

## Article

# Deploying Virtual Quality Gates in a Pilot-Scale Lithium-Ion Battery Assembly Line

Xukuan Xu <sup>1,\*</sup>, Simon Stier <sup>2</sup>, Andreas Gronbach <sup>2</sup> and Michael Moeckel <sup>1</sup>

<sup>1</sup> Faculty of Engineering and Computer Science, Aschaffenburg University of Applied Sciences, Würzburger Str. 45, 63743 Aschaffenburg, Germany; michael.moeckel@th-ab.de

<sup>2</sup> Fraunhofer Institute for Silicate Research ISC, Neunerpl. 2, 97082 Würzburg, Germany; simon.stier@isc.fraunhofer.de (S.S.); andreas.gronbach@isc.fraunhofer.de (A.G.)

\* Correspondence: xukuan.xu@th-ab.de

## Abstract

Pilot production is a critical transitional phase in the process of new product development or manufacturing, aiming at ensuring that products are thoroughly validated and optimized before entering full-scale production. During this stage, a key challenge is how to leverage limited resources to build data infrastructure and conduct data analysis to establish and verify quality control. This paper presents the implementation of a cyber–physical system (CPS) for a lithium battery pilot assembly line. A machine learning-based predictive model was employed to establish quality control mechanisms. Process knowledge-guided data analysis was utilized to build a quality prediction model based on the collected battery data. The model-centric concept of ‘virtual quality’ enables early quality judgment during production, which allows for flexible quality control and the determination of optimal process parameters, thereby reducing production costs and minimizing energy consumption during manufacturing.

**Keywords:** lithium-ion battery; machine learning; virtual quality gates; quality management



Academic Editor: Sheng S. Zhang

Received: 4 June 2025

Revised: 16 July 2025

Accepted: 21 July 2025

Published: 25 July 2025

**Citation:** Xu, X.; Stier, S.; Gronbach, A.; Moeckel, M. Deploying Virtual Quality Gates in a Pilot-Scale Lithium-Ion Battery Assembly Line. *Batteries* **2025**, *11*, 285. <https://doi.org/10.3390/batteries11080285>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The pilot production phase refers to the small-scale trial production conducted before entering full-scale manufacturing [1]. During this stage, aspects such as product design, manufacturing technique, equipment configuration, process parameters, and quality control measurements are validated and adjusted within a practical production environment [2]. The increasingly fierce market competition demands that manufacturing companies swiftly adapt to new technologies, new processing materials, and material or energy cost fluctuations. This imposes higher requirements on pilot production. To stay competitive, companies need to conduct pilot production efficiently while thoroughly investigating the interrelationships between quality parameters and process parameters during manufacturing. This is essential for determining the optimal process parameters and establishing a quality control system for mass production. Data-driven quality analysis offers solutions for handling such a challenge.

Once the product design, manufacturing techniques, and manufacturing equipment have been established, another critical objective during pilot production is to identify the quality-related process parameters. For each quality indicator, acceptable tolerance ranges must be defined according to the specified quality requirements and standards. Various studies have introduced quality assurance methods for individual processes within

battery production. For example, defect analysis of electrode coatings, optical inspection of electrode surface particles using optical camera systems after pre-treatment, and online detection systems for separator manufacturing have been extensively documented [3,4]. However, these quality assurance schemes targeting single processes do not automatically constitute a comprehensive quality management system that ensures overall product quality.

Due to the complexity of the LIB production chain, traditional quality management tools such as Statistical Process Control (SPC) and Process Capability Indices quickly encounter limitations when applied [5]. Thus, the concept of Quality Gates (QGs) has been introduced to address the challenges posed by complex interrelationships [6]. The core idea of this concept is to systematically divide the defined manufacturing chain into different quality-related decision points. At these QGs, the intermediate quality of the product or the predicted final quality indicators are inspected to ensure that the required quality attributes have been met before proceeding to a further manufacturing step [7]. To construct a valid machine learning (ML) model capable of predicting the quality characteristics of a product, readily available data from the production environment can be utilized. This forms the foundation for advanced quality management, enabling both quality enhancement and high process transparency at minimal cost. The maturity of digitalization solutions enables the timely aggregation and processing of process information during production, allowing decision-making related to quality control to be implemented in a data-driven manner [8].

Unlike data analyses that focus on pursuing high-accuracy predictions, data-driven quality analysis emphasizes the development of transparent and interpretable ML models. In this context, approaches such as ‘explainable ML’, ‘informed ML’, and similar methodologies have been widely discussed [9]. Not all algorithms are inherently interpretable. In this article, we follow the terminology set by [10] and adopt the term ‘interpretable’ for all forms of explainable ML models. One direction addresses informed ML, which includes domain-specific knowledge as constraints in the ML optimization loop. One example is physic-informed or physic-constrained ML [11], which allows us to ensure physical consistency for model results [12]. Interpretable ML models have been applied, for instance, in healthcare [13], process engineering, e.g., biomethane production [14], or lithium-ion battery (LIB) manufacturing [15]. In process engineering, the lack of interpretability can greatly exacerbate the difficulty of conducting comprehensive and in-depth process analyses and may challenge the reliability of data-driven predictions. These issues are magnified in the context of small data volumes, where data resources available for analysis are limited [16].

The use of ML for data analysis is, nevertheless, necessary in small-data contexts, especially when conventional data analysis methods are failing in multidimensional parameter spaces. In the pilot production phase, data analysis is often constrained by the lack of data resources. Even when available data volumes are abundant, cost-effective data analysis is necessary to reduce the total investment. Domain-specific feature engineering can be applied to address the challenges associated with limited data: Dawson-Elli et al. illustrate how the use of physically meaningful features leads to improved model accuracy [17]. Such feature selection allows for dimensionality reduction and contributes to the efficient utilization of limited data resources. It also enhances plausibility checks, as features can be designed to relate to meaningful quantities.

This paper presents the implementation of a cyber–physical system (CPS) for a lithium battery pilot assembly line. A machine learning (ML) based predictive model was employed to establish quality control mechanisms. The paper is structured as follows. Section 2 provides a brief overview of the assembly line for LIBs established in the laboratory. This assembly line is used as a case study to introduce the establishment and components of a quality evaluation system based on the concept of virtual quality gates. The deployment of

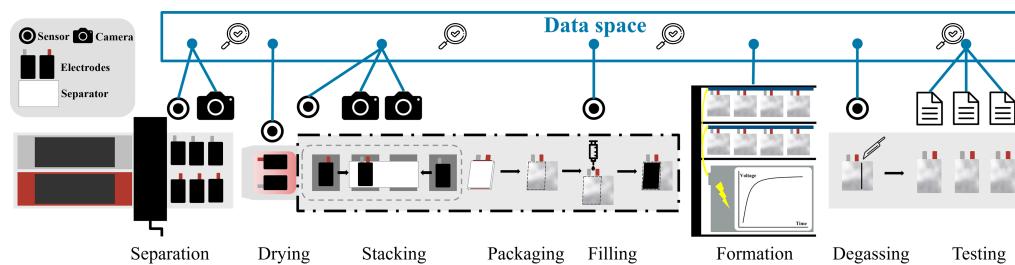
this virtual quality gate system, along with the actual data obtained during the project, is illustrated in Section 3. Section 4 summarizes the contribution of this paper.

## 2. Virtual Quality Gate for LIB Cell Assembly

This section focuses on LIB cell production with an emphasis on cell assembly. It outlines the procedure for establishing a quality control system based on quality gates for a pilot production line.

