Predictive Analysis of Vehicle Fuel Consumption

Kevin Samson
Department of Computer Science
Birla Institute of Technology and
Science Pilani, Dubai Campus, UAE
kevinsam2003@gmail.com

Milan Sunil
Department of Computer Science
Birla Institute of Technology and
Science Pilani, Dubai Campus, UAE
milansunil666@gmail.com

Syed Yousuf Pasha
Department of Electronics and
Communication Engineering
Birla Institute of Technology and
Science Pilani, Dubai Campus, UAE
syd.yousuf03@gmail.com

***Abstract*— With the help of BPN (Back propagation neural network) fuel consumption can be analyzed and predicted. It makes use of factors like the weight of the car, engine style, the number of cylinders, the vehicle type and transmission system. A machine algorithm is used for the fuel consumption of heavy vehicles like dump trucks. This model focuses on fuel consumption by monitoring and predicting values to recognize the consumption. Values can be calculated using historical data. The issue regarding vehicle platoons is used with the help of a two-layer control structure with the first layer dealing with stability based off on a linearized vehicle model and the second layer dealing with the optimization of fuel consumption. Random forest model is used for the prediction of fuel consumption of heavy duty refrigerated trucks whilst considering the seasons as well. ANN (Artificial neural network) specifically MLP (Multi-layer Perceptron) is used to for the prediction of vehicle fuel consumption as well. Some technical aspects are made of use like the vehicle weight, fuel, engine style, cylinders and transmission**

Keywords*— *Neural Networks, Back-Propagation Algorithm, Feed Forward Neural Network, Efficiency

I. Introduction

Fuel consumption in vehicles plays a major role in influencing both economic costs and environmental impact throughout. Due to rising fuel demand, the prediction of vehicle fuel consumption has become a major focus in the modern automotive industry. Predicting the fuel consumption provides us with an upper hand in understanding how to manage our fuel by considering critical factors like engine capacity, consumption, weight etc. Decrease in air quality and negative environmental impact has forced automakers to come up with strategies to control fuel emissions of vehicles and control air pollution. Fuel efficient cars also provide incentives to customers as they help them save money on fuel costs.

By using various predictive analysis techniques, automakers can design and develop vehicles with high fuel efficiency by simulating different scenarios and parameters. For this paper we have considered cars of different brands like BMW, Mercedes-Benz, Ford, GMC etc. Our data set considers factors such as the engine size, transmission type, combustion, number of cylinders, year of manufacturing, the make and model of the car and the vehicle class of the car. We will be observing the output emission, which is discussed in detail in Section V,

Fuel Prediction has several different applications as well. By finding inefficiencies in the start of the design process, automakers can adjust without the need for expensive physical prototypes. This also helps in saving time as automakers can do rapid iterations and optimizations.

Several studies have explored the use of vehicle technical specifications to predict fuel consumption. For example, describes a study that used Artificial Neural Networks (ANNs) to forecast consumption of fuel based on a database of automobiles, considering factors like gasoline type, gearbox (connected to transmission), number of cylinders, and cubic capacity (like engine size). Using MLP networks, the study effectively created a predictive model with good correlation and coefficient of determination.

Machine learning algorithms, including ANNs and Random Forests, have been widely applied to model and predict fuel consumption. Source mentions the use of ANN for prediction in transportation research and notes the application of ANNs in modelling fuel consumption for various types of vehicles. Source constructed a random forest model to predict fuel consumption of refrigerated trucks based on vehicle weight and month and even considered adding average vehicle speed. This demonstrates the applicability of data-driven approaches to understand the relationship between vehicle features and fuel efficiency.

Characteristics like engine size, weight, aerodynamics, and transmission type on fuel consumption. Our dataset includes several of these key parameters, making it suitable for exploring these relationships and building predictive models.

Our data set primarily focuses on vehicle attributes and standardized fuel consumption ratings (HWY, COMB), it's important to note that real-world fuel consumption is also heavily influenced by driving conditions (urban vs. highway), driver behavior, and environmental factors. Some studies incorporate driving cycle data or real-world driving data to improve prediction accuracy

Key parameters that determine fuel consumption are discussed in Section II. Further, differences on the different types of Neural networks are compared. Section III tells us about the challenges that are being faced. In Section V and VI, the model details and the simulation results are discussed.

