



## Development and Prediction of Kuala Terengganu Driving Cycle via Long Short-Term Memory Recurrent Neural Network

Nurru Anida Ibrahim<sup>1</sup>, Arunkumar Subramaniam<sup>1</sup>, Paul Walker<sup>2</sup>, Siti Norbakyah Jabar<sup>1,3,4</sup>, Salisa Abdul Rahman<sup>1,3,4\*</sup>

<sup>1</sup> Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, Kuala Nerus 21030, Terengganu, Malaysia

<sup>2</sup> School of Mechanical and Mechatronic Engineering, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia

<sup>3</sup> Renewable Energy and Power Research Interest Group (REPRIG), Universiti Malaysia Terengganu, Kuala Nerus 21030, Terengganu, Malaysia

<sup>4</sup> Energy Storage Research Group (ESRG), Universiti Malaysia Terengganu, Kuala Nerus 21030, Terengganu, Malaysia

Corresponding Author Email: [salisa@umt.edu.my](mailto:salisa@umt.edu.my)

<https://doi.org/10.18280/ijtdi.070205>

### ABSTRACT

**Received:** 12 April 2023

**Accepted:** 15 May 2023

#### Keywords:

*driving cycles, k-means clustering, LSTM, RNN, energy consumption, fuel economy, hybrid electric vehicles*

Driving cycle is as representation of traffic behaviour in an area or city. It plays a fundamental role in the design of vehicles and to test the performance of the vehicles. This paper studies a driving cycle development method based on k-means clustering and driving cycle prediction based on Long Short-Term Memory (LSTM) by Recurrent Neural Network (RNN). The objectives of this paper are to develop a Kuala Terengganu Driving Cycle (KTDC) by using k-means clustering, to develop a prediction of future KTDC, and lastly to analyse the energy consumption and emissions of KTDC. Firstly, the driving data is collected in five different routes in Kuala Terengganu city at go-to-work times. Then the data is divided into micro-trips and the driving features are extracted. The features are used to develop a driving cycle using k-means clustering approach. The prediction is developed after the training of neural networks by using LSTM network approach. Finally, the energy consumption and emissions of KTDC is analysed by using AUTONOMIE software.

## 1. INTRODUCTION

As the global population grows, so does the need for transportation, and as a result, pollution and other critical environmental problems will only get worse. Exhaust emissions from automobiles, poor driving techniques and traffic hazard all contribute to pollution in the streets. According to the Department of Environmental Malaysia (2017), motor vehicles are responsible for 70.4% of all air pollution. Modern cities have a strong desire for automobiles that are more environmentally friendly and efficient. Engineers, analysts, and technologists are designing a hybrid vehicle to address this issue. In terms of reducing fuel consumption and reducing carbon emissions, the hybrid car is the most viable [1].

A driving cycle is a speed-time profile that depicts how people drive in a specific city or region [2-4]. The driving cycle describes how a vehicle behaves on the road and is used by a wide range of people, from those who build activity control frameworks to those who decide how cars should be executed. For the certification of emission standards, it is also used in the emission testing of vehicles [5]. Among its many users are automobile manufacturers as well as environmentalists and traffic engineers alike.

All this while, there a few standard driving cycles that have been used in the global automotive industry such as the Worldwide Harmonized Light Vehicle Test Cycle (WLTC),

New European Drive Cycle (NEDC), Federal Test Procedure (FTP), and Japanese modes, which originated from the United States, Europe, Japan, China, and India. Nonetheless, some regions, such as Toronto, Canada [6], and Singapore [7], have established their own local driving cycles. In an effort to manage emissions and fuel consumption, governing bodies typically employ drive cycles in their legislation. Vehicle manufacturers and automotive suppliers often utilize drive cycles for product evaluation objectives. Up until today, Malaysia's government, vehicle manufacturers and suppliers still use NEDC for the legislative and evaluation purposes [8].

As stated in the study [9], vehicle design is heavily dependent on driving cycles, and if vehicle makers are solely concerned with optimizing a vehicle for one particular driving cycle, they may end up with a non-robust and sub-optimal design for another driving cycle, which could have a negative impact on a vehicle's overall performance. A local drive cycle will continue to be necessary even after the switch to WLTC because it will help determine realistic fuel consumption estimates for Malaysian actual driving.

Thus, this paper discussed about the development and prediction of Kuala Terengganu Driving Cycle (KTDC) via deep learning approach. In this study, the development and the prediction of the driving cycle are by using k-means clustering method and the Long Short-Term Memory (LSTM) network. Recurrent neural networks (RNNs) of the LSTM network type can learn order dependence in sequence prediction issues. The

objectives of this paper are to characterize and construct the driving cycle of Kuala Terengganu city during the peak hour along the main routes, to develop the prediction of the future KTDC using LSTM network, and to examine the energy consumption and emissions of KTDC using conventional, hybrid electric vehicle (HEV), and plug-in hybrid electric vehicle (PHEV) powertrain.