### 2.1. LIB Cell Production

The production of lithium-ion battery cells involves three main process steps: electrode manufacturing, cell assembly, and cell finishing. In this article, the cell assembly and the cell finishing are focused. Depending on the cell type, the cell assembly line may contain different manufacturing steps and machining methods. An assembly line for pouch cells in our laboratory is shown in Figure 1. It also indicates the sensors integrated into the production line for the monitoring of process parameters. The data collected by these sensors are transmitted to a unified data space. This ontology-based data space serves as one of the foundations for data-driven process analysis. Start with the roll material sheets; a separation step is carried out with a pouching tool. The components of a pouch, the anode and the cathode, are separated into individual groups. To control the rest of the moisture, the anodes and cathodes are dried in a post-drying step. Note that the post-drying is relative to the first drying after coating in electrode manufacturing. After the post-drying step, cell production requires a controlled atmosphere. This is usually achieved in a laboratory-scale production line with a glove box. As depicted in the figure, multiple anodes and cathodes are stacked into the Z-folded separator in the glove box during the stacking step. In a subsequent process, this stack is encapsulated and filled with electrolytes. In the cell finishing stage, the cell is activated by the first charge and discharge procedure. The gas generated during formation is removed, and the cell is finally sealed and ready for the end-of-line test (EoL-test). The complexity of battery cell production arises from the unintentional multidisciplinary interactions of the process parameters. Strongly coupled parameters influence the electrochemical reaction by acting directly or indirectly on certain characteristics of the battery, which further affects the final quality of the LIB.



**Figure 1.** A pilot LIB cell assembly line from separation to the end-of-line test; the magnifiers in data space indicate the presence of quality gates.

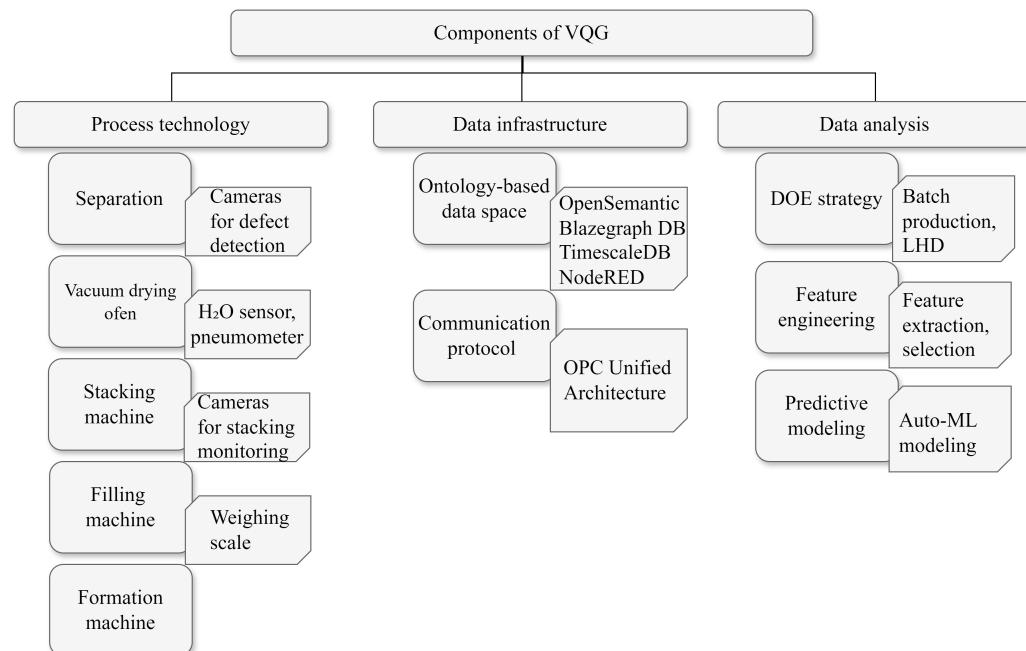
### 2.2. Establishing Quality Control Measures for Pilot Production Lines

As presented in Figure 2, building a quality control system for the pilot production line requires driving back from the final quality requirements to the production side. At the beginning, the product's final quality indicators (FQI) must be determined. For LIBs, the identified FQIs include cell discharging capacity, internal resistance, and open-circuit voltage (OCV). In addition to defining quality indicators directly, other production key performance indicators (KPIs) such as cost, material usage, or energy consumption can be incorporated to impose constraints on process parameters.



**Figure 2.** Driving back logic for the construction of a quality control system.

After establishing the FQIs, the next step is to identify quality-related process parameters. In this work, we introduce a correlation matrix approach based on expert opinions to identify quality-related process parameters. On this basis, technical components for data collection—such as sensors, data transmission systems, and data storage solutions—should be integrated into the production line. Afterward, initial production can be conducted to test and calibrate the readiness of these components, ensuring that the production line is fully prepared for trial production. Once the production line is ready for trial production, the next step is to conduct the trial runs and gather sufficient data for battery quality analysis. ML models for the prediction of specified FQIs can be developed with these data. These models are the key components of the virtual quality gate (VQG) quality control system, which can forecast the quality of intermediate products at various stages of production. For the LIB cell assembly line in this project, Figure 3 shows all the technical components used for deploying VQG in quality inspection.



**Figure 3.** Components of virtual quality gates for LIB assembly production.

### 3. Implementation of the VQG Based Quality Control for a Pilot Production Line

The small-data scenario brought about by pilot production presents new challenges for data analysis. Data analysis methods typically assume a sufficient amount of data, with some even requiring big data to be implemented. These methods may not yield meaningful results when applied to small data [18]. In response, this section demonstrates how data-driven predictive modeling can be implemented to LIB cell production in a small-data context. Detailed information, including materials, cell setup, chemical composition, and related cell characteristics, is summarized in the tables provided in the Appendix A.

### 3.1. Process Knowledge Organization

Before entering systematic dataset acquisition, the first step is to acquire and organize existing priori process knowledge, which refers mainly to the interconnections between process parameters and target parameters. Two main sources of process knowledge are the literature and expert knowledge by process engineers. A classical representation of process knowledge is given by the Design Structure Matrix (DSM) [19]. Typically, 0 indicates no correlation while 1 indicates full or strong correlation [20]. The intensity of the correlation can be quantitatively differentiated between 0 and 1 based on expert opinion.

