



Data-Based Interpretable Modeling for Property Forecasting and Sensitivity Analysis of Li-ion Battery Electrode

Kailong Liu¹ · Qiao Peng² · Kang Li³ · Tao Chen⁴

Received: 29 June 2021 / Accepted: 22 November 2021 / Published online: 29 January 2022
© The Author(s) 2022

Abstract

Lithium-ion batteries have become one of the most promising technologies for speeding up clean automotive applications, where electrode plays a pivotal role in determining battery performance. Due to the strongly-coupled and highly complex processes to produce battery electrode, it is imperative to develop an effective solution that can predict the properties of battery electrode and perform reliable sensitivity analysis on the key features and parameters during the production process. This paper proposes a novel tree boosting model-based framework to analyze and predict how the battery electrode properties vary with respect to parameters during the early production stage. Three data-based interpretable models including AdaBoost, LPBoost, and TotalBoost are presented and compared. Four key parameters including three slurry feature variables and one coating process parameter are analyzed to quantify their effects on both mass loading and porosity of battery electrode. The results demonstrate that the proposed tree model-based framework is capable of providing efficient quantitative analysis on the importance and correlation of the related parameters and producing satisfying early-stage prediction of battery electrode properties. These can benefit a deep understanding of battery electrodes and facilitate to optimizing battery electrode design for automotive applications.

Keywords Li-ion batteries · Battery electrode · Battery modeling · Data analysis

Abbreviations

| | |
|----------|-----------------------------------|
| AdaBoost | Adaptive boosting |
| AUC | Area under curve |
| CG | Comma gap |
| CRISP | Cross-industry standard process |
| DT | Decision tree |
| EVs/HEVs | Electric/hybrid electric vehicles |
| KNN | K-nearest neighbors |
| KPI | Key performance indicator |
| Li-ion | Lithium-ion |
| LPBoost | Linear programming boosting |
| MC | Mass content |

| | |
|------------|------------------------|
| StLR | Solid-to-liquid ratio |
| SVM | Support vector machine |
| TotalBoost | Total boosting |

1 Introduction

Global challenges including dramatic climate change and depleting fossil fuel reserves have spurred the acceleration of sustainable transportation technologies. Due to the advantages in terms of high energy density and low self-discharging rates, lithium-ion (Li-ion) batteries have become one of the most promising energy storage devices in automotive applications such as electric vehicles (EVs) and hybrid electric vehicles (HEVs) [1]. However, the performance of Li-ion batteries such as capacity, service life, and energy and power densities are heavily dependent on their electrode properties, which are largely determined by the related production stages. To optimize battery performance, it is vital to understand the correlation between production parameters and battery electrode property variables [2].

Unfortunately, battery electrodes are produced under many intermediate substages with numerous

Academic editor: Lei Zhang

✉ Kailong Liu
kliu02@qub.ac.uk; Kailong.Liu@warwick.ac.uk

¹ WMG, The University of Warwick, Coventry, UK

² Queen's University Belfast, Belfast, UK

³ School of Electronic and Electric Engineering, University of Leeds, Leeds, UK

⁴ School of Electrical Engineering, Southeast University, Nanjing, China

strongly-coupled variables or parameters [3]. As multi-disciplinary knowledge including electrical, mechanical, and chemical information is often involved in producing battery electrode, the analysis on the importance and correlation between the intermediate process parameters and the battery electrode variables is challenging and usually reliant on engineering expertise and the trial-and-error method [2, 4]. These conventional methods are extremely laborious and time-consuming, and even lead to deteriorated quality control and difficulty in product rating at earlier stages [3]. In this context, it is imperative to devise an effective data-based strategy to perform reliable sensitivity analysis on the aforementioned aspects.

With the rapid development of machine learning and intelligent computing techniques, data-based modeling solutions have become one of many popular tools for the final battery production management [5, 6]. A great deal of research has been conducted for internal battery states estimation [7, 8], battery aging prognostics [9, 10], battery fault diagnostics [11], cell equalization management [12], charging control [13, 14], and energy management [15, 16]. Usually, effective battery management can be obtained based on suitable data-based models development [17]. However, these works mainly focus on the macro performance improvement of batteries while little has been done on enhancing their related production process considering microscopic properties of battery electrodes [18]. As battery electrodes play a vital role in determining the final performance of battery cells, it is also worthwhile to design effective data-based models for efficient analysis and prediction of battery electrode properties.

In comparison with the in-situ operation management of batteries with fruitful data-based modeling strategies, fewer existing works have been done on deriving data-based models to benefit battery production management. Among limited research regarding the monitoring [19] and adjustment [20] of battery production, a suitable data-based model to analyze and forecast battery intermediate properties is especially crucial. For example, on the basis of cross-industry standard process (CRISP), the data-based models using both linear regression and neural network models are proposed in Ref. [21] to predict battery properties and to identify the dependency of its production chain. By using a recursive feature elimination method, a data-based model is derived in Ref. [22] to identify improvement potentials and to analyze the relevant elements within the drying stage of battery production. After designing a statistical data-based solution to analyze the fluctuations of battery production, their effects on battery capacity are evaluated in Ref. [23]. The data-driven methodologies from the most existing research are mainly utilized to perform property prediction during battery production, while little has been done on designing data-based solutions, especially using interpretable framework,

for in-depth analysis of the effects of process parameters on battery properties at key production stages [24, 25]. It should be known that the parameters from early production stages such as mixing and coating are crucial to determine the properties of battery electrodes [26]. To achieve smarter battery production management and optimize battery electrode properties, it is vital to perform efficient sensitivity analyses of battery electrode properties with respect to its mixing and coating specifications.

Given the aforementioned consideration, a boosting-tree-based interpretable machine learning framework is proposed for the sensitivity analysis of key production process parameters and the prediction of battery electrode porosity qualities. This paper focuses on the effects of mixing and coating key parameters on final battery electrode properties. The key works done in this paper are summarized as follows: (1) to quantify the importance and correlation of four key parameters from both mixing and coating via a well-designed interpretable machine learning framework; (2) to classify and predict battery electrode porosity at early production stages via effective data-based models; (3) to evaluate and compare performance of typical AdaBoost model and two other improved tree boosting-based models (LPBoost and TotalBoost) for battery electrode property classification. All these developments can help battery engineers to effectively produce high-performance batteries, further benefiting smart battery production management for automotive applications.

2 Battery Electrode Properties and Data Curation

In this section, the fundamentals of battery electrode production and two key electrode properties are introduced, followed by the description of battery electrode data curation.

2.1 Fundamentals of Battery Electrode Production

As a key element of Li-ion batteries, electrode generally consists of anode and cathode types. To produce high-quality battery electrodes, a highly complicated production process is generally required [27]. As shown in Fig. 1, after preparing electrode component formulation such as the electrode additive, active material, and polymeric binder, a mixing process is deployed to mix these components based on the formulation within a mixer such as soft blender to generate homogenous slurries. Then, a coating process is performed to coat slurries onto the surface of the collector foils. In general, copper foil is used for anode electrodes, while aluminum foil is often adopted for cathode electrodes. Then the intermediate product from coating stage would be dried by an oven with a predefined temperature and calendered by

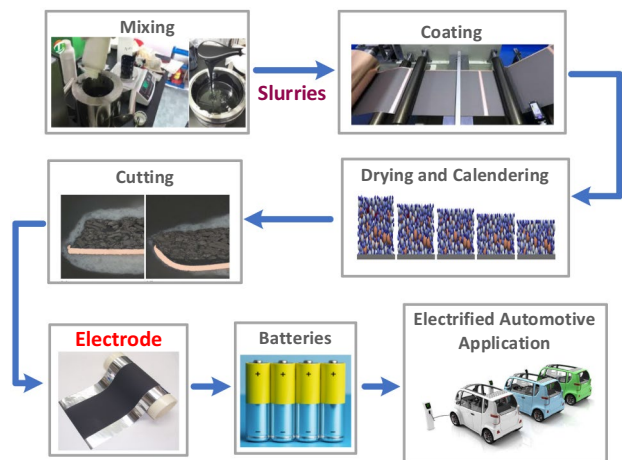


Fig. 1 Typical processes of battery electrode production

applying mechanical pressure through two cylindrical rolls. Finally, a cutting process is performed to cut electrodes into proper shapes for coin or pouch battery cells. It should be noted that the whole production process of battery electrodes involves electrical, mechanical, and chemical operations. Specific equipment are usually required at all intermediate production stages [26].

