Article

A Machine-Learning-Based Approach to Analyse the Feature Importance and Predict the Electrode Mass Loading of a Solid-State Battery

Wenming Dai [1], Yong Xiang [2,*], Wenyi Zhou [3] and Qiao Peng [4,*]

1 College of Control Science and Engineering, Zhejiang University, Hangzhou 310000, China
2 Advanced Energy Technology Research Institute, University of Electronic Science and Technology, Chengdu 611731, China
3 Yixin Semiconductor (Jiaxing) Co., Ltd., Zhejiang 200540, China
4 Queen's Business School, Queen's University Belfast, Belfast BT7 1NN, UK
* Correspondence: qiao.peng@qub.ac.uk (Q.P.); xiang@uestc.edu.cn (Y.X.)

Abstract: Solid-state batteries are currently developing into one of the most promising battery types for both the electrification of transport and for energy storage applications due to their high energy density and safe operating behaviour. The performance of solid-state batteries is largely determined by the manufacturing process, particularly in the production of electrodes. However, efficiently analysing the effects of key manufacturing features and predicting the mass loading of electrodes in the early stages of battery manufacturing remain a major challenge. In this study, a machine-learning-based approach is proposed to effectively analyse the importance of manufacturing features and accurately predict the mass loading of electrodes. Specifically, the importance of four key features during the manufacturing process of solid-state batteries is first quantified and analysed using a machine-learning-based method to analyse the importance of features. Then, four effective machine-learning-based regression methods, including decision tree, boosted decision tree, support vector regression and Gaussian process regression, are used to predict the mass loading of the electrodes in the mixing and coating stages. The comparative results show that the developed machine-learning-based approach is able to provide a satisfactory prediction of the electrode mass loading of a solid-state battery with 0.995 R^2 while successfully quantifying the importance of four key features in the early manufacturing stages. Due to the advantages of its data-driven nature, the developed machine-learning-based approach can efficiently assist engineers in monitoring/predicting the electrode mass loading of solid-state batteries and analysing/quantifying the importance of manufacturing features of interest. This could benefit the production of solid-state batteries for further energy storage applications.

Keywords: battery manufacturing; feature importance analysis; electrode mass loading prediction; machine learning; regression model

1. Introduction

Solid-state batteries are becoming a promising next-generation battery type in energy storage applications due to their higher energy density and safer operation behaviour when compared with lithium-ion batteries [1,2]. In addition, solid-state batteries are able to charge faster, making them more suitable for electric vehicles and other applications, where sustainable development goals are important [3]. The manufacturing process, especially for the electrode production stage, plays an important role in determining the solid-state battery performance [4]. In this context, to ensure high-performance solid-state batteries and widen their applications, it is vital to analyse/quantify the effects of manufacturing feature terms and monitor/forecast electrode key properties such as mass loading during the early manufacturing stages.

However, solid-state batteries have complicated energy storage resources, containing many electrochemical dynamics, during their production [5]. To date, the trial-and-error strategy is the main solution, which has been widely utilized to analyse the effects of key manufacturing feature terms on the mass loading of produced electrodes, which is can easily lead to high costs and can take a long time to complete [6,7]. In light of this, developing an efficient solution to analyse the importance of manufacturing feature terms of interest and predict the electrode mass loading of solid-state batteries is a challenging but essential target when attempting to widen the applications of solid-state batteries.

With the rapid development of artificial intelligence, machine learning and computational platforms [8], machine-learning-based methods have become a powerful solution to management issues in batteries [9,10]. Numerous machine-learning-based solutions have been developed to estimate batteries' internal states [11–14], forecast batteries' future ageing dynamics [15–17] and remaining useful life (RUL) [18,19], diagnose battery faults [20–22] and optimize battery charging [23–26] and energy management [27–30]. In summary, based on a well-developed machine learning method, an effective management solution can be obtained to improve the operational performance of batteries. However, these methods mainly focus on exploring the macrodynamics rather than microdynamics of batteries, such as their manufacturing feature terms. To date, there are still limited studies focusing on analyses of battery electrode manufacturing feature terms through the development of related machine-learning-based solutions. For instance, a machine learning method using random forest is proposed in [31] to capture the interdependency among key feature terms of battery manufacturing and the electrical behaviours of the associated battery cells. Based on the typical cross-industrial standard process, a machine learning approach using neural networks is developed to quantify the dependency among key battery components. In [32], a neural-network-based machine learning approach is designed to capture battery products' capacity for highly efficient battery manufacturing. To solve the imbalance that usually occurs in the battery manufacturing process, an RUBoost-based machine learning method is developed to predict battery capacity and quantify the formulation terms in [33]. For solid-state batteries, it should be noted that the corresponding manufacturing process, especially for electrode production, plays a key role in the solid-state battery properties. To ensure high-performance solid-state battery production, it is necessary to design an effective machine-learning-based approach to predict the produced electrode properties and analyse how feature terms at the early manufacturing stages will affect the electrode properties.

Based on the above discussion, to benefit high-performance solid-state battery production, this study develops a machine-learning-based approach to accurately forecast the electrode mass loading of solid-state batteries and quantify the importance of related manufacturing feature terms in determining electrode mass loading. This study makes the following contributions to the field: (1) three effective machine-learning-based feature analysis methods, including minimum redundancy–maximum relevance (MRMR), the F-test and RreliefF, are developed to quantify and analyse the importance and effects of four key manufacturing feature terms on determining electrode mass loading; (2) after analysing feature importance, four efficient machine-learning-based regression methods, including decision tree (DT), boosted decision tree (BDT), support vector regression (SVR) and Gaussian process regression (GPR), are developed to predict the electrode mass loading a of a solid-state battery during the early manufacturing stage. Due to its data-driven characteristics, the designed machine-learning-based approach could accurately forecast battery electrode mass loading at the mixing and coating stages and successfully quantify the importance of manufacturing feature terms, paving the way to monitoring electrode mass loading and analysing the effects of early-stage manufacturing feature terms, further benefitting the production of high-performance solid-state batteries.

The rest of this paper is organized as follows: Section 2 introduces four key manufacturing feature terms from mixing and coating stages to produce an electrode for the solid-state battery. Section 3 describes the machine-learning-based feature importance

analysis methods, the machine-learning-based regression methods and the related prediction performance indicators. Then, the feature importance analysis results of four key manufacturing feature terms and the electrode mass loading prediction results are provided and discussed in Section 4. The conclusion of this study is given in Section 5.

