A Data-driven Energy Estimation based on the Mixture of Experts Method for Battery Electric Vehicles

Patrick Petersen[a], Thomas Rudolf[b] and Eric Sax[c]
FZI Research Center for Information Technology,
Haid-und-Neu-Straße 10-14, 76131 Karlsruhe, Germany

Keywords: Battery Electric Vehicle, Energy Estimation, Machine Learning.

Abstract: Battery electric vehicles (BEVs) are an immediate solution to the reduction of greenhouse gas emissions. However, BEVs are limited in their range by the battery capacity. An accurate estimation of BEV's range and its energy consumption have become a significant factor in eliminating customers "range anxiety". To overcome range anxiety, advanced algorithms can predict the remaining capacity, estimate the range and inform the driver. Algorithms need to consider various influencing factors for their range estimation. A crucial part for an accurate range estimation is the energy consumption modeling itself. Thus, machine learning-based approaches are highly investigated which are able to learn nonlinear relations between relevant features and the energy consumption. In this paper, we propose a data-driven approach for the energy estimation of BEVs by utilizing ensemble learning to achieve a feature-specific estimation. In this paper, we trained neural networks on different road types independently. We improve the overall estimation by combining models via the mixture of experts method compared to a monolithic trained neural network. The results demonstrate that specialized neural networks for the energy estimation of BEVs are beneficial for the energy estimation. This approach contributes to reducing range anxiety and therefore helping toward elevated adoption of BEVs.

1 INTRODUCTION

The electrification of the automotive industry has become the solution for future sustainable transport and contributes to the reduction of air pollution and the dependency on fossil fuels (Arif et al., 2021). Especially BEVs have become popular as alternative for conventional internal combustion engine vehicles. The reason for this is that BEVs offer zero local emissions in combination with reduced noise pollution (Mahmoudzadeh Andwari et al., 2017; Sanguesa et al., 2021; Hua et al., 2021). Due to their simpler design with regard to required parts for building them, they are easier and cheaper to build, easier to maintain and moreover more efficient than combustion vehicle counterparts (Sanguesa et al., 2021). Despite the environmental benefits of BEVs, customers are still reluctant to fully accept and adept the electrification trend in the transport sector. The reason for this customer behavior is the limited charging infrastructure and lower driving range of BEVs (Eisel et al., 2016; Thorgeirsson et al., 2020). This causes one of the major concerns of drivers, called "range anxiety", which is defined as the anxiety that drivers experiences when worrying about whether the battery of their BEVs runs out of power before arriving at the destination or before a suitable charging point is reached (Noel et al., 2019). In consequence, automakers are expanding the range of BEVs by incorporating higher density batteries and advanced technology to reduce charging times as well as support the deployment of appropriate charging infrastructure. However, drivers still do not fully trust the displayed range of their vehicle, which leads to a safety margin of roughly 27% (Yuan et al., 2018). To achieve greater customer acceptance of sustainable electrified mobility and trust in the displayed range, an accurate estimation of the remaining range of BEVs is essential to eliminate range anxiety and increase the everyday usability. An accurate estimation of the energy consumption is also important for the whole energy management strategy of BEVs (Rudolf et al., 2021). However, determining the energy consumption and estimating range remains is a non-trivial task. The actual energy consumption of a BEV is influenced by many factors such as driving

[a] https://orcid.org/0000-0003-3203-5470
[b] https://orcid.org/0000-0002-6020-2611
[c] https://orcid.org/0000-0003-2567-2340

Petersen, P., Rudolf, T. and Sax, E.
A Data-driven Energy Estimation based on the Mixture of Experts Method for Battery Electric Vehicles.
DOI: 10.5220/0011081000003191
In *Proceedings of the 8th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2022)*, pages 384-390
ISBN: 978-989-758-573-9; ISSN: 2184-495X

style, road topology, weather and traffic conditions. Thus, for an accurate estimation, advanced algorithms are required to take all these factors into account and try to reduce uncertainties for the estimation (Kruppok et al., 2018).