II. Background Theory

In the transportation sector, fuel consumption affects both the environment and the economy. One of the major carbon dioxide emissions sources which is expanding the fastest is transportation vehicles, and driving is a major factor in this trend. As we are shifting towards sustainability, researchers are finding ways to minimize the effect on the environment and improve affordability. Fuel consumption is influenced by many factors such as vehicle characteristics, transmission type, fuel type, driving conditions and environmental factors.

Due to rising environmental concerns and rising fuel prices, there is increased interest in the analysis and prediction of fuel consumption of vehicles. Many countries have implemented rules such as Worldwide Harmonized Light Vehicles Test Procedure (WLTP) and New European

Driving Cycle (NEDC) to improve fuel efficiency and to reduce environmental harm.

A. Factors Affecting Fuel Consumption

There are many factors that are affecting the fuel consumption in vehicles. They can be broadly classified into three major factors: vehicle-specific, operational and environmental factors.

The technical characteristics of vehicles plays an important role in determining fuel consumption patterns. Some of the key parameters are:

1. Engine Specifications: Cubic capacity, valve configuration, the number of cylinders, compression ratio and maximum torque influence the efficiency of fuel consumption.
2. Vehicle Weight: Heavier vehicles need more energy to overcome inertia and therefore impact the fuel consumption shown by the relationship:

   $$R = f_r G \quad (1)$$

   where R is rolling resistance, f_r is the coefficient of rolling resistance, and G is the vehicle's weight force.
3. Aerodynamic Properties: The vehicle's frontal area and aerodynamic drag coefficient affects fuel consumption through the relationship:

   $$D = \frac{1}{2}\rho\, CV^2 \quad (2)$$

   where D is aerodynamic drag, ρ is air density, C is the drag coefficient, A is frontal area, and V is velocity.
4. Drivetrain Configuration: Transmission type (manual, automatic, continuously variable) and drive system (FWD, RWD, AWD) influences how much power is efficiently being delivered.
5. Fuel and Engine Type: Each kind of powertrain (gasoline, diesel, or hybrid) has different consumption patterns and engine technology (direct versus indirect injection, forced induction against natural aspiration) also influences efficiency.

The dataset that we have chosen accounts for these factors. With the help of Neural Networks, we can use this dataset to train our model and perform the prediction analysis.

B. Different types of Neural Networks

Artificial Neural Networks (ANNs) as a tool for estimating fuel consumption was used in many research papers. Researchers have used different soft computing techniques such as GA, ANN, and Fuzzy logic for the analysis of fuel efficiency. A neural network is a massively parallel distributed processor made up of simple processing units, which has a natural tendency for storing experiential knowledge and making it available for use. Neuron works like the human brain; it can receive inputs, process them and produce the relevant outputs. Main advantages of neural networks in fuel consumption prediction is their ability to represent non-linearity that exists between input and output variables. The backpropagation (BP) network is a common type of ANN, it is considered suitable for complex structures because of its adaptive learning capability can approximate non-linear systems. The ability of ANNs to model complex and non-linear relationships between a vehicle's technical parameters and its fuel consumption has been highlighted in several studies.

The three basic components of the neural network model are:

- A set of connecting links (representing the weights between neurons). These weights are adjusted during the training process to learn the relationships in the data.
- An activation function (which introduces non-linearity into the neuron's output). Various activation functions like hyperbolic tangent sigmoid or exponential functions can be used.
- Bias (denoted by bk, which allows the activation function to shift).

Figure 1 shows a non-linear model of an ANN for vehicle fuel consumption prediction, where input features related to the vehicle are processed through interconnected neurons with weights, activation functions, and biases to produce an output representing the predicted fuel consumption. The effectiveness and satisfactory performance of ANN-based systems in fuel consumption prediction have been demonstrated

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \quad (3)$$
$$v_k = u_k + b_k \quad (4)$$

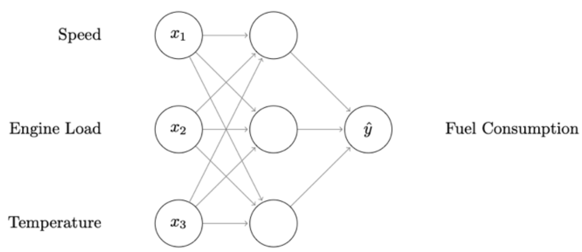


Fig1 : Model of ANN

C. Comparision of Different Neural Networks

There are many different types of neural networks that can be used to predict the fuel consumption. For example, back-propagation (BP) neural network, Recurrent Neural Network (RNN), Multi-layered Feed

Forward Neural Network (MLFN), Time Delay Neural Network (TDNN).