2. Key Manufacturing Feature Terms

Electrode manufacturing is a key and complex process within the production line for solid-state batteries, involving numerous disciplines, including material, mechanical and electrical engineering [34]. Furthermore, the produced electrode properties such as mass loading significantly affect final the performance of solid-state batteries, which must be well monitored and controlled to produce a high-performance solid-state battery.

Figure 1 illustrates the main stages to manufacture an electrode of solid-state batteries. During electrode manufacturing, after preparing suitable materials, the first main stage is mixing to generate slurries that contain active material, conductive additive material and solvent. After that, the mixed slurry needs to be coated on the surface of the metal foil through a coating machine. Generally, anode coating will use copper foil, while cathode coating will adopt aluminium foil. During the coating stage, the ratio of the coating machine will be fixed, while CG will be adjusted to determine the thickness of the coating product. Then, the coating product will be dried by using ovens at a predefined temperature, followed by a calendaring stage to evaporate the residual solvent. After cutting the coating product to a suitable size, the electrode of the solid-state battery can be obtained. It should be noted that lots of manufacturing terms are involved within the electrode production line. Some feature terms, specifically for those from mixing and coating stages (the early manufacturing stages), play a key role in affecting electrode mass loading, which must be well analysed and monitored to ensure good electrode performance.
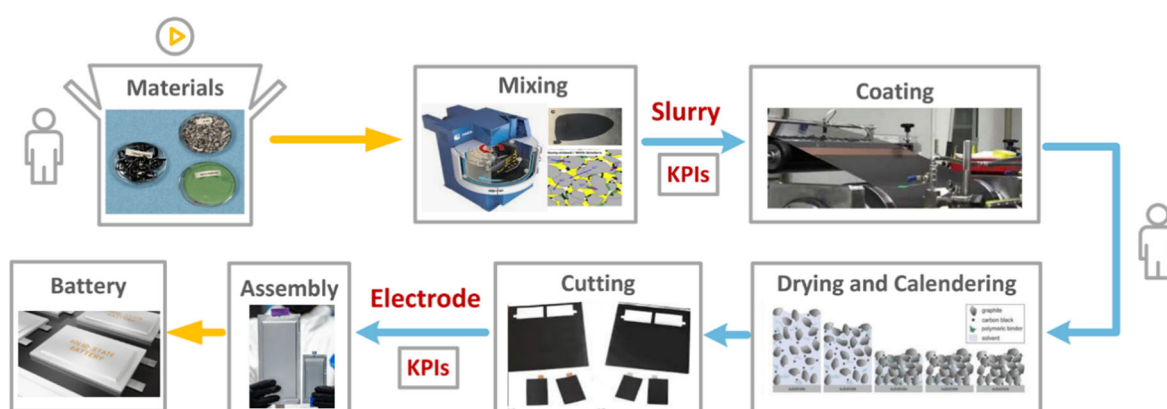

Figure 1. Main stages to manufacture electrode of solid-state batteries.

In this study, three key feature terms, including active material–mass content (AM-MC), solid-to-liquid ratio (S-LR), viscosity from the mixing stage and a key feature term named comma-gap (CG), from the coating stage are utilized as the explored feature terms of interest. All of these feature terms come from the early manufacturing stages, which brings challenges for electrode mass loading prediction. Theoretically, CG stands for a gap value between comma and coating rolls, affecting both the weight and thickness of the coating. S-LR stands for the mass ratio between a solid component (i.e., active material, conductive additive material and binder) and slurry mass (i.e., solvent and solid component). A well-proven dataset [35] from the Franco Laboratoire de Reactivite Chimie-des-Solides (LRCS) is adopted here to ensure that the machine-learning-based models used can be trained well. More information on the design of experiments and the introduction of the data can be obtained in [35] for readers' interest. Based on this dataset, the feature importance of the above-mentioned four feature terms can be quantified, while suitable machine-learning-

based regression models can be built to predict the electrode mass loading of the solid-state battery in the early manufacturing stages.

3. Machine Learning Methodology

In this section, three typical machine learning methods are first introduced to analyse the importance of feature terms of interest. Then, four regression methods are derived to predict the electrode mass loading of solid-state batteries. Afterwards, three performance indicators are used to quantify and compare the electrode mass loading prediction performance via the developed regression models.

3.1. Feature Importance Analysis Method

To quantify the importance and analyse the effects of these feature terms of interest on the determination of electrode mass loading of solid-state batteries, three effective feature selection methods, including minimum redundancy–maximum relevance (MRMR), F-test and RReliefF, are utilized in this study, respectively.

3.1.1. Minimum Redundancy–Maximum Relevance

Minimum redundancy–maximum relevance (MRMR) is an efficient feature selection method as it selects feature terms with high correlations with the output of interest and low correlations with each other [36]. If the interested feature terms are continuous, the correlation with relevance in MRMR can be calculated by using the F-statistic, while the correlation with redundancy will be obtained by using the Pearson correlation coefficient. In this context, the importance of the feature terms could be determined one by one through adopting the greedy search to maximize target functions corresponding to both relevance and redundancy.

In this study, two classical target functions, including the Mutual Information Difference (MID) and the Mutual Information Quotient (MIQ), are adopted to represent the difference and the quotient between relevance and redundancy. Through using MID and MIQ, MRMR is capable of ensuring that the quantified feature importance not only refers to the one that gives the minimum correlation among inputted feature terms but also provides the large correlations with the target output.

3.1.2. F-Test

The F-test is one of the classic statistical tests providing an F-score through computing variance ratios [37]. In this study, the F-test in a one-way analysis of variance (ANOVA) is adopted to calculate variance's ratio among groups and a group's variance for a feature. A larger value of the F-score represents that the distance within the group is smaller, while the distance between the groups is larger.

The ANOVA could determine each feature term's value to the target output by adopting the F-test to quantify the mean of various groups. Then, each feature term can be scored and ranked to reflect which term plays the most important role in determining the target output. For the ANOVA, the number of feature terms and the F-ration will be obtained. The F-score of each feature term can be obtained by:

$$F^2 = \frac{\sum \left(\overline{x_j} - \overline{x} \right)^2 \cdot N_j}{n - 1} \tag{1}$$

where N_j stands for the number of group j within the set; $\overline{x_J}$ and $\overline{x}$ represent the mean values of the group and feature term, respectively. n denotes the number of groups. Finally, the feature importance value under the F-test can be obtained through dividing distances among groups by the distance within the group. The larger the result, the more important that feature's contribution to the target output.

