Energy Proceedings
Vol 31, 2023
ISSN 2004-2965

Virtual Sensor: A Machine Learning Based Performance Prediction Tool for Next Generation Intelligent Internal Combustion Engines

Qifan Wang[1], Zhentao Liu[1*], Yu Zhang[2], Jiahong Fu[2]

1 Power Machinery and Vehicular Engineering Institute, College of Energy Engineering, Zhejiang University, Hangzhou 310027, PR China

2 Mechanical Engineering Department, Zhejiang University City College, Hangzhou 310015, PR China

ABSTRACT

With the continuous promotion of the carbon peak emissions and carbon neutralization strategies, higher demands are placed on engine economic performance. Virtual sensors as an online information collection technology can be used to control various performance indicators of engines. Here is an example of ISFC to represent the engine performance prediction. In this paper, the feasibility of three machine learning methods, Artificial Neural Network (ANN), Random Forest (RF), and Support Vector Regression (SVR), for predicting fuel consumption applications are explored. Firstly, a calibrated engine one-dimensional (1D) model is constructed. Then, the 1D model generates a dataset with engine load, engine speed and spark time, and indicative specific fuel consumption (ISFC) as an output, for the training of machine learning methods. The performance of different algorithms was compared using the coefficient of determination (R^2), the root-mean-square error (RMSE), and the mean absolute percentage error (MAPE) as evaluation metrics. By comparing test dataset prediction and map prediction, RF has a large prediction error at boundary operation conditions and ANN sometimes has a relative error of more than 10%. SVR performs well in each statistical index and map prediction, and therefore it is an algorithm that can be used by virtual sensors.

Keywords: carbon peak emissions; carbon neutralization; virtual sensor; machine learning methods

NONMENCLATURE

Abbreviations	
ABDC	After the bottom dead center
ANN	Artificial Neural Network
ATDC	After the top dead center
BBDC	Before the bottom dead center
BTDC	Before the top dead center
CAD	Crank angle degree
ISFC	Indicated Specific Fuel Consumption
MAPE	The mean absolute percentage error
ML	Machine learning
OBD	On Board Diagnostics
RF	Random Forest
RMSE	The root-mean-square error
R^2	The coefficient of determination
SVR	Support Vector Regression
1D	One-dimensional

1. INTRODUCTION

In the current global context of carbon peak emissions and carbon neutralization, research to improve the energy efficiency of engines is receiving increasing attention [1]. The need to improve the performance of electronic control systems for better control of engines implies an increasing number and accuracy of sensors. Virtual sensors are used to replace traditional measurement instruments and devices using existing computers with specially designed modular hardware combined with efficient and flexible software and corresponding algorithms [2]. With the dramatic increase in the amount of data and the continuous development of machine learning methods, many researchers have carried out the development of virtual sensors based on machine learning methods [3]. Unlike the traditional physical model-based inspection methods, the data-driven performance metrics inspection methods based on machine learning methods are based on real-time measurement data from on-board OBD (On Board Diagnostics) devices. The objective is to explore and establish mapping relationships between individual vehicle operating condition data and desired performance metrics, and then use the data to calibrate the model [3]. In the current research of virtual sensor algorithms, the ANN model was used for NO_x [4], HC, CO_2, and CO [5] emissions detection to control engine emissions. Fuel consumption rate is an important indicator of engine economy. In the actual driving environment, the factors affecting vehicle fuel consumption are complex, such as vehicle driving environment, vehicle characteristics, driver behavior, road structure, traffic conditions, etc. The impact is mainly reflected in the real-time operating status data of the vehicle [6]. By constructing fuel consumption models through regression analysis methods in machine learning methods, it is possible to describe their physical meaning without the need for accuracy. This paper discusses the accuracy and applicability of three machine learning methods commonly used for engine performance prediction, namely ANN, SVR, and RF [7]. The expectation is to find methods to accurately establish the relationship between state parameters and vehicle fuel consumption rate. The overall design of the virtual sensor for fuel consumption rate prediction can be seen in Fig. 1.