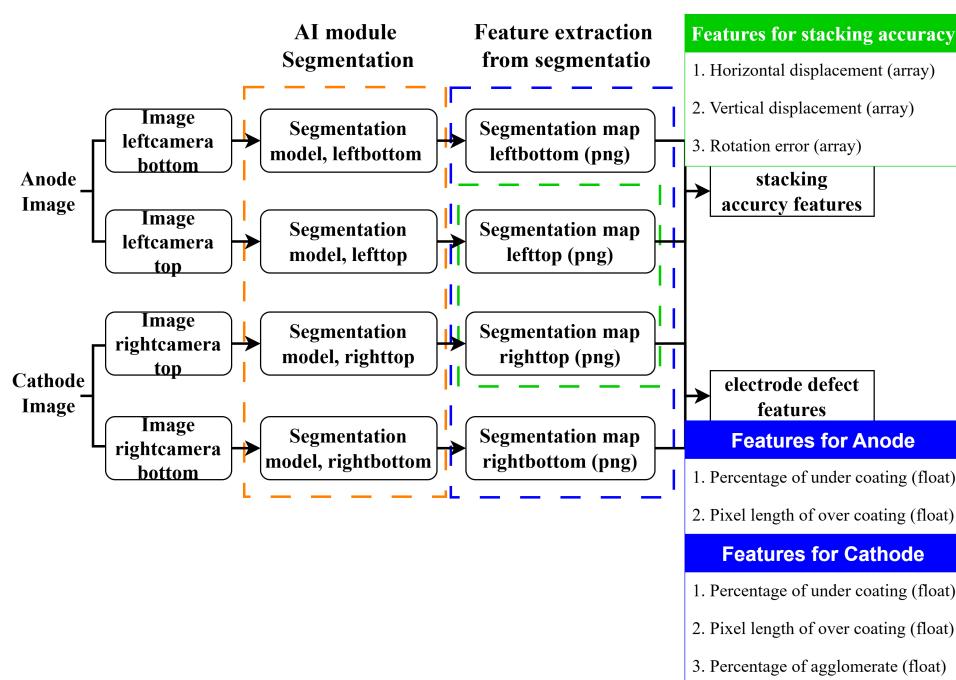
A DSM for the pilot LIB assembly production line in this work is provided in our repository. The LIB process parameters that require further study through data analysis are determined by this DSM and the project's KPIs. A plausible range of values for process parameters, as well as the acceptable range for the quality parameters, can be determined and included in the DSM. This provides basic information for Design of Experiments (DOE) planning. Moreover, the integration of expert knowledge, coupled with feature selection, effectively restricted the scale of the parameter space, thereby enabling the application of data-driven methods in scenarios with limited data resources [21].

### 3.2. Ontology-Based Data Space for Data Acquisition

The data space primarily comprises schema-less databases (NoSQL), including Document Stores and Triple/Quad-Stores (Graph Databases) [22,23]. The schema-less approach emphasizes the connections between data structures, eliminating the need for predefining each instance. Additionally, it can tolerate sparse or missing data, enabling the schema-less (NoSQL) approach to adapt more quickly and dynamically to changes and scalability requirements in pilot production compared to traditional SQL solutions. OpenSemanticLab software stack supports the annotation of process data as semantic documents, which are then mapped to a triple store (Blazegraph DB) and made available for dataset queries through SPARQL [24]. Continuous time-series data generated during production is collected using TimescaleDB. Data automatically captured by sensors during production is transmitted directly via an OPC-UA server client architecture. As an interactive web dashboard, NodeRED is employed for semi-automated data collection, such as data gathered from OPC-UA devices like weight sensors or camera images [25]. Automatically generated self-descriptions in the Resource Description Framework (RDF) are seamlessly integrated into the background of the data space while storing data. This enables the development of a publicly accessible platform that enhances the discovery and tracking of linked information and data. Furthermore, raw data generated from the LIB assembly process is regularly archived as RDF/JSON-LD exports on Zenodo under a CC-BY license. The platform <https://kiprobatt.de> (accessed on 20 July 2025) provides public access to the data space and related datasets.

### 3.3. Constructing Data-Efficient IPFs Through Feature Extraction

The constrained amount of data with pilot production requires minimizing the complexity of the parameter space. Cutting or merging unimportant process parameters with feature selection or dimensionality reduction strategies is necessary. The raw process data collected during the LIB assembly, such as images captured during the electrode stacking process or time-series data from LIB cycling tests, presents high complexity (high dimensionality). This complexity poses challenges for data processing and analysis, which contrasts with the requirements of small data scenarios. Therefore, an important step in data processing is to perform feature extraction to obtain clear and concise intermediate process features. Figure 4 demonstrates the feature extraction workflow using the image data collected for electrode defect monitoring as an example.



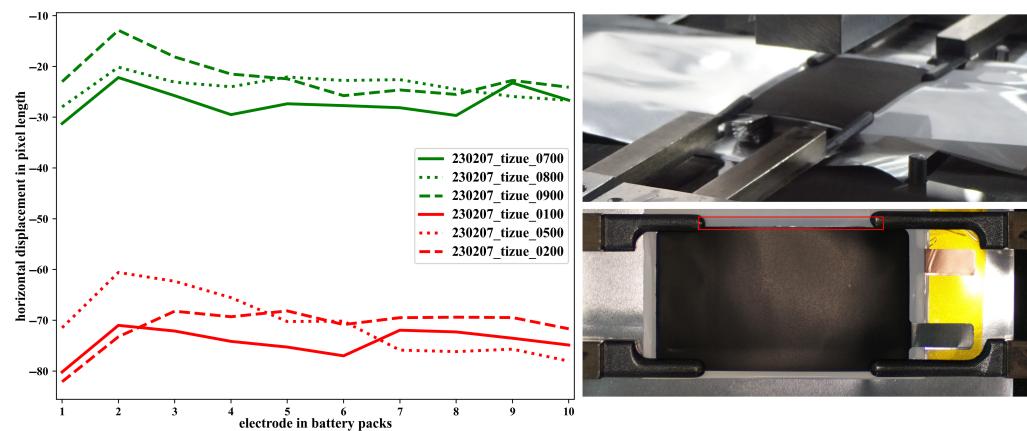
**Figure 4.** Data flow of the feature engineering process from electrode image to process features.

In this workflow, the raw data collected from LIB assembly are electrode images in JPEG format. Trained segmentation models are adopted to perform semantic segmentation on newly acquired electrode images. Semantic segmentation refers to the process of labeling each pixel in an image to classify the image into multiple segments. It is considered a pixel-level classification technique in image processing, providing higher accuracy than image classification or object detection. For anodes, apart from the irrelevant background, the goal of segmentation is to categorize the coating area and the conductive sheet made of copper. For cathodes, the background, the coating area, the conductive sheet made of aluminum, and the material-agglomerate of the coating material are considered.

Additionally, the positioning of the separators during the stacking process also needs to be determined through image recognition. However, the segmentation map remains too complex and voluminous for the quality analysis supported by the volume of pilot production data. Therefore, simplified numeric features are extracted for use in subsequent data analysis processes. For electrode sheets in the lithium battery assembly process, coating defects, including both overcoating and undercoating, are identified by the correlation matrix as strongly related to battery performance. Additionally, agglomerates appearing on the cathode are also extracted as numeric features. These feature extraction processes effectively reduce the dimensionality by compressing an image into a single numeric value. However, the segmentation map remains overly complex and voluminous in terms of the data volume arising from pilot production. As a result, simplified numeric features are desired in subsequent data analysis.

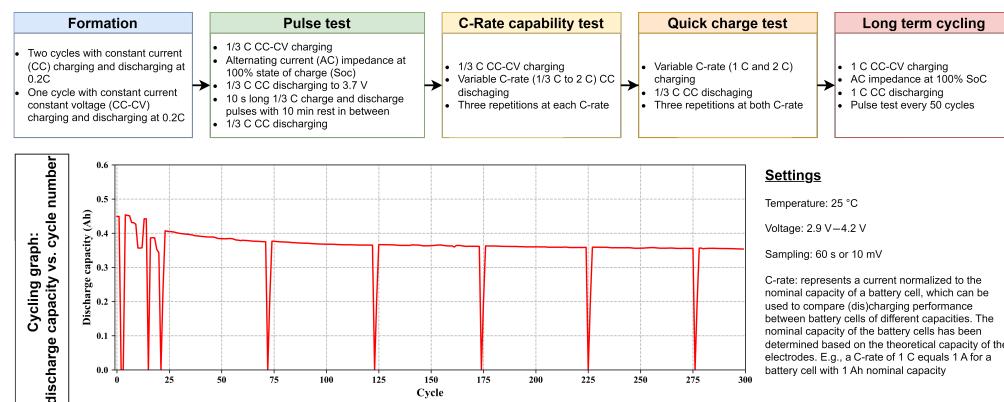
For electrodes in the LIB assembly process, coating defects, including both overcoating and undercoating, and agglomerates appearing on the cathodes are identified by the correlation matrix as strongly related to battery performance. Such feature extraction effectively reduces the dimensionality by compressing each image into a single numeric value. Stacking accuracy in the LIB assembly process is another process factor that correlates with battery performance. The LIB cells assembled on this pilot production line consist of five cathodes and six anodes. Therefore, stacking features should describe the entire cell pack. Features for stacking accuracy are defined by the relative displacement (horizontal and vertical) of each cathode with respect to the anode directly beneath it. For the entire cell

pack, the final coating characteristics and stacking accuracy can be regarded as array-type data. Compared to the initial image data, array-type data is more suitable for analysis when dealing with smaller data volumes. Figure 5 demonstrates LIBs with varying stacking accuracy based on the horizontal displacement feature extracted from the images.



