Interpretable Data-Driven Capacity Estimation of Lithium-ion Batteries

Yixiu Wang [*] **Anurakt Kumar** [**] **Jiayang Ren** [*] **Pufan You** [***]
Arpan Seth [****] **R. Bhushan Gopaluni** [*] **Yankai Cao** [*]

[*] *Department of Chemical and Biological Engineering, University of British Columbia, Vancouver, BC, V6T 1Z3, Canada (e-mail: bhushan.gopaluni@ubc.ca; yankai.cao@ubc.ca)*
[**] *Indian Institute of Technology, Kharagpur, 721302, India*
[***] *Department of Statistics, University of Manitoba, Winnipeg, MB, R3T 2N2, Canada*
[****] *Evonik Corporation, Trexlertown, Pennsylvania 18195, USA*

Abstract: Battery degradation poses a significant challenge for the usage of Lithium-ion batteries, making accurate capacity estimation crucial for efficient operation. Data-driven approaches hold promise for addressing this task, yet their complex structures often lead to overfitting and obscure the decision-making process. The objective of this work is to build a robust and interpretable model for capacity estimation. We propose the utilization of a robust decision tree-based ensemble model, extremely randomized trees (ERT), to accurately estimate battery capacity based on the features extracted from the partial charging curve. The random splits in the tree construction process enhance the model's generalization ability. Given that the combination of multiple decision trees reduces interpretability, we further employ SHAP to interpret the contributions of each feature to the ERT model's predictions. The effectiveness of the proposed method is validated on a large cycling dataset of Lithium-ion batteries.

Keywords: Lithium-ion batteries, Capacity estimation, Data-driven modeling, Extremely randomized trees, Interpretability, SHAP

1. INTRODUCTION

With the continuous advancement of modern technology, Lithium-ion batteries have been seamlessly integrated into every facet of our lives because of their high energy and power density. They power the cell phones we hold, an array of electronic devices, and even electric cars, underscoring their indispensable role (Bresser et al. (2018)). Despite their central role in our technology-driven world, batteries are not without limitations. The issue of battery degradation is a well-known challenge, typically characterized as a gradual decline in performance over time, This not only impacts the longevity of the battery but also its overall efficiency (Han et al. (2019)). When assessing battery performance, a key determinant is the battery capacity, which measures the amount of electricity a fully charged battery can discharge. However, due to the uncertainty in battery operating conditions for practical applications, a complete discharge process is seldom available for obtaining battery capacity (Wang et al. (2022)). Therefore, to optimize energy utilization and ensure proper equipment functionality, it is crucial to precisely estimate battery capacity using voltage and current measurements from the partial charging curve.

Numerous research efforts have been devoted to the task of battery capacity estimation in the last decades, which can be broadly categorized into two main types, namely model-based and data-driven methods. The former revolves around the utilization of either equivalent circuit models (Hu et al. (2012)) or electrochemical models (Doyle et al. (1993)) to simulate the dynamic characteristics of batteries. Subsequently, recursive adaptive filters such as the Kalman filter (Plett (2004)) and particle filter (Schwunk et al. (2013)) are employed to identify model parameters that are associated with capacity. While these methods can attain a high level of accuracy, they lack feasibility for online applications due to computational intensity.

Contrastively, data-driven methods do not rely on in-depth expert knowledge about batteries, focusing instead on the relationship between inputs (i.e., features) and outputs (i.e., capacity). A wide array of machine learning techniques, including support vector regression (SVR) (Wang et al. (2018)), Gaussian process regression (GPR) (Yang et al. (2018)), and deep neural networks (Roman et al. (2021)), have been employed for this task, achieving commendable results in capacity estimation. In recent years, ensemble models based on decision trees (Kotsiantis (2013)), such as random forest (RF) (Breiman (2001)) and light gradient boosting machine (LightGBM) (Ke et al. (2017)), have gained widespread attention across various tasks due to their ability to maintain high accuracy while being easy to implement. However, these models suffer from the risk of overfitting, as decision trees recursively find the optimal split at each node, and data noise can make it challenging for the model to develop appropriate tree structures. To address this issue, extremely random-

ized trees (ERT) (Geurts et al. (2006)) adopt a method of randomly selecting features and thresholds to construct decision trees, thereby enhancing the model's generalization ability against data noise.