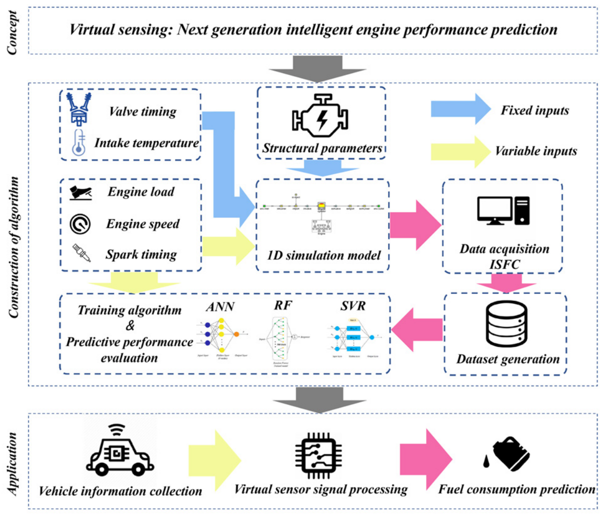

Fig. 1. The overall design of the virtual sensor.

2. METHODOLOGY

2.1 Engine setup

In this research, a single-cylinder 500cc naturally aspirated SI gasoline engine was chosen to construct a 1D GT-Power model. The stroke and bore of the engine are 86.07mm and 86mm, respectively. The compression ratio is 9.5, and connecting rod length is 175 mm. The valve phasing and other important parameters can be seen in Table. 1. The numerical model was calibrated by the reference [8]. In this paper, “EngCylCombSITurb” was selected to predict turbulent combustion, which can be used to obtain SI engine performance. To obtain the inputs data set for the ML method, three kinds of engine input parameters were chosen engine speed, load (controlled by the intake pressure), and spark timing, as shown in Table 2. The engine speed was set to 1000 to 3800 RPM with an interval of 400 RPM. Intake pressure was set from 0.5 to 1 bar with an interval of 0.1 bar. Spark Timing was set to -40 to 0 CAD ATDC with an interval of 4 CAD.

Table 1. Engine specifications.

Engine type	Single-cylinder 4-stroke SI gasoline engine
Injection method	Port fuel injection
Stroke × Bore	86.07 mm × 86 mm
Intake valve open	9 CAD BTDC
Intake valve close	84 CAD ABDC
Exhaust valve open	55 CAD BBDC
Exhaust valve close	38 CAD ATDC
Compression ratio	9.5
Connecting rod length	175 mm

Table 2. Engine input parameters set.

Title	Range	Step
Engine speed	1000~3800 RPM	400 RPM
Intake pressure	0.5~1 bar	0.1 bar
Spark timing	-40~0 CAD ATDC	4 CAD

2.2 Machine learning method

2.2.1 ANN method

ANN, designed based on the structure of the human brain information processing system, is a machine learning model consisting of multiple layers of connected neurons. The structure of an ANN model consists of a combination of input, output, and hidden layers. The different layers are connected by neurons or nodes. The ANN model is constructed in a training stage and a validation stage. The input and output layers are derived from the system model, and the hidden layers can process the input data during the training stage. By adjusting the connection weights between neurons, the ANN model can predict the output results with the smallest possible error with the system model during the validation stage. the ANN model structure in this paper is “3-8-1”.

2.2.2 SVR method

SVR is a model derived from SVM, which has a structure similar to ANN. The structure of SVR has an input layer, hidden layer, and output layer. By learning the training dataset of the input layer, the parameters of the hidden layer can be obtained automatically. By using the kernel function in SVR model, feature vectors of sample data can be mapped from low to high dimensions. Then, the hyperplane which brings all the data of a set to the closest distance to the plane can be found. In this research, RBF (Radial basis function) was chosen as the kernel function.

2.2.3 RF model

Random Forest is an idea based on ensemble learning. RF uses a bootstrap redrawing technique to randomly select *n* samples from the original training set with back, thus forming a new training subset, and randomly selecting *m* ($m<M$) features. and then, select the best features from *M* features as the basis for decision tree splitting, repeatedly, until *N* decision trees are obtained, and the set of these trees is the final training model. The random forest combines each decision tree, and the modeling of each tree depends on the samples extracted independently each time. The prediction of the classification error depends on the classification ability of each decision tree and the correlation between each tree. The number of the decision tree is 500.