**Figure 5.** Horizontal displacement between electrodes in the battery packs; green: cells with high stacking accuracy; red: cells with poor stacking accuracy. Photos on the right side show electrodes being stacked and the relative displacement between a cathode and an anode in the LIB cell (ID 230207-tizue-0500).

As can be seen in the figure, the horizontal displacement feature clearly distinguishes LIBs with different stacking accuracies. On the quality side, the battery's performance, obtained from cyclic tests as time-series data, also needs to be compressed to suit small data scenarios. The cycling procedure and the discharge capacity of an example battery cell during the cycling test are provided in Figure 6. Electrochemical features closely related to cell quality were extracted as FQIs for assessing the performance of the produced LIB cells. The process parameters and quality indicators specifically considered in this work are shown in Table 1. Process features refer to input variables derived from process parameters, which are directly applicable for data analysis. For each produced LIB cell, the values of its process features are collected to construct a final dataset for data analysis. Correlation analysis and the development of predictive models for quality control were conducted for some of these quality indicators.



**Figure 6.** Cycling procedure in EoL-test (top) and visualization of the output capacity of an example battery cell.

**Table 1.** Process parameters, corresponding process features, and final quality indicators.

Process Step	Process Parameter	Process Feature	FQIs
Separation	Coating defect	Under- and overcoating (anode, cathode)	—
Drying	Residual humidity	Drying time	—
Stacking	Stacking accuracy	Horizontal, vertical displacement	—
Filling	Electrolyte amount	Electrolyte amount	—
Cycling test	—	—	Discharging capacity, cell resistance, open-circuit voltage after formation

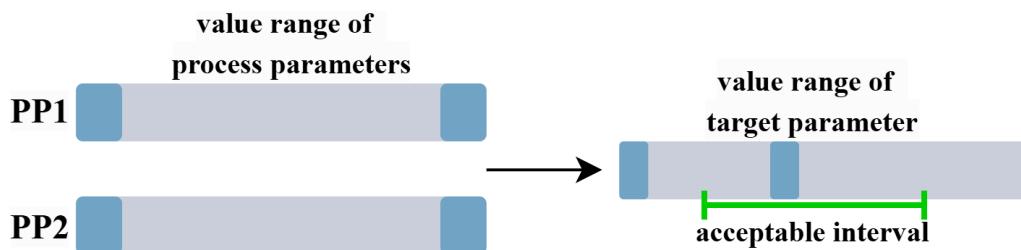
### 3.4. DOE Plan with Limited Data Volume

Purpose-built datasets for ML modeling with DOE may address two possible directions [26]:

1. Finding the optimal values of process variables that give rise to an optimal response
2. Exploring the defined parameter space around the optimal values to generate knowledge for monitoring, anomaly detection, and overall quality control

Prior to the realization of stable large-scale production, research and pilot-scale production primarily intend to acquire process expertise, optimize process parameters, or develop a quality control system for the production line. Accordingly, the prepared dataset should not be limited to a parameter space centered around a presumed optimum, but rather be designed to facilitate comprehensive exploration of the entire parameter space. Moreover, the concept of flexible production, which emphasizes increasing the adaptability and variability of the manufacturing process, necessitates a wide exploration of the parameter space during pilot production to identify potentially acceptable sets of process parameters. Therefore, this work focuses on the second direction in dataset construction.

Figure 7 presents a representative example that demonstrates the advantages of exploring a wider range of process parameters in pilot production. The blue regions represent the edge regions within the value ranges of PP1 and PP2, referring to extreme values (maximum or minimum) of the process parameters. When considered individually, an extreme value of a single PP may cause the target variable to fall outside the acceptable range. However, since PPs have a combined effect on the target variable, it is possible that certain combinations involving extreme values—or a single extreme value paired with appropriate settings of other PPs—can still result in acceptable outcomes. These acceptable PP sets do not have to be the optimal solutions within the entire parameter space. Instead, they provide more possibilities and greater flexibility to cope with potential changes.



**Figure 7.** Taking process parameters (PP) in the edge region may also lead to acceptable target parameter.

When the total available data resources are fixed, the chosen DOE strategy for an expanded parameter space should distribute the data points with higher efficiency. Therefore, traditional statistical DOE strategies may no longer be applicable in this context. Space-filling design and iterative sampling have shown advantages, particularly in small-data

scenarios [27]. Iterative active learning DOE strategies are more flexible than Space-filling strategies: they can continuously generate additional data points besides existing data. In contrast, LHD requires that the amount of data volume be specified at the beginning, which is not compatible with additional data generation [28]. However, iterative structure is not always the optimal choice. It has to start with an initial amount of data for subsequent iterations. If the model trained with the initial dataset does not drive active learning correctly, then the results of the iterations can be disastrous [29]. Moreover, iterative schemes are difficult to apply when data acquisition is time-consuming. For example, it may take over a month for the entire LIB cycling procedure to obtain a complete quality assessment [30]. In such a case, the time cost of an iterative scheme that considers one new data point at a time is unacceptable. Even if it iterates with batches, this time delay cannot be managed within the project's timeline.

In this work, the Latin Hypercube Design (LHD) strategy was chosen for data planning. Two independent batches of production were carried out, and the related process as well as quality parameters were collected. LHD strategy was adopted as the data generation scheme for these two batches. According to the selected DOE plan, the first batch was expected to contain 70 cells, and the second batch was expected to contain 30 cells. Finally, the actual number of data points available for analysis was 67 for the first batch and 21 for the second batch. The dataset generated in this work is attached in the repository: [https://github.com/xinchengxxc/KIproBatt\\_dataset](https://github.com/xinchengxxc/KIproBatt_dataset) (accessed on 20 July 2025).

### 3.5. Feature Selection

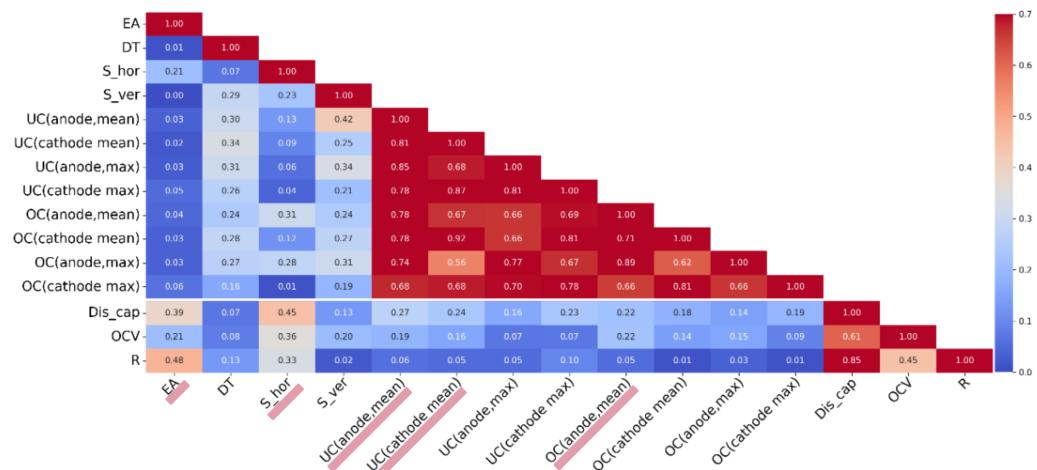
Table 1 presents the process parameters considered according to the DSM, along with the corresponding process features used in data analysis. The key quality indicators considered in data analysis are also listed.