Although data-driven models have achieved remarkable success in estimating the capacity of batteries, the complex structure of these models often makes the decision-making process difficult to understand. This opacity may raise concerns about the reliability and safety of data-driven models, especially in practical applications involving batteries. Interpretable machine learning (Murdoch et al. (2019)) is an emerging field aimed at enhancing trust in models by elucidating the contribution of each feature to model predictions. Originating from game theory, the Shapley value is proposed to fairly allocate the contributions of each player in a game (Shapley (1953)). When the Shapley value is represented as an additive feature attribution method, it evolves into the SHapley Additive exPlanations (SHAP) model (Lundberg and Lee (2017)). The SHAP model integrates the fairness principles of game theory with local interpretability, offering a novel approach to explain and understand complex data-driven models (Agarwal and Das (2020)).

In this work, we propose the utilization of ERT to estimate battery capacity based on features extracted from charging curves. Subsequently, the SHAP model is employed to interpret and quantify the contribution of each feature to the estimated capacity. The remainder of the paper is organized as follows. Section 2 provides the cycling dataset studied in this work. Section 3 introduces the proposed methodology, including the feature extraction and the proposed models. The capacity estimation results and discussion are reported in Section 4 and finally, some conclusions are drawn in Section 5.

2. CYCLING DATASET

To validate the effectiveness of the proposed capacity estimation method, a large cycling dataset comprising 66 battery cells (Zhu et al. (2022)) is used in this study. These battery cells, all featuring the same cathode material ($LiNi_{0.86}Co_{0.11}Al_{0.03}O_2$), were cycled under various temperatures and different charging current rates. Table 1 lists the detailed cycling conditions. As shown in the table, the cycling temperatures include 25 °C, 35 °C, and 45 °C, with charging current rates ranging from 0.25 C to 1 C. It is important to note that the current rates were determined based on the battery's nominal capacity, where 1 C corresponds to 3.5 A.

A complete cycle consists of both charging and discharging processes. Specifically, in the charging phase, the cell cycling is first performed in a constant current (CC) mode with a specific current rate until the voltage reaches 4.2 V, followed by a constant voltage (CV) mode at 4.2 V until the current decreases to 0.05 C. During the discharging process, a constant current of 1 C is employed until the voltage drops to 2.65 V. After repeated charging and discharging cycles, we can observe the change in the capacity of all battery cells with the number of cycles, as illustrated in Fig. 1. It is noticeable that the batteries exhibit varying degradation characteristics, with the cycle number ranging from 50 to 800 when the capacity drops to around 2500 mAh. Generally, higher charging current rates tend to accelerate battery degradation. Another finding is that, even under identical working conditions, the batteries still show inconsistencies in their cycling characteristics, suggesting that relying solely on the cycle number for capacity estimation is insufficient. More measurements from the charging and discharging processes need to be considered for accurate capacity estimation.

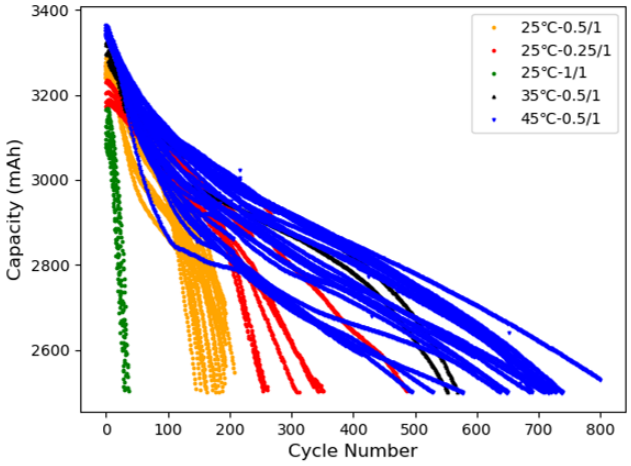

Fig. 1. Capacity change vs cycle number of all batteries over varying cycling temperature and charge/discharge rates from Table 1