As two key processes to produce battery electrodes, mixing and coating are complicated with many strongly-coupled parameters, and these process parameters directly determine the properties of battery electrodes, further affecting the relevant battery performance. In this context, a reliable solution is desired to forecast properties of battery electrode as well as analyze the sensitivity of related process parameters on the electrode properties at the early production stage and improve battery product performance. To achieve this, a tree boosting model-based interpretable machine learning framework is designed to forecast two properties of battery electrode, while both importance as well as correlations of some key process parameters are also analyzed and quantified. Specifically, two electrode property Key Performance Indicators (KPIs) including the electrode mass loading with unit of mg/cm^2 and porosity with unit of % are adopted. Mass loading is directly related to the final battery capacity while porosity is crucial for dealing with highly porous electrodes. The effects of three mixing parameters including the mass content (MC) of active materials, solid-to-liquid ratio (StLR), and viscosity (Vis), as well as one coating parameter namely the comma gap (CG) on these two electrode properties are investigated. Here StLR represents the mass ratio between slurry solid and mass, while Vis affects the shear rate of coating stage, and CG is the gap between coating comma and coating roll. As the electrode mass loading is directly related to the final battery capacity and measured by a high-precision scale for weight while electrode porosity

would deal with highly porous electrodes and be affected by other properties such as coating thickness and coating density, both of these two properties are worthy of exploration.

2.2 Battery Electrode Data Curation

Data curation is a key step for modeling activities. Without the loss of generality, the battery electrode experimental dataset from Franco Laboratoire-de-Reactivite-et-Chimie-des-Solides is adopted. Detailed experimental information and data explanation can be referred to Ref. [28]. The original data set has a size of 656 samples in total. For each sample, three slurry parameters after mixing and one coating parameter are used to generate one related mass loading and porosity observation of battery electrode. To fully investigate the prediction/classification performance of designed tree boosting-based models, both battery electrode mass loading and porosity are labeled with five classes (very low, low, medium, high, and very high). Specifically, for the electrode mass loading, very low and low represent the ranges of (0, 15] and (15, 25], respectively. Medium stands for the range of (25, 35], while high and very high refer to the ranges of (35, 45] and (45, 60], respectively. For the electrode porosity, very low refers to the range of (0, 47.5], low refers to the range of (47.5, 50], medium reflects the range of (50, 52.5], while high and very high refer to the ranges of (52.5, 55], and (55, 70], respectively. After setting battery electrode mass loading and porosity with predefined class labels, the tree boosting model-based interpretable machine learning framework for prediction/classification of both electrode properties as well as sensitivity analyses of process parameters can be designed.

3 Ensemble Learning Technologies

In this section, the fundamentals of AdaBoost are first introduced, followed by the descriptions of another two improved boosting techniques including LPBoost and TotalBoost. Then the framework of using tree boosting-based model to predict battery electrode properties and analyze related process parameters of interest is designed. Some indicators are also given to evaluate the performance of established models.

3.1 AdaBoost, LPBoost, and TotalBoost

Boosting is one of the most popular solutions to derive ensembled tree-based models. The key idea of boosting is to sequentially train different weak hypothesis, while the training dataset's distribution would be also changed dynamically based on the performance of previously trained weak hypothesis.

Adaptive boosting (AdaBoost) is a typical and effective boosting solution for prediction applications [29]. Let training dataset TD includes M observations as: $TD = \{(x_1, y_1), (x_2, y_2), \dots, (x_M, y_M)\}$. Here x_m ($m = 1, 2, \dots, M$) reflects the input vector of corresponding battery process parameters, y_m ($m = 1, 2, \dots, M$) stands for the preset classification labels with a total number of C , $L(x)$ stands for a weak hypothesis that would provide an output result related to x , the detailed process to establish AdaBoost-based classification model is summarized in Workflow 1.

Workflow 1: Process to establish AdaBoost-based model

1. Initialize the training observations' weights as:

$$w_m = 1/M, m = 1, 2, \dots, M.$$

2. Suppose D is the number of all weak learners $L^{(d)}$ (x). For $d = 1$ to D :

- a) Fit $L^{(d)}(x)$ to TD based on the weight w_m .
- b) Compute the error $er^{(d)}$:

$$er^{(d)} = \sum_{m=1}^M w_m \cdot I(L^{(d)}(x_m) \neq y_m) / \sum_{m=1}^M w_m$$

where $I(\cdot)$ means a zero-one judgement with a rule:

$$I(L^{(d)}(x_m) \neq y_m) = 1$$

- c) Compute the update factor $f^{(d)}$ of weights: $f^{(d)} = \log[(1 - er^{(d)})/er^{(d)}] + \log(K - 1)$
- d) For $m = 1, 2, \dots, M$, update weight w_m :

$$w_m \leftarrow w_m \cdot \exp[f^{(d)} \cdot I(L^{(d)}(x_m) \neq y_m)]$$

e) Renormalize w_m .

3. Output the predicted class result $\tilde{y}(x)$ as:

$$\tilde{y}(x) = \arg \max_c \sum_{d=1}^D f^{(d)} \cdot I(L^{(d)}(x_m) = y_c)$$

where y_c is the preset classification label ($c = 1, 2, \dots, C$). $I(\cdot)$ represents another zero-one judgement with a rule of $I(L^{(d)}(x_m) = y_c) = 1$, $\arg \max_c$ would output the class that has the largest counted number from results of all trained weak hypotheses.

Based upon AdaBoost, another two effective boosting-based strategies including the linear programming boosting (LPBoost) and total boosting (TotalBoost) are also explored. It should be noted that both LPBoost and TotalBoost have the similar establishment workflow as AdaBoost, but these two improved solutions would be self-terminating and produce ensembles with small weights. More information regarding the process to establish LPBoost-based model and TotalBoost-based model are described in Workflow 2 and

Workflow 3, respectively. To be specific, LPBoost adopts the weighted linear combination of weak hypothesis, so that a weak hypothesis can be added in each iteration with the adjustment of previous weak hypothesis' weights [30]. TotalBoost realizes the classification through maximizing the minimal margin [31]. More details of these two boosting solutions can be found in Ref. [32].