## 2. METHODOLOGY

Figure 1 depicts a flowchart and study activities on how to create a KTDC along the major roads that Kuala Terengganu residents use to commute to work, namely Routes A, B, C, D, and E. KTDC's inputs are second-by-second speeds. These routes were chosen because of its high volume of traffic in Kuala Terengganu city. Ten sets of data were gathered along the selected routes at three different peak hours which is 7.30, 8.00, and 8.30 a.m. The on-board measuring method equipped with the Global Positioning System (GPS) sensor will be utilized in this study. The data gathered then will be stored and managed in MATLAB. Then the parameters for each run will be extracted for the development of the driving cycle. The driving cycle is developed by using *k*-means clustering approach. Then, the prediction of the future driving cycle will be taking place by using LSTM network of deep learning.

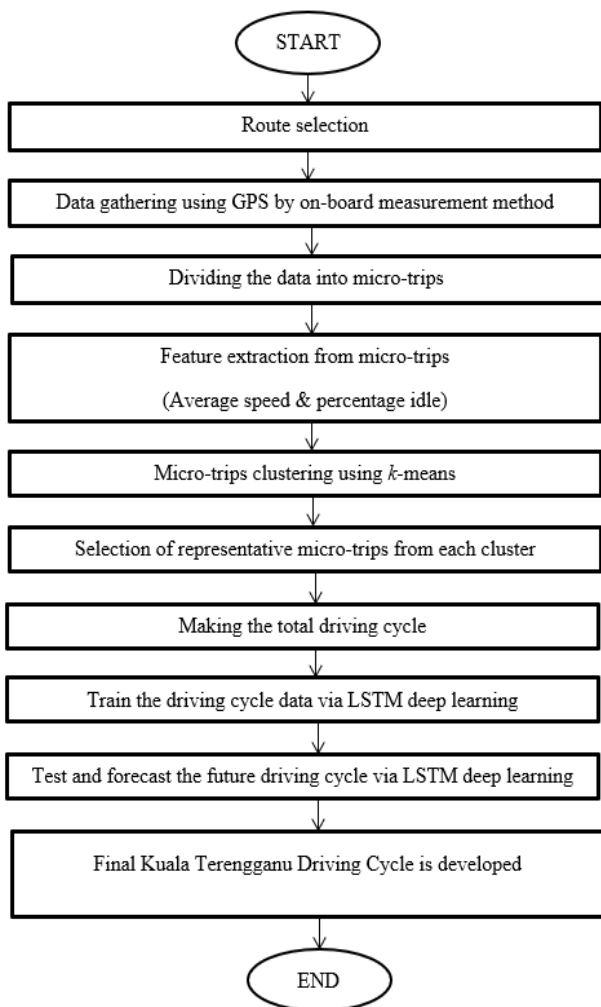


Figure 1. Flow chart of research activities

## 2.1 Data collection

Five main routes from Kampung Wakaf Tembesu to Wisma Persekutuan namely as Route A-E are depicted in Figure 2 (a), (b), (c), (d), and (e) as retrieved in the Google Maps. Kampung Wakaf Tembesu as the beginning point due to its population and Wisma Persekutuan as the destination point since most of the government sectors' offices are located there. According to the Malaysian Ministry of Works, five routes with the worst peak-hour traffic were chosen to illustrate typical urban driving conditions in Kuala Terengganu [10]. In this study, speed-time data are gathered along designated routes utilizing GPS-based on-board measurement.