3. METHODOLOGY

3.1 Feature Extraction

As demonstrated by Chen et al. (2017), extracting valuable features as a first step of building a data-driven model for capacity estimation, significantly impacts the model's performance. Compared to the discharge process, which can be uncertain due to varying operational conditions and loads, the charging process of a battery is often more regular, thus attracting considerable attention. In this study, as suggested by the work of (Li et al. (2018)), we utilize the partial CC charging curve to extract relative charging capacity for capacity estimation.

Fig. 2 illustrates the voltage change over time in a battery's CC charging process. For feature extraction, it is essential first to fix a specific voltage range with a lower voltage bound V_l and an upper voltage bound V_u. It's worth mentioning that our motivation for utilizing only a segment of the charging curve is to accommodate the real-world situation where batteries rarely commence charging from 0% State of Charge (SOC). Subsequently, we discretize the voltage range from V_l to V_u at equal voltage intervals ΔV. This operation ensures that different batteries have the same number of sampling points across all cycles, thus maintaining a fixed feature length. The time t_j to reach a discrete voltage point V_j is obtained through linear interpolation. The relative capacity Q_j at V_j is then calculated using the following formula:

$$Q_j = I \cdot (t_j - t_0) \tag{1}$$

where I is the constant charging current and t_0 is the charging time to reach V_l.

Table 1. Cycling conditions of battery cells for dataset generation

Cell Specifications	Cycling Temperature (°C)	Charging/Discharging Current Rate (C)	Number of Cells
Anode: Graphite/Si		0.25/1	7
Cathode: $LiNi_{0.86}Co_{0.11}Al_{0.03}O_2$	25	0.5/1	19
Type: 18650		1/1	9
Cutoff Voltage: 2.65 - 4.2 V	35	0.5/1	3
Nominal Capacity: 3.5 Ah	45	0.5/1	28

For the task of capacity estimation, each cycle is considered as one sample. In this context, the input feature of each sample is $\mathbf{x}_i = [Q_1, \cdots, Q_k]$, where k represents the number of discrete points, which can be calculated as $(V_u - V_l)/\Delta V$. The output is the capacity for that cycle, experimentally determined by integrating the current over time during the complete discharge process. From the dataset we utilized in this work, we generated a total of 22638 samples for training and testing various regression models.

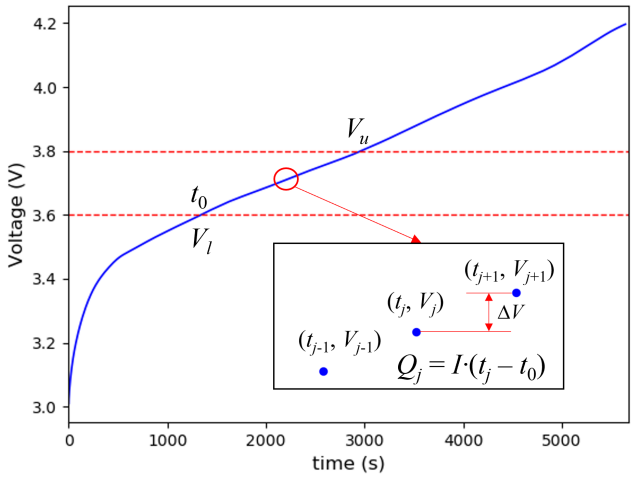

Fig. 2. Illustration of feature extraction from partial CC charging curve

3.2 Extremely Randomized Trees

Extremely Randomized Trees (ERT) are a popular ensemble learning technique, especially effective for complex supervised machine learning tasks. To facilitate understanding of ERT, we start with an introduction to the decision tree, which forms the basis of an ERT model. The decision tree is a machine learning method that is used for classification as well as regression tasks, as shown in Fig. 3. In the tree, for a branch node tb, $\mathbf{a}_{tb}$ is composed of binary variables, each representing whether a feature participates in the split, and b_{tb} is the threshold for splitting. Moreover, to ensure that only one feature participates in the split at each node, the sum of all elements in $\mathbf{a}_{tb}$ is forced to be exactly 1. For regression tasks, the prediction at a leaf node is typically the average of labels of all samples that fall into that leaf.

