Cilt 4(2) 99-107 2015 DOI: 10.17100/nevbiltek.210935 URL: http://dx.doi.org/10.17100/nevbiltek.210935

[9] Sheth, S., Patel, K. H., & Patel, H. (2016). Design of Automatic Fuel Filling System Using a Mechatronics Approach. CAD/CAM, Robotics and Factories of the Future, 785–795. doi:10.1007/978-81-322-2740-3_76

[10] Machine Learning and Edge Cloud Computing, Advanced Engineering Informatics, Volume 45,2020,101101,ISSN 1474-0346,https://doi.org/10.1016/j.aei.2020.101101.

[11] Emami Javanmard, M., Ghaderi, S. F., & Hoseinzadeh, M. (2021). Data mining with 12 machine learning algorithms for predict costs and carbon dioxide emission in integrated energy-water optimization model in buildings. Energy Conversion and Management, 238, 114153. doi:10.1016/j.enconman.2021.114153

[12] Mahfuzur Rahman, Izlin Ismail, & Shamshul Bahri (2020). Analyzing consumer adoption of cashless payment in Malaysia. Digital Business, Volume 1, Issue 1, September 2020, 100004

[13] M. Singh and R. K. Dubey, "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, Jan. 2023, doi: 10.1109/TIV.2021.3102400.

[14] R. Kale, Nuzhat F. Shaikh (2021). Forecasting Models for Carbon Emission: A survey and Discussion. International Journal of Mechanical Engineering. Vol. 6 No. 3 December, 2021

[15] R. Kale, Nuzhat F. Shaikh, K. Wankhade (2023). “Survey Paper on Smart Automation System for Calculating the Vehicular Carbon Emission using gadget CariQ and its Applications” International conference on Recent Trends in Science, Technology, Engineering & Mathematics (MICRT-STEM 2023), February 24, 2023. ISBN-13: 978-93-5967-432-2.

[16] R. Kale, Nuzhat F. Shaikh (2024). IoT-Enabled Forecasting of Vehicle based Carbon Emissions and Smart Fuel Management: Optimizing Efficiency and Performance. International Journal of Intelligent Systems and Applications in Engineering (IJISAE), February 24, 2024, 12(19s), 161–167, ISSN:2147-6799214,

[17] Dhotre, D.R., Dahake, R., Choubey, N., Khandare, A., Patil, M. and Jadhav, A., 2024. Multi-Class Classification of Brain Disease using Machine Learning-Deep Learning approaches and Ranking based Similar Image Retrieval from Large Dataset. International Journal of Intelligent Systems and Applications in Engineering, 12(1s), pp.771-782.

[18] Dhotre, D.R. and Bamnote, G.R.,. Multilevel haar wavelet transform and histogram usage in content based image retrieval system. In 2017 International Conference on Vision, Image and Signal Processing (ICVISP) (pp. 82-87). IEEE, September, 2017.