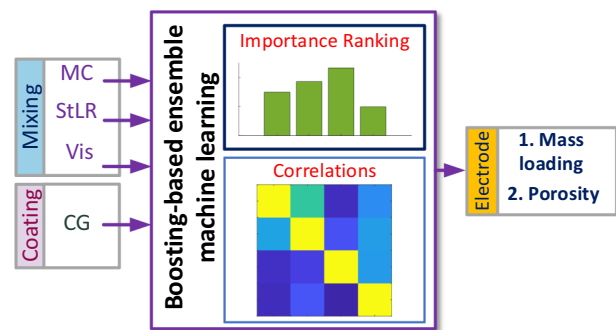


Fig. 2 Detailed structure of tree boosting-based model for battery electrode properties prediction and parameter analyses

Workflow 2: Process to establish LPBoost-based model

1. Suppose the accuracy parameter $\epsilon > 0$, the initial weight distribution wd^1 is a uniform distribution, parameter $\hat{r}_1 = 1$.

2. The main processes are similar as AdaBoost, but for $d = 1$ to D , the weights of previous weak hypothesis would be adjusted:

- a) Obtain hypothesis h^d and set a vector u^d with the terms $u_m^d = h^d(x_m)y_m$ ($m = 1, 2, \dots, M$);
- b) Set $\hat{r}_{d+1} = \min(\hat{r}_d, wd^d \cdot u^d)$;
- c) Update the distributions to wd^d by solving a LP problem:

$$[wd^{d+1}, r_d^*] \in \frac{\min}{wd, r} r$$

$$\text{s.t. } wd \cdot u^t \leq r, \text{ for } 1 \leq t \leq d;$$

d) If $r_d^* \geq \hat{r}_{d+1} - \epsilon$, then set $DD = d$ and terminate.

e) Weight w_m realizes the margin r_{DD}^* .

Workflow 3: Process to establish TotalBoost-based model

1. Suppose the accuracy parameter $\epsilon > 0$, the initial weight distribution w^d is an uniform distribution, parameter $\hat{r}_1 = 1$.
2. The main processes are similar as AdaBoost, but for $d = 1$ to D , the weight distribution of the previous weak hypothesis is adjusted as follows:
 - a) Obtain hypothesis h^d and set a vector \mathbf{u}^d with the terms $u_m^d = h^d(x_m)y_m$ ($m = 1, 2, \dots, M$);
 - b) Set $\hat{r}_{d+1} = \min(\hat{r}_d, w^d \cdot \mathbf{u}^d)$;
 - c) Update the distributions to w^d by solving a LP problem:

$$w^{d+1} = \arg \max_{w^d} \Delta(w^d, w^d)$$
 s.t. $w^d \cdot \mathbf{u}^t \leq \hat{r}_{d+1} - \epsilon$, for $1 \leq t \leq d$;
 - d) If the above solution becomes infeasible or w^{d+1} has a zero, then set $DD = d$ and terminate;
 - e) Weight w_m realizes the hard margin r_{DD}^* over the hypothesis.

3.2 Framework of Designing Tree Boosting-Based Model for Analyzing Battery Electrode Properties

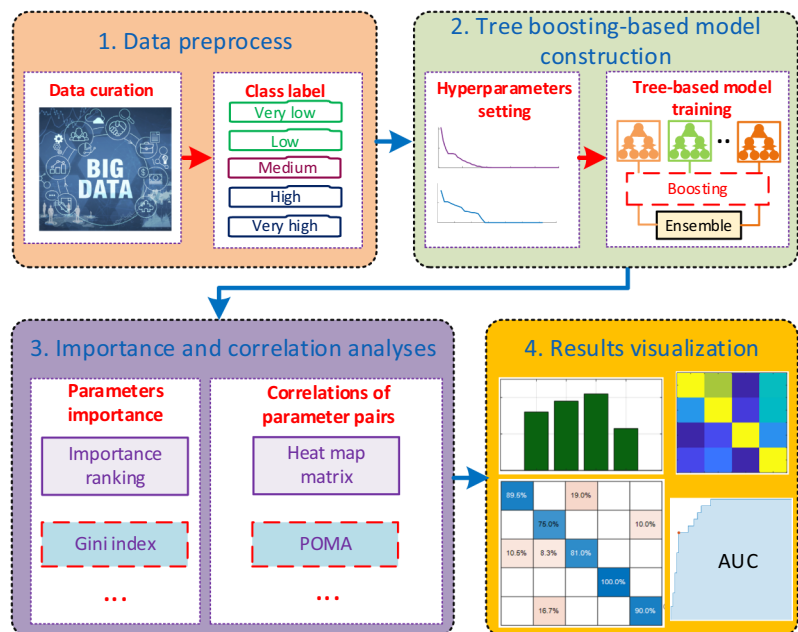
To well predict the properties of battery electrode and effectively analyze the sensitivity of mixing as well as coating related parameters, a novel interpretable machine learning framework with a tree boosting-based model structure as

shown in Fig. 2 is designed. To be specific, three mixing parameters (MC, StLR, and Vis) and one coating parameter (CG) are utilized as the inputs to the model, while the relevant manufactured battery electrode properties including electrode mass loading and porosity are used as outputs. As illustrated in Fig. 3, the detailed framework of using tree boosting-based technique to carry out the sensitivity analysis and forecast/classify the qualities of relevant electrode mass loading as well as porosity can be summarized with four key parts as follows:

Part 1. Data preprocess The raw data related to battery electrode production will be first preprocessed to remove outliers and add the missing values. Through setting the thresholds of manufactured electrode mass loading and porosity as [5 mg/cm², 60 mg/cm²] and [30%, 70%], respectively, the outliers outside these thresholds would be removed, further providing a robust result under uncertainties during battery production. Afterward, the output observations will be set with the relevant classification labels. According to the predefined rules in Sect. 2.2, five labels including very low, low, medium, high, and very high are set to reflect the qualities of both battery electrode mass loading and porosity.

Part 2. Tree boosting-based model construction To establish effective tree boosting-based model, the hyperparameters of AdaBoost, LPBoost, and TotalBoost-based models need to be determined. It should be known that for all these three boosting solutions, decision tree is usually selected as their weak hypothesis. Two main hyper-parameters require to be preset: the number of ensembled decision tree (N) and their learning rates (r). In theory, large N leads to the improved prediction accuracy, but too many decision

Fig. 3 Framework of designing tree boosting-based model to predict battery electrode properties and perform sensitivity analyses of parameters



trees will also cause overfit and increase the computational effort. In order to determine a suitable N , an iteration strategy through comparing learner weights via various numbers of utilized weak hypothesis is adopted. For learning rate r , it reflects a decay rate of each hypothesis's weight, further affecting the performance of each decision tree. As suggested by Ref. [33], r could be set as 0.1 for the general prediction/classification applications. After these two hyper-parameters are set, all these tree boosting-based models can be well-trained with the process as illustrated in their related workflow.

Part 3. Importance and correlation analyses To quantify how important the battery production process parameters would affect electrode properties, Gini index that represents the variation of impurity due to the splits of each parameter is utilized. For the tree-based classification, in theory, impurity could stand for how good a potential split is for decision tree's nodes [33]. The larger Gini index value a process parameter can obtain, the greater effect that this parameter could impose on battery electrode. Besides, to carry out the correlation analysis of each battery production parameter pair, predictive-measure-of-association (PMOA) is utilized. Supposing two parameters of interest are p_i and p_j , the PMOA value to reflect their correlation is calculated by:

$$PMOA_{ij} = \frac{\min(OP_l, OP_r) - 1 + OPl_{ij} + OPr_{ij}}{\min(OP_l, OP_r)} \quad (1)$$

where l and r represent the left child and right child of nodes; OP_l and OP_r mean the observation proportions of $p_i < y$ and $p_i \geq y$, respectively; OPl_{ij} is the observation proportion under the condition of $p_i < y$ and $p_j < z$, while OPr_{ij} reflects the observation proportion under the condition of $p_i \geq y$ and $p_j \geq z$. The process of obtaining PMOA is to investigate all the potential splits with the best case that is obtained during decision tree's training stage. In this context, PMOA has the ability to quantify the similarity between different rules for splitting observation.