3.1.3. RReliefF

RReliefF is an efficient machine-learning-based approach to quantify feature importance by penalizing the feature term that provides various values to neighbours with similar target outputs while rewarding the feature term that gives different values to neighbours with different target outputs [38]. Supposing W_{do} represents the weight of providing various values for the target output, W_{di} stands for the weight of providing various values for the feature term, and $W_{do \wedge di}$ denotes the weight of providing various target outputs and various values for the feature term, thus giving two nearest neighbours. All weights in RReliefF can be calculated by:

$$\begin{cases} W_{do}^i = W_{do}^{i-1} + \Delta_o(x_l, x_r) \cdot d_{lr} \\ W_{di}^i = W_{di}^{i-1} + \Delta_i(x_l, x_r) \cdot d_{lr} \\ W_{do \wedge di}^i = W_{do \wedge di}^{i-1} + \Delta_o(x_l, x_r) \cdot \Delta_i(x_l, x_r) \cdot d_{lr} \end{cases} \tag{2}$$

Afterwards, RReliefF can calculate the importance weight W_{ft} of feature terms through updating all intermediate weights as:

$$W_{ft} = \frac{W_{do \wedge di}}{W_{do}} - \frac{W_{di} - W_{do \wedge di}}{m - W_{do}} \tag{3}$$

3.2. Regression Method

After developing a feature importance analysis method to quantify the importance of different feature terms, the prediction method needs to be adopted to forecast the electrode mass loading based on the feature terms of interest. In this study, four classical and effective prediction methods named decision tree, boosted decision tree, support vector machine and Gaussian process regression are utilized.

3.2.1. Decision Tree

Decision tree (DT) is an algorithm that uses a tree model to solve regression problems. Each leaf in the tree model represents a predicted value, which is the mean value of the outputs of all training set factors contained in this leaf. The DT is a binary tree, which means that except for leaf nodes, all other nodes have only two child nodes. In addition, all inputs that fall on the same leaf have the same output [39]. Figure 2 describes a simple structure example of DT. Specifically, when using DT to solve regression tasks, a certain feature of the samples is tested from the root nodes, and then the samples are assigned to the child nodes according to the testing results. It is ensured that each child node corresponds to a value of this feature. By analogy, all samples are tested and distributed in a recursive manner until they reach the leaf nodes.
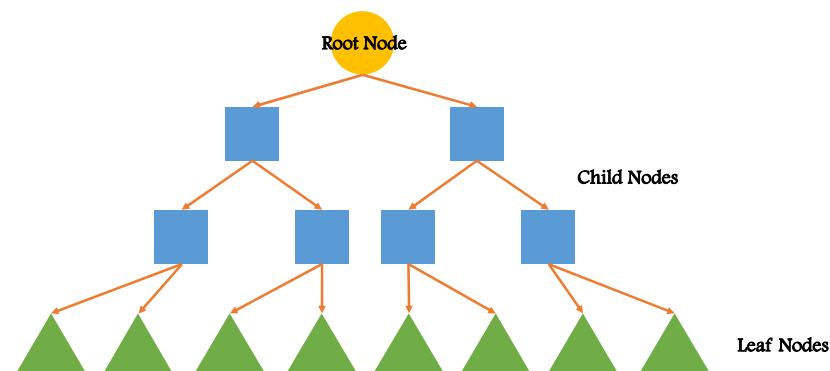

Figure 2. A simple structure example of DT.

3.2.2. Boosted Decision Tree

The working mechanism of the boosting algorithm is to first train the training set through a weak learner A with a set of initial weights. Then, the weights of the training samples are updated according to the performance of the error rates of this weak learner A, so that the weights of the training samples with higher learning error rates will become greater. Then, a weak learner B is used to train the training set based on the adjusted weights, and these samples with greater weights will receive more attention. This process is repeated until the number of weak learners reaches the pre-specified number T (T = 30 in this study). Finally, these T weak learners are integrated through the predefined strategy to obtain a final strong learner. In this study, the medium decision trees are employed as the weak learners, and the AdaBoost algorithm is utilized as the integration strategy. When dealing with practical problems, the boosted decision tree (BDT) model can handle different types of predictors and allows for missing values. This method can fit complex nonlinear relationships and can automatically manage the mutual influence between predictors. Thus, the BDT model usually has stronger predictive ability than most traditional tree methods.

3.2.3. Support Vector Regression

Support vector machine (SVM) is a machine learning classification model, and its principle is to seek the optimal solution in data mining through establishing an optimal hyperplane, so that the support vectors of the two (or more) classification sets are the farthest away from this classification plane [40]. The traditional methods usually simplify problems by reducing the dimensionality of the sample, while the SVM increases the dimensionality of the sample point. Then, it maps the sample points to a high-dimensional or even infinite-dimensional space through a kernel function to deal with linear and nonlinear problems. As an application of SVM to regression problems, support vector regression (SVR) is an approach to estimate the function mapped from input objects to output numbers based on training data. Regression and classification are essentially the same in a sense. Similar to SVM, the principle of SVR is to find a regression hyperplane, so that all the data in a dataset are closest to the plane. When considering data distributed in two dimensions (x-axis and y-axis are used to define the location of the points), SVR tries to find the best line within a predefined or threshold error value. It classifies the data into two types, one type above the error line and one type below the error line. For failed lines, the error is estimated as the difference between the predicted and actual values and represented by ε, as illustrated in Figure 3.
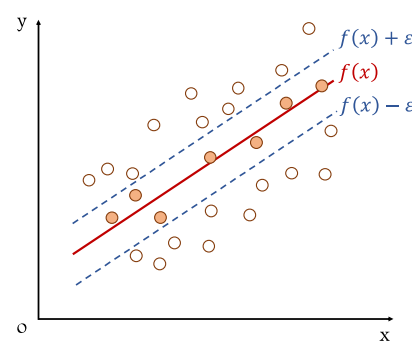

Figure 3. A simple structure example of SVR.