Current estimation algorithms utilize data-driven methods such as machine learning to accurately estimate the energy consumption of BEVs. Often, a single monolithic neural network architecture with respect to the available data features is used. However, over the last couple of decades, research has shown that combining multiple machine learning models have a beneficial impact on the predictive performance compared to a single model (Sagi and Rokach, 2018).

Ensemble learning has already been used for energy estimation BEVs (Chung et al., 2019; Ullah et al., 2021). Each model of the ensemble approach is trained on the same data, thus each model is able to give an overall estimation in every upcoming situation. The final estimation result is computed by combining the results of each individual model estimations. However, no detailed studies have yet been performed how ensemble learning can benefit if each individual model is specialized for certain conditions e.g. road types or driver-styles. Thus, the goal of this paper is to extend the mentioned related work by presenting a methodology on how specialized neural networks can be utilized in an ensemble learning manner for the energy estimation of BEVs.

The remaining sections are set out as follows: A literature review on related works is presented in section 2. In section 3 we present our methodology for utilizing an ensemble neural network for the energy consumption estimation specialized on road types. In section 4, we show and discuss results of our implemented energy consumption estimation method and provide results and discussion. Finally, section 5 gives concluding remarks and an outlook.

2 ENSEMBLE LEARNING

Ensemble learning is a field of extensive research, showing potential performance increment for estimation applications. In general, the primary concept of ensemble learning is “the wisdom of the crowd” meaning that, based on the implemented ensemble learning methods, the predictions of several base models are combined for a final prediction. Ensemble learning can be either homogeneous or heterogeneous. A homogeneous ensemble is a collection of base models of the same type built on different data sets, whereas a heterogeneous ensemble is a collection of base models of different types built on the same data set. In the following we want to give a brief introduction to four common ensemble learning methods (bagging, boosting, stacking and mixture of experts) (Zhang and Ma, 2012). Describing their individual approach on how base models are used for improving the final prediction. Figure 1 shows an overview of the common ensemble learning methods.

Bagging

Bootstrap aggregating, or more commonly known by the acronym bagging, belongs to the homogeneous ensemble learning methods (Breiman, 1996). It creates individual base models $M_1,\dots,M_n$ for its ensemble by training each model on a random distributed subset of the data. However, sampling is done with replacements, thus samples are likely to appear more than once. The final output $y(x)$ is retrieved by combining the outputs of each base models $M_1(X),\dots,M_n(X)$ (see Figure 1a). For classification, this can be done by a simple majority vote for class $c \in C$ (see the following equation):

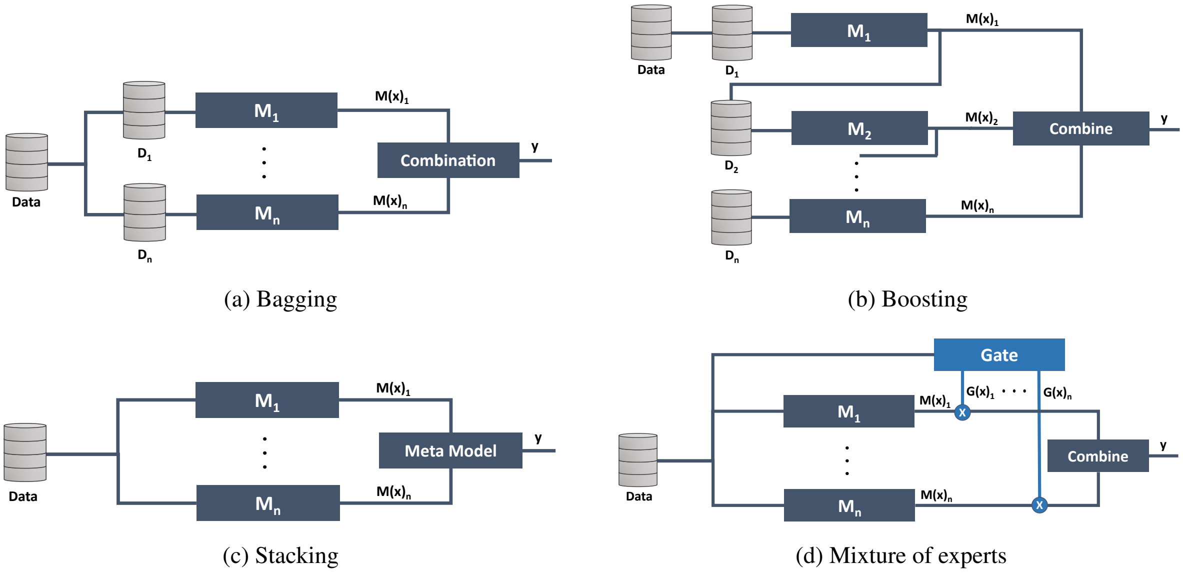
$$y(x) = \arg\max_{c\in C} \sum_{i=1}^{n} 1\left(M_i(x) = c\right) \tag{1}$$