Part 4. Results visualization After obtaining the Gini index for each parameter of interest, the relevant feature importance ranking can be generated. After using Eq. (1), the PMOA values of all parameter pairs can be obtained and shown as a 4×4 heat map. Both of these visualization results could give information for users to directly understand the importance and correlations of parameters that could affect battery electrode properties.

3.3 Performance Indicators

To further quantify and investigate the prediction/classification performance, confusion matrix (CM) is adopted as a key performance indicator. Supposing positive stands for an interested class while negative relates to other classes, four

basic elements including the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) could be obtained. Afterwards, the positive predictive value $PPV(C_i)$ as well as false discovery rate $FDR(C_i)$ of interested class can be obtained:

$$\begin{cases} PPV(C_i) = TP(C_i) / [TP(C_i) + FP(C_i)] \\ FDR(C_i) = FP(C_i) / [TP(C_i) + FP(C_i)] \end{cases} \quad (2)$$

Then a popular performance indicator for the accuracy of classification results namely micro F1 score ($microF1$) can be calculated as:

$$microF1 = (TP_{all} + TN_{all}) / N_{total} \quad (3)$$

where TP_{all} and TN_{all} represent all correct classifications, and N_{total} means the total amount of observations.

Besides, the receiver operating characteristic (ROC) curve as well as its area under curve (AUC) value are adopted to explore the results of electrode properties prediction/classification. It should be known that ROC curve is a statistical plot to reflect the diagnostic ability of a classification model under the case of varying its discrimination threshold [34]. The AUC could give the degree or measure of separability of classes.

4 Results and Discussions

In this section, on completion of all steps within the parts from Sect. 3.2, the forecast and sensitivity analysis tests by designing proper tree boosting-based models are carried out to quantify both importance and correlations of four input battery production parameters of interest, while the battery electrode mass loading as well as porosity qualities will be also classified. Without the loss of generality, all designed tree boosting-based models (AdaBoost, TotalBoost, and LPBoost) are evaluated through using the five-fold cross-validation. It should be known that for different folds of cross-validation, these quantified importance and correlation values would present the same trends without large difference.

4.1 Case Study 1: Li-ion Battery Electrode Mass Loading

4.1.1 Model Training

The first case study focuses on the prediction of battery electrode mass loading and the relevant sensitivity analysis of four process parameters. To evaluate if the training process of all three tree boosting-based models could converge as well as to avoid overfitting of their training process, a test of using tree stumps with only 1 maximum split as weak

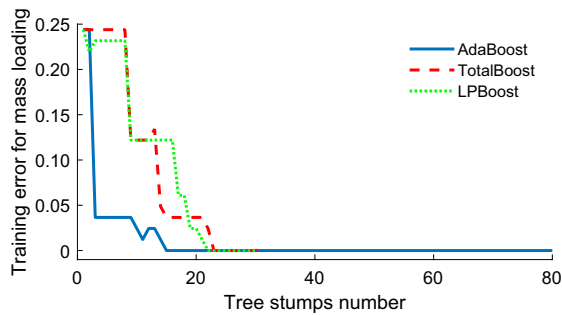


Fig. 4 Training error via tree stumps number for electrode mass loading case

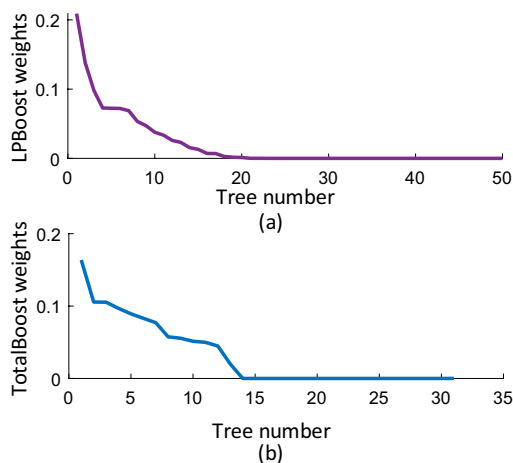


Fig. 5 Weights of LPBoost and TotalBoost via ensemble tree number for electrode mass loading case

hypothesis is carried out. Figure 4 illustrates the related training error via the increased number of tree stumps. It is evident that the training errors of all these three tree boosting-based models could converge to 0 after using 23 tree stumps, indicating that reasonable convergence results can be achieved for the battery electrode mass loading case. Then, the number of ensemble decision tree (N) for AdaBoost is set as a default value 50. As the weights of both LPBoost and TotalBoost cases are decreased via the number of ensemble weak hypothesis, to determine N of them, their hypothesis weights after compacting the corresponding weak hypothesis (decision tree) are illustrated in Fig. 5. Obviously, both LPBoost and TotalBoost-based models present clear decrease trajectories as the number of ensemble decision trees increases. Here, the weights of LPBoost and TotalBoost become negligible after using 19 and 14 decision trees, respectively, indicating that a satisfactory convergence of model training can be achieved. In this context, the N of LPBoost model and TotalBoost model are

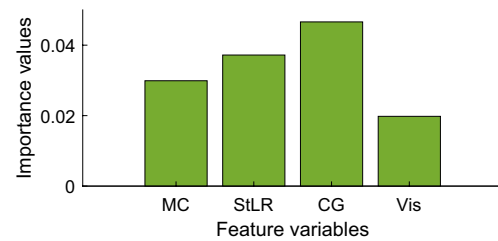


Fig. 6 Importance ranking of battery manufacturing parameters for electrode mass loading case

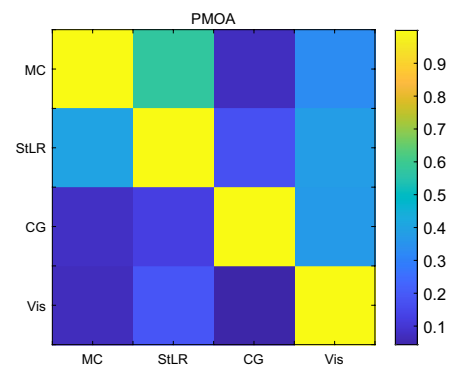


Fig. 7 PMOA-based correlations of interested parameter pairs for electrode mass loading case

set as 19 and 14 for the case study of battery electrode mass loading, respectively.

4.1.2 Parameter Sensitivity Analysis

After setting the hyper-parameters of all tree boosting-based models, the sensitivity analyses of how the process parameters of interest affect battery electrode mass loading can be conducted. After calculating the Gini index values of MC, StLR, CG, and Vis, their importance ranking is quantified, as illustrated in Fig. 6. It is shown that CG and Vis achieve the highest and lowest importance ranking among four interested parameters, indicating that CG and Vis play the most and least effects on determining the mass loading property of electrode, respectively. This conclusion is reasonable as the coating weight as well as thickness that directly determine battery electrode mass loading are significantly affected by CG in theory.