As listed in Table 1, the discharging capacity refers to the cell discharging capacity obtained during the first long-term cycling test (the 25th cycle); the OCV is defined as the OCV obtained after formation procedure (after the 3rd cycle); the cell resistance is derived from the first pulse test (the 4th cycle). These FQIs were identified as the target parameters to be predicted. As for the process features, it should be noted that the features regarding the coating defect should apply to the pouch cell as a whole rather than to each electrode sheet in the cell. Thus, the process feature for coating defect for each pouch cell is characterized by a summary of the features of the electrodes it contains, presented as an average or maximum. Feature correlation analysis and modeling are performed solely on the training dataset with data points from the first batch. Data from the second batch serves as an additional independent test set for the evaluation of the trained predictive models.

Before starting the modeling, data-driven feature selection was conducted to quantitatively analyze the interrelations between quality data and process features. In this study, we used Spearman correlation coefficient [31] and the r-Boruta feature selection method, which is suitable for scenarios with small data sets, to perform feature selection [32].

The heatmap based on the Spearman correlation coefficient is separated into three parts. The first block (upper triangular region) records the inter-correlations between process parameters, the second block (lower rectangle area) records the correlations between process parameters and FQIs, and the third block (right triangular region) records the inter-correlations between FQIs. As shown in Figure 8, it is evident that the features related to coating defects extracted from the images exhibit high intercorrelations (dark area in the triangle on the left). Highly correlated features need to be screened for considerations of maintaining independence between input features and to reduce data demands for modeling. Therefore, the mean series features for coating defect, which have a higher

correlation with the quality parameters, were retained, while the max series features were excluded.



**Figure 8.** r-Boruta feature selection with discharging capacity as target parameter, ‘EA’: ‘Electrolyte Amount’, ‘DT’: ‘Drying time’, ‘S\_hor’: ‘stacking\_horizontal’, ‘S\_ver’: ‘stacking\_vertical’, ‘UC(anode,mean)’: ‘under coating\_anode\_mean’, ‘UC(cathode mean)’: ‘under coating\_cathode\_mean’, ‘UC(anode,max)’: ‘under coating\_anode\_max’, ‘UC(cathode max)’: ‘under coating\_cathode\_max’, ‘OC(anode,mean)’: ‘over coating\_anode\_mean’, ‘OC(cathode mean)’: ‘over coating\_cathode\_mean’, ‘OC(anode,max)’: ‘over coating\_anode\_max’, ‘OC(cathode max)’: ‘over coating\_cathode\_max’, ‘Dis\_cap’: ‘Cap\_out(Cyc25)’, ‘OCV’: ‘OCV(Cyc4)’, ‘R’: ‘DC(Cyc4)’.

On this basis, model-based correlation analysis methods are employed to further filter out process features with low correlation to the target parameters. r-Boruta is a wrapper-based feature selection algorithm built upon Random Forest [33]. For each original feature, r-Boruta generates a corresponding ‘shadow feature’ by randomly shuffling the feature values (column-wise), assuming that these shadow features have no correlation with the target parameter. Based on this assumption, r-Boruta constructs a selection criterion to identify and retain only those features that are significantly associated with the target. Only when the importance of a real feature surpasses that of all shadow features is the significance of this feature acknowledged. However, with small sample sizes, using a standard p-percentage in Boruta may risk over-deleting variables due to chance correlations. In a modified approach, r-Boruta calculates p-percentage by setting it based on the maximum correlation with the response variable, reducing the risk of over-deletion when sample sizes are small. Table 2 records the parameters inside r-Boruta and the results of the feature selection with discharging capacity as the target FQI.

**Table 2.** r-Boruta feature selection with discharging capacity as target parameter.

r-Boruta Hyperparameter	Selected Process Features
verbose = 2	Electrolyte amount
max_iter = 70	Horizontal displacement (stacking)
perc = 42.5	Under_coating_mean (anode)
—	Over_coating_mean (anode), Over_coating_mean (cathode)

By examining Table 2 and Figure 8, it can be observed that the feature selection results of r-Boruta align with those of the Spearman correlation coefficient: the Spearman correlation coefficients between the process parameters selected by r-Boruta and the discharging capacity are all above 0.2.

### 3.6. Establishment of the Predictive Models

Auto-sklearn [34] was employed to build predictive models for the three quality indicators independently. Auto-sklearn with the same parameter settings is adopted to model both the selected features and the original features. All modeling tasks were performed on a workstation equipped with an Intel Core i7-10700 CPU running at 2.90 gigahertz. Considering that the dataset is a small-scale dataset, a modeling time of 400 s was set for each modeling task to prevent overfitting. Table 3 presents the parameter settings used in Auto-sklearn for model establishment.

**Table 3.** Auto-sklearn parameter settings for regression modeling.

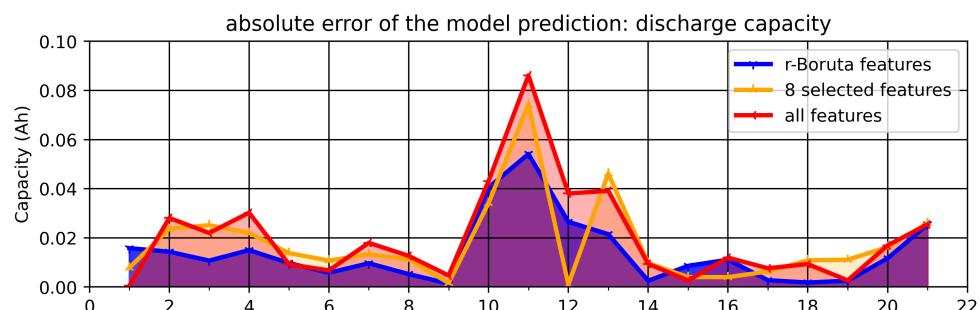
Auto-Sklearn Parameter Setting	Value/Description
Total time allocated for the task	400 s
Maximum time per model training run	40 s
Initial configurations via metalearning	25 configurations
Memory limit	4096 MB
Resampling strategy	Cross-validation
Resampling strategy arguments	folds: 5
Number of parallel jobs	2

Table 4 presents the R-squared scores and root mean square error (RMSE) for each model on the test dataset with cell discharge capacity, cell resistance, and OCV as target parameter respectively.