For regression, the final output $y(x)$ is retrieved via averaging each individual output, as seen in the following equations:

$$y(x) = \frac{1}{n}\sum_{i=1}^{n} M_i(x) \tag{2}$$

Boosting

Boosting revolves around the idea of creating a “strong” learner from a set of “weak” homogeneous learners (Schapire, 1990; Freund and Schapire, 1999). A weak learner is a model which achieves an accuracy slightly better than random guessing. On the other hand, a strong learner refers to a model which produces close to perfect outputs. For applying the boosting method, models are trained sequentially (see Figure 1b). The first weak learner is trained on a random subset from the data. Every following weak learner is trained on the previous output, thus trying to correct its predecessors. This is done by redistribution of the previous weights of the preceding weak learners, to focus the resources of the following models on tougher data for the output. A strong learner uses weighted aggregation of the particular outputs of each weak learner to get the final output. Adaptive boosting (AdaBoost) is a popular representative for the ensemble learning boosting method. It uses a weighted majority vote, which weights $w_1,\dots,w_n$ are based on the premise of giving the more accurate weak learner



(a) Bagging

(b) Boosting

(c) Stacking

(d) Mixture of experts

Figure 1: Graphical illustration of the four common ensemble learning methods.

outputs $M_1(X),\dots,M_n(X)$ a greater influence for the final output $y(x)$. The following equation shows the basic formula for AdaBoost:

$$y(x) = \text{sign}\left(\sum_{i=1}^{n} w_i M_i(x)\right) \tag{3}$$

Stacking

Stacking (or stacked generalization) relies on a meta model, which tries to improve the final output $y(x)$ by inducing which base models $M_1(X),\dots,M_n(X)$ produce an accurate output (Smyth and Wolpert, 1997). In contrast to bagging and boosting, the stacking method often uses different learning algorithms such as decision trees and neural networks within the ensemble, thus following a heterogeneous approach for the ensemble. Similar to cross-validation, the whole data is split in k subsets, one for validation and the remaining $k-1$ for training. Each base model is then trained on a different set of $k-1$ subsets and afterwards validated on the unseen k^{th} subsets. The output of each base model is then used as the training data for a meta model, which learns on how to combine the outputs of each base model (see Figure 1c). Thus, resulting in learning weights $w_1,\dots,w_n$ for the outputs of each base model. After finished training for the meta model all the previously trained base models are discarded to retrain on the combined entire training data (Rincy and Gupta, 2020). The following equation demonstrates the combination rule for the meta model output $y(x)$:

$$y(x) = \sum_{i=1}^{n} w_i M_i(x) \tag{4}$$

Mixture of Experts

Mixture of experts relies on the idea of divide-and-conquer as the problem space is "divided" and then "conquered" by combining experts for the output (see Figure 1d) (Sagi and Rokach, 2018). It trains models $M_1,\dots,M_n$ on separated subsets of the training data, which are derived by separating different regions of the feature space (Jacobs et al., 1991). Hence, each model becomes an expert on its individual feature space. However, there is no general rule on how to create the subsets, and they are even allowed to overlap. After each model becomes an expert for its subspace of the problem, it is responsible for its specialized subset. An additional learning model is then used as a "gate" to learn and apply a weighted combination rule for the outputs of each expert $M_1(X),\dots,M_n(X)$. Thus, the weights $w_1,\dots,w_n$ are determined by the gate and the final output $y(x)$ is aggregated by the following equation:

$$y = \sum_{i=1}^{C} w_i(x) M_i(x) \tag{5}$$

3 METHODOLOGY

Our proposed methodology for the application of ensemble learning for energy estimation is based on the mixture of experts method. Thus, experts are combined to improve the estimation of energy usage for a given trip. These experts apply a machine learning technique on real-world driving-data which is enriched by external data sources such as data from navigation providers (e.g. Google or HERE) (Petersen et al., 2019). The input data for the experts is aggregated by GPS information, based on links from the navigation provider. This is reasonable due to common eco-routing systems which rely on road segments or links for their routing algorithms (Kucukoglu et al., 2021). In this study, data from the navigation provider HERE is used, which distinguishes 5 different road types. For each road type, an expert model is trained by selecting corresponding link-wise data. Thus, due to the nature of this methodology, we call this approach Mixture of Road Energy Experts (MORE). An illustration of MORE is given in Figure 2.

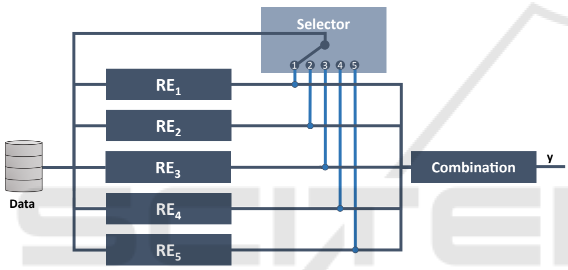


Figure 2: Graphical illustration of the postulated methodology MORE. Consisting of 5 road experts (RE) and a road-based selector for combining the estimations.

It is important to note, that in contrast to the standard mixture of experts method, we don't incorporate weights for each expert. Since the approach is a link-wise estimation method, based on routing information, the composition of the trip is known. Thus, it is known to the algorithm when to apply which road expert $RE_1, \dots, RE_5$ for a link to benefit from its expertise. In the remaining part of this section, we will give a description of the available data, presenting the selected features as well as the energy estimation model.

3.1 Description of Real-world Driving Data

Data-driven approaches such as MORE learn the relationship between different real-world driving signals such as velocity and the energy consumption. In this study, data from 152 real-world test drives of a *Porsche Taycan* were used. The test drives were conducted under real-world conditions in Europe, mainly in Germany, and consists of an accumulated length of approximately 7500 km. All signals on the Controller Area Network (CAN) were sampled at 10 Hz. As mentioned in the beginning of this section, the existing CAN data was enriched by traffic and road data from the navigation provider HERE and segmented based on the link information. Yielding roughly 40000 road segments as the data input for the models used in MORE.

3.2 Machine Learning-based Energy Consumption Model

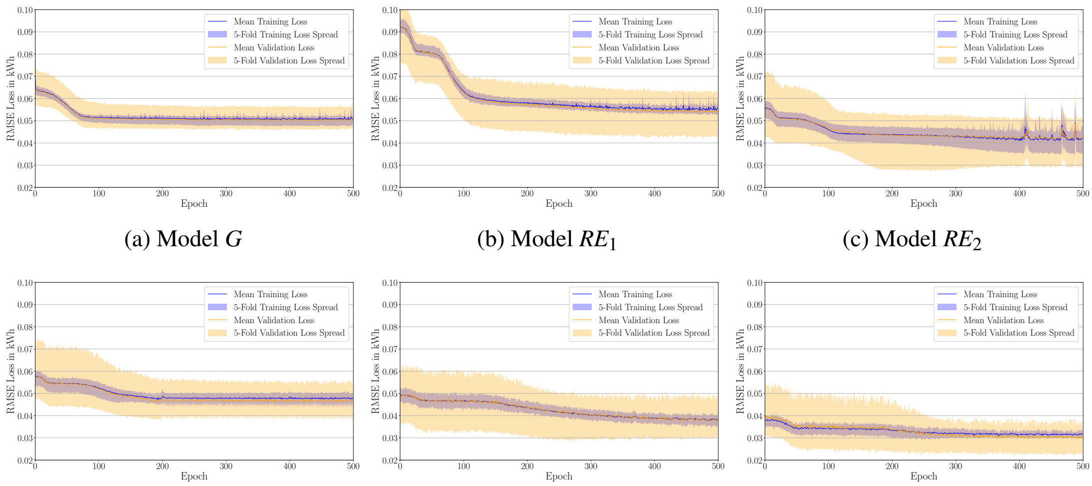
The proposed methodology consists of five expert models for the estimation of the energy consumption on different road types. Research has shown that Multilayer Perceptron (MLP) architectures achieve high accuracy when used in ensembles for time series estimation (Mahalakshmi et al., 2016), thus it was chosen as an initial expert model type. Even tough newer and more complex models for time series estimation exist, MLPs yield sufficient estimation accuracy and can be used in a variety of different circumstances. Thus, MLPs can be used as a baseline to investigate the MORE approach. For the input of the MLPs various statistical features (e.g. mean and standard deviation) were calculated on the recorded signals for each road segment. By applying a correlation-based feature selection, 6 relevant features were deduced to have a high impact on the energy consumption. Table 1 gives an overview of the selected features as the input for the expert models of MORE.

