Electric Vehicles Driving Range and Energy Consumption Investigation: A Comparative Study of Machine Learning Techniques

Abdollah Amirkhani[1], Arman Haghanifar[2], Mohammad R. Mosavi[2]
[1]Dept. of Automotive Engineering, Iran University of Science and Technology, Tehran 16846-13114, Iran
[2]Dept. of Electrical Engineering, Iran University of Science and Technology, Tehran 16846-13114, Iran
amirkhani@iust.ac.ir, arman.haghanifar@ieee.org, m_mosavi@iust.ac.ir

***Abstract*— Electric vehicles are the next generation of cars which are pollutant-free, resulting in the elimination of many environmental and healthcare problems caused by fossil-fueled vehicles. On the other hand, mass production and wide adoption of these vehicles are facing significant barriers; long battery charging time and limited trip distance per charge are the most important ones to mention. Due to the development of fast DC chargers, the former problem is resolved to a certain extent, while the latter is still a topic of interest. In this article, using a publicly available dataset, driving range estimation of a specific electric vehicle model is scrutinized. At first, multiple regression models are trained based on the features like the average speed, type of the route and driving style; then the driving range prediction accuracy is investigated. On the next step, sensitivity analysis is performed on the energy consumption rate, and the results are discussed. At the final phase, the effect of each feature on the energy consumption rate is highlighted, and the deviation between experimental rates and the factory-defined rates are explained in details.**

***Keywords*— Electric vehicles, driving range estimation, energy consumption, machine learning, deep learning**

I. Introduction

The use of electric vehicles (EVs) is the most comprehensive solution to reduce the environmental contamination of the air. Hence, governments are encouraging people to purchase and use these vehicles instead of the internal combustion engine powered cars [1]. According to reports, the global sales of EVs in 2018 has grown 72% compared to 2017, with market share rising to 2.1% [2]. Despite the aforementioned benefits and the presence of giant companies in the electric vehicle industry, the little market share may seem strange; but this lack of general adoption comes for many reasons, including high purchase cost, long charging time compared to the fossil-fueled cars and the limited distance per charge are the most noticeable reasons to refer to [3].

Among the other reasons, especially in the developing countries, battery technology [4] and the lack of a universal network of charging stations [5] are to be mentioned. The purpose of the following research is to concentrate on the problem of driving range limitation per charge. As mentioned in [6], the main challenge of EVs is to determine and increase the trip distance precisely. To understand the importance of this issue, the concept of energy consumption rate (ECR) should be followed at first; vehicle manufacturers set a specific factor as the ECR at production which is explained with the kilowatt-hours per hundred kilometers. For example, if the ECR parameter for a vehicle is defined as 12 kilowatt-hours per hundred kilometers, and the total battery capacity is 30 kilowatt-hours, then it can cover the distance of 250 kilometers [7].

The numbers in this range seem to be sufficient; however, the car manufacturer calculates the ECR parameter under certain circumstances, meaning that it may differ from the actual value. The real world energy consumption can increase up to 60% more than the factory-defined value, leading to the range anxiety issue [7] which is of utmost importance to the drivers, according to a survey [8]. In reality, the use of auxiliary equipment, air temperature, traffic volume, type of the road, driving patterns and the driving speed noticeably affect the ECR [9]. Therefore, achieving high precisions in trip distance estimation is considered as a complex problem with many influencing concepts changing it. As mentioned in [10], the number of control parameters are considered to be 43, divided into five different groups.

This paper is structured as follows: Section II delivers a brief overview about the research studies on electric vehicle energy consumption and the affecting parameters. In Section III, the data preparation step is explained. Section IV is about the algorithm selection and the process of model preparation/ examination. The results and their comparison are mentioned in Section V. Finally, Section VI is dedicated to the conclusion.