3.2.4. Gaussian Process Regression

Gaussian process regression (GPR) is a typical regression method based on Bayesian theory [41,42]. The probability distribution of GPR can be described through mean function $m(j)$ and kernel function $k(j, j')$ as:

$$G(j) \sim GPR(m(j), k(j, j'))$$ (4)

with:

$$\begin{cases} m(j) = E(G(j)) \\ k(j,j') = E[(m(j) - f(j'))(m(j) - f(j'))] \end{cases} \tag{5}$$

Here, $m(j)$ is usually set as zero to simply the computation process. In this context, for a prediction, the prior distribution of output in a regression case is defined as:

$$y \sim N(0, k(j,j')) \tag{6}$$

Supposing the training set j and testing set j' own the same Gaussian distribution, then the testing output y' presents the joint prior distributions with the training output y as [43]:

$$\begin{bmatrix} y \\ y' \end{bmatrix} \sim N(0, \begin{bmatrix} k(j,j) & k(j,j') \\ k(j,j')^T & k(j',j') \end{bmatrix}) \tag{7}$$

Based on this joint prior distribution, the predicted output y' related to the input j' can be obtained through calculating the conditional distribution $p(y'|j, y, j')$:

$$p(y'|j, y, j') \sim N(y'|\overline{y}', cov(y')) \tag{8}$$

with

$$\begin{cases} \overline{y}' = k(j,j')^T [k(j,j)]^{-1} y \\ cov(y') = k(j',j') - k(j,j')^T [k(j,j)]^{-1} k(j,j') \end{cases} \tag{9}$$

where $\overline{y}'$ represents the mean predicted value, while $cov(y')$ stands for the corresponding variance value.

In real application of GPR, the kernel function $k(j,j')$ needs to be carefully selected. In this study, a squared exponential (*SE*) kernel in the following form with effective performance is utilized:

$$k_{SE}(j,j') = \sigma_{SE}^2 exp\left(-\frac{\|j - j'\|^2}{2\sigma_m^2}\right) \tag{10}$$

where σ_{SE} and σ_m represent two hyperparameters to affect both the amplitude and length of the *SE* kernel.

3.3. Modelling Process

Figure 4 illustrates the designed model structure and relevant process. To be specific, four machine-learning-based regression models, including DT, BDT, SVR and GPR, are constructed using CG, S-LR, AM-MC, and viscosity as inputs to predict electrode mass loading. Subsequently, three feature selection methods, namely MRMR, F-test and RReliefF, are employed to assess the importance of these four features. This approach unveils the decision-making process of the model and offers valuable insights for real battery production.
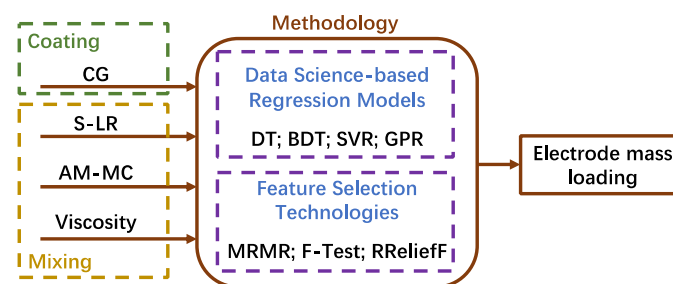

Figure 4. Model structure and process.

3.4. Prediction Performance Indicators

To quantitatively reflect the solid-state battery electrode mass loading prediction performance of machine-learning-based regression models, three classical prediction performance indicators [44] are utilized:

(1) Mean absolute error (MAE): Supposing TS represents the total amount of all samples, Y_j is the real battery electrode mass loading value, while $\hat{Y}_j$ stands for the predicted electrode mass loading value from the machine-learning-based regression model. MAE can then be defined by:

$$\text{MAE} = \frac{1}{TS}\sum_{j=1}^{TS}|Y_j - \hat{Y}_j| \tag{11}$$

(2) Root mean square error (RMSE): As another typical performance indicator, RMSE is an effective prediction performance indicator, defined as:

$$\text{RMSE} = \sqrt{\frac{1}{TS}\sum_{j=1}^{TS}(Y_j - \hat{Y}_j)^2} \tag{12}$$

(3) R^2: supposing $\overline{Y}$ stands for the mean value of all predicted electrode mass loading, R^2 can be defined as:

$$R^2 = 1 - \sum_{j=1}^{TS}(Y_j - \hat{Y}_j)^2 / \sum_{j=1}^{TS}(Y_j - \overline{Y})^2 \tag{13}$$

In real solid-state battery electrode mass loading prediction, when the predicted value is close to the real value, MAE and RMSE should become close to 0, while R^2 should be close to 1.

4. Result and Discussion

In this section, the feature importance analysis of four battery manufacturing terms is carried out to quantify the effects of CG, S-LR, AM-MC and viscosity on determining the electrode mass loading of the solid-state battery. Afterwards, four machine-learning-based regression models with DT, BDT, SVR and GPR are established to predict battery electrode mass loading in the early manufacturing stage.

4.1. Feature Importance Analysis Results

To quantify the importance of four manufacturing feature terms for all samples, the MRMR-based machine learning method is first utilized. Figure 5 quantifies and plots the importance values of four manufacturing feature terms of interest (CG, S-LR, AM-MC, viscosity). It can be obviously seen that CG is the most important feature term, with the highest importance value of 1.01, while S-LR provides the second-highest importance value of 0.185, which is five-times smaller than CG. In comparison, the importance values of AM-MC and viscosity are smaller than S-LR, meaning that their effect on determining solid-state battery electrode mass loading decreases accordingly.

Next, an F-test is adopted to quantify the importance of these four manufacturing terms, and the corresponding results are shown in Figure 6. It can be seen that although the values of the F-test are significantly larger than those of MRMR, the importance trend of these four manufacturing terms is similar. To be specific, CG gives the most important contribution to determine electrode mass loading, with a value of over 80. S-LR and AM-MC present the second and third most important contribution, with values of 21.732 and 8.121, respectively. Viscosity gives the smallest contribution to determine electrode mass loading.
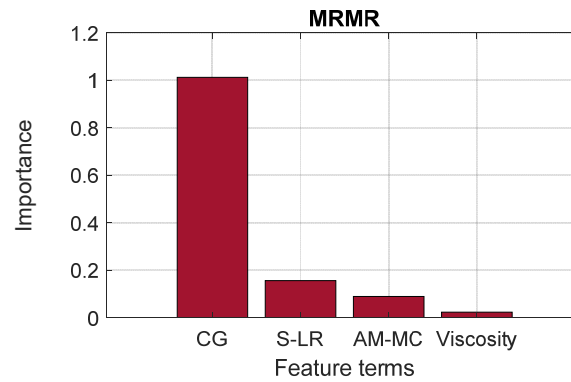
Figure 5. Quantified feature importance values of CG, S-LR, AM-MC, viscosity through using MRMR method.
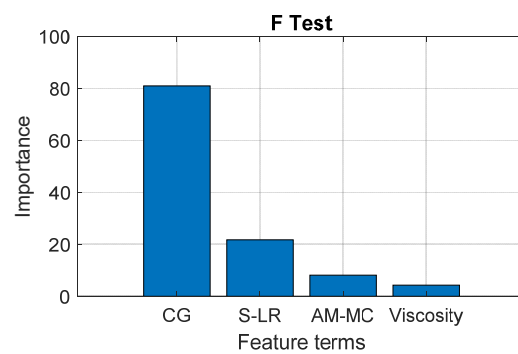