To quantify the correlations of each process parameter pair for electrode mass loading case, the PMOAs of all pairs derived from four parameters (MC, StLR, CG, and Vis) are calculated and visualized with a heat map matrix, as shown in Fig. 7. Quantitatively, the largest value of PMOA is achieved for the pair of MC and StLR, but this value is just

Table 1 Performance indicator of all these three tree boosting-based models for electrode mass loading case

| Tree boosting-based models | <i>microF1</i> values (%) |
|----------------------------|---------------------------|
| AdaBoost | 89.7 |
| LPBoost | 90.2 |
| TotalBoost | 92.7 |

around 0.7. Besides, most other PMOA values are between 0.1 and 0.6. In light of this, no strong correlations of process parameter pairs exist for electrode mass loading case.

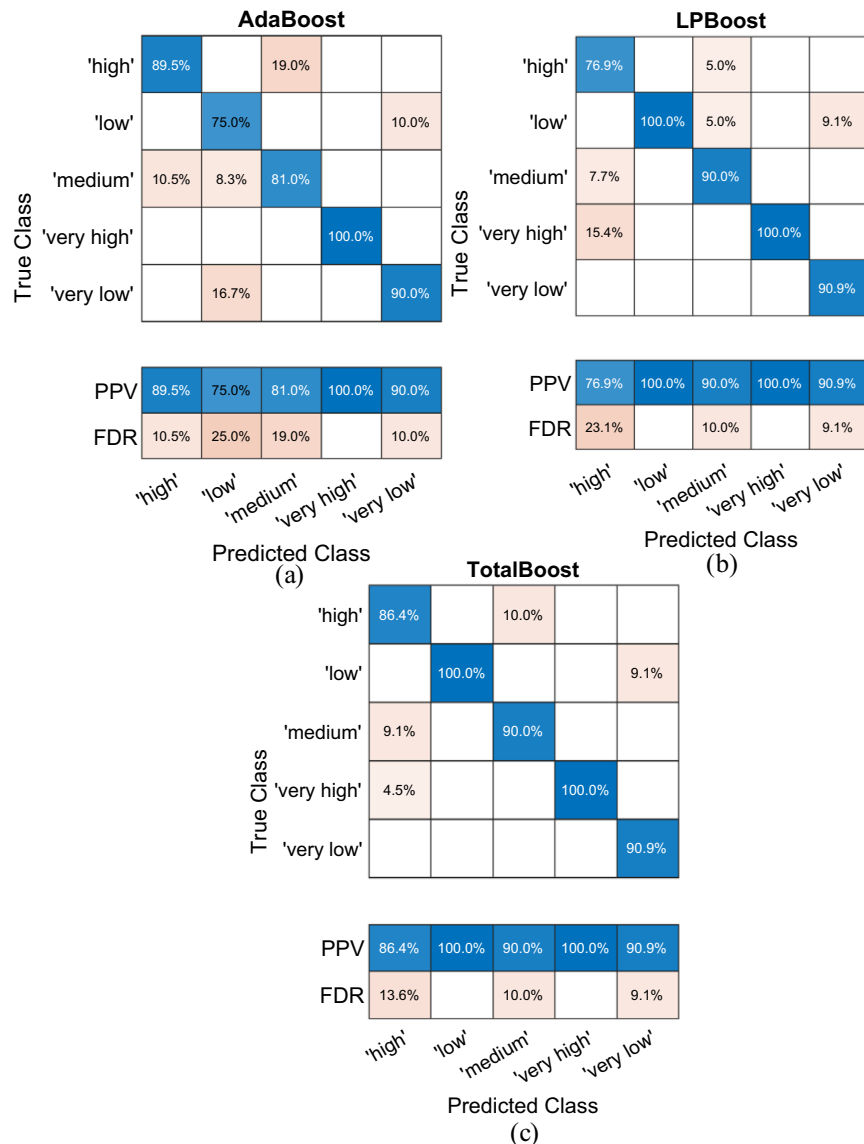
4.1.3 Classification Performance Evaluation

After well establishing the tree boosting-based models, the *microF1* values of battery electrode mass loading prediction for AdaBoost model, LPBoost model, as well as TotalBoost model cases are illustrated in Table 1. The corresponding

confusion matrix and ROC curves are also shown in Figs. 8 and 9, respectively.

It is shown that battery electrode mass loading can be well predicted through using these four parameters at the early prediction stage as the *microF1* values for all models are larger than 89%. Quantitatively, TotalBoost-based model presents the best classification result with a *microF1* value of 92.7%, which is 2.8% and 3.3% larger than that of LPBoost and AdaBoost cases, respectively. From Fig. 9, the AUC value of TotalBoost case is the highest one with 0.97, which is 4.3% and 3.2% larger than that of AdaBoost and LPBoost cases, respectively. In light of this, the designed TotalBoost model presents the most competent performance for classification/prediction of battery electrode mass loading case.

Fig. 8 Confusion matrix of battery electrode mass loading classification under different tree-based models



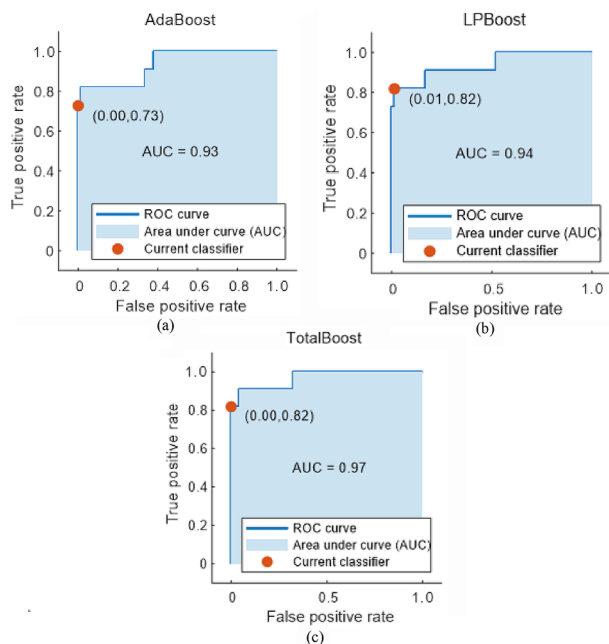


Fig. 9 ROC curves of battery electrode mass loading classification under different tree-based models

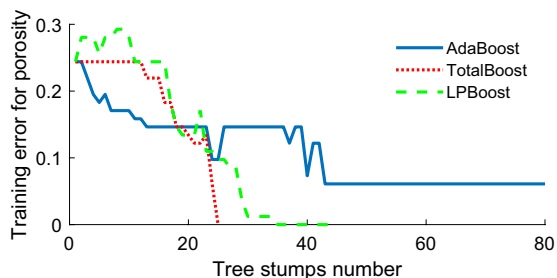


Fig. 10 Training error via tree stumps number for electrode porosity case

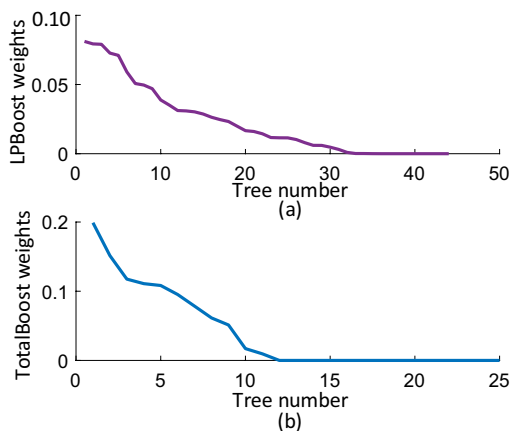


Fig. 11 Weights of LPBoost and TotalBoost via ensemble tree number for electrode porosity case

4.2 Case Study 2: Li-ion Battery Electrode Porosity

4.2.1 Model Training

The second case study focuses on the Li-ion battery electrode porosity. As illustrated in Fig. 10, the training errors of both TotalBoost and LPBoost cases converge to 0 after the number of tree stumps becomes larger than 33, while the error of AdaBoost case could also converge to 0.06 after using 42 tree stumps. Therefore, all these three tree boosting-based models are capable of achieving reliable convergence results for battery electrode porosity case. Without the loss of generality, the hyper-parameter N of AdaBoost is also set as 50. To further determine N of LPBoost and TotalBoost cases, their learner weights after compacting weak hypothesis (decision tree) are shown in Fig. 11. It is evident that the weight of LPBoost and TotalBoost would become negative after 32 and 12 decision trees are utilized, respectively. In light of this, N of LPBoost and TotalBoost cases is, respectively, set as 32 and 12 for battery electrode porosity prediction and analysis case.