II. Background

Recently, data-driven approaches have been expanded as an efficient means of measuring driving range and predicting the

978-1-7281-5350-6/19/$31.00 ©2019 IEEE

TABLE I. A SAMPLE DATA RECORD OF ELECTRIC VEHICLE FUELING

Model	Manufacturer	Power (kW)	Fuel date	Odometer	Distance (km)	Quantity (kWh)	Tire type
Model S	Tesla	225	09.6.2017	88514	67.5	11.30	winter tires
Motor-way roads	**City streets**	**Country roads**	**Driving style**	**Consumption (kWh/100km)**	**A/C**	**Park heating**	**Avg speed (km/h)**
no	yes	no	fast	16.74	on	off	48

consumption. The reason is their cost-effectiveness and reliability in comparison with the traditional methods, which is due to the decrease in the implementation costs brought by the internet of things developments. By reducing the cost of installing sensors and transmitting data from the vehicles, a high volume of data is extracted from the in-vehicle network and goes into the cloud so that machine learning algorithms can be applied on them to provide various useful services [11].

Factors affecting ECR in EVs have been fewer studied than the hybrid and gasoline-fueled cars, because of their lower market-penetration rates [7]. Many research studies conducted so far have investigated the role of vehicle components [12], while there is a drastic difference between the results of vehicle manufacturers and the real-world values [13]. Some other studies investigated the distinction between driving range outlined by the manufacturer and the driving range resulted by the real road data [14]. Thus, the environmental variables and the trip conditions are expected to be considered in the research studies. In this regard, various variables have been investigated in terms of their influence on estimating the energy consumption rate; in [15] the speed parameter has been focused, different driving patterns have been analyzed in [16-17], authors of [18] used GPS data and the factory information to propose their model. In [19], route characteristics, such as altitude deviation and traffic volume, and weather conditions, such as temperature, daytime, and humidity have been investigated.

A common issue among most research works is the limited size of dataset. As an example, in [20], only 60 trip records on two different routes passed by Mitsubishi i-MiEV were used as data. Authors of [21] had examined 100 simulated electric vehicle trips as their input data. In the research conducted in [9], 169 driving records by Nissan D21 Pickup were used as the model input. In [22] the model was trained by nearly 800 kilometers trip record passed by the Nissan Leaf. Another research using Nissan Leaf and Ford Connect EV models had utilized a 14400-kilometers dataset [19]. There was only one research work [7] who had a dataset of 250000 kilometers recorded by drivers with Citroen C-Zero, Peugeot Ion and Mitsubishi i-MiEV models. Most research studies have obtained experimental data which despite having many advantages, still is faced by some major disadvantages; being time-consuming and the use of only one vehicle are among the most critical issues.

In the following research, we firstly extract data from an open-source reference. Based on this dataset, four different learning models, which were popular among the aforementioned research studies, are trained, and their accuracy scores on driving range estimation before the trip are compared. In the next step, a classifier model is proposed in order to detect the energy consumption rate deviation from the factory-specified rates. Finally, the effect of each input parameter, such as the route type, on the energy consumption change is investigated, and the causes are explained.

III. DATASET

To conduct this research, we utilized publicly accessible dataset. Among the available datasets, we selected the SpritMonitor [23] as the best option, having one of the most extensive vehicle fueling data records. SpritMonitor is a crowd-platform where users can submit their vehicle model and features and insert a new record for each of their fuelings. Every new fueling record includes information such as the odometer, the distance traveled since previous charging, the amount of consumed fuel since the last charging, the tire type, the fuel type, and so on. A sample row of data is illustrated in Table I. Among the available models, the highest number of records are related to some of the globally popular cars. Table II demonstrates some of the EVs with the largest number of fueling data reports, recorded by different users (in other words, different vehicles).