Considering a regression task with a given dataset $(\mathbf{X}, \mathbf{y})$ containing n samples $(\mathbf{x}_i, y_i)$, $i \in \{1, \cdots, n\}$. Each sample comprises p features $\mathbf{x}_i \in \mathbb{R}^p$ and a continuous label $y_i \in \mathbb{R}^1$, and the decision tree establishes a map from $\mathbf{x}_i$ to y_i. More specifically, for a data sample i, as it traverses

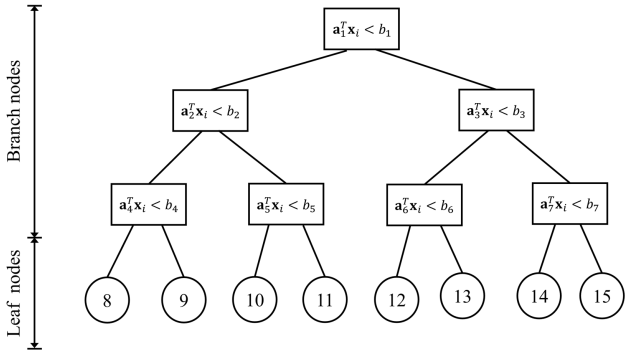

Fig. 3. The structure of a decision tree

a branch node tb, it is assigned to the left child node if $\mathbf{a}_{tb}^T\mathbf{x}_i < b_{tb}$, otherwise to the right child node. In this manner, sample i will eventually be assigned to a leaf node, and the prediction of this leaf node is given to the sample.

Traditional decision tree algorithms, like CART (Breiman (1984)), recursively find the optimal split at each branch node until the number of samples at a branch node is less than the minimum allowed leaf size or the maximum depth is reached. For a regression task, mean squared error is a common criterion to evaluate the loss, which is given by

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

where n is the number of samples passing through the branch node and $\hat{y}$ is the predicted value. For each branch node, it tries to split the samples into two subsets. In other words, CART finds the optimal splitting rule at each branch node to divide the data samples into the two subsets with the lowest possible variance.

Building upon this, ERT adds additional randomness. Unlike traditional decision trees, each tree within ERT randomly selects its splitting rules. To be more specific, at each branch node of the tree, ERT randomly chooses K features and generates random splitting thresholds for these features. Subsequently, for each randomly generated splitting rule, ERT computes the MSE of the post-split data and keeps the splitting rule that minimizes the MSE. Due to this random splitting approach, ERT is more efficient than CART in building decision trees, especially for datasets with a large number of features. Additionally, this randomness reduces the variance of the model, thereby strengthening its generalization ability on unseen testing samples. It is worth mentioning that, unlike RF, ERT constructs all its decision trees using the original dataset, as random splitting allows for the creation of entirely independent trees even on the same dataset. After constructing M randomly generated decision trees, ERT outputs the average of the predictions from all trees, which

can be expressed as:

$$\hat{y}_i = \frac{1}{M} \sum_{m=1}^{M} f^m(\mathbf{x}_i) \tag{3}$$

where $f^m : \mathbb{R}^p \to \mathbb{R}^1$ denotes the output from the m-th decision tree model.

3.3 SHAP

The decision-making process of a single decision tree is fairly straightforward to understand. However, when numerous decision trees are combined, as in the case of ERT, the overall decision-making process becomes much more complex and less interpretable. Therefore, there's a need for a tool to elucidate the predictions made by ERT. SHAP offers a promising solution in this regard. By using SHAP, we gain a more granular insight into how different features contribute to the model's predictions, thus enabling a clearer interpretation and justification of complex models.

SHAP quantitatively measures the influence of each feature on the model's prediction, accounting for not just the value of the feature itself but also its interaction with other features. The foundation of SHAP is the Shapley value, a concept from game theory that provides a fair allocation of a coalition's payoff to its participants.