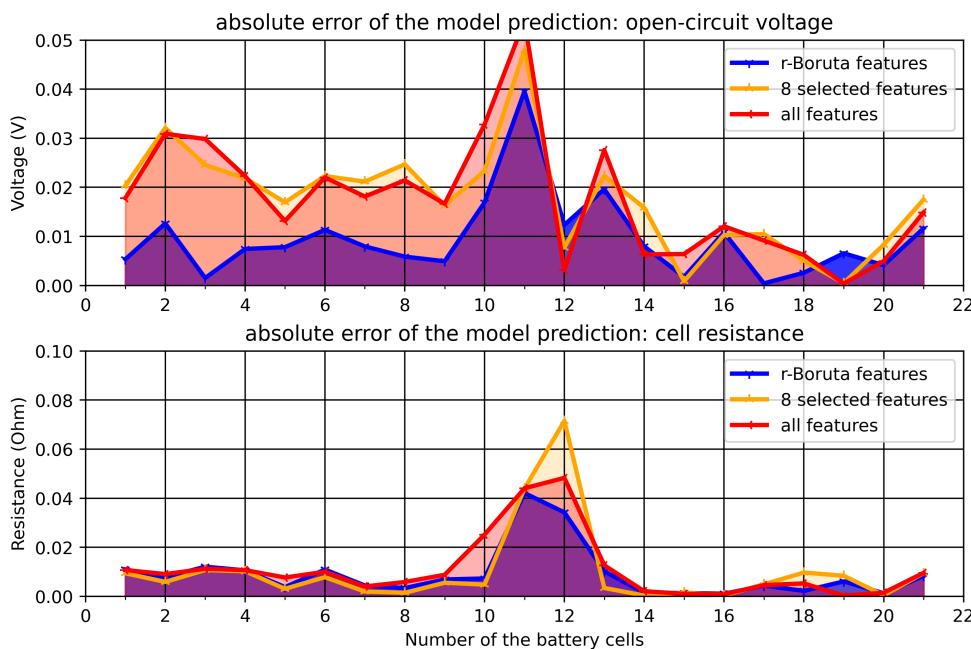
**Table 4.** Performance of the predictive models on the test set.

ML Models	RMSE (Test)	R <sup>2</sup> Score (Test)
Discharge Capacity		
with 5 r-Boruta features	0.019	0.664
with 8 selected features	0.024	0.460
with all features	0.028	0.292
Open-Circuit Voltage		
with 3 r-Boruta features	0.012	0.502
with 8 selected features	0.021	<0
with all features	0.065	<0
Cell Resistance		
with 6 r-Boruta features	0.013	0.585
with 8 selected features	0.020	0.167
with all features	0.017	0.366

Figure 9 illustrates the absolute error in the prediction of models with different process features for different target parameters. The performance of trained models on the test data indicates that the feature selection process effectively enhances the generalizability of the predictive models.



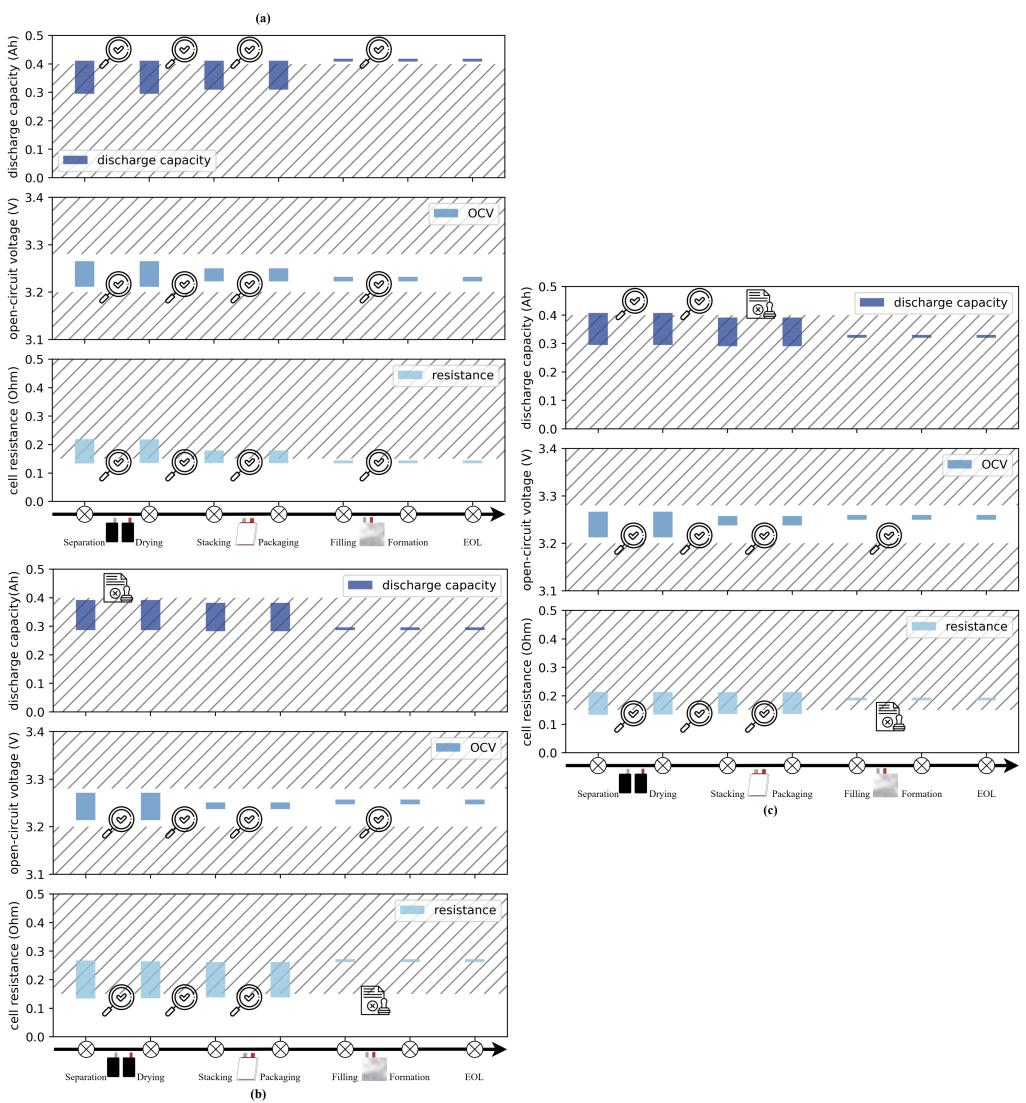
**Figure 9. Cont.**



**Figure 9.** Absolute error in the prediction of models with different features, from top to bottom are the predictive models for discharge capacity, cell resistance, and OCV as quality indicators.

### 3.7. Deployment of the VQG for Early Rejection of Low-Performance Cell

The manufacture of LIB is energy intensive and therefore efforts are made to reduce energy consumption in line with the greenhouse gas reduction targets set by the European Union [35]. There is a need for material reduction and inline recycling to increase material efficiency. The early elimination of insufficient cells by avoiding unnecessary production steps contributes to both of these goals. In the conventional VQG framework, quality control is conducted through in-process threshold-based virtual quality gates [36]. Product properties at intermediate process steps are assessed with respect to predefined value ranges of process parameters. If an intermediate process feature falls outside its tolerances, the product is labeled unqualified and rejected. By utilizing machine learning predictive models, it is feasible to forecast LIB final quality indicators during the production process while assessing a broader range of process characteristics. However, by employing ML-based predictive models, it becomes possible to predict the FQIs during production, evaluating a larger collection of process features simultaneously. Such data-driven frameworks replace the threshold-based VQG with model predictions. At a specific stage of the process, an ML model can create VQGs by adopting currently available process features for the predictions of final quality indicators. For newly manufactured LIB cells, the trained model can predict potential future FQIs based on provided process parameter values. However, the final quality outcome is still influenced by the process parameter settings applied in subsequent production steps. The quality control in production is conducted with the following steps. First, quality criteria of the FQIs must be defined. As shown in Figure 10, cells with a discharge capacity below 0.4 Ah or an internal resistance above 0.15  $\Omega$  are classified as inferior. For the OCV after formation, the acceptable value range is between 3.2 V and 3.275 V. The quality criteria were assessed by experts based on the obtained data. Then, model-based quality prediction is performed at the end of each process step to forecast the potential FQIs of the product.