4.2.2 Parameter Sensitivity Analysis

For the sensitivity analysis of battery electrode porosity, after calculating the Gini index values of MC, StLR, CG, and Vis, the corresponding importance ranking is visualized in Fig. 12. Quantitatively, the quantified importance values of StLR and Vis are 0.051 and 0.049, which are higher than that from other parameters. This implies that StLR and Vis are the two most important process parameters to determine battery electrode porosity. In contrast, MC shows the minimum Gini index with 0.032, which means that it provides the lowest effects on the classification/prediction of battery electrode porosity.

To quantify the correlations of each parameter pair for porosity case, the PMOAs of all pairs are calculated and visualized with a heat map matrix, as illustrated in Fig. 13. Obviously, the PMOA of MC and StLR pair gives the largest value around 0.9, indicating that there exists relatively strong correlations between these two process parameters. This result is expected as the mass ratio of slurry solid and mass present strong correlations with the electrode porosity in theory. For other parameter pairs, their PMOAs are all lower than 0.6, indicating that no strong correlations of these parameter pairs exist in determining the battery electrode porosity.

4.2.3 Classification Performance Evaluation

After using all three well-trained tree boosting-based models to forecast/classify battery electrode porosity, their relevant confusion matrices and *microF1* values are illustrated in

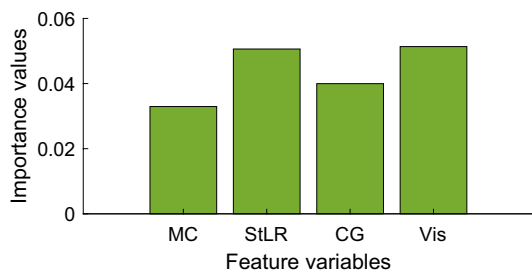


Fig. 12 Importance ranking of battery manufacturing parameters for electrode porosity case

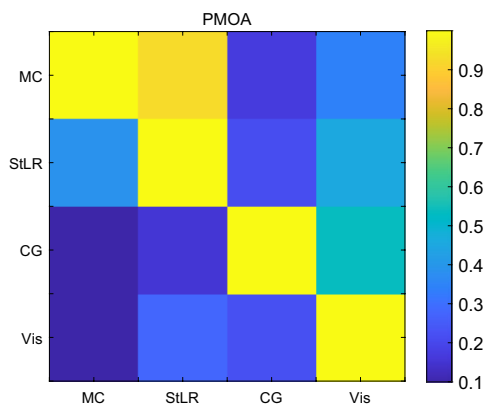


Fig. 13 PMOA-based correlations of parameter pairs for electrode porosity case

Fig. 14 and Table 2, respectively. Quantitatively, AdaBoost-based model presents the worst classification result with 71.2% *microF1*, while TotalBoost-based model achieves the best result of 74.1% *microF1*, which is 2.1% better than that of LPBoost. Figure 15 illustrates the ROC curves of battery electrode porosity classification results through various tree-based models. It can be noted that the AUC values of all models are higher than 0.9. Quantitatively, AUC of TotalBoost-based model presents the largest value with 0.94, which is 2.2% and 3.3% larger than that of LPBoost-based model and AdaBoost-based model, respectively. Therefore, the proposed tree boosting-based interpretable machine learning framework is able to forecast/classify the battery electrode porosity with satisfactory AUC values at the early mixing and coating stages, while TotalBoost-based model also shows the most competent performance among these three adopted boosting techniques. Besides, it can be noticed that the prediction performance of “high” label gets worse than others for both electrode mass loading and porosity cases. This is mainly because the natures of “high” and “very high” classes determined by battery manufacturer are

similar, while some observations with “high” label are classified into “very high” class.

4.3 Discussion of Performance and Benefits

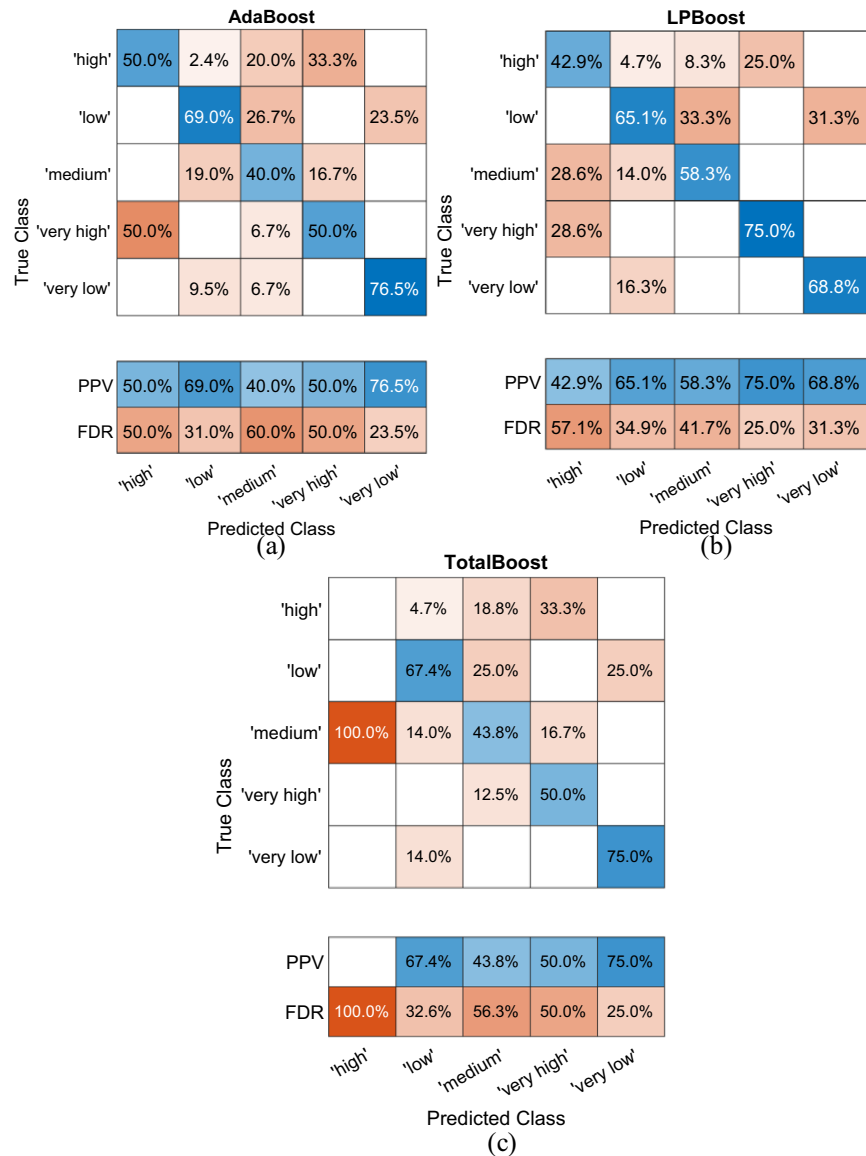
4.3.1 Comparisons with Other Approaches

To further evaluate the performance of designed boosting-tree-based models, another three widely utilized existing approaches including the decision tree (DT), k-nearest neighbors (KNN), and support vector machine (SVM) are compared here. Specifically, DT is a single tree. KNN is a typical instance learning-based approach relying on the instance for classification. SVM is a kernel-based classification method through mapping inputs to high-dimensional space [35]. Without the loss of generality, five-fold cross-validation is utilized for these comparisons, and their performance indicators for both electrode mass loading and porosity are illustrated in Table 3. It can be noticed that DT presents the worst results, while SVM gives the best results of 89.3% *microF1* and 68.3% *microF1* for electrode mass loading and porosity cases, respectively. However, all these results are worse than boosting-tree-based models. In this context, due to the benefits of boosting solution, the proposed framework provides competent predictions of battery electrode mass loading and porosity.