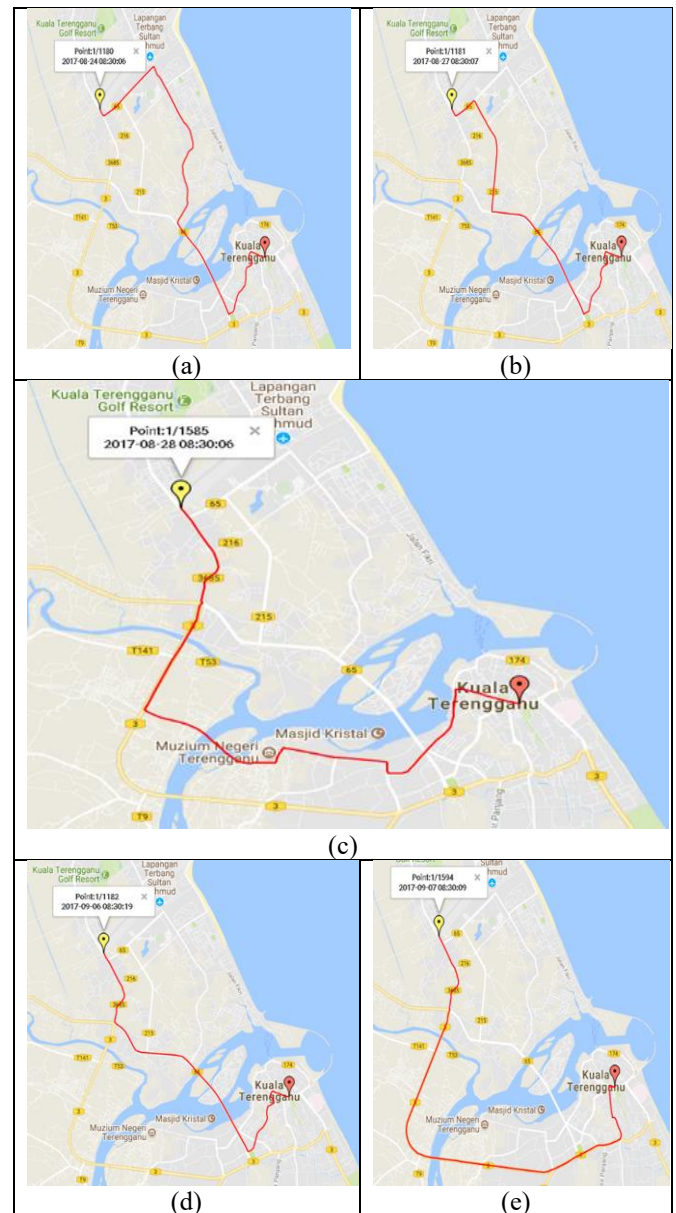


Figure 2. (a) Route A; (b) Route B; (c) Route C; (d) Route D; and (e) Route E

Ten runs were conducted at 7.30, 8.00, and 8.30 am along the chosen road to collect data. For the real-time driving data collection method, there are three main methods of collecting data which chase car technique, on-board measurement technique and hybrid technique. Chase car technique is when instrumented cars record second-by-second speed data as they follow target vehicles. In contrast, on-board measurement is

when speed and time data are collected along a defined route utilising a vehicle's on-board real-time logging system. Also known as a combination of on-board measuring and circulation driving, the hybrid method combines the two strategies [11]. For KTDC, the on-board measuring technique will be employed for data gathering because it is better suited for KT drivers' erratic behaviour to prevent risks like accidents and unexpected loss of control.

## 2.2 Micro-trips clustering

Micro-trips are essential to the development of a driving cycle. Micro-trip is the distance travelled between two successive zero-velocity time points [12]. Micro trips begin with idle time and end with a deceleration that approaches zero. It is necessary to split up the complete data set into a variety of micro-trips. This process allows for the collection of many micro trips for all the data that has been gathered. The micro-trips are then divided into categories based on the flow of traffic, including congested, moderate, and clear. To divide or cluster the micro-trips, an unsupervised learning algorithm known as  $k$ -means clustering approach will be utilized.

This clustering method is easy and uncomplicated, as it will be used to divide a given data set into a specified number of clusters which  $k$  clusters are assumed. Clustering in  $N$ -dimensional Euclidean space,  $R_N$ , is the division of  $n$  points into  $K$  groups based on some features of two metric. Let the set of  $n$  points  $\{x_1, x_2, \dots, x_n\}$  be presented by set  $S$ , and  $K$  clusters be represented by  $C_1, C_2, \dots, C_k$ . Then:

$$\begin{aligned} C_i \cap C_j &= \emptyset \text{ for } i=1, \dots, K \\ C_i \cap C_j &= \emptyset \text{ for } i=1, \dots, K, j=1, \dots, K \text{ and } i \neq j \\ \text{and: } \bigcup_{i=1}^k C_i &= S \end{aligned} \quad (1)$$

By maximising a specified measure, the  $k$ -means method attempts to address the clustering problem. First, pick a value for  $k$  first. The levels of traffic flow-heavy, moderate, and light-that are present in this study decides the value of  $k$ . In Step 2, initialise the  $k$  cluster nodes. The whole  $N$  data are assigned to the closest cluster centre in step three to ascertain their class memberships. Re-estimating the  $k$ -cluster centres in the fourth phase is done under the assumption that the memberships found in the prior stage are correct. If none of the  $N$  data had their membership status altered during the previous iteration, stop the programme. Otherwise, move on to Step 3 [13].