2.3 Data process

To evaluate the performance of the ML method, the training datasets (80%, i.e., 422) and validation datasets (20%, i.e., 106) were divided from all the 528 1D simulation experimental data. To establish the machine learning models used to predict ISFC, steady-state data sets were collected. At 1 bar intake pressure, ISFC prediction under different spark timings with certain speeds (i.e., 1000, 2600, and 3800 RPM) was analyzed. Additionally, at -20 CAD ATDC spark timing, ISFC under different speeds with certain loads (i.e., 0.6, 0.8, and 1.0 bar) was analyzed. Besides, at 2600 RPM, combustion phasing prediction under different loads with certain spark timings (i.e., -40, -20, and 0 CAD ATDC) was analyzed. Besides, some statistical metrics such as the coefficient of determination (R^2 or R-squared), the root-mean-square error (RMSE), and the mean absolute percentage error (MAPE) evaluated the modeling performance in the following section.

3. RESULT AND DISCUSSION

3.1 Comparisons of ISFC prediction

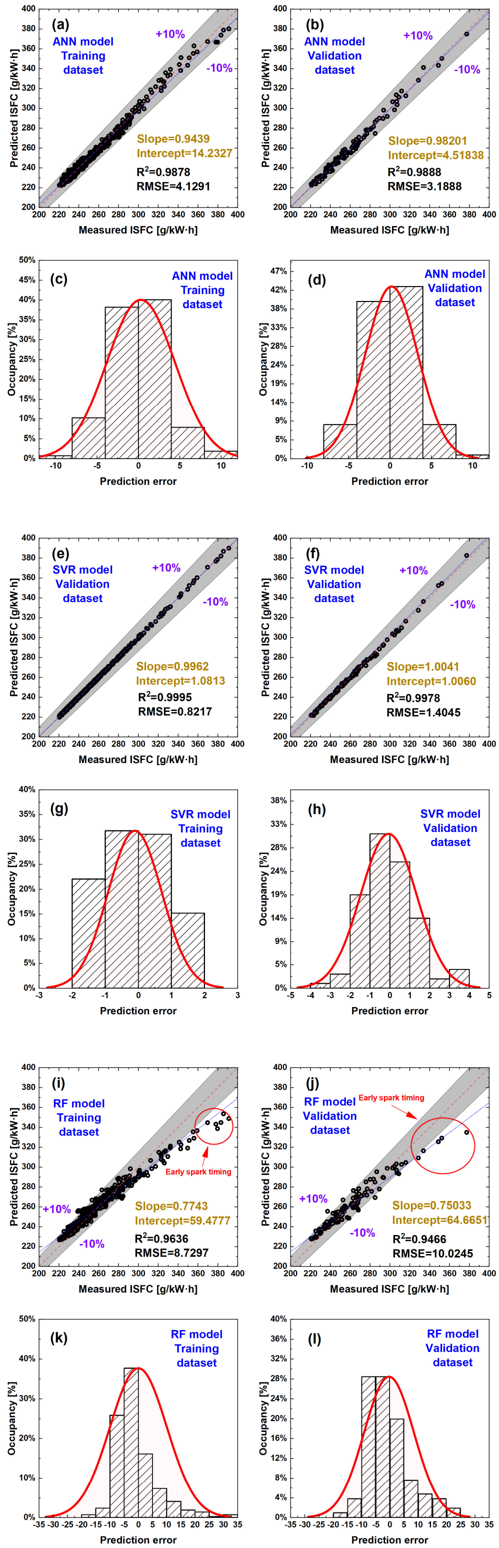

Fig. 2. Comparison of predicted ISFC with the experiment for both the training and the validation datasets