Table 1: Input data for the expert models.

Feature	Description
v_{mean}	Mean speed of the ego vehicle
m_{neg}	% of negative slope
v_{base}	Base speed of road participants
l	Length of the road segment
RT	Road type of the current road segment
T_{out}	Outside temperature

4 EXPERIMENT AND DISCUSSION

In this study, we evaluate the performance of MORE and its individual road experts $RE_1, \dots, RE_5$ against the performance of a general monolithic model G which is trained on all the training data. The original data is split into training and testing by applying an 80/20 split. For verifying the robustness of the models, we conducted a 5-fold cross-validation experiment on the training data. Figure 4 illustrated the 5-fold cross-validation procedure. For evaluation



(a) Model G (b) Model RE_1 (c) Model RE_2

(d) Model RE_3 (e) Model RE_4 (f) Model RE_5

Figure 3: Training and validation loss for applied 5-fold cross-validation.

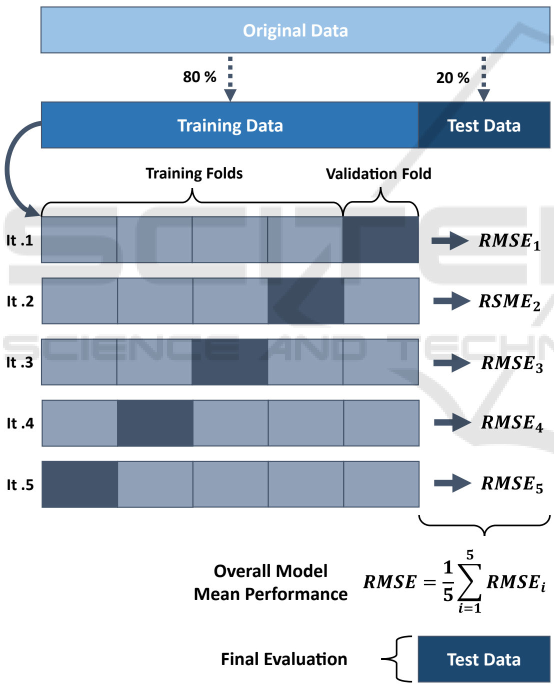


Figure 4: 5-Fold cross-validation for the training and test data.

of each model, we used the root mean squared error (RMSE) (see equation 6).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{6}$$

Each model was trained for 500 epochs on its dedicated data. Thus, road experts RE_n were trained and validated on the subset of road type n. As a result, model G had the most data for the training and validation process compared to the individual road experts.

Table 2 give an overview of the composition of the data in regard to the road types.

Table 2: Proportions of road types on the total distance.