## 2.3 KTDC development

Micro-trip clustering is the suggested method for developing the KTDC. Before micro-trips can be clustered, driving characteristics must be extracted. Many driving characteristics can be analyzed from micro-trips, including average speed, average running speed, average acceleration and average deceleration, and root mean square acceleration. Idling, acceleration, deceleration, and cruising modes are all included in the four driving modes [14]. However, for the development purpose, only average speed, and percentage of idle will be considered. Since they will have the biggest effect on the emission, these two parameters were selected [15]. After going through the  $k$ -means clustering process, the micro-trips are then divided into three distinct groups as seen in the Figure 3. The three groups-clear traffic, moderate traffic, and

heavy traffic-each have unique traits and signify different traffic conditions.

Then, the representative micro trips are identified to generate the driving cycle for each cluster. The selection of representative micro-trip depends on the distance to the cluster centre as the closest will be chose. The representative of micro-trips for each group are depicted in Figure 4 (a), (b), and (c). The micro-trips will then be merged to form the complete Kuala Terengganu driving cycle.

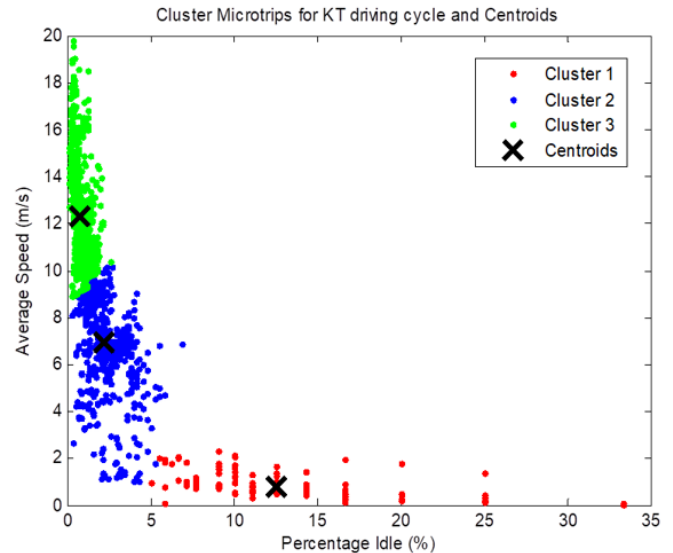


Figure 3. Clustering micro-trips

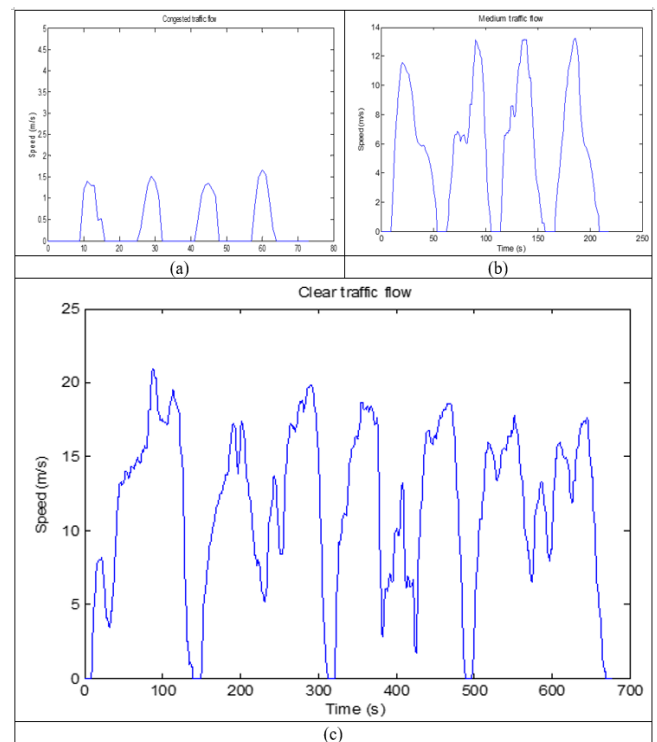


Figure 4. Micro-trips clustering. (a) Cluster one; congested traffic flow, (b) Cluster two; medium traffic flow, (c) Cluster three; clear traffic flow

## 2.4 KTDC prediction

Recurrent neural network (RNN) is a category of neural network designs designed to model sequential input. It is

particularly beneficial for several natural languages processing applications, including sentiment analysis, text categorization, speech-to-text, machine translation, etc. A specific type of recurrent neural network design known as the Long Short-Term Memory network (LSTM) has shown to have enormous potential in several natural language processing applications. By using memory cells to retain data, it is primarily intended to comprehend long-term dependencies. The entry and exit of information from the cell are also controlled by a set of gating mechanisms. Due to its ability to overcome the problem of the gradient disappearing, this architecture is superior to the traditional recurrent neural network [16].

The number of LSTM cells is denoted by  $N$ . A single LSTM cell consists of the input gate  $i_t$ , the forget gate  $f_t$ , and the output gate  $o_t$ , as well as the hidden state  $h_t$  and the memory cell  $c_t$ . The input gate regulates the amount of the current input that is contributed to the cell's state. The forget gate regulates how much of the prior long-term state gets read into the current state of the cell. And lastly, output gate determines which portions of the long-term state will be included in the subsequent hidden state.