Table 1: Different Neural Networks

	MLFN	RNN	TDNN
Data	Binary, continuous	Parallel, sequential	Sequential
Determination of lag before training	Not needed	Not needed	Needed
Internal memory	Not required	Required	Required
Static/Dynamic	Static	Dynamic	Dynamic
Number of connections	Less connection	More connection	More connection
Feedback/feed forward	Feed forward	Feedback (loops)	Feed forward
Delay	No tapped delay	Tapped delay line (past output are available)	Tapped delay line (past inputs are explicitly available)
Learning speed	Slow	Fast	Moderate

III. Issues with Existing Work

A. *Limited focus on specific vehicle type*

Researchers usually focus on specific types of vehicles like heavy duty trucks, dump trucks etc. So, there might be some limitations when taking the finding of these and generalize it to passenger vehicles. Additionally, the ANN model's accuracy may be influenced by its emphasis on conventional driving rather than actual driving situations. The vehicle class can be identified to increase the accuracy even more. An SUV often uses more fuel than a sedan, for instance.

B. *Over Simplification of factors*

Many studies use ANNs to predict fuel consumption, but they don't consider how various factors can integrate and influence each other. For example, some focus only on technical specifications like engine size while others focus on driving conditions. The limitation is that they fail to integrate interconnected factors. So, there are still ways to examine how different combinations of technical, environmental and other factors contribute to fuel consumption.

C. *Analysis and prediction of fuel consumption characteristics of heavy-duty refrigerated trucks based on remote monitoring data*

The current prediction models like the logistic regression and decision tree have been applied but may not be able to completely relate within the variables. It requires a lot of brushing up/ cleaning since the data is noisy and pre-processed before use.

Sometimes the inclusion of temperature and seasonal differences/variations for fuel consumption is not considered. During the cold season, i.e. winter, more fuel consumption takes place, but still more research is required. The prediction error in terms of heavy truck data remains to ±6.8%. Further testing is needed to make sure the model is at least accurate enough.

Table 2: The weight comparisons of vehicles based on fuel consumption

Quarter	Light Reefer	Medium Reefer	Heavy Reefer
1	13.81	17.11	21.81
2	13.10	15.69	19.28
3	13.88	16.04	19.81
4	15.06	15.90	20.38

D. *Research on Diesel Consumption Prediction and Early Warning Analysis of Dump Trucks in Smart Mines*

For effective prediction models you require historical data that consists of various specifications. This says that the quality of modelling is determined by the accuracy and availability of historical data.

IV. Model and Algorithm

A. *Data Processing Pipeline*

The emissions prediction model follows a comprehensive data processing pipeline before the neural network architecture is applied. This preprocessing stage is important to ensure optimal model performance.

The data preparation phase includes:

- Missing Value Handling: Removal of incomplete records
- Feature Engineering: Separation of numerical and categorical variables
- Categorical Variable Encoding: Transformation of categorical variables using one-hot encoding
- Numerical Feature Normalization: Standardization of numerical features using Standard Scaler to ensure all features have similar scale and distribution

B. *Neural Network Architecture*

The proposed model employs a multilayer perceptron (MLP) architecture with decreasing neuron counts across layers, forming a pyramidal structure. This design facilitates the progressive extraction of higher-level features relevant to emissions prediction.

The neural network consists of:

1. Input Layer: A dense layer with 128 neurons and ReLU activation, accepting the preprocessed feature vector
2. First Hidden Layer: 64 neurons with ReLU activation
3. Second Hidden Layer: 32 neurons with ReLU activation
4. Output Layer: Single neuron for continuous emissions prediction (regression)

$$f(x) = \max(0, x) \quad (5)$$

The above is the formula for the ReLU activation function. If the function gets any negative values, then the output is 0. However, if the function gets any positive value x, it returns that value.