In the context of machine learning, we consider the "coalition" as the set of all features used by the model, and the "payoff" as the prediction output by the model. The SHAP value for a feature j is calculated as follows:

$$\phi_j(f) = \sum_{S' \subseteq P \setminus j} w_j(S') \left[f(S' \cup j) - f(S') \right] \tag{4}$$

where $\phi_j(f)$ represents the Shapley value of the feature j under the model f. The set P denotes all input features, and S' is a subset of these features excluding j. The term $S' \cup j$ symbolizes the union of the subset S' with feature j, essentially assessing the impact of adding j to the subset S'. The weight $w_j(S')$ is calculated as $|S'|! \cdot (|P| - |S'| - 1)!/|P|!$, where $|S'|$ is the size of the subset S' and $|P|$ is the total number of features. This weight factors in the number of permutations in which j can be combined with the features in S', given all possible combinations of features. Finally, $f(S' \cup j) - f(S')$ represents the marginal contribution of feature j when it is added to the subset S'.

4. RESULTS AND DISCUSSIONS

4.1 Performance of Capacity Estimation

In this work, we utilize ERT to construct a data-driven model for estimating the capacity of Lithium-ion batteries. Specifically, we extract features from the constant current charging voltage curve in the voltage range from 3.65 V to 3.85 V, and discretize the voltage at 0.002 V intervals, thus resulting in 100 features. We normalize the capacity based on the nominal capacity, making the output a percentage. The model is trained on a randomly selected 80% of the samples (18110 samples), and its performance is evaluated on the remaining 20% of the samples (4528 samples) using root mean squared error (RMSE) and coefficient of determination (R^2) as performance metrics. Typically, an RMSE close to 0 and R^2 close to 1 suggests a good match between the measured and predicted data.

In order to evaluate the effectiveness of our proposed ERT model, we compared it to a variety of models including RF, LightGBM, CatBoost, and linear regression. The linear regression model was chosen for comparison because linear regression assumes a linear relationship between features, which contrasts with the assumptions of tree-based models, which assume a non-linear relationship between features. Through this comparison, we aim to better understand the relationships between features and their impact on model performance. In addition, models such as SVR and GPR are also tried to fully evaluate the performance of different algorithms on the battery capacity estimation task.

All models are trained using the same training dataset, and their performances on the testing samples are summarized in Table 2. The results show that linear regression has the poorest performance with an RMSE of 2.24% and R^2 of 0.8589, implying a nonlinear relationship between the features and capacity. The proposed ERT model achieves the best performance with an RMSE of 0.28% and R^2 of 0.9978. It is noteworthy that the ERT model shows a significant improvement in accuracy compared to other decision tree-based ensemble models, indicating that ERT is a more effective and robust model for this task.

Table 2. Performance comparison of different modelling methods on test dataset

Model	RMSE (%)	R^2
ERT	**0.28**	**0.9978**
RF	0.32	0.9971
LightGBM	0.50	0.9929
CatBoost	0.58	0.9905
Linear Regression	2.24	0.8589
SVR	1.90	0.8985
GPR	0.39	0.9957

Fig. 4 (a) provides a comparison of observed capacity versus predicted capacity from the ERT model. As shown in the figure, it is evident that the proposed model accurately estimates the capacity for most testing samples, as demonstrated by most points lying near the reference line. Fig. 4 (b) further displays the testing errors across all samples, highlighting that the errors for a vast majority of samples remain within 10mAh, thereby confirming the superior performance of our proposed model.

We further analyze the impact of different selected voltage ranges on the model's performance, as depicted in the graph. In this study, we maintain a constant voltage segment length of 0.2V, hence using the mean value to denote the selected voltage range. For instance, a mean of 3.7 V refers to the range from 3.6 V to 3.8 V. As illustrated in Fig. 5, the accuracy tends to initially increase and then decrease with the rising mean value, with a performance at higher voltage ranges generally inferior to that at lower ones, and the optimal performance achieving at 3.75 V. A significant observation is that mean values from 3.7V to 3.85V all show remarkably excellent results, suggesting that the mid-voltage ranges are all viable for the task of capacity estimation. This broad range of effective choices bolsters the practical feasibility of the proposed method.