TABLE II. LIST OF EVS WITH THE MOST RECORDED FUELING DATA [23]

Model	Number of records	Number of users
Mitsubishi i-MiEV	6555	21
Volkswagen e-Golf	7950	71
Opel Ampera	7634	24
Citroen Saxo	2859	9
Kia Soul	3713	24
Nissan Leaf	6871	102
Renault ZOE	13619	225
Tesla Model S	9079	114
BMW i3	5396	84

1	manufacturer	version	power(kW)	fuel_date	distance(km)	quantity(kWh)	tire_type	in-city	motor_way	roads	driving_style	ecr(kWh/100km)	a/c	park_heating	avg_speed(km/h)	ecr_deviation
2	Volkswagen	e-Golf	85	08.02.2019	80		Winter tires	0	0	1	Normal	17.8	0	1	53	1
3	Volkswagen	e-Golf	85	06.02.2019	50	12.29	Winter tires	0	0	1	Normal	15.5	0	1	47	-1.3
4	Volkswagen	e-Golf	85	05.02.2019	43	8.68	Winter tires	0	1	1	Normal	18	0	1	58	1.2
5	Volkswagen	e-Golf	85	04.02.2019	44	1.5	Winter tires	0	1	1	Normal	16.1	0	1	43	-0.7
6	Volkswagen	e-Golf	85	04.02.2019	76	14.44	Winter tires	0	1	0	Normal	19	0	1	76	2.2
7	Volkswagen	e-Golf	85	03.02.2019	15	6.84	Winter tires	1	0	0	Normal	16.1	0	1	23	-0.7
8	Volkswagen	e-Golf	85	01.02.2019	63	7.1	Winter tires	0	1	0	Fast	18.3	0	1	80	1.5
9	Volkswagen	e-Golf	85	31.01.2019	85	15.3	Winter tires	0	1	1	Fast	18	0	1	47	1.2
10	Volkswagen	e-Golf	85	31.01.2019	71	12.77	Winter tires	0	0	1	Fast	19.4	0	1	45	2.6

FIGURE 1. EXTRACTED DATASET FILE HEADER

It is important to note that the average speed feature is not an arbitrary parameter to submit for each record, hence not provided for all the input data records. Given the primary differences between different vehicle models, like the previous research studies, any machine learning model should be trained separately for each electric vehicle model. Therefore, among the vehicles with the highest volume of data, and even though the Renault ZOE has the most recorded data, the Volkswagen e-Golf was selected for the following study. The reason was the ratio of the number of records to the number of users (the vehicles, actually) because as this ratio is larger, the dataset will be more one-handed. Also, there were a small number of outliers in the data recorded by users driving Volkswagen e-Golf in comparison with other vehicle records. The cleaned dataset after the outlier removal includes 3328 trips with a total distance of 140163 kilometers.

IV. METHODOLOGY

We extracted the data by using crawlers and stored it in a CSV file, the heading lines of which is given in Fig. 1. In this dataset, we inserted a new column called "ecr_deviation," which is based on the difference between the energy consumption rate recorded at each trip and the value announced by the manufacturer, which is stated as 16.8 kilowatts per hundred kilometers. Among the features, the "a/c" is the inside vehicle air-conditioning and the "park_heating" is the car's heating system; both are considered as auxiliary loads. Hence we expect them to increase vehicle energy consumption when being used during the trip.

As illustrated in Fig. 1, there are some missing values which make us eliminate the entire record when any of the variables are blank, and imputing methods such as mean value substitution is not applicable. To develop the driving range estimator, several popular regression models are compared; linear regression, simple multilayer perceptron (MLP), deep multilayer perceptron (Deep MLP), random forest (RF), and adaptive boosting (AdaBoost); where the latter two models are ensemble methods. After imputing, the useless columns, such as model and the version of the vehicle, are eliminated and the "distance (km)" column is labeled as the target.

The K-Fold cross-validation is applied to validate the models; half of the dataset was used as the training-set, and the other half was used as the test-set. Then, this train-test-method was repeated ten times ($k = 10$) to give us the final accuracy scores as the average of the results of the ten iterations.