Figure 6. Quantified feature importance values of CG, S-LR, AM-MC, viscosity through using F-test method.

Furthermore, after using the RReliefF-based machine learning approach, the importance of these four manufacturing terms is also quantified and plotted in Figure 7. According to Figure 6, although the importance values of CG and viscosity are positive, while S-LR and AM-MC are negative, the quantified effects on determining electrode mass loading are similar to those of MRMR-based and F-test-based machine learning methods. Specifically, CG gives the largest quantified importance value of 0.1236, while S-LR provides the second-largest quantified importance value with −0.0837. In comparison, viscosity presents the smallest quantified importance value of 0.0126. The trend of the quantified importance values of all three machine-learning-based approaches is similar, indicating that the importance of CG, S-LR, AM-MC and viscosity can be successfully quantified and analysed using our designed feature importance analysis methods.
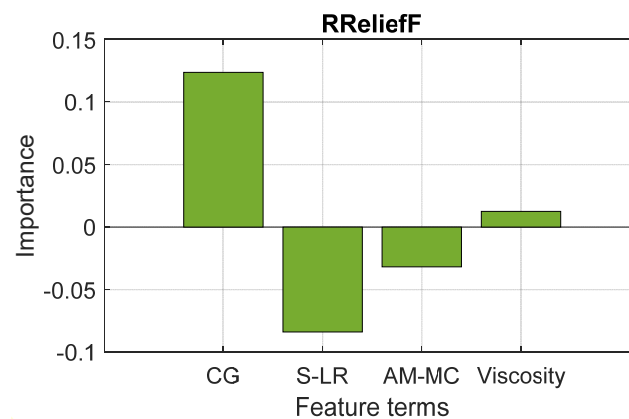

Figure 7. Quantified feature importance values of CG, S-LR, AM-MC, viscosity through using RReliefF method.

4.2. Mass Loading Prediction Results

Next, after developing a machine-learning-based approach to quantify and analyse the importance of four key manufacturing terms (CG, S-LR, AM-MC and viscosity), the machine-learning-based regression models by using four different technologies, including DT, BDT, SVR and GPR, are carried out to predict the electrode mass loading of the solid-state battery in the early manufacturing stage, respectively.

We first focus on the electrode mass loading prediction by using four manufacturing feature terms (CG, S-LR, AM-MC and viscosity) as inputs via various machine-learning-based regression models. After carrying out five-fold cross-validation, the electrode mass loading prediction results of using DT-, BDT-, SVR- and GPR-based regression models are shown in Figure 7, while their related prediction performance indicators are quantified in Table 1. According to Figure 8, it can be seen that all these four machine-learning-based regression models are capable of capturing most of the electrode mass loading sample points, indicating the effectiveness of the machine-learning-based regression model in predicting electrode mass loading in the early manufacturing stages. Specifically, the DT-based regression model provides the worst prediction result, with an RMSE value of 1.919 and a R^2 value of 0.978, which is 21.5% and 1.2% worse than those from the BDT-based regression model. In comparison, the GPR-based regression model gives the best result for electrode mass loading predictions, whose MAE and RMSE values are 0.883 and 1.089, respectively, which are 3.5% and 30.3% better than those of the SVR-based regression model. It can be concluded that four electrode manufacturing feature terms (CG, S-LR, AM-MC and viscosity) present strong relationships with the electrode mass loading. Through using more complicated regression models with kernel functions like SVR and GPR, the electrode mass loading prediction performance could be further enhanced.

Table 1. Performance indicators for electrode mass loading prediction by using different machine-learning-based regression models.

Models	DT	BDT	SVR	GPR
MAE	1.423	1.229	0.915	0.883
RMSE	1.919	1.506	1.145	1.089
R^2	0.978	0.989	0.992	0.995

To further explore the electrode mass loading prediction results of using machine-learning-based regression models, the prediction versus true plots for DT-, BDT-, SVR- and GPR-based regression cases are shown in Figure 9. Theoretically, the farthest observations in the prediction versus true plots are able to allow the prediction line for this sample. The better prediction performance a model gives, the closer the observations from that model to the perfect prediction line. According to Figure 8, it can be noted that for all four machine-learning-based regression models, most of the observations are close to the perfect prediction line, while only a few observations are far from the perfect prediction line. This is mainly due to the overfitting phenomenon and can be enhanced by giving more high-quality electrode manufacturing data observations.

4.3. Comparison and Discussion

To further explore the effectiveness of the designed approach, three other traditional algorithms, including the linear regression (LR) model, the random forest (RF)-based model and the AdaBoost-based ensemble model, are adopted for further comparisons. To be specific, the LR model is a linear traditional algorithm to capture the relationships between inputs and responses. RF and AdaBoost are all tree-based ensemble models. Table 2 shows the results of electrode mass loading prediction by using all these traditional algorithms after five-fold cross-validation. Obviously, the LR model gives the worst results, with an R^2 of 0.851, indicating that the underlying relationships between the investigated manufacturing parameters and the electrode mass loading are highly non-linear. In contrast, both the RF-based model and the AdaBoost-based model perform well in predicting the