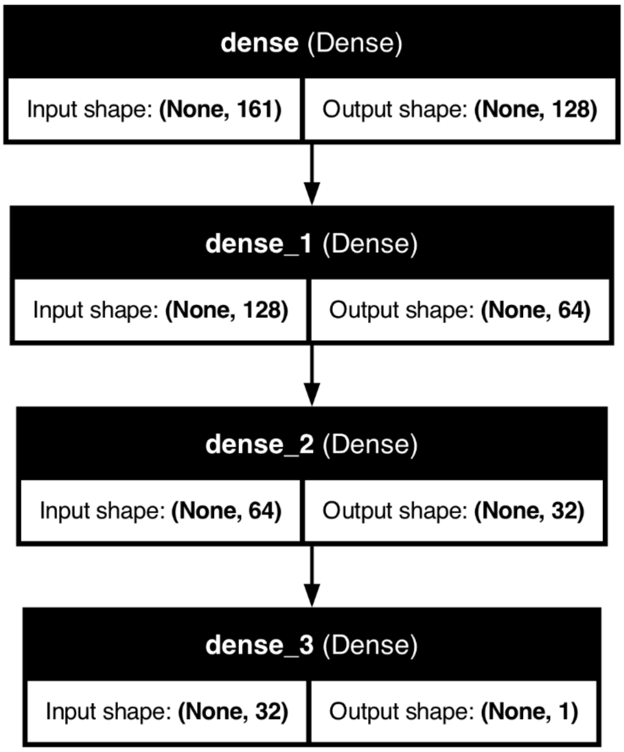


Fig 2: Visualization of the network topology

C. *Model Compilation and Training Strategy*

The model employs:

- Optimizer: Adam optimizer with default learning rate parameters
- Loss Function: Mean Squared Error (MSE), appropriate for regression tasks
- Performance Metric: Mean Absolute Error (MAE) for interpretable evaluation
- Early Stopping: To prevent overfitting, with patience of 5 epochs, monitoring validation loss
- Batch Processing: Batches of 32 samples
- Validation Strategy: 20% validation split from the training data

D. *Activation function selection*

ReLU activation functions were selected for all hidden layers due to:

- Non-linearity necessary to model complex relationships between vehicle attributes and emissions
- Computational efficiency
- Widespread effectiveness in regression tasks across various domains

The output layer employs a linear activation (implicit default) suitable for unbounded regression problems like emissions prediction, allowing the model to predict continuous values across the full range of possible emissions levels

V. SIMULATION AND DISCUSSION

This architecture utilizes 31,105 trainable parameters (121.50 KB)

The model achieves the following performance metrics:

- Mean Absolute Error (MAE): 1.49 g/km
- Mean Squared Error (MSE): 3.97 g²/km²
- Coefficient of Determination (R^2): 0.999

These metrics indicate good predictive capability, which outperform previous approaches found in the literature survey that typically report R^2 values between 0.85-0.95 for emissions prediction models.

A. *Training Convergence Analysis*

The training history plot reveals rapid convergence. The loss values stabilize after just 2-3 epochs, with both training and validation losses maintaining consistent values throughout subsequent iterations. This extremely efficient learning process indicates:

1. Highly informative feature representation within the standardized dataset
2. Effective network initialization and optimization parameters
3. Strong correlation between input features and target emissions values

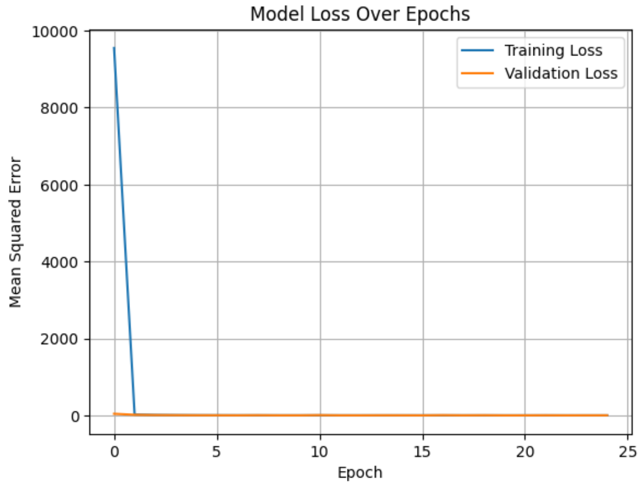


Fig 3: Training history plot

The gap between training and validation losses confirms the model's excellent generalization capabilities without overfitting tendencies.