V. RESULTS

The best architecture for the deep MLP neural network is according to Fig. 2, which was obtained experimentally by trial-and-error. In this model, the Rectifier Linear Unit (ReLU) as the activation function for the hidden layers, and the Adaptive Moment Estimation (Adam) as the optimizer had resulted in achieving the best possible scores. The shallow MLP is composed by only adding a 10-neuron hidden layer to the input and output layers. Because the neural network is used as the regressor here, the activation function of the output layer must be set as "linear" for both mentioned models. The RF and AdaBoost algorithms are based on the voting system between a set of base estimator models; which is defined as the decision tree for both of them. For RF the base estimator population of 200, and for AdaBoost, the population of 50 resulted in the best scores. The mentioned hyper-parameter values were obtained by the Grid Search Cross-Validation (GridSearchCV) method.

To measure the performance of the trained models, two criteria were defined, the first is the Mean Absolute Error explained as follows:

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{1}$$

Where y_i is the distance for the i-th record, $\hat{y_i}$ is the estimated distance for the i-th record, and n is the total number of dataset records. As MAE becomes lower, the estimated value is closer to the real value. The second criterion is the R2 score which is as follows:

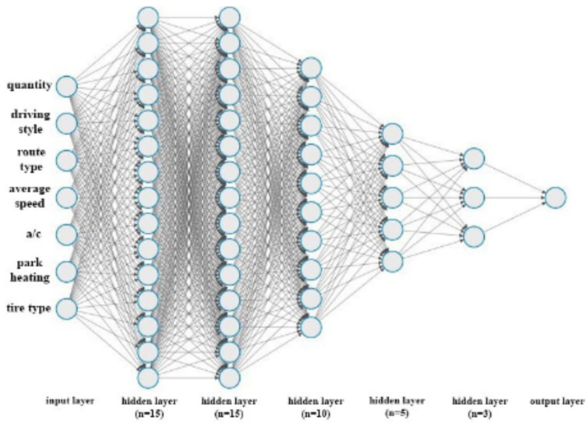

FIGURE 2. OPTIMAL ARCHITECTURE FOR THE DEEP MLP NETWORK, CONSISTED BY FIVE HIDDEN LAYERS

TABLE III. MODEL RESULTS FOR ESTIMATING THE DRIVING RANGE

Model	Average MAE value	Best MAE value	Average r2 score	Best r2 score	Test-set size	Parameters
Linear Regression	12.224	11.689	0.87	0.88	0.5	-
MLP	9.861	7.22	0.89	**0.93**	0.5	activation = "ReLU", solver = "Adam"
Deep MLP	**5.58**	**5.115**	**0.91**	0.92	0.5	activation = "ReLU", solver = "Adam", batch_size = 16
Random Forest	7.776	7.268	0.9	0.92	0.5	n_estimators = 200
AdaBoost	16.93	12.68	0.84	0.87	0.5	n_estimators = 50

$$R^2 = 1 - \frac{SS_{res}=\sum_i (y_i - \hat{y}_i)^2}{SS_{tot}=\sum_i (y_i - \bar{y})^2} \qquad (2)$$

Where $\bar{y}$ stands for the average of all the distance values of the dataset. As R2 score approaches one, the performance is better.

As illustrated in Table III, the best average MAE is obtained via deep MLP which is 5.58 kilometers. To be more specific, we need to look at the average and standard deviation of trip distance parameter in the dataset, where these two parameters are 42.1 and 53.3 kilometers, respectively. To have an excellent model, the MAE should be lower than the 10% of the standard deviation (5.3 kilometers), and the models with errors smaller than its 20% (10.6 kilometers) are considered as good ones.

The linear regressor fitted line based on only the consumed energy amount, or the "quantity," and also the fitted line based on only the average speed value, or the "avg_speed," are represented in Fig. 3 and Fig. 4, respectively. As seen, the driving range has a high correlation with the quantity. However, this correlation is not noticeable with the average speed.