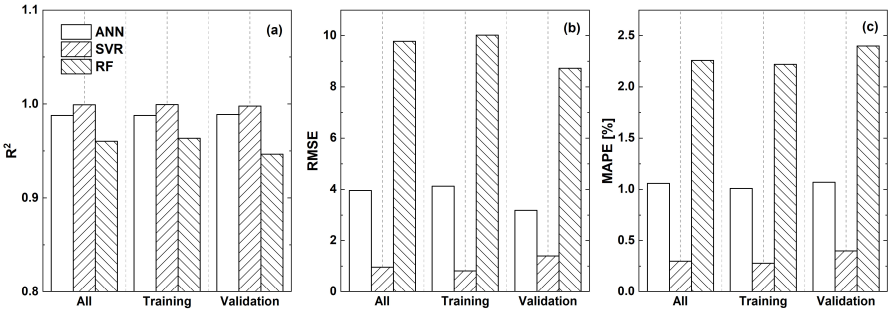

Fig. 3 Comparison of ISFC prediction ability for ANN, SVR and RF model

Fig. 2 and Fig. 3 show the comparison between the measured ISFC data and different machine learning predicted results, and the distribution of error is also given. For the training dataset, the R^2 of ANN, SVR, and RF were 0.9878, 0.9995, and 0.9636. The RMSE of these models were 4.1291, 0.8217, and 8.7297. As for the validation dataset, the R^2 of ANN, SVR, and RF were 0.9888, 0.9978, and 0.9466. The RMSE of these models were 3.1888, 1.4045, and 10.0245. For all datasets, the R^2 of ANN, SVR, and RF were 0.9878, 0.9992, and 0.9604. The RMSE of these models were 3.9582, 0.9673, and 9.7784. SVR predicted ISFC accurately the R^2 for both the training and test dataset were all very close to 1. As for the ANN model, the R^2 of the validation dataset is better than the training dataset, indicating the great generalization ability of ANN. The prediction performance of RF is a little poorer because many points predicted by RF are away from the line slope = 1, especially when the spark timing is too early. As for the MAPE, SVR shows the best performance, which is below 0.5%; the MAPE of ANN is around 1%; the MAPE of RF is above 2%, showing the worst prediction performance. It can be seen that the prediction errors of ANN were mainly concentrated between -5 and 5, and the distributions of the training and validation datasets are similar, while the prediction errors of SVR were concentrated between -2 and 2, and the error distribution of the validation dataset was larger, with some errors reaching 4. It can be seen that the error distribution of RF is the worst, mainly concentrated between -10 and 10, with some errors exceeding 10.

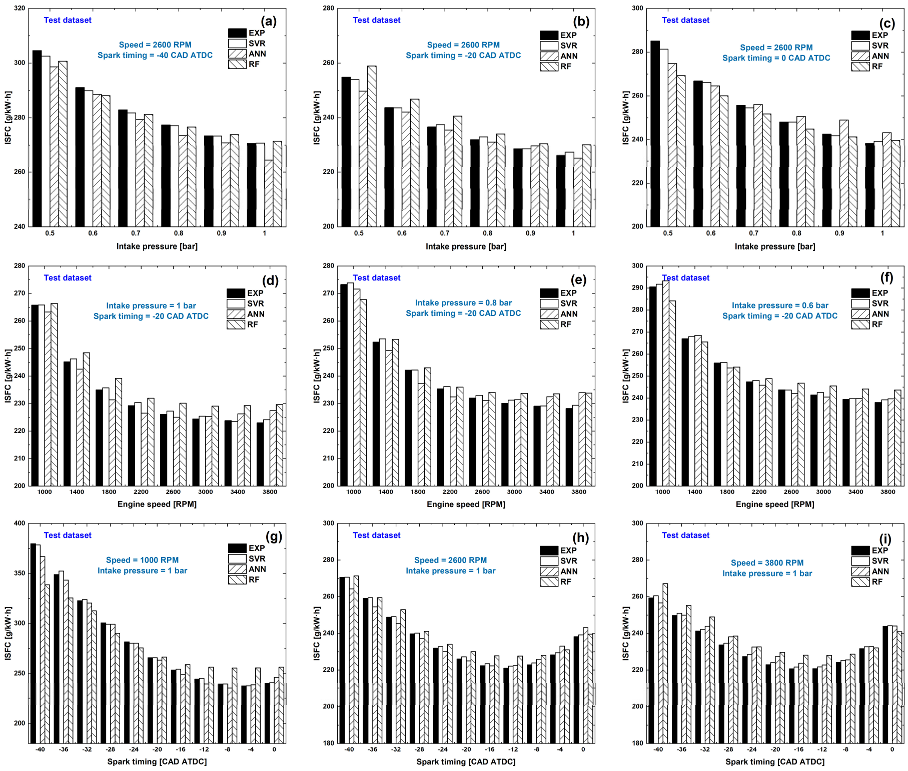

Fig. 4. Effect of engine variables on the ISFC

Therefore, from the perspective of statistical metrics, the SVR model shows the best performance in fuel consumption prediction.