electrode mass loading, with an R^2 above 0.992, which is similar to the GPR-based approach we developed. In this context, the designed GPR-based approach and ensemble models can provide a competent performance for early predictions of battery electrode mass loading.
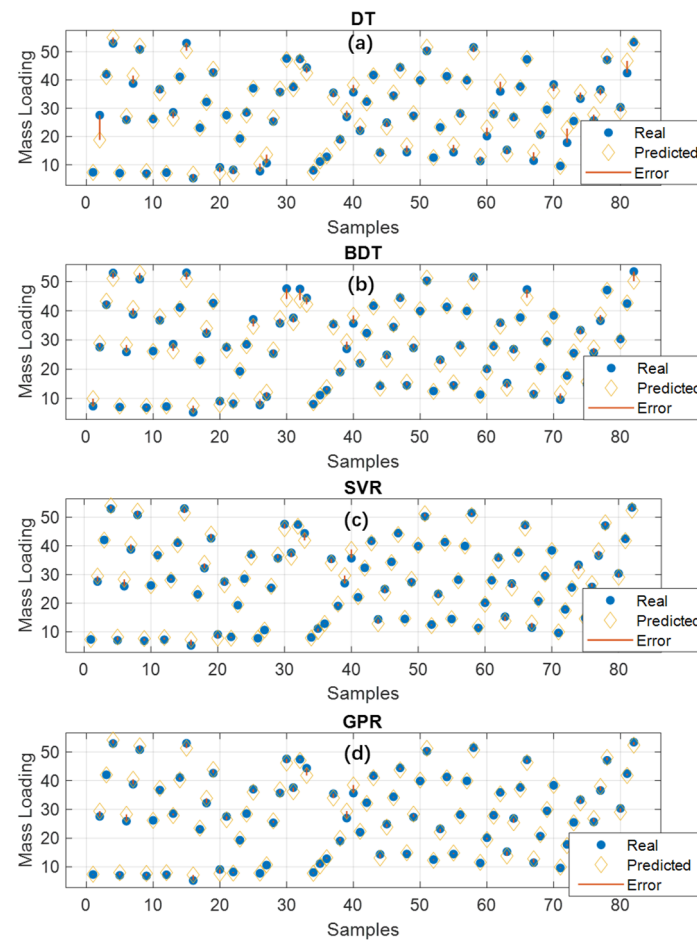

Figure 8. Battery electrode mass loading prediction by using various machine-learning-based regression models: (**a**) DT, (**b**) BDT, (**c**) SVR, (**d**) GPR.

Table 2. Performance indicators for mass loading prediction by using traditional algorithms.

Models	LR	RF	AdaBoost	GPR-Based Method
MAE [mg/cm^2]	2.132	0.889	0.891	0.883
RMSE [mg/cm^2]	3.271	1.091	1.094	1.089
R^2	0.851	0.992	0.991	0.995

Due to its data-driven nature, this machine-learning-based approach can be easily adopted by engineers in real-world applications, as limited professional knowledge of the battery manufacturing mechanism is required. The designed modelling process can be conveniently extended to other data analysis applications of battery manufacturing if appropriate data are available. It should be noted that this is of great importance for battery manufacturers to better understand their manufacturing line and make reasonable battery property predictions in the early manufacturing stage. It is also convenient to extend this machine-learning-based approach for analysing other battery cases such as the composite electrode that comprises both the solid electrolyte and the electrode. Moreover, the developed method, as supported by [45], is flexible and free from limitations of the battery mechanism. Analysing different combinations of solid electrolytes and electrodes can also

be performed via the developed machine-learning-based solution when the corresponding data are collected.
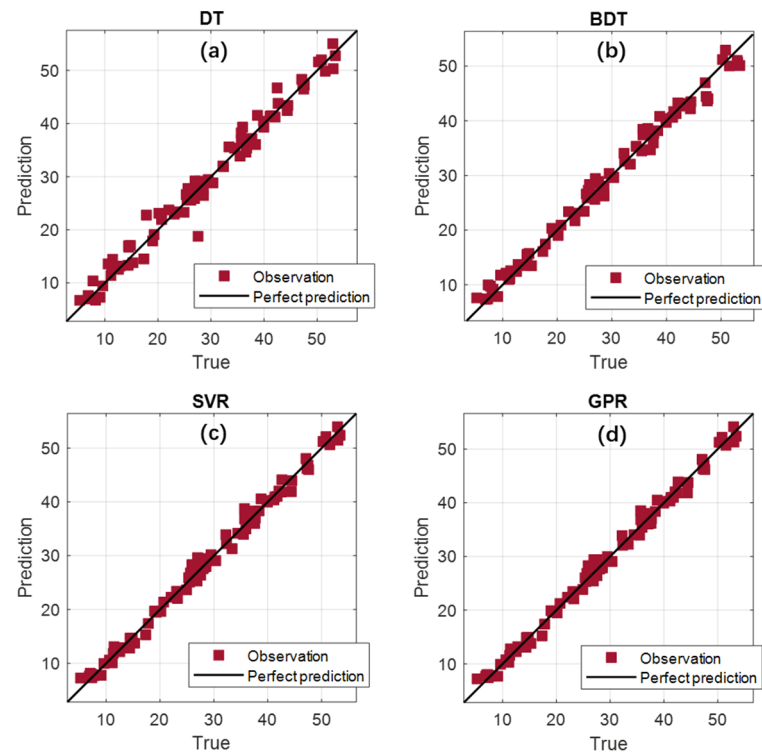

Figure 9. Prediction versus true plots by adopting various machine-learning-based regression models: (**a**) DT, (**b**) BDT, (**c**) SVR, (**d**) GPR.

5. Conclusions

The solid-state battery is developing into a promising technology that enables safer operation of energy storage systems, with its performance largely determined by the charge of the electrode mass. However, it is difficult to effectively analyse the importance of the strongly coupled manufacturing characteristics and predict the load on the electrode mass in the early stages of battery manufacturing. As suggested and requested by battery manufacturers, this study proposes a machine-learning-based approach that combines the advantages of effective feature selection and a user-friendly regression model to predict electrode mass loading and quantify the importance and impact of manufacturing parameters in the mixing and coating stages. The comparative analysis shows that CG and S-LR are the first and second most important contributors to the determination of electrode mass loading of a solid-state battery, while viscosity is the least important contributor to the prediction of electrode mass loading. The developed GPR-based regression model provides the best performance in predicting the electrode mass loading, with an RMSE and R^2 of 1.089 and 0.995, respectively. Due to its data-driven nature, the developed machine-learning-based approach can efficiently assist engineers in monitoring/predicting the mass loading of solid-state battery electrodes in early manufacturing stages and analysing/quantifying the significance of manufacturing features of interest. Additionally, it can be conveniently extended to other applications of data analysis in battery manufacturing, as no professional knowledge of the battery manufacturing mechanism is required. This could further advance the development of high-performance solid-state batteries for broader energy storage applications. As obtaining available data from battery production requires specialised equipment and is time consuming, we will conduct further experiments in the future to generate more relevant data related to other characteristics, such as drying temperature and pressure, and also consider the composite electrode to further improve the sensitivity analysis of characteristics for high-performance Li-ion battery production.