Using the sigmoid as an activation function, each gate produces a value between zero and one.

$$\begin{aligned} i_t &= \delta(W_i v_t + U_i h_{t-1} + b_i) \\ f_t &= \delta(W_f v_t + U_f h_{t-1} + b_f) \\ o_t &= \delta(W_o v_t + U_o h_{t-1} + b_o) \end{aligned} \quad (2)$$

where the sigmoid function is denoted by  $\delta(\cdot)$ . Specifically, the weight matrices for the gates are  $W_i, W_f, W_o, U_i, U_f, U_o$ . While weights for the bias are denoted by  $b_i, b_f, b_o$ .

Generally, at the time step  $t$ , the LSTM architecture takes as input. The driving vector  $v_t \in R^D$ , the prior hidden state  $h_{t-1} \in R^M$  and the last memory cell vector  $c_{t-1}$  are typically inputs to the LSTM architecture at the time step  $t$ . On the other hand,  $h_t$  and  $c_t$  are calculated as follows:

$$\begin{aligned} g_t &= \tan h(W_g v_t + U_g h_{t-1} + b_g) \\ c_t &= f_t \odot c_{t-1} + i_t \odot g_t \\ h_t &= o_t \odot \tan h(c_t) \end{aligned} \quad (3)$$

where, the notation  $\odot$  denotes the multiplication component. While  $\tan h(\cdot)$  applies to the nonlinear activation function,  $W_g$  and  $U_g$  are the weight vectors, and  $b_g$  is the bias component.

The prediction process begins with the driving data load as the input. It is necessary to standardize the training data such that it has a zero mean and unit variance for a better fit and to avoid the training from splitting. The parameters used in training must be used in the prediction of the test data. To predict the values of subsequent time steps in a series, the outputs must be produced after the training sequences with values moved by one time step. To put it another way, at each time step in the input sequence, the LSTM network gains the ability to anticipate the value of the next time step Predictors are extracted from training sequences with the final time step removed. A LSTM regression network is then constructed. In this project, the LSTM layer is 200 hidden units and train the LSTM network with the specified training options. After the training process, the forecast future time steps will take place. After the forecasted data is developed, the root mean square error (RMSE) is calculated to compare between the forecast and test data. RMSE stands for residual value standard deviation or prediction errors. Residuals quantify the deviation

of data points from the regression line, while RMSE quantifies the spread of residuals [17]. Figure 5 shows the flow chart of the prediction of the KTDC.