4.3.2 Further Discussions

According to the predictions of both electrode mass loading and porosity, TotalBoost-based model presents the best results while AdaBoost-based model is the worst one for both these two cases. This is mainly due to the fact that compared with AdaBoost, both LPBoost and TotalBoost have the abilities to automatically adjust weights. Here the TotalBoost is also able to realize the classification by maximizing the minimal margin, further generalizing well for battery electrode property predictions. The statistical metrics obtained from the designed tree boosting model-based framework presents extensive connection to the battery production management strategy. For example, as battery manufacturing line owns numerous strongly-coupled feature parameters, monitoring all these feature parameters all the time is inefficient and would cause large cost. The battery manufacturer can utilize the feature importance and correlation information from the proposed approach to readjust the monitored features and better understand dependency among these features. Moreover, according to the prediction results of electrode properties, battery manufacturer is able to reoptimize related parameters to improve the qualities of battery product at early manufacturing stage, further enhancing battery manufacturing efficiency.

Fig. 14 Confusion matrix of battery electrode porosity classification under different tree-based models



5 Conclusions

Electrode is a crucial part for determining the performance of batteries and related automotive applications. In this study, an effective tree boosting-based interpretable machine learning framework is designed for effective sensitivity analysis of process parameters and forecast/classify the battery electrode mass loading and porosity for the first time.

Some conclusions can be drawn as follows:

- (1) StLR is a key process parameter to determine both electrode mass loading and porosity. CG is important for electrode mass loading while Vis is crucial for electrode porosity case.

Table 2 Performance indicator of all these three tree boosting-based models for electrode porosity case

| Boosting-based models | microF1 values (%) |
|-----------------------|--------------------|
| AdaBoost | 71.2 |
| LPBoost | 72.6 |
| TotalBoost | 74.1 |

- (2) With the largest PMOA value, there is a relatively strong correlation between MC and StLR pair for both electrode mass loading and porosity cases.
- (3) Electrode mass loading can be well captured by using process parameters of MC, StLR, CG, and Vis, while more other process parameters should be considered to

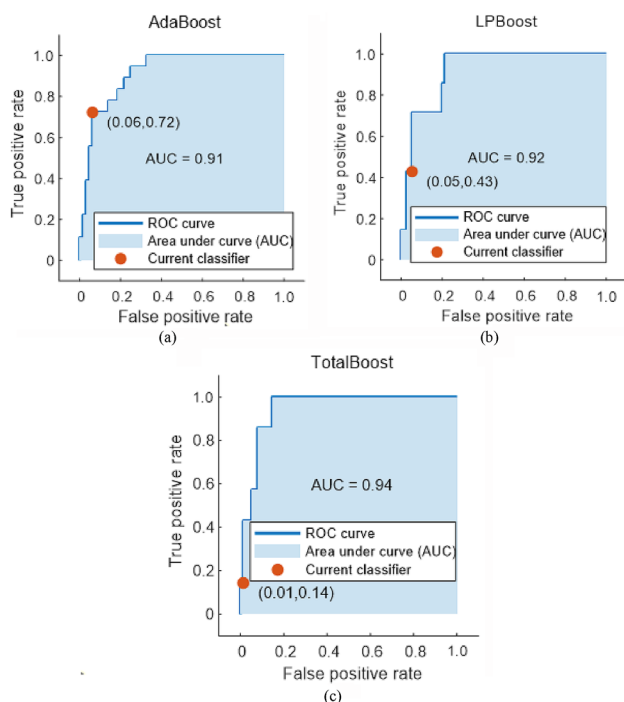


Fig. 15 ROC curves of battery electrode porosity classification under different tree-based models

improve the quality classification of battery electrode porosity.

Due to the advantage of data-driven nature and sensitivity analysis ability, the designed tree boosting model-based framework can be utilized to analyze more battery manufacturing parameters when related data are available, further benefitting the understanding of battery electrode properties and wider applications of battery-based automotive applications.

Acknowledgements This work was supported by the EPSRC under grant EP/R030243/1 and the High Value Manufacturing Catapult project under Grant No. 8248 CORE.

Declarations

Conflict of interest On behalf of all the authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

Table 3 Performance indicator of DT, KNN, and SVM for both electrode mass loading and porosity cases

| Existing approaches | <i>microF1</i> values (%) |
|---------------------|---------------------------|
| DT (mass loading) | 79.5 |
| KNN (mass loading) | 84.7 |
| SVM (mass loading) | 89.3 |
| DT (porosity) | 61.2 |
| KNN (porosity) | 65.4 |
| SVM (porosity) | 68.3 |

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Hu, J., Wei, Z., He, H.: An online adaptive internal short circuit detection method of lithium-ion battery. *Automot. Innov.* **4**(1), 93–102 (2021)
- Kwade, A., Haselrieder, W., Leithoff, R., Modlinger, A., Dietrich, F., Droeder, K.: Current status and challenges for automotive battery production technologies. *Nat. Energy* **3**(4), 290–300 (2018)
- Liu, Y., Zhang, R., Wang, J., Wang, Y.: Current and future lithium-ion battery manufacturing. *Science*. **2**, 102332 (2021)
- Lombardo, T., Duquesnoy, M., El-Bouysidy, H., et al.: Artificial intelligence applied to battery research: hype or Reality? *Chem. Rev.* **6**, 9381 (2021)
- Finegan, D.P., Cooper, S.J.: Battery safety: data-driven prediction of failure. *Joule* **3**(11), 2599–2601 (2019)
- Li, Y., Liu, K., Foley, A.M., Zülke, A., Berecibar, M., Maury, E.N., Mierlo, J.V., Hoster, H.E.: Data-driven health estimation and lifetime prediction of lithium-ion batteries: a review. *Renew. Sustain. Energy Rev.* **113**, 109254 (2019)
- Che, Y., Foley, A., El-Gindy, M., Lin, X., Hu, X., Pecht, M.: Joint estimation of inconsistency and state of health for series battery packs. *Automot. Innov.* **4**(1), 103–116 (2021)
- Li, G., Liu, C., Wang, E., Wang, L.: State of charge estimation for Lithium-Ion battery based on improved cubature Kalman filter algorithm. *Automot. Innovat.* **3**, 1–12 (2021)
- Liu, K., Shang, Y., Ouyang, Q., Widanage, W.D.: A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery. *IEEE Trans. Ind. Electron.* **68**(4), 3170–3180 (2020)
- Tang, X., Liu, K., Wang, X., Gao, F., Macro, J., Widanage, W.D.: Model migration neural network for predicting battery aging trajectories. *IEEE Trans. Transp. Electrification*. **6**(2), 363–374 (2020)
- Hu, X., Zhang, K., Liu, K., Lin, X., Dey, S., Onori, S.: Advanced fault diagnosis for Lithium-Ion battery systems: a review of fault mechanisms, fault features, and diagnosis procedures. *IEEE Ind. Electron. Mag.* **14**(3), 65–91 (2020)
- Ouyang, Q., Wang, Z., Liu, K., Xu, G., Li, Y.: Optimal charging control for lithium-ion battery packs: a distributed average tracking approach. *IEEE Trans. Ind. Inf.* **16**(5), 3430–3438 (2019)
- Liu, K., Zou, C., Li, K., Wik, T.: Charging pattern optimization for lithium-ion batteries with an electrothermal-aging model. *IEEE Trans. Ind. Inf.* **14**(12), 5463–5474 (2018)