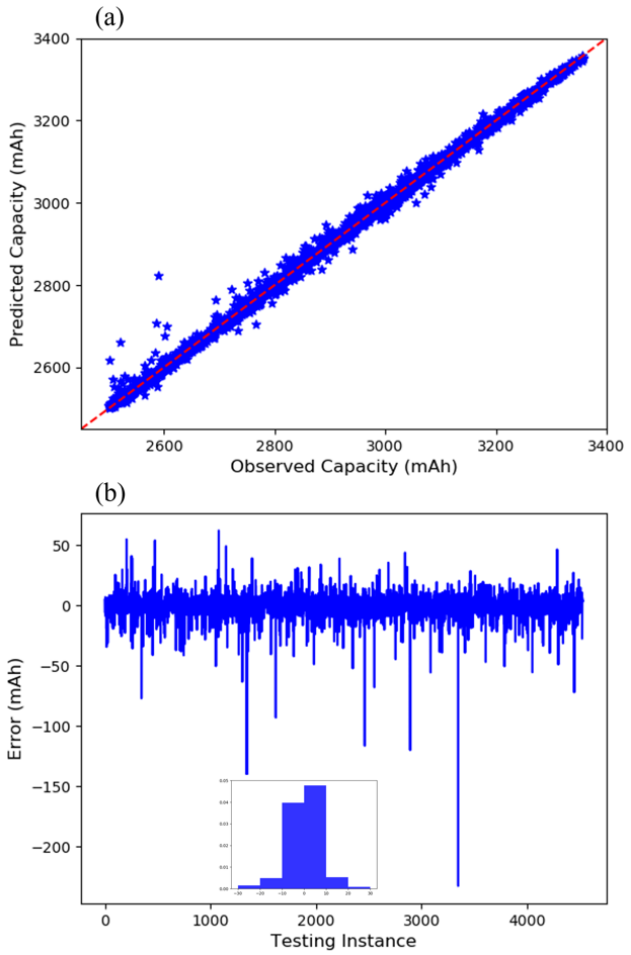

Fig. 4. Testing results of the ERT model: (a) observed capacity versus predicted capacity, (b) residual error over all testing samples

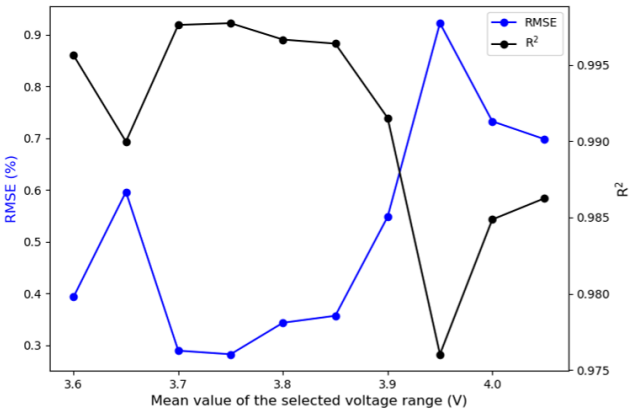

Fig. 5. ERT model performance over selected voltage ranges

4.2 SHAP Analysis

After obtaining an ERT model that accurately estimates battery capacity, we utilize SHAP for interpreting the model's predictions. Firstly, Fig. 6 provides a global explanation of the model, showcasing the average influence of each feature on the predictions across all samples. As illustrated in the figure, the indices of the 10 most significant features are all above 80, corresponding to the relative capacity after a 0.16V increase in voltage during the charging process. This indicates that longer intervals of charging offer more effective data for estimating capacity. For these features, a larger feature value corresponds to a greater contribution, suggesting a positive correlation between the features and capacity. Moreover, the similar levels of contribution from the top 10 features also hint at potential strong collinearity among them, demonstrating the ERT model's exceptional performance in handling feature collinearity.