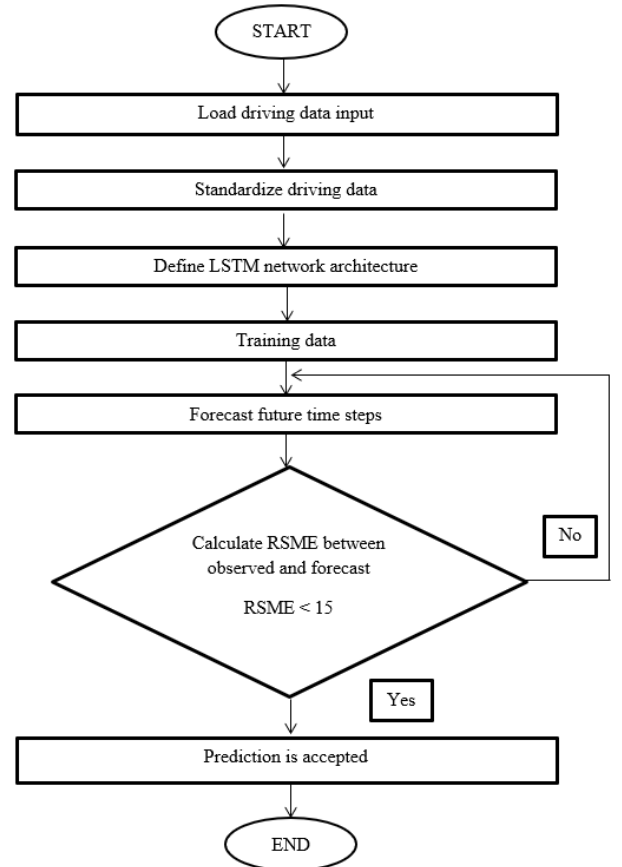


Figure 5. Flow chart of KTDC prediction process

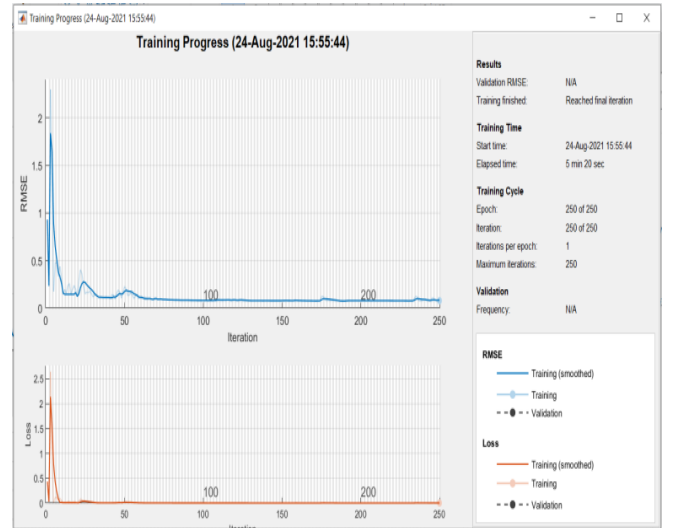


Figure 6. Training progress for KTDC prediction

Figure 6 depicts the training procedure for the created KTDC. In order to prevent gradients from inflating, the gradient threshold for this project is set at 1, and the training option is 250 epochs. Every 125 epochs, the learn rate is reduced by a factor of 0.20, starting with an initial value of 0.005.

### 3. EXPERIMENTAL RESULTS

#### 3.1 Kuala Terengganu Driving Cycle (KTDC)

After collecting all driving cycle data along five distinct routes at 7:30, 8:00, and 8:30 a.m., the KT driving cycle can be developed to its completion. Figure 7 depicts the final driving cycle of the KT. The entire distance is 8.8 kilometers and there are a total of 12 micro-trips. Tabulated in Table 1 are the assessment parameters of the KT driving cycle. The defined driving cycle led to the following findings: The predominant speed range was greater than 10 km/h. This is due to the fact that KT key routes are frequently backed up with a heavy traffic. Micro-trips with a wider speed range are therefore longer than those with a narrower speed range. This is because a vehicle moves more faster and stops less frequently in a free-flowing environment. The created KT driving cycle reported an average speed of 35.15 km/h, showing that vehicles are moving more slowly and that more micro-trips take place at lower speeds than the average. As a result of the frequent stops along the road, fuel consumption and emissions are increased during this time.

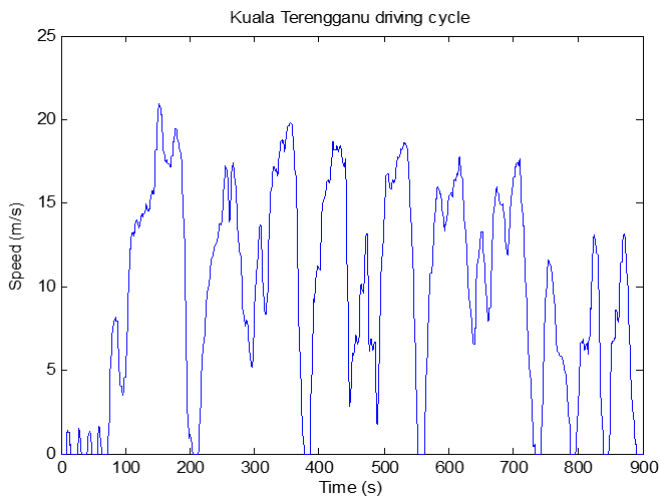


Figure 7. Kuala Terengganu driving cycle

Table 1. The parameters of developed KTDC

Parameters	KTDC
Distance travelled (km)	8.8
Total time (s)	953
Average speed (km/h)	35.15
Average running speed (km/h)	40.55
Average acceleration ( $m/s^2$ )	0.59
Average deceleration ( $m/s^2$ )	0.56
Root mean square ( $m/s^2$ )	0.73
Percentage idle (%)	12.10
Percentage cruise (%)	5.11
Percentage acceleration (%)	40.51
Percentage deceleration (%)	42.29

#### 3.2 Prediction of Kuala Terengganu Driving Cycle (KTDC)

Figure 8 shows the plot of forecast and observed data of KTDC after the training process. Figure 9 in the other hand shows the RSME between speed and time of the forecast and test data. From the figure, it shows that the RSME is 4.4344. Lower RMSE values suggest a better match. RMSE is a good indicator of how well the model predicts the response, and it

is the most relevant criterion for fit if prediction is the primary aim of the model.