14. Liu, K., Li, K., Zhang, C.: Constrained generalized predictive control of battery charging process based on a coupled thermoelectric model. *J. Power Sources* **347**, 145–158 (2017)
15. Wang, G., Gao, Q., Yan, Y., Wang, Y.: Thermal management optimization of a Lithium-Ion battery module with graphite sheet fins and liquid cold plates. *Automot. Innov.* **3**(4), 336–346 (2020)
16. Liu, K., Hu, X., Yang, Z., Xie, Y., Feng, S.: Lithium-ion battery charging management considering economic costs of electrical energy loss and battery degradation. *Energy Convers. Manag.* **195**, 167–179 (2019)
17. Hu, X., Feng, F., Liu, K., Zhang, L., Xie, J., Liu, B.: State estimation for advanced battery management: key challenges and future trends. *Renew. Sustain. Energy Rev.* **114**, 109334 (2019)
18. Duquesnoy, M., Boyano, I., Ganborena, L., Cereijo, P., Ayerbe, E., Franco, A.: Machine learning-based on assessment of the impact of the manufacturing process on battery electrode heterogeneity. *Energy AI* **2**, 100090 (2021)
19. Knoche, T., Surek, F., Reinhart, G.: A process model for the electrolyte filling of lithium-ion batteries. *Procedia CIRP* **41**, 405–410 (2016)
20. Schünemann, J.H., Dreger, H., Bockholt, H., Kwade, A.: Smart electrode processing for battery cost reduction. *ECS Trans.* **73**(1), 153–159 (2016)
21. Schnell, J., Nentwich, C., Endres, F., Kollenda, A., Distel, F., Knoche, T., Reinhart, G.: Data mining in lithium-ion battery cell production. *J. Power Sources* **413**, 360–366 (2019)
22. Thiede, S., Turetskyy, A., Loellhoeffel, T., Kwade, A., Kara, S., Herrmann, C.: Machine learning approach for systematic analysis of energy efficiency potentials in manufacturing processes: a case of battery production. *CIRP Ann.* **69**(1), 21–24 (2020)
23. Hoffmann, L., Grathwol, J.K., Haselrieder, W., et al.: Capacity distribution of large Lithium-Ion battery pouch cells in context with pilot production processes. *Energ. Technol.* **8**(2), 1900196 (2020)
24. Kornas, T., Knak, E., Daub, R., Bühner, U., et al.: A multivariate KPI-based method for quality assurance in Lithium-ion battery production. *Procedia CIRP* **81**, 75–80 (2019)
25. Liu, K., Hu, X., Meng, J., Guerrero, J.M., Teodorescu, R.: RUBoost-based ensemble machine learning for electrode quality classification in Li-ion battery manufacturing. *IEEE/ASME Trans. Mechatron.* **2**, 1023 (2021)
26. Wenzel, V., Nirschl, H., Nötzel, D.: Challenges in Lithium-Ion battery slurry preparation and potential of modifying electrode structures by different mixing processes. *Energ. Technol.* **3**(7), 692–698 (2015)
27. Liu, K., Wei, Z., Yang, Z., Li, K.: Mass load prediction for lithium-ion battery electrode clean production: a machine learning approach. *J. Clean. Prod.* **289**, 125159 (2021)
28. Cunha, R.P., Lombardo, T., Primo, E.N., Franco, A.A.: Artificial intelligence investigation of NMC cathode manufacturing parameters interdependencies. *Batteries Supercaps* **3**(1), 60–67 (2020)
29. Zhou, Y., Mazzuchi, T.A., Sarkani, S.: M-AdaBoost-a based ensemble system for network intrusion detection. *Expert Syst. Appl.* **162**, 113864 (2020)
30. Li, Y., Shi, H., Han, F., Duan, Z., Liu, H.: Smart wind speed forecasting approach using various boosting algorithms, big multi-step forecasting strategy. *Renew. Energy* **135**, 540–553 (2019)
31. Warmuth, M.K., Liao, J., Rätsch, G.: Totally corrective boosting algorithms that maximize the margin. In: *Proceedings of the 23rd international conference on Machine learning*, pp. 1001–1008 (2006)
32. Chen, H., Lin, Z., Tan, C.: Random subspace-based ensemble modeling for near-infrared spectral diagnosis of colorectal cancer. *Anal. Biochem.* **567**, 38–44 (2019)
33. Chen, J., Lian, Y., Li, Y.: Real-time grain impurity sensing for rice combine harvesters using image processing and decision-tree algorithm. *Comput. Electron. Agric.* **175**, 105591 (2020)
34. Narkhede, S.: Understanding auc-roc curve. *Towards Data Sci.* **26**, 220–227 (2018)
35. Zhang, Y., Lu, S., Zhou, X., et al.: Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: decision tree, k-nearest neighbors, and support vector machine. *Simulation* **92**(9), 861–871 (2016)



Kailong Liu received the Ph.D. degree in Electrical Engineering from Queen's University Belfast, UK in 2018. He is an assistant professor in the Warwick Manufacturing Group, University of Warwick, UK. His research is on AI for electrochemical energy storage and transportation electrification applications. Dr. Liu has published over 40 high impact Q1 journal papers and authored 5 patents, with total citation of over 2000, and an H-index of 25. He is on editorial boards of some journals of his area including *Renewable and Sustainable Energy Reviews*, *IEEE/CAA Journal of Automatica Sinica*, and *Control Engineering Practice*.



Qiao Peng is a Ph.D. candidate at Queen's University Belfast, UK. She obtained the B.S. degree at Nanjing University of Information Science & Technology, China and the M.S. degree at Queen's University Belfast, in 2016 and 2017, respectively. Her research interests include the applications of machine learning, data analysis, modelling and prediction technologies in many fields such as battery management, and credit union performance assessment.



Kang Li is the director and chair professor of the Smart Energy System Group at the University of Leeds, UK. He serves as the executive editor or editorial board member of 5 internationally renowned academic journals including *IET/IFAC*. He is engaged in cutting-edge scientific research in areas such as control, energy-saving and emission reduction in high-energy-consumption manufacturing industries, electric power and new energy, intelligent transportation and energy storage systems. He has published more than 350 papers, of which more than 10 articles have won awards.



Tao Chen is currently an assistant professor in Southeast University, a visiting scholar in Tsinghua University/Alibaba Lab, and also affiliated with Tampere University, Finland. He received the Ph.D. degree from University of Michigan, and worked as a Postdoc in Virginia Tech. His research interests are about renewable energy and power data analytics.