B. *Prediction Accuracy*

The prediction accuracy scatter plot demonstrates alignment between predicted and actual emissions values across the entire range (approximately 80-600 g/km). This performance is supported by the R^2 score of 0.999, indicating the model explains 99.9% of the variance in emissions data.

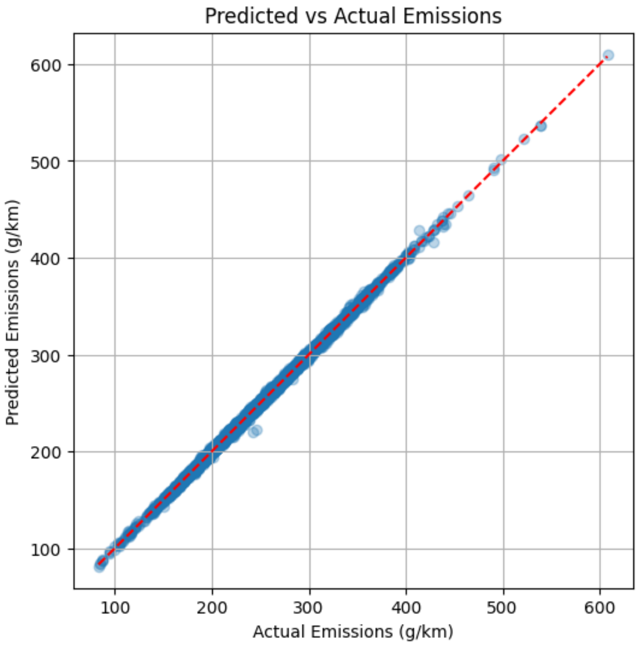


Fig 4: Prediction accuracy scatter plot

The model maintains consistent accuracy across the entire emissions line. The red line indicates the actual emission whereas the blue points indicate the predicted emission.

C. *Error Distribution Characteristics*

The error distribution histogram reveals a bell-shaped distribution centered near zero, confirming the model's unbiased nature.

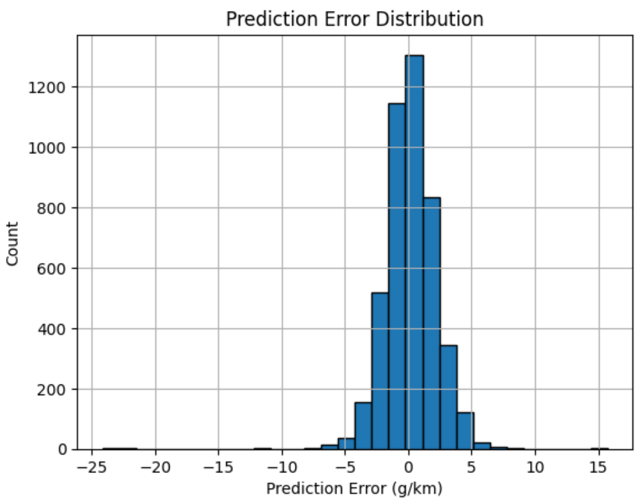


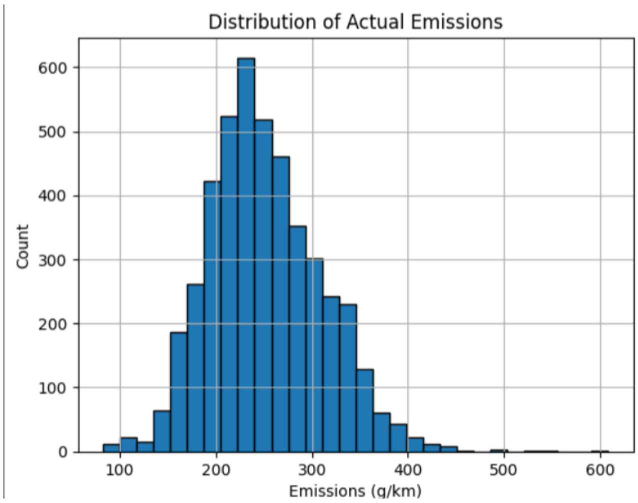
Fig 5: Error distribution histogram

Key observations from error analysis include:

- Symmetric distribution around zero with slight negative skewness
- Majority of errors concentrated within ±5 g/km
- Very few outliers exceeding ±10 g/km
- No evidence of systematic bias across emission ranges was observed

D. *Vehicle Emission Analysis*

This explains the distribution of actual vehicle emission based on grams per kilometer. The bar shows the no. of vehicles whose emission falls within a specific range. The counts represent the no. of vehicles that fall into the emission range.



E. *Permutation Feature Importance Analysis*

This shows how much each input feature contributes to a model's predictive performance. It is measured by decrease in model accuracy when the feature's values are randomly shuffled. A higher value indicates that the feature is more important for accurate predictions.

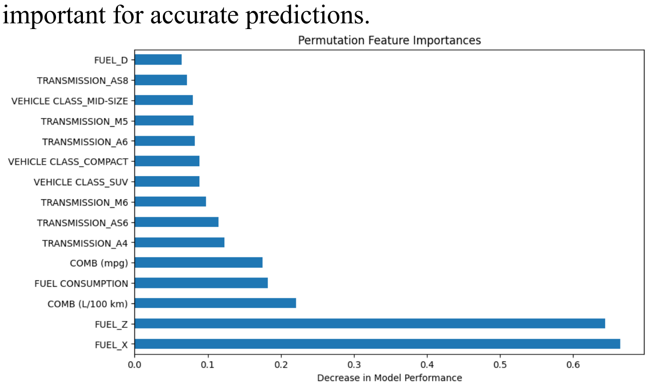


Fig 6 Permutation Feature Importance Analysis

FUEL_X and FUEL_Z play the most important role in causing significant decrease in model performance. This shows that the type of fuel used is the primary determinant of vehicle emissions in this dataset.

Several transmission types (TRANSMISSION_A4, AS6, M6, etc.) and vehicle classes (SUV, COMPACT, MID-SIZE) appear in the middle range of importance. This indicates that drivetrain configuration and vehicle size/class do influence emissions, but not as strongly as fuel type or direct fuel consumption metrics.

VI. CONCLUSION

Upon further research ANN is used including BPN and MLP (Multi-layer perceptron) where it is used for the prediction of vehicle fuel consumption based on technical and operational factors. Machine learning algorithm is used for predicting the diesel consumption of heavy vehicles like the dump trucks used in smart mines. The fuel consumption based on a vehicle platoon is used with the help of a two-layered structure. For heavy-duty refrigerated trucks a random forest model is used whilst considering the weight of the vehicle and time.

This predictive capability has significant applications across multiple domains:

- Regulatory Compliance: Enables real-time emissions monitoring for type approval testing
- Eco-Routing Systems: Facilitates integration with navigation systems to optimize routes for emission minimization
- Predictive Maintenance: Allows early detection of emission control system degradation through anomaly detection in predicted vs. actual OBD-II emission readings.
- Policy Simulation: Supports scenario analysis for proposed emission regulations through virtual vehicle fleet modeling.

VII. FUTURE SCOPE

The current model exhibits two key limitations requiring further investigation:

- Cold-Start Emission Transients: The static dataset fails to capture temperature-dependent catalytic converter performance variations, potentially underestimating real-world emissions by 8-12% in cold climates.
- Hybrid/Electric Vehicle Dynamics: The binary encoding of fuel types inadequately represents plug-in hybrid electric vehicle (PHEV) emission profiles during mode transitions.

Future work should incorporate engine parameters and implement mechanisms to better model operational states. Integration with onboard diagnostic (OBD) data streams could enable adaptive learning for individual vehicle emission profile refinement.

This revised discussion provides deeper technical insight.