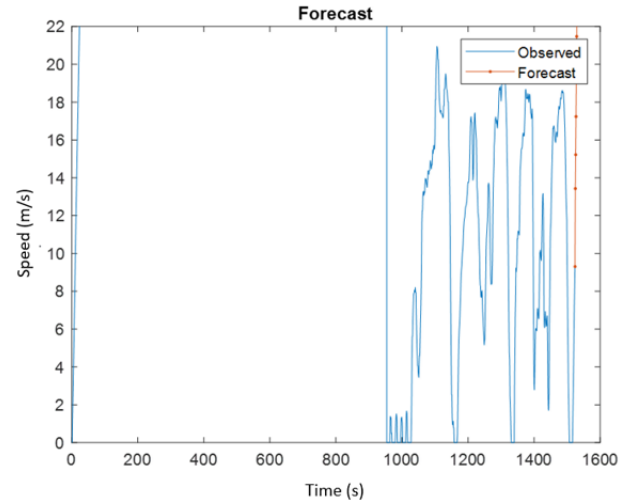


Figure 8. Forecast of Kuala Terengganu driving cycle

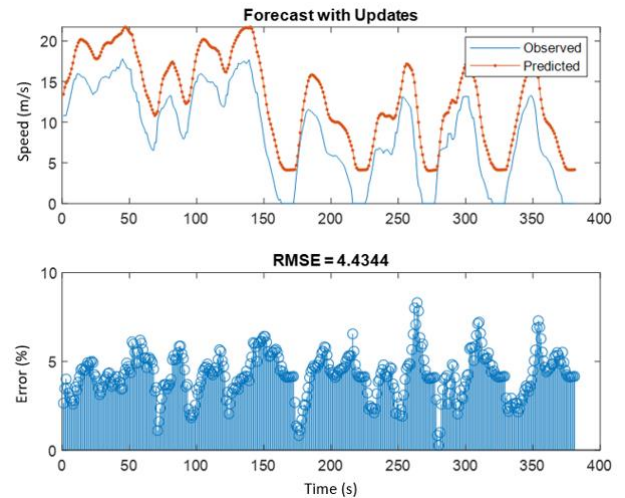


Figure 9. Comparison between forecast and test data

#### 3.3 Energy consumption and emission analysis of KTDC

For the design, modelling, and analysis of automotive control systems, there is a tool called AUTONOMIE. It is a MATLAB-based forward simulation program with Simulink models and MATLAB data and configuration files. After the establishment of the driving cycle, the fuel analysis, including fuel consumption, fuel efficiency, and emissions, may be calculated using the AUTONOMIE software version v1210. Table 2 illustrates the fuel consumption and emissions (carbon monoxide,  $CO_2$ ) analysis of the KTDC using three different vehicles powertrain which conventional engine vehicles, hybrid electric vehicles (HEV), and plug-in hybrid electric vehicles (PHEV).

Table 2. Comparison of PHEV, HEV, and conventional vehicle engine

	PHEV	HEV	Conventional
Fuel economy (mile/gallon)	212.84	53.78	32.2
Fuel consumption (l/100km)	1.11	4.37	7.3
$CO_2$ emissions (g/mile)	1234	1234	1234

The table clearly demonstrates that the PHEV is superior to the conventional engine and HEV in terms of fuel consumption, emissions, and fuel economy. This is because of the power split type PHEV's design, which consists of a single engine, two motor-generators (MGs), and numerous planetary gears [18]. As a result, it will cut down on pollutants and energy use.

#### 4. CONCLUSION

The KT driving cycle is achieved effectively, and it can be inferred that the suggested method can be used to generate a KT driving cycle for a PHEV powertrain to address difficulties with exhaust emissions and fuel consumption problems. The KT driving cycle is successfully developed by grouping micro-trips using the *k*-means algorithm. The prediction also has been successfully done by using LSTM by RNN deep learning with the RMSE is 4.4344. The data are gathered along Routes A, B, C, D, and E during Go-to-Work times of 7.30 a.m., 8.00 a.m., and 8.30 a.m. However, the results are specifically designed for Kuala Terengganu city only and cannot be represent as Malaysia Driving Cycle. For the future work, it is recommended that all states and cities in Malaysia should be considered in order to develop an accurate Malaysian driving cycle. Malaysia's government, vehicle manufacturer and suppliers such as Proton, Perodua, HICOM Automobile Manufactures, and others can get huge benefits from the development of Malaysia Driving Cycle especially for accurate legislative and evaluation purposes.

#### DATA AVAILABILITY STATEMENT

The data used for this study are available on request through Faculty of Ocean Engineering Technology and Informatics institutional account for data access link provided by authors.

#### ACKNOWLEDGEMENTS

The authors would like to be obliged to Universiti Malaysia Terengganu for providing financial assistance under Matching Grant 2022 (MG1+3/ 2022/ 53495/ UMT) and Faculty of Ocean Engineering Technology and Informatics, UMT for all their technical and research support for this work to be successfully completed.