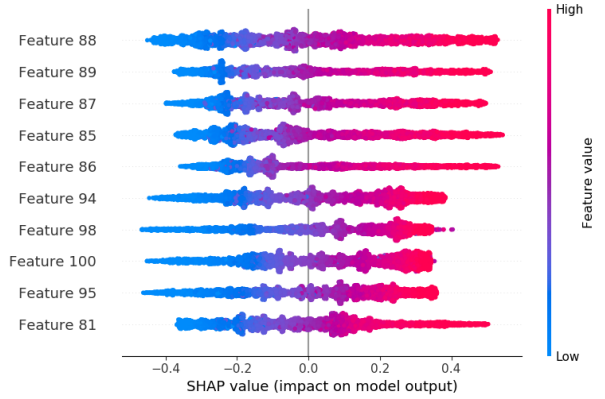

Fig. 6. Global interpretation of the ERT model

To further illustrate the contribution of each feature to the prediction of individual samples, Fig. 7 displays the waterfall plots for two specific samples, with observed capacities being 95.45% (3340.70 mAh) and 82.38% (2883.18 mAh), respectively. These plots delineate the contribution of each feature from the base prediction to the final prediction. Notably, the base prediction is the mean predicted value for all training samples made by the ERT model. As depicted in Fig. 7 (a), when the final prediction significantly exceeds the base prediction, all features contribute positively to the result, and the sum of these contributions leads to the final prediction. In contrast, Fig. 7 (b) shows the individual feature contributions when the final prediction is close to the base prediction, with some contributions being positive and others negative. In this way, we are able to have a better understanding of the decision-making process for battery capacity estimation of a complex ERT model, which will benefit both the monitoring and early warning of battery health status.

5. CONCLUSIONS

In this study, we propose to adopt the ERT model for battery capacity estimation, followed by an explanation of the model using SHAP. By constructing decision trees through random splits, ERT reduces the possibility of overfitting to some extent, thus yielding more robust and effective results. Experimental results confirm the superiority of our proposed ERT model, achieving an RMSE of 0.28% and R^2 of 0.9978, significantly outperforming other decision tree-based ensemble models. Furthermore, using SHAP, we interpret the ERT model's predictions from both a global perspective and at the individual sample level. The success of this work promotes the development of interpretable data-driven models for capacity estimation.

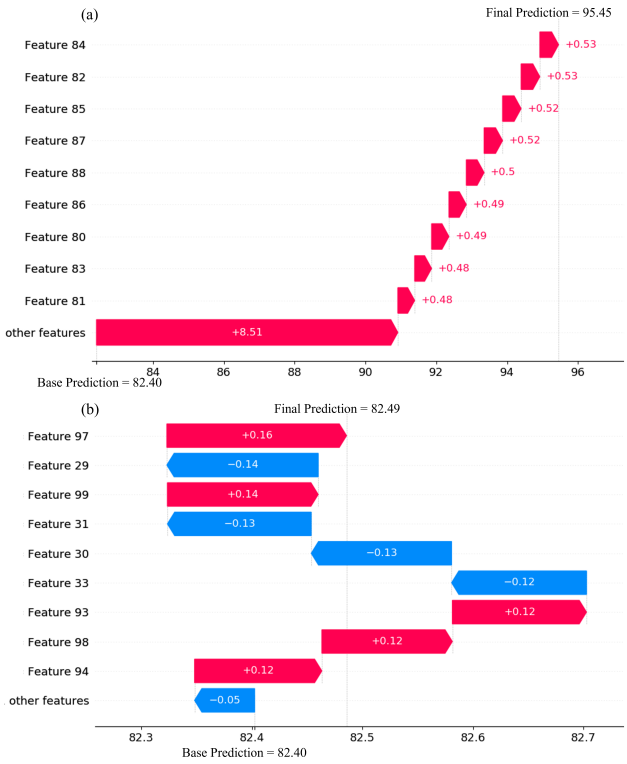

Fig. 7. Interpretation of the ERT model for 2 samples with observed capacities of 95.45% (a) and 82.38% (b)

ACKNOWLEDGEMENTS

Yankai Cao acknowledges funding from New Frontiers in Research Fund under grant NFRFE-2022-00663 and the NSERC Discovery Program under grant RGPIN-2019-05499. Bhushan Gopaluni would like to acknowledge the financial support from NSERC Discovery grant. We gratefully acknowledge the computing resources provided by SciNet (www.scinethpc.ca) and Digital Research Alliance of Canada (alliancecan.ca).