Road Type	$\overline{v}$ in data	Percentage of data
1	90 km h^{-1}	41 %
2	60 km h^{-1}	35 %
3	57 km h^{-1}	17 %
4	45 km h^{-1}	4 %
5	28 km h^{-1}	3 %

HERE defines the road types as ”used to classify roads depending on the speed, importance and connectivity of the road” (HERE, 2022). A lower number indicates a higher priority, thus higher speed and importance of that road type. Figure 3 shows the mean training and validation loss, as well as the loss spread across the folds for each model. It can be seen that during the process, no model underfitted or overfitted. However, the spread of the validation loss is much higher for the road expert models $RE_1, \ldots, RE_5$ compared to the general monolithic model G. This is due to the decreased data for the cross-validation approach for the road expert models. Table 3 summarizes the best mean validation loss of the cross-validation approach.

The results indicate that besides RE_1 each road expert yield a smaller RMSE for its individual road type. For a final evaluation of the combined road experts as our proposed MORE we compare it to the general monolithic model G on the test data which was not used for the cross-validation. Table 4 compares the two models on the whole test data, as well as a detailed comparison of each road type within the test data.

Table 3: Cross-validation results of each model.

Model	RMSE [kWh]
G	0.052
RE_1	0.056
RE_2	0.045
RE_3	0.048
RE_4	0.041
RE_5	0.034

Table 4: Comparison of general monolithic model G and MORE on test data.

Data	RMSE_G [kWh]	RMSE_{MORE} [kWh]
All test data	0.053	0.049
Road type 1	0.048	0.045
Road type 2	0.048	0.044
Road type 3	0.051	0.047
Road type 4	0.055	0.052
Road type 5	0.052	0.051

It can be seen that MORE has an overall better performance on each road type. Each individual road expert RE_n of MORE has an better estimation performance on its dedicated part of the test data compared to the general monolithic model G. In total, MORE has an 7.5% improvement over the performance of model G for the whole test data. It illustrates the potential on how ensemble learning can improve the energy estimation for BEVs.

5 CONCLUSION

This paper presents a data-driven approach for the energy estimation of BEVs based on ensemble learning, utilizing the mixture of experts method to specialize models on specific road types. It is found that, the proposed method MORE with 5 specialized road experts improves the RMSE of the energy estimation by roughly 7.5% compared to the estimation of a monolithic model. The results show that, energy estimation benefits from utilizing an ensemble neural network approach. However, testing this concept in live operation on a BEV may yield additional insights on the applied advantages of MORE.

The research shown in this paper could be extended in the future in different aspects. Different specializations for the mixture of experts method should be investigated, e.g. for different driver styles. Further work could incorporate advanced methods for a robust and reliable classification of different driver-styles, which will be used for the experts. A further study could assess the impact of individual features for each specialized neural network due to their importance for the energy estimation, e.g. on different road types or for different driver-styles. Simultaneously investigating different combinations of neuronal network architectures (e.g. RNN, CNN or Transformer) might also optimize the overall accuracy and data efficiency of utilizing heterogeneous models to benefit from their individual traits.

REFERENCES

Arif, S. M., Lie, T. T., Seet, B. C., Ayyadi, S., and Jensen, K. (2021). Review of Electric Vehicle Technologies, Charging Methods, Standards and Optimization Techniques. *Electronics*, 10(16):1910.

Breiman, L. (1996). Stacked regressions. *Machine Learning*, 24(1):49–64.

Chung, Y. W., Khaki, B., Li, T., Chu, C., and Gadh, R. (2019). Ensemble machine learning-based algorithm for electric vehicle user behavior prediction. *Applied Energy*, 254(April):113732.

Eisel, M., Nastjuk, I., and Kolbe, L. M. (2016). Understanding the influence of in-vehicle information systems on range stress – Insights from an electric vehicle field experiment. *Transportation Research Part F: Traffic Psychology and Behaviour*, 43:199–211.

Freund, Y. and Schapire, R. E. (1999). A Short Introduction to Boosting. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence - Volume 2*.

HERE (2022). FunctionalClassType. https://developer.here.com/documentation/routing/dev_guide/topics/resource-type-functional-class.html [Online; Accessed: 2022-01-22].