References

1. Sun, Y.-K. Promising all-solid-state batteries for future electric vehicles. *ACS Energy Lett.* **2020**, *5*, 3221–3223. [CrossRef]
2. Liu, K.; Wang, Y.; Lai, X. *Data Science-Based Full-Lifespan Management of Lithium-Ion Battery: Manufacturing, Operation and Reutilization*; Springer Nature: Berlin/Heidelberg, Germany, 2022.
3. Chakraborty, S.; Kumar, N.M.; Jayakumar, A.; Dash, S.K.; Elangovan, D. Selected Aspects of Sustainable Mobility Reveals Implementable Approaches and Conceivable Actions. *Sustainability* **2021**, *13*, 12918. [CrossRef]
4. Yu, X.; Chen, R.; Gan, L.; Li, H.; Chen, L. Battery safety: From lithium-ion to solid-state batteries. *Engineering* **2023**, *21*, 9–14. [CrossRef]
5. Janek, J.; Zeier, W.G. Challenges in speeding up solid-state battery development. *Nat. Energy* **2023**, *8*, 230–240. [CrossRef]
6. Liu, K.; Peng, Q.; Li, K.; Chen, T. Data-based interpretable modeling for property forecasting and sensitivity analysis of Li-ion battery electrode. *Automot. Innov.* **2022**, *5*, 121–133. [CrossRef]
7. Yang, Y.; Okonkwo, E.G.; Huang, G.; Xu, S.; Sun, W.; He, Y. On the sustainability of lithium ion battery industry–A review and perspective. *Energy Storage Mater.* **2021**, *36*, 186–212. [CrossRef]
8. Peng, Q.; Liu, W.; Zhang, Y.; Zeng, S.; Graham, B. Generation planning for power companies with hybrid production technologies under multiple renewable energy policies. *Renew. Sustain. Energy Rev.* **2023**, *176*, 113209. [CrossRef]
9. Wei, Z.; Liu, K.; Liu, X.; Li, Y.; Du, L.; Gao, F. Multi-level Data-driven Battery Management: From Internal Sensing to Big Data Utilization. *IEEE Trans. Transp. Electrif.* **2023**, *9*, 4805–4823. [CrossRef]
10. Liu, K.; Wei, Z.; Zhang, C.; Shang, Y.; Teodorescu, R.; Han, Q.L. Towards Long Lifetime Battery: AI-Based Manufacturing and Management. *IEEE/CAA J. Autom. Sin.* **2022**, *9*, 1139–1165. [CrossRef]
11. Zhang, M.; Fan, X. Review on the state of charge estimation methods for electric vehicle battery. *World Electr. Veh. J.* **2020**, *11*, 23. [CrossRef]
12. Noura, N.; Boulon, L.; Jemeï, S. A review of battery state of health estimation methods: Hybrid electric vehicle challenges. *World Electr. Veh. J.* **2020**, *11*, 66. [CrossRef]
13. Xie, Y.; Li, W.; Hu, X.; Tran, M.-K.; Panchal, S.; Fowler, M.; Zhang, Y.; Liu, K. Coestimation of SOC and three-dimensional SOT for lithium-ion batteries based on distributed spatial–temporal online correction. *IEEE Trans. Ind. Electron.* **2022**, *70*, 5937–5948. [CrossRef]
14. Wang, Y.; Zhang, X.; Liu, K.; Wei, Z.; Hu, X.; Tang, X.; Chen, Z. System identification and state estimation of a reduced-order electrochemical model for lithium-ion batteries. *eTransportation* **2023**, *18*, 100295. [CrossRef]
15. Hu, T.; Ma, H.; Liu, K.; Sun, H. Lithium-ion Battery Calendar Health Prognostics Based on Knowledge-data-driven Attention. *IEEE Trans. Ind. Electron.* **2022**, *70*, 407–417. [CrossRef]
16. Liu, K.; Peng, Q.; Sun, H.; Fei, M.; Ma, H.; Hu, T. A Transferred Recurrent Neural Network for Battery Calendar Health Prognostics of Energy-Transportation Systems. *IEEE Trans. Ind. Inform.* **2022**, *18*, 8172–8181. [CrossRef]
17. Hu, T.; Ma, H.; Sun, H.; Liu, K. Electrochemical-Theory-Guided Modelling of the Conditional Generative Adversarial Network for Battery Calendar Ageing Forecast. *IEEE J. Emerg. Sel. Top. Power Electron.* **2022**, *11*, 67–77. [CrossRef]
18. Liu, K.; Peng, Q.; Che, Y.; Zheng, Y.; Li, K.; Teodorescu, R.; Widanage, D.; Barai, A. Transfer learning for battery smarter state estimation and ageing prognostics: Recent progress, challenges, and prospects. *Adv. Appl. Energy* **2022**, *9*, 100117. [CrossRef]
19. Ren, L.; Dong, J.; Wang, X.; Meng, Z.; Zhao, L.; Deen, M.J. A data-driven auto-cnn-lstm prediction model for lithium-ion battery remaining useful life. *IEEE Trans. Ind. Inform.* **2020**, *17*, 3478–3487. [CrossRef]
20. Kosuru, V.S.R.; Venkitaraman, A.K. A Smart Battery Management System for Electric Vehicles Using Deep Learning-Based Sensor Fault Detection. *World Electr. Veh. J.* **2023**, *14*, 101. [CrossRef]
21. Wang, G.; Zhao, G.; Xie, J.; Liu, K. Ensemble Learning Based Correlation Coefficient Method for Robust Diagnosis of Voltage Sensor and Short-Circuit Faults in Series Battery Packs. *IEEE Trans. Power Electron.* **2023**, *38*, 9143–9156. [CrossRef]
22. Gu, X.; Li, J.; Liu, K.; Zhu, Y.; Tao, X.; Shang, Y. A Precise Minor-Fault Diagnosis Method for Lithium-Ion Batteries Based on Phase Plane Sample Entropy. *IEEE Trans. Ind. Electron.* **2023**. *Early Access*. [CrossRef]
23. Hu, T.; Ma, H.; Liu, H.; Sun, H.; Liu, K. Self-attention-based machine theory of mind for electric vehicle charging demand forecast. *IEEE Trans. Ind. Inform.* **2022**, *18*, 8191–8202. [CrossRef]