On the next step, the "ecr_deviation" column was selected as the target variable, and a couple of classifiers models are trained to classify it into two categories; positive if the ECR is higher than the manufacturer-defined ECR ($ecr_{deviation} \geq 0$) and negative if vice-versa ($ecr_{deviation} < 0$). Four classifiers were applied; support vector machine (SVM), simple MLP classifier, deep MLP classifier, and the RF classifier. The criterion was selected as the accuracy score, and the best confusion matrix was also extracted. The results are represented in Table IV and Table V.

As it is reported in Table IV, the deep MLP network gives out the best average accuracy score. The common point between regression and classification tasks is the closeness of the deep MLP and the RF results, both performing much better than other algorithms. Finally, we took a look at the effect of various environmental features on the ECR deviation. In terms of tire type, we distinguished that in winters the ECR deviation is positive in most cases; whereas in summer trips most were negative. The impact of auxiliary systems (the "park_heating" and "a/c" features) is also noticeable on the increase in consumption; in 79% of the cases, when both the auxiliary

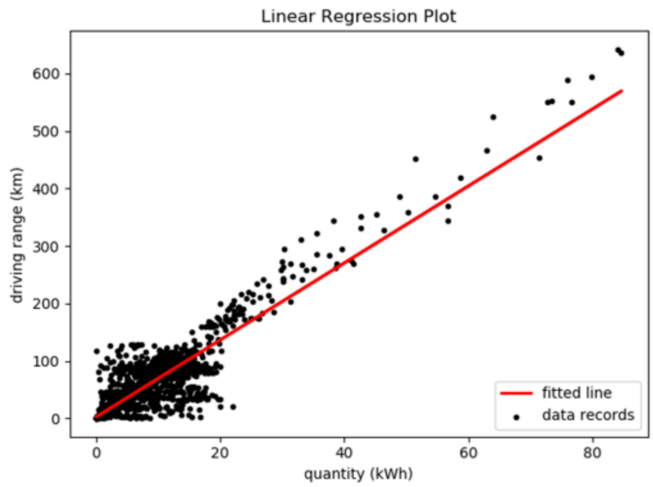

FIGURE 3. LINEAR REGRESSION MODEL RESULT FOR THE DRIVING RANGE IN TERMS OF THE CONSUMED ENERGY AMOUNT

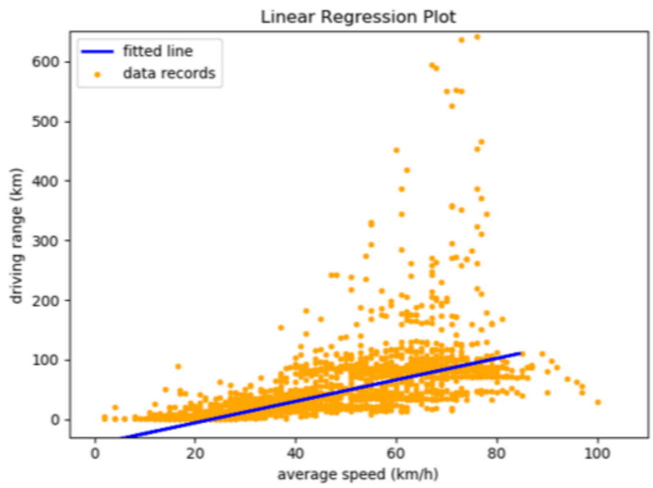

FIGURE 4. LINEAR REGRESSION MODEL FOR THE DRIVING RANGE IN TERMS OF THE AVERAGE SPEED

TABLE IV. MODEL RESULTS FOR DETECTING THE ECR DEVIATIONS

Model	Average accuracy	Best accuracy	Test-set size	Parameters
SVM	88.418	92.36	0.25	c = 100, gamma = 0.12, kernel = “rbf”
MLP	91.417	92.495	0.25	activation = “ReLU”, solver = “Adam”
Deep MLP	**92.24**	**94.23**	0.25	activation = “ReLU”, solver = “Adam”, batch_size = 16
Random Forest	90.719	94	0.25	n_estimators = 50

TABLE V. THE CONFUSION MATRIX FOR THE BEST ACCURACY SCORE, OBTAINED FROM THE DEEP MLP (94.23%)

	predicted positive	*predicted negative*
positive	598	20
negative	28	186

systems are switched off, the energy consumption is lower than the factory-defined amount. These interpretations totally confirm the results of the previous studies [10, 14, 24].