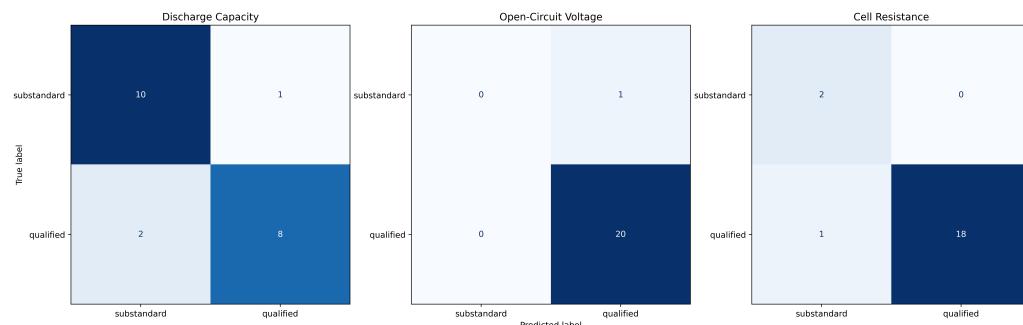


**Figure 10.** Early quality assessment based on model prediction. The expectation of the range of possible outcomes (bar plot) is shown. Subplot (a): a qualified battery cell with ID KI-230718-0400 monitored throughout the production process; Subplot (b): a defective battery cell with ID KI-230720-0300 should be rejected after separation; Subplot (c): rejection of a battery cell with ID KI-230720-0300 after stacking as quality specifications cannot be reached any more.

Figure 10 presents a visualization of VQG-based quality control for three cells produced on the pilot production line. The cell with ID KI-230718-0400 (subplot (a)) is a qualified product. The predicted FQI values remain within the acceptable quality range at each stage during the production process. In terms of discharge capacity, for instance, the predicted optimal discharge capacity (i.e., the highest achievable value) consistently meets the quality criteria for discharge capacity across all stages from separation to post-filling. Therefore, this cell is classified as qualified. Subplot (b) shows an example of a defective cell (ID KI-230720-0300). Based on discharge capacity, the trained model predicts that its optimal achievable discharge capacity after separation falls below the established quality criteria, meaning it can be classified as defective right after the separation stage. Additionally, based on cell resistance, this cell can also be identified as defective after the filling stage. For a product to be considered qualified, each predefined FQI must meet the specified quality criteria. Subplot (c) on the far right shows another example of a defective cell (ID KI-230720-0300). After the stacking stage, its predicted optimal discharge capacity falls below the specified quality criteria. Therefore, this cell should be classified as defective.

after the stacking process. If the predicted FQI range falls entirely outside the preset quality criteria, the battery manufacturing process cannot be successfully completed with any production procedures. This triggers an early rejection of the battery immediately after this process. However, if producing lower-performance secondary batteries is justified by different business objectives, the battery can be labeled as a secondary product and continue through the production process.

Taking substandard cells as positives, Figure 11 demonstrates the predictive capability of the three prediction models after filling from the perspective of identifying substandard cells. All FQIs need to be considered in the final determination of whether a battery is qualified. A battery shall be deemed qualified only if all its FQIs fall within their specified thresholds. Table 5 provides the precision and recall for the VQG-based quality control at each production stage. Attention should be put on precision at each stage, which represents the proportion of predicted substandard batteries that are truly substandard.



**Figure 11.** Assessment of the trained predictive models for discharge capacity, OCV, and cell resistance with confusion matrices.

**Table 5.** Precision and recall of the VQG-based quality control.

Production Stage	Separation	Stacking	Filling
Precision	80%	88%	83%
Recall	36%	73%	91%

The precision values for this VQG-based quality control at each stage of the production have all exceeded 80%. This performance is considered acceptable considering the limited data resource [37]. In general, the recall value reflects the proportion of substandard batteries that are correctly identified as substandard. However, in the VQG-based quality control system, a defective battery does not necessarily indicate that it was defective at earlier stages of production. Some unqualified batteries may have become defective due to inappropriate process parameters in subsequent stages (e.g., during the filling process), which caused them to fall out of qualified thresholds. The recall values in Table 4 increase as the production progresses, reaching 91% after the filling stage. This suggests that the VQG system is capable of identifying the vast majority of defective batteries by the end of the production process.

#### 4. Conclusions

This article presents the development and implementation of a CPS for quality control in the pilot production of lithium-ion batteries. The proposed system consists of process sensors, an ontology-based data space, and a data infrastructure, all integrated into a pilot assembly line that supports the creation of a machine learning-based predictive model for the early detection and evaluation of battery quality. The impact of limited data volume on data-driven solutions was mitigated by preparing the dataset using appropriate DOE strategies. The proposed ontology-based data space enables a transparent overview of

the entire data flow and offers an effective way to address challenges caused by potential errors in the data during the pilot production stage. The VQG-based quality control system ensures real-time, data-driven quality control by systematically identifying and analyzing quality-related process parameters. The article focuses on the common issue of data volume limitations in pilot production environments. Selecting appropriate DOE strategies for experiment/production planning, as well as performing dimensionality reduction and feature selection during data analysis, can enhance the efficiency of data analysis, enabling modeling tasks to be accomplished with small data. Results indicate that the proposed VQG framework successfully predicts end-of-line quality, which can improve flexibility and energy consumption efficiency for the adaptation of variations. As a subsequent step, the VQG quality control system developed in this study will be tested and improved in a more industrially relevant production environment with an expanded scale.

**Author Contributions:** Investigation, A.G.; Methodology, X.X. and S.S.; Resources, A.G.; Software, X.X. and S.S.; Supervision, M.M.; Writing—original draft, X.X.; Writing—review and editing, M.M. All authors will be updated at each stage of manuscript processing, including submission, revision, and revision reminder, via emails from our system or the assigned Assistant Editor. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the German Federal Ministry of Education and Research (BMBF), grant no. 03XP0309A-C.

**Institutional Review Board Statement:** Not applicable

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The dataset used in the paper as well as the DSM for process correlations are posted at: [https://github.com/xinchengxxc/KiproBatt\\_dataset](https://github.com/xinchengxxc/KiproBatt_dataset) (accessed on 20 July 2025). The project platform <https://kiprobatt.de> (accessed on 20 July 2025) and the related dataset provide public access to the results.