24. Li, W.; Xie, Y.; Hu, X.; Tran, M.-K.; Fowler, M.; Panchal, S.; Zheng, J.; Liu, K. An internal heating strategy for lithium-ion batteries without lithium plating based on self-adaptive alternating current pulse. *IEEE Trans. Veh. Technol.* **2022**, *72*, 5809–5823. [CrossRef]

25. Xie, Y.; Li, W.; Song, Z.; Chen, B.; Liu, K.; Yang, R.; Zhang, Y. A health-aware AC heating strategy with lithium plating criterion for batteries at low temperatures. *IEEE Trans. Ind. Inform.* **2023**, *20*, 2295–2306. [CrossRef]

26. Zhu, T.; Cruden, A.; Peng, Q.; Liu, K. Enabling extreme fast charging. *Joule* **2023**, *7*, 2660–2662. [CrossRef]

27. Zhu, C.; Du, L.; Guo, B.; Fan, G.; Lu, F.; Zhang, H.; Liu, K.; Zhang, X. Internal Heating Techniques for Lithium-ion Batteries at Cold Climates: An Overview for Automotive Applications. *IEEE Trans. Transp. Electrif.* **2022**, *9*, 5012–5027. [CrossRef]

28. Li, J.; Liu, K.; Zhou, Q.; Meng, J.; Ge, Y.; Xu, H. Electrothermal Dynamics-Conscious Many-Objective Modular Design for Power-Split Plug-in Hybrid Electric Vehicles. *IEEE/ASME Trans. Mechatron.* **2022**, *27*, 4406–4441. [CrossRef]

29. Zhang, F.; Xiao, L.; Coskun, S.; Pang, H.; Xie, S.; Liu, K.; Cui, Y. Comparative study of energy management in parallel hybrid electric vehicles considering battery ageing. *Energy* **2022**, *264*, 123219. [CrossRef]

30. Shang, Y.; Chen, G.; Peng, Q.; Zhu, T.; Liu, K. An intelligent preheating approach based on high-gain control for lithium-ion batteries in extremely cold environment. *IEEE Trans. Ind. Electron.* **2023**, *71*, 4697–4706. [CrossRef]

31. Niri, M.F.; Liu, K.; Apachitei, G.; Ramirez, L.R.; Lain, M.; Widanage, D.; Marco, J. Machine learning for optimised and clean Li-ion battery manufacturing: Revealing the dependency between electrode and cell characteristics. *J. Clean. Prod.* **2021**, *324*, 129272. [CrossRef]

32. Yuan, Y.; Kong, X.; Hua, J.; Pan, Y.; Sun, Y.; Han, X.; Yang, H.; Li, Y.; Liu, X.; Zhou, X.; et al. Fast grading method based on data driven capacity prediction for high-efficient lithium-ion battery manufacturing. *J. Energy Storage* **2023**, *73*, 109143. [CrossRef]

33. 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.* **2021**, *27*, 2474–2483. [CrossRef]

34. Liu, K.; Niri, M.F.; Apachitei, G.; Lain, M.; Greenwood, D.; Marco, J. Interpretable machine learning for battery capacities prediction and coating parameters analysis. *Control Eng. Pract.* **2022**, *124*, 105202. [CrossRef]

35. Cunha, R.P.; Lombardo, T.; Primo, E.N.; Franco, A.A. Artificial intelligence investigation of NMC cathode manufacturing parameters interdependencies. *Batter. Supercaps* **2020**, *3*, 60–67. [CrossRef]

36. Özyurt, F. A fused CNN model for WBC detection with MRMR feature selection and extreme learning machine. *Soft Comput.* **2020**, *24*, 8163–8172. [CrossRef]

37. Dhanya, R.; Paul, I.R.; Akula, S.S.; Sivakumar, M.; Nair, J.J. F-test feature selection in Stacking ensemble model for breast cancer prediction. *Procedia Comput. Sci.* **2020**, *171*, 1561–1570. [CrossRef]

38. Zhang, B.; Li, Y.; Chai, Z. A novel random multi-subspace based ReliefF for feature selection. *Knowl.-Based Syst.* **2022**, *252*, 109400. [CrossRef]

39. Liu, K.; Hu, X.; Zhou, H.; Tong, L.; Widanalage, D.; Marco, J. Feature analyses and modeling of lithium-ion battery manufacturing based on random forest classification. *IEEE/ASME Trans. Mechatron.* **2021**, *26*, 2944–2955. [CrossRef]

40. Liu, K.; Peng, Q.; Teodorescu, R.; Foley, A.M. Knowledge-guided data-driven model with transfer concept for battery calendar ageing trajectory prediction. *IEEE/CAA J. Autom. Sin.* **2023**, *10*, 272–274. [CrossRef]

41. Tagade, P.; Hariharan, K.S.; Ramachandran, S.; Khandelwal, A.; Naha, A.; Kolake, S.M.; Han, S.H. Deep Gaussian process regression for lithium-ion battery health prognosis and degradation mode diagnosis. *J. Power Sources* **2020**, *445*, 227281. [CrossRef]

42. Liu, K.; Tang, X.; Teodorescu, R.; Gao, F.; Meng, J. Future Ageing Trajectory Prediction for Lithium-ion Battery Considering the Knee Point Effect. *IEEE Trans. Energy Convers.* **2021**, *37*, 1282–1291. [CrossRef]

43. Williams, C.K.I.; Rasmussen, C.E. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006.

44. Šeruga, D.; Gosar, A.; Sweeney, C.A.; Jaguemont, J.; Mierlo, J.V.; Nagode, M. Continuous modelling of cyclic ageing for lithium-ion batteries. *Energy* **2021**, *215*, 119079. [CrossRef]

45. Ayerbe, E.; Berecibar, M.; Clark, S.; Franco, A.A.; Ruhland, J. Digitalization of battery manufacturing: Current status, challenges, and opportunities. *Adv. Energy Mater.* **2022**, *12*, 2102696. [CrossRef]