There is a significant point about the effect of route type on the rates of consumption; in 80% of the trips on urban streets, the ECR deviation is negative. On the face of it, this negativity is strange considering street features such as traffic congestion, but at the root of it, and according to [9], in-city roads are more efficient than the intercity ways. The reason lies in the use of the regenerative braking system (RBS), which converts the energy of every brake load into the electricity, thus reduces the battery discharge amount.

VI. CONCLUSION

To have a cursory glance over this research study, we estimated the driving range capability based on the data records composed of the quantity of consumed energy, the average speed, and the environmental features, such as route type and driving style. Among the developed models, the deep multilayer perceptron hit the best average mean absolute error of 5.58 kilometers. Although the linear regression did not obtain the best results, the mean absolute error scores of this model were acceptable. With the help of a much simpler model which is not comparable to the neural networks, in terms of complexity, learning time, and the dataset volume required to learn, we reasonably estimated the driving range of the EVs. After the trip distance prediction, we classified the energy consumption rate deviation between the real-world rates and the manufacturer’s; reaching at the best average accuracy score of 91.417% by applying the deep multilayer perceptron model. According to the Table V, we find that the false negative rate (FNR=3.24%) is much lower than the false positive rate (FPR=13.08%), which indicates that our model is more efficient in terms of detecting the trips with higher energy consumption rate deviations. This is more important to us; because running out of fuel before reaching the destination (range anxiety) is more important than consuming less energy than the factory-defined amount. Also, the impact of each environmental feature was investigated on the energy consumption rate deviation, and the reasons were discussed, which were consistent with the previous research studies.

REFERENCES

[1] Y. Huang, H. Wang, A. Khajepour, H. He, and J. Ji, “Model predictive control power management strategies for HEVs: A review,” *J. Power Sources*, vol. 341, pp. 91–106, Feb. 2017.

[2] M. Bohlsen, “EV Company News for The Month of January 2019,” *Seeking Alpha*, 2019.

[3] J. Hong, S. Park, and N. Chang, “Accurate remaining range estimation for Electric vehicles,” in *2016 21st Asia and South Pacific Design Automation Conference (ASP-DAC)*, pp. 781–786, 2016.

[4] J. Axsen, K. S. Kurani, and A. Burke, “Are batteries ready for plug-in hybrid buyers?,” *Transp. Policy*, vol. 17, no. 3, pp. 173–182, May 2010.

[5] E. Narassimhan and C. Johnson, “The role of demand-side incentives and charging infrastructure on plug-in electric vehicle adoption: analysis of US States,” *Environ. Res. Lett.*, vol. 13, no. 7, p. 074032 Jul. 2018.

[6] G. Wager, J. Whale, and T. Braunl, “Driving electric vehicles at highway speeds: The effect of higher driving speeds on energy consumption and driving range for electric vehicles in Australia,” *Renew. Sustain. Energy Rev.*, vol. 63, pp. 158–165, Sep. 2016.

[7] G. M. Fetene, S. Kaplan, S. L. Mabit, A. F. Jensen, and C. G. Prato, “Harnessing big data for estimating the energy consumption and driving range of electric vehicles,” *Transp. Res. Part D Transp. Environ.*, vol. 54, pp. 1–11, Jul. 2017.