Hua, X., Thomas, A., and Shultis, K. (2021). Recent progress in battery electric vehicle noise, vibration, and harshness. *Science Progress*, 104(1):003685042110052.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E. (1991). Adaptive Mixtures of Local Experts. *Neural Computation*, 3(1):79–87.

Kruppok, K., Kriesten, R., and Sax, E. (2018). Calculation of route-dependent energysaving potentials to optimize EV's range. In Bargende, M., Reuss, H.-C., and Wiedemann, J., editors, *18. Internationales Stuttgarter Symposium*, pages 1349–1363. Springer Fachmedien Wiesbaden, Wiesbaden.

Kucukoglu, I., Dewil, R., and Cattrysse, D. (2021). The electric vehicle routing problem and its variations: A literature review. *Computers & Industrial Engineering*, 161:107650.

Mahalakshmi, G., Sridevi, S., and Rajaram, S. (2016). A survey on forecasting of time series data. In *2016 International Conference on Computing Technologies and Intelligent Data Engineering, ICCTIDE 2016*. Institute of Electrical and Electronics Engineers Inc.

Mahmoudzadeh Andwari, A., Pesiridis, A., Rajoo, S., Martinez-Botas, R., and Esfahanian, V. (2017). A review of Battery Electric Vehicle technology and readiness levels. *Renewable and Sustainable Energy Reviews*, 78(May):414–430.

Noel, L., Zarazua de Rubens, G., Sovacool, B. K., and Kester, J. (2019). Fear and loathing of electric vehicles: The reactionary rhetoric of range anxiety. *Energy Research & Social Science*, 48:96–107.

Petersen, P., Thorgeirsson, A. T., Scheubner, S., Otten, S., Gauterin, F., and Sax, E. (2019). Training and validation methodology for range estimation algorithms. In *Proceedings of the 5th International Conference on Vehicle Technology and Intelligent Transport Systems, VEHITS 2019, Heraklion, Crete, Greece, May 3-5, 2019*, pages 434–443. SciTePress.

Rincy, T. N. and Gupta, R. (2020). Ensemble Learning Techniques and its Efficiency in Machine Learning: A Survey. In *2nd International Conference on Data, Engineering and Applications (IDEA)*, pages 1–6, Bhopal, India. IEEE.

Rudolf, T., Schurmann, T., Schwab, S., and Hohmann, S. (2021). Toward Holistic Energy Management Strategies for Fuel Cell Hybrid Electric Vehicles in Heavy-Duty Applications. *Proceedings of the IEEE*, 109(6):1094–1114.

Sagi, O. and Rokach, L. (2018). Ensemble learning: A survey. *WIREs Data Mining and Knowledge Discovery*, 8(4).

Sanguesa, J. A., Torres-Sanz, V., Garrido, P., Martinez, F. J., and Marquez-Barja, J. M. (2021). A Review on Electric Vehicles: Technologies and Challenges. *Smart Cities*, 4(1):372–404.

Schapire, R. E. (1990). The strength of weak learnability. *Mach. Learn.*, 5(2).

Smyth, P. and Wolpert, D. (1997). Stacked Density Estimation. In *Advances in Neural Information Processing Systems 10, [NIPS Conference, Denver, Colorado, USA, 1997]*, pages 668–674. The MIT Press.

Thorgeirsson, A. T., Scheubner, S., Funfgeld, S., and Gauterin, F. (2020). An Investigation Into Key Influence Factors for the Everyday Usability of Electric Vehicles. *IEEE Open Journal of Vehicular Technology*, 1(Xx):348–361.

Ullah, I., Liu, K., Yamamoto, T., Zahid, M., and Jamal, A. (2021). Electric vehicle energy consumption prediction using stacked generalization: An ensemble learning approach. *International Journal of Green Energy*, 18(9):896–909.

Yuan, Q., Hao, W., Su, H., Bing, G., Gui, X., and Safikhani, A. (2018). Investigation on Range Anxiety and Safety Buffer of Battery Electric Vehicle Drivers. *Journal of Advanced Transportation*, page 12.

Zhang, C. and Ma, Y. (2012). *Ensemble Machine Learning*. Springer US, Boston, MA.