3.2 Comparisons of steady-state prediction

To investigate whether these models can learn the in-cylinder combustion law when operating conditions change, the following discussions will focus on the comparison of the test dataset prediction performance. Fig. 4 shows the effect of engine variables on the ISFC. As Fig. 4(a), (b), (c) show, the load has an impact on the ML prediction performance, the prediction error of all ML methods became larger when the load was 0.5. The fuel consumption rate was high because of the low flammable mixture velocity, caused by low intake pressure. Since such high fuel consumption working conditions were rare, machine learning was not effective. Fig. 4(d), (e), (f) show the effect of engine speed on the ISFC. It can be seen that with the speed increasing, the fuel consumption gradually decreased because the heat loss in the cylinder was less, the flame propagation speed was faster, and the thermal efficiency was improved. The effect of spark timing on the ISFC can be seen in Fig. 4(g), (h), and (i). To different engine operation conditions at full engine load, there existed a Most Brake Torque (MBT) spark timing, indicating the best power performance and efficiency for the engine. All three machine learning methods can capture the characteristics of the MBT. When engine speed was 1000 RPM, the MBT for SVR, ANN and RF were -4, -8, and -8 CAD ATDC, respectively, while the actual MBT was -4 CAD ATDC. When engine speed was 2600 RPM, the MBT for all models were -12 CAD ATDC, which were the same as the actual MBT. The MBT for all models were -16, -12, and -12 CAD ATDC, respectively at 3800 RPM engine speed, while the actual MBT was -16 CAD ATDC. This shows that small prediction errors can affect the machine learning judgment of MBT.

3.3 Comparisons of engine map prediction

To reflect the relationship between ISFC and engine inputs, the engine map of ISFC is constructed by using the different machine learning models with the relative error, as shown in Fig. 5 and Fig. 6. It can be seen in Fig. 5(a), (d), (g), and (j), that at early spark timing and low engine speed, the fuel consumption rate was very high. With the spark timing delayed, the ISFC decreased. When retarded after MBT, ISFC increased slightly. Fig. 5(b), (e),

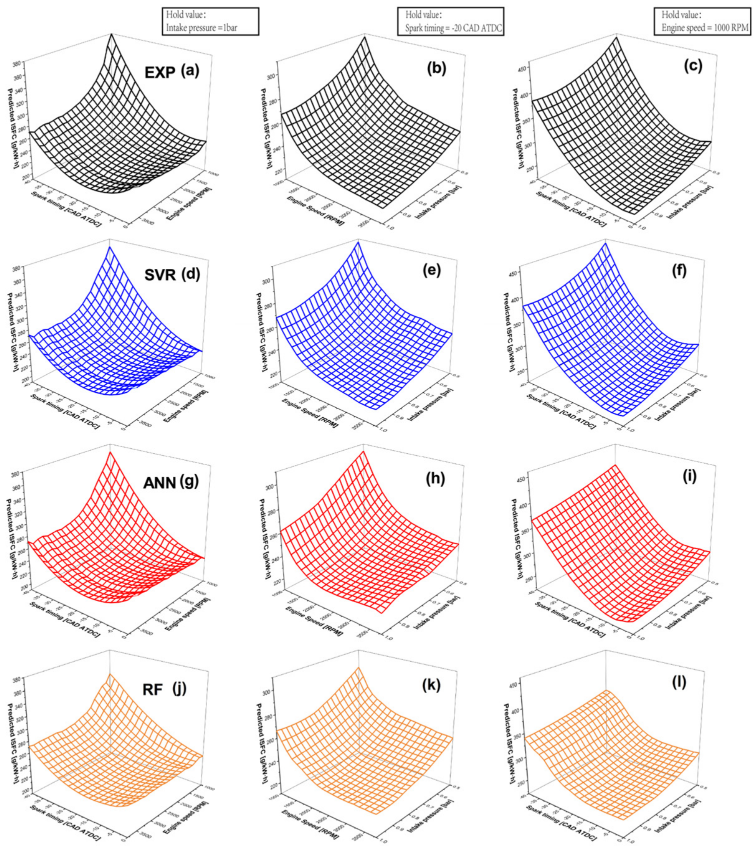

Fig. 5. ISFC versus different engine variables combination.

(h), and (k) show when the engine speed and load increased, the ISFC decreased, the all the ML methods could learn this trend. Fig. 5(c), (f), (i), and (l) show that at -40 CAD ATDC spark timing, the ANN and RF model were not sensitive to the effect of engine load changes on ISFC. The prediction error also can be seen in Fig. 6(f) and (i), at the early spark timing, the relative error for ANN and RF can reach more than 10%. This error shows the limitation of ANN and RF application for virtual sensors design algorithm. It can be seen in Fig. 6, RF has a large relative error while predicting the ISFC at 1000 RPM and -40 CAD ATDC spark timing. Generally, SVR model predicted well for different operation conditions.