[8] B. W. Lane, J. Dumortier, S. Carley, S. Siddiki, K. Clark-Sutton, and J. D. Graham, “All plug-in electric vehicles are not the same: Predictors of preference for a plug-in hybrid versus a battery-electric vehicle,” *Transp. Res. Part D Transp. Environ.*, vol. 65, pp. 1–13, Dec. 2018.

[9] X. Wu, D. Freese, A. Cabrera, and W. A. Kitch, “Electric vehicles’ energy consumption measurement and estimation,” *Transp. Res. Part D Transp. Environ.*, vol. 34, pp. 52–67, Jan. 2015.

[10] M. Smuts, B. Scholtz, and J. Wesson, “A critical review of factors influencing the remaining driving range of electric vehicles,” in *2017 1st International Conference on Next Generation Computing Applications (NextComp)*, pp. 196–201, 2017.

[11] B. Zheng, P. He, L. Zhao, and H. Li, “A Hybrid Machine Learning Model for Range Estimation of Electric Vehicles,” in *2016 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2016.

[12] M. Duke, D. Andrews, and T. Anderson, “The feasibility of long range battery electric cars in New Zealand,” *Energy Policy*, vol. 37, no. 9, pp. 3455–3462, Sep. 2009.

[13] H. Huo, Z. Yao, K. He, and X. Yu, “Fuel consumption rates of passenger cars in China: Labels versus real-world,” *Energy Policy*, vol. 39, no. 11, pp. 7130–7135, Nov. 2011.

[14] S. A. Birrell, A. McGordon, and P. A. Jennings, “Defining the accuracy of real-world range estimations of an electric vehicle,” in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 2590–2595, 2014.

[15] Y. Kondo, H. Kato, R. Ando, T. Suzuki, and Y. Karakama, “To what extent can speed management alleviate the range anxiety of EV?,”

in *2013 World Electric Vehicle Symposium and Exhibition (EVS27)*, pp. 1–8, 2013.

[16] A. Braun and W. Rid, "Assessing driving pattern factors for the specific energy use of electric vehicles: A factor analysis approach from case study data of the Mitsubishi i–MiEV minicar," *Transp. Res. Part D Transp. Environ.*, vol. 58, pp. 225–238, Jan. 2018.

[17] H. He, C. Sun, and X. Zhang, "A Method for Identification of Driving Patterns in Hybrid Electric Vehicles Based on a LVQ Neural Network," *Energies*, vol. 5, no. 9, pp. 3363-80, 2012.

[18] J. G. Hayes, R. P. R. de Oliveira, S. Vaughan, and M. G. Egan, "Simplified electric vehicle power train models and range estimation," in *2011 IEEE Vehicle Power and Propulsion Conference*, pp. 1–5, 2011.

[19] C. De Cauwer et al., "A Data-Driven Method for Energy Consumption Prediction and Energy-Efficient Routing of Electric Vehicles in Real-World Conditions," *Energies*, vol. 10, no. 5, p. 608, May 2017.

[20] J. Felipe, J. C. Amarillo, J. E. Naranjo, F. Serradilla, and A. Diaz, "Energy Consumption Estimation in Electric Vehicles Considering Driving Style, in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pp. 101–106, 2015.

[21] X. Yuan, C. Zhang, G. Hong, X. Huang, and L. Li, "Method for evaluating the real-world driving energy consumptions of electric vehicles," *Energy*, vol. 141, pp. 1955–1968, Dec. 2017.

[22] D. Jiménez-Bermejo, J. Fraile-Ardanuy, S. Castaño-Solis, J. Merino, and R. Álvaro-Hermana, "Using Dynamic Neural Networks for Battery State of Charge Estimation in Electric Vehicles," *Procedia Comput. Sci.*, vol. 130, pp. 533–540, Jan. 2018.

[23] https://spritmonitor.de

[24] S. Sautermeister, M. Falk, B. Baker, F. Gauterin, and M. Vaillant, "Influence of Measurement and Prediction Uncertainties on Range Estimation for Electric Vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 8, pp. 2615–2626, Aug. 2018.