#### REFERENCES

[1] Murgovski, N., Johannesson, L., Sjöberg, J., Egardt, B. (2012). Component sizing of a plug-in hybrid electric powertrain via convex optimization. *Mechatronics*, 22(1): 106-120. <https://doi.org/10.1016/j.mechatronics.2011.12.001>

[2] Zhao, X., Zhao, X.M., Yu, Q., Ye, Y.M., Yu, M. (2020). Development of a representative urban driving cycle construction methodology for electric vehicles: A case study in Xi'an. *Transportation Research Part D: Transport and Environment*, 81: 102279. <https://doi.org/10.1016/j.trd.2020.102279>

[3] Borlaug, B., Holden, J., Wood, E., Lee, B., Fink, J., Agnew, S., Lustbader, J. (2020). Estimating region-

specific fuel economy in the United States from real-world driving cycles. *Transportation Research Part D: Transport and Environment*, 86: 102448. <https://doi.org/10.1016/j.trd.2020.102448>

[4] Kaymaz, H., Korkmaz, H., Erdal, H. (2019). Development of a driving cycle for Istanbul bus rapid transit based on real-world data using stratified sampling method. *Transportation Research Part D: Transport and Environment*, 75: 123-135. <https://doi.org/10.1016/j.trd.2019.08.023>

[5] Arun, N.H., Mahesh, S., Ramadurai, G., Nagendra, S.M.S. (2017). Development of driving cycles for passenger cars and motorcycles in Chennai, India. *Sustainable Cities and Society*, 32: 508-512. <https://doi.org/10.1016/j.scs.2017.05.001>

[6] Amirjamshidi, G., Roonda, M.J. (2015). Development of simulated driving cycles: Case study of the Toronto waterfront area. *Transportation Research Part D*, 34: 255-266.

[7] Ho, S.H., Wong, Y.D., Chang, V.W.C. (2014). Developing Singapore driving cycle for passenger cars to estimate fuel consumption and vehicular emissions. *Atmospheric Environment*, 97: 353-362. <https://doi.org/10.1016/j.atmosenv.2014.08.042>

[8] Abas, M.A., Rajoo, S., Abidin, S.F.Z. (2018). Development of Malaysian urban drive cycle using vehicle and engine parameters. *Transportation Research Part D: Transport and Environment*, 63: 388-403. <https://doi.org/10.1016/j.trd.2018.05.015>

[9] Nyberg, P., Frisk, E., Nielsen, L. (2016). Driving cycle equilance and transformation. *IEEE Transactions on Vehicular Technology*, IEEE, 66(3): 1963-1974. <https://doi.org/10.1109/TVT.2016.2582079>

[10] Malaysia (2019). Road Traffic Volume Malaysia (RTVM): Ministry of Works Malaysia, Highway Planning Division.

[11] Galgamuwa, U., Perera, L., Bandara, S. (2015). Developing a general methodology for driving cycle construction: Comparison of various established driving cycles in the world to propose a general approach. *Journal of Transportation Technologies*, 5(04): 191-203. <https://doi.org/10.4236/jtts.2015.54018>

[12] Wang, Q., Huo, H., He, K., Yao, Z.L., Zhang, Q. (2008). Characterization of vehicle driving patterns and development of driving cycles in Chinese cities. *Transportation Research Part D: Transport and Environment*, 13(5): 289-297. <https://doi.org/10.1016/j.trd.2008.03.003>

[13] Maulik, U., Bandyopadhyay, S. (2000). Genetic algorithm-based clustering technique. *Pattern Recognition*, 33(9): 1455-1465. [https://doi.org/10.1016/S0031-3203\(99\)00137-5](https://doi.org/10.1016/S0031-3203(99)00137-5)

[14] Anida, I.N., Salisa, A.R. (2019). Driving cycle development for Kuala Terengganu city using k-means method. *International Journal of Electrical and Computer Engineering*, 9(3): 1780-1787. <https://doi.org/10.11591/ijece.v9i3.pp1780-1787>

[15] Fotouhi, A., Montazeri-Gh, M. (2013). Tehran driving cycle development using the k-means clustering method. *Scientia Iranica*, 20(2): 286-293. <https://doi.org/10.1016/j.scient.2013.04.001>

[16] Hammou, B.A., Lahcen, A.A., Mouline, S. (2020). Towards a real-time processing framework based on improved distributed recurrent neural network variants

- with fastText for social big data analytics. *Information Processing & Management*, 57(1): 102122. <https://doi.org/10.1016/j.ipm.2019.102122>
- [17] Stephanie Glen, (2022). RMSE: Root Mean Square Error. *Statistics How To*. <https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/>.
- [18] Son, H., Park, K., Hwang, S., Kim, H. (2017). Design methodology of a power split type plug-in hybrid electric vehicle considering drivetrain losses. *Energies*, 10(4): 437. <https://doi.org/10.3390/en10040437>