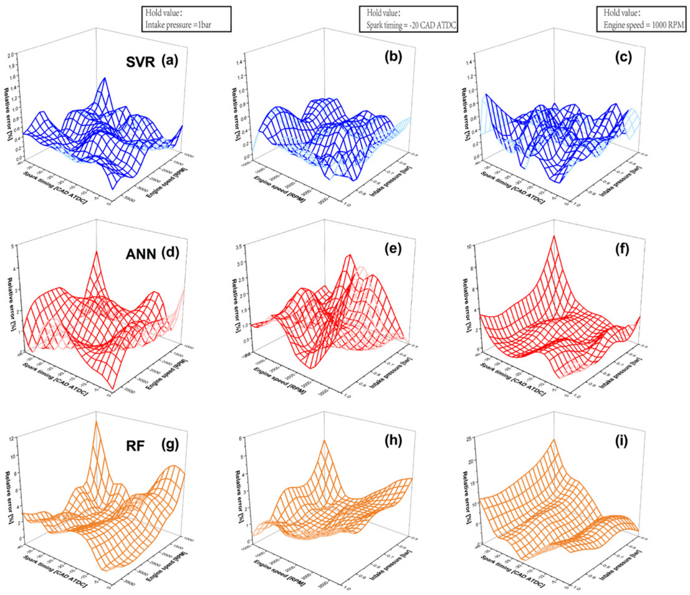

Fig. 6. Prediction relative error of engine map for different machine learning models.

4. CONCLUSIONS

ISFC is an indicator for the engine economic performance, which is increasingly important as carbon neutrality and carbon peaking strategies are implemented. The virtual sensor for the next generation intelligent engine needs high accuracy and a good generalization ability algorithm for predicting ISFC precisely. Here is an example of ISFC to represent the engine performance prediction. In this study, the prediction performance of ANN, SVR, and RF was compared. R^2(all) of ANN, SVR and RF were 0.9878, 0.9992 and 0.9604. The choice of machine learning method affects the judgment of MBT, the SVR can be well aligned with the actual value under various operating conditions. Meanwhile, the consistently low relative error of SVR in map prediction reflects its potential to predict ISFC well in the virtual sensor for different operating conditions. The approach paves the way for using ML as a virtual sensor for fuel consumption prediction and control. Future work will simplify the modeling approach and improve the training speed.

ACKNOWLEDGEMENT

Thanks for the equipment and the right to use the software provided by Power Machinery and Vehicular Engineering Institute, Zhejiang University.

CONTACT INFORMATION

Corresponding author: Zhentao Liu
Professor
College of Energy Engineering
Zhejiang University
liuzt@zju.edu.cn

REFERENCE

[1] Qu J, Feng Y, Zhu Y, Zhou S, Zhang W. Design and thermodynamic analysis of a combined system including steam Rankine cycle, organic Rankine cycle, and power turbine for marine low-speed diesel engine waste heat recovery. Energy Conversion and Management. 2021; 245: 114580.

[2] Mentink P, Seykens X, Valdivieso DE. Development and application of a virtual NOx sensor for robust heavy duty diesel engine emission control. SAE International Journal of Engines. 2017; 10(3): 1297-304.

[3] Bellone M, Faghani E, Karayiannidis Y. Comparison of CNN and LSTM for Modeling Virtual Sensors in an Engine. SAE International Journal of Advances and Current Practices in Mobility. 2020; 2(2020-01-0735): 2632-9.

[4] Yap WK, Karri V. Comparative analysis of artificial neural networks and dynamic models as virtual sensors. Applied Soft Computing. 2013; 13(1): 181-8.

[5] Yap WK, Karri V. ANN virtual sensors for emissions prediction and control. Applied Energy. 2011; 88(12): 4505-16.

[6] Chen Z, Yang C, Chen A. Estimating fuel consumption and emissions based on reconstructed vehicle trajectories. Journal of Advanced Transportation. 2014; 48(6): 627-41.

[7] Liu J, Huang Q, Ulishney C, Dumitrescu CE. Machine learning assisted prediction of exhaust gas temperature of a heavy-duty natural gas spark ignition engine. Applied Energy. 2021; 300: 117413.

[8] Yang R, Sun X, Liu Z, Zhang, Y, Fu J. A numerical study into the importance of equivalence ratio measurement accuracy for spark ignition engines. American Society of Mechanical Engineers International Mechanical Engineering Congress and Exposition. 2021, Virtual, Online, IMECE2021-70992.