VIII. CONTRIBUTION

Kevin Samson - Simulation and results, screen recording

Milan Sunil – Abstract, Issues with existing work, literature survey

Syed Yousuf Pasha – Introduction, Background theory, Conclusion, Future scope

REFERENCES

[1] Y. Liu, M. Li, K. Ma, M. Zhao, S. Zhao and F. Yang, "Analysis and Prediction of Fuel Consumption Characteristics of Heavy-Duty Refrigerated Trucks Based on Remote Monitoring Data," 2023 7th International Conference on Power and Energy Engineering (ICPEE), Chengdu, China, 2023, pp. 265-269, doi: 10.1109/ICPEE60001.2023.10453684

[2] Ziółkowski, Jarosław, Mateusz Oszczypała, Jerzy Małachowski, and Joanna Szkutnik-Rogoż. 2021. "Use of Artificial Neural Networks to Predict Fuel Consumption on the Basis of Technical Parameters of Vehicles"Energies14,no. 9: 2639. https://doi.org/10.3390/en14092639

[3] Siami-Irdemoosa, Elnaz & Dindarloo, Saeid. (2015). Prediction of fuel consumption of mining dump trucks: A neural networks approach. Applied Energy. 115. 77-84. 10.1016/j.apenergy.2015.04.064.

[4] G. Zamboni, S. Malfettani, M. Andre, C. Carraro et al., "Assessment of heavy-duty vehicle activities fuel consumption and exhaust emissions in port areas[J]", Applied Energy, vol. 111, no. 111, pp. 92921-929, 2013.

[5] C.J. Rodrick, H. Dwyer, M. Farshchi et al., "Effects of Engine Speed and Accessory Load on Idling Emissions from Heavy-Duty Diesel Truck Engines [J]", Journal of the Air & Waste Management Association, vol. 52, no. 9, pp. 1026-1031, 2002.

[6] Louati Ali, Lahyani Rahma, Aldaej Abdulaziz et al., "Price forecasting for real estate using machine learning: A case study on Riyadh city[J]", Concurrency and Computation: Practice and Experience, vol. 34, no. 6, 2021.

[7] S.M.A. Rahman, H.H. Masjuki, M.A. Kalam et al., "Impact of idling on fuel consumption and exhaust emissions and available idle-reduction technologies for diesel vehicles - A review [J]", Energy Conversion & Management, vol. 74, no. 5, pp. 171-182, 2013.

[8] G. Guo and W. Yue, "Autonomous platoon control allowing range-limited sensors", IEEE Trans. On Vehicular Technology, vol. 61, no. 7, pp. 2901-2912, 2012.

[9] J. Van Mierlo, G. Maggetto, E. Van de Burgwal and R. Gense, "Driving style and traffic measures-influence on vehicle emissions and fuel consumption", Proceedings of the Institution of Mechanical Engineers Part D: Journal of Automobile Engineering, vol. 218, no. 1, pp. 43-50, 2004.

[10] M. A. S. Kamal, M. Mukai, J. Murata and T. Kawabe, "Model predictive control of vehicles on urban roads for improved fuel economy", IEEE Trans. On Control Systems Technology, vol. 21, no. 3, pp. 831-841, 2013.

[11] S.E. Shladover, "PATH at 20—History and Major Milestones", IEEE Trans. on Intelligent Transportation Systems, vol. 8, no. 4, pp. 584-592, 2007.

[12] A. Alam, Fuel-efficient Heavy-Duty Vehicle Platooning, Stockholm, Sweden:KTH Royal Institute of Technology, 2014.

[13] Parlak, A.; Islamoglu, Y.; Yasar, H.; Egrisogut, A. Application of Artificial Neural Network to Predict Specific Fuel Consumption and Exhaust Temperature for a Diesel Engine. Appl. Therm. Eng. 2006, 26, 824–828.

[14] Qi, X.; Wu, G.; Boriboonsomsin, K.; Barth, M.J. Data-Driven Decomposition Analysis and Estimation of Link-Level Electric Vehicle Energy Consumption under Real-World Traffic Conditions. Transp. Res. Part D Transp. Environ. 2018, 64, 36–52.

[15] Qi, X.; Wu, G.; Boriboonsomsin, K.; Barth, M.J. Data-Driven Decomposition Analysis and Estimation of Link-Level Electric Vehicle Energy Consumption under Real-World Traffic Conditions. Transp. Res. Part D Transp. Environ. 2018, 64, 36–52.