REFERENCES

Agarwal, N. and Das, S. (2020). Interpretable machine learning tools: A survey. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1528–1534. IEEE.

Breiman, L. (1984). *Classification and Regression Trees*. Routledge.

Breiman, L. (2001). Random forests. *Machine learning*, 45, 5–32.

Bresser, D., Hosoi, K., Howell, D., Li, H., Zeisel, H., Amine, K., and Passerini, S. (2018). Perspectives of automotive battery r&d in china, germany, japan, and the usa. *Journal of Power Sources*, 382, 176–178.

Chen, K., Zheng, F., Jiang, J., Zhang, W., Jiang, Y., and Chen, K. (2017). Practical failure recognition model of lithium-ion batteries based on partial charging process. *Energy*, 138, 1199–1208.

Doyle, M., Fuller, T.F., and Newman, J. (1993). Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell. *Journal of the Electrochemical society*, 140(6), 1526.

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63, 3–42.

Han, X., Lu, L., Zheng, Y., Feng, X., Li, Z., Li, J., and Ouyang, M. (2019). A review on the key issues of the lithium ion battery degradation among the whole life cycle. *ETransportation*, 1, 100005.

Hu, X., Li, S., and Peng, H. (2012). A comparative study of equivalent circuit models for li-ion batteries. *Journal of Power Sources*, 198, 359–367.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.

Kotsiantis, S.B. (2013). Decision trees: a recent overview. *Artificial Intelligence Review*, 39, 261–283.

Li, Y., Zou, C., Berecibar, M., Nanini-Maury, E., Chan, J.C.W., Van den Bossche, P., Van Mierlo, J., and Omar, N. (2018). Random forest regression for online capacity estimation of lithium-ion batteries. *Applied energy*, 232, 197–210.

Lundberg, S.M. and Lee, S.I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

Murdoch, W.J., Singh, C., Kumbier, K., Abbasi-Asl, R., and Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences*, 116(44), 22071–22080.

Plett, G.L. (2004). Extended kalman filtering for battery management systems of lipb-based hev battery packs: Part 3. state and parameter estimation. *Journal of Power sources*, 134(2), 277–292.

Roman, D., Saxena, S., Robu, V., Pecht, M., and Flynn, D. (2021). Machine learning pipeline for battery state-of-health estimation. *Nature Machine Intelligence*, 3(5), 447–456.

Schwunk, S., Armbruster, N., Straub, S., Kehl, J., and Vetter, M. (2013). Particle filter for state of charge and state of health estimation for lithium–iron phosphate batteries. *Journal of Power Sources*, 239, 705–710.

Shapley, L.S. (1953). *A Value for n-Person Games*, volume 28 of *Contributions to the Theory of Games (Volume II)*, 307–317. Princeton University Press.

Wang, Y., Zhu, J., Cao, L., Gopaluni, B., and Cao, Y. (2022). Online capacity estimation of lithium-ion batteries by partial incremental capacity curve. In *2022 IEEE Vehicle Power and Propulsion Conference (VPPC)*, 1–6. IEEE.

Wang, Z., Zeng, S., Guo, J., and Qin, T. (2018). Remaining capacity estimation of lithium-ion batteries based on the constant voltage charging profile. *PloS one*, 13(7), e0200169.

Yang, D., Zhang, X., Pan, R., Wang, Y., and Chen, Z. (2018). A novel gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve. *Journal of Power Sources*, 384, 387–395.

Zhu, J., Wang, Y., Huang, Y., Bhushan Gopaluni, R., Cao, Y., Heere, M., Mühlbauer, M.J., Mereacre, L., Dai, H., Liu, X., et al. (2022). Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation. *Nature communications*, 13(1), 2261.