**Acknowledgments:** This work was supported in part by the German Federal Ministry of Education and Research (BMBF), grant no. 03XP0309A-C. The project platform <https://kiprobatt.de> (accessed on 20 July 2025) and the related dataset provide public access to the results.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

CPS	Cyber–Physical System
ML	Machine Learning
SPC	Statistical Process Control
QG	Quality Gates
LIB	Lithium-ion Battery
EoL-test	End-of-Line test
FQI	Final Quality Indicators
OCV	Open-circuit Voltage
KPI	Key Performance Indicator
VQG	Virtual Quality Gate
DSM	Design Structure Matrix
DOE	Design of Experiments
RDF	Resource Description Framework
LHD	Latin Hypercube Design
RMSE	Root Mean Square Error
PP	Process Parameter

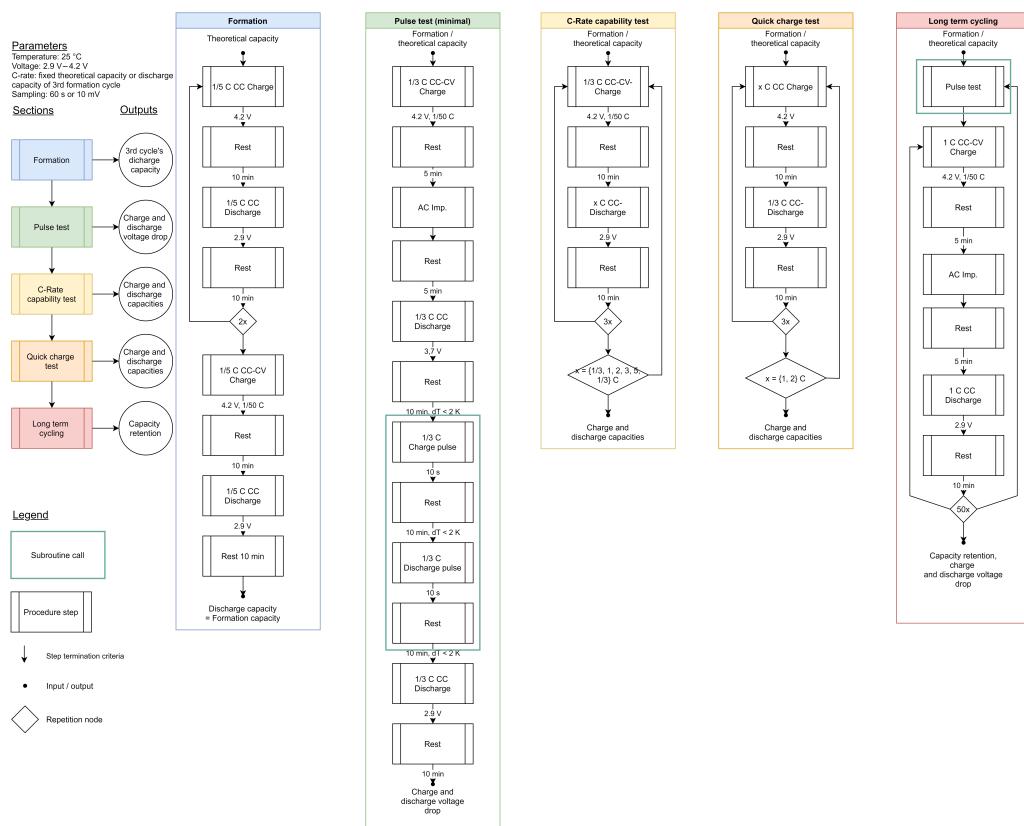
## Appendix A

**Table A1.** Battery component setup and materials.

Component	Description
Active Material	Graphite 94%, 3.5 mAh/cm <sup>2</sup> (Theoretical capacity)
Additive	Carbon Black 1%
Binder	CMC 2%, SBR 3%
Electrolyte	LP57-like electrolyte (1M LiPF <sub>6</sub> in EC:EMC 3:7 by wt) + 2% VC
Current Collector	Copper foil, 10 µm
Separator	Celgard 2320-0810M-E
Cathode Composition	NMC622 95.5%, CB/Graphite 3%, PVDF ~1.5%; 3.1 mAh/cm <sup>2</sup> , 15 µm Al foil
Electrode Density	Anode: 1.5 g/cm <sup>3</sup> ; Cathode: 3.0–3.2 g/cm <sup>3</sup>
Electrode Thickness	Not controlled
Cell Format	Pouch cell, full cell
Electrochemically Active Area	150 cm <sup>2</sup>

**Table A2.** Capacity measurement and testing conditions.

Metric	Value/Description
Theoretical Areal Capacity	Anode: 3.5 mAh/cm <sup>2</sup>   Cathode: 3.1 mAh/cm <sup>2</sup>
Theoretical Specific Capacity	Anode: 372 mAh/g   Cathode: 184 mAh/g
Areal Mass Loading	Anode: 0.0094 g/cm <sup>2</sup>   Cathode: 0.0169 g/cm <sup>2</sup>
Electrochemically Active Area	150 cm <sup>2</sup>
Initial Capacity (theoretical)	450 mAh (theoretical: 465 mAh)
Initial Capacity (mAh/g)	Anode: 319   Cathode: 177
Capacity Achieved (mAh/g)	Anode: 258   Cathode: 143
Test Conditions	1C rate, 135 cycles
Total Cycles Tested	300
Capacity Retention	81%
Testing Temperature	25 °C
Characterization Techniques	Charge/discharge test, Electrochemical Impedance Spectroscopy (EIS), Rate capability, Quick charge, Pulse test



**Figure A1.** Detailed procedure of lithium battery charge/discharge cycle test.

## References

- Chinnathai, M.K.; Alkan, B.; Vera, D.; Harrison, R. Pilot to Full-Scale Production: A Battery Module Assembly Case Study. *Procedia CIRP* **2018**, *72*, 796–801. [[CrossRef](#)]
- Almgren, H. Pilot Production and Manufacturing Start-Up: The Case of Volvo S80. *Int. J. Prod. Res.* **2000**, *38*, 4577–4588. [[CrossRef](#)]
- Zhu, G.; Wang, Y.; Yang, S.; Qu, Q.; Zheng, H. Correlation between the Physical Parameters and the Electrochemical Performance of a Silicon Anode in Lithium-Ion Batteries. *J. Mater.* **2019**, *5*, 164–175. [[CrossRef](#)]
- Mohanty, D.; Hockaday, E.; Li, J.; Hensley, D.K.; Daniel, C.; Wood, D.L. Effect of Electrode Manufacturing Defects on Electrochemical Performance of Lithium-Ion Batteries: Cognizance of the Battery Failure Sources. *J. Power Sources* **2016**, *312*, 70–79. [[CrossRef](#)]
- Rogers, A.J.; Hashemi, A.; Ierapetritou, M.G. Modeling of Particulate Processes for the Continuous Manufacture of Solid-Based Pharmaceutical Dosage Forms. *Processes* **2013**, *1*, 67–127. [[CrossRef](#)]
- Schnell, J.; Reinhart, G. Quality Management for Battery Production: A Quality Gate Concept. *Procedia CIRP* **2016**, *57*, 568–573. [[CrossRef](#)]
- Aumi, S.; Corbett, B.; Clarke-Pringle, T.; Mhaskar, P. A Study in Advanced Process Control. *AICHE J.* **2013**, *59*, 2852–2861. [[CrossRef](#)]
- Hebert, J. Predicting Rare Failure Events Using Classification Trees on Large Scale Manufacturing Data with Complex Interactions. In Proceedings of the 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 5–8 December 2016; pp. 2024–2028. [[CrossRef](#)]
- Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-Informed Machine Learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440. [[CrossRef](#)]
- Roscher, R.; Bohn, B.; Duarte, M.F.; Garcke, J. Explainable Machine Learning for Scientific Insights and Discoveries. *IEEE Access* **2020**, *8*, 42200–42216. [[CrossRef](#)]
- Zhao, W.L.; Reichstein, M.; Zhang, Y.; Zhou, S.; Wen, Y.; Lin, C.; Li, X.; Qiu, G.Y. Physics-Constrained Machine Learning of Evapotranspiration. *Geophys. Res. Lett.* **2019**, *46*, 14496–14507. [[CrossRef](#)]
- Kashinath, K.; Mustafa, M.; Albert, A.; Wu, J.-L.; Jiang, C.; Esmaeilzadeh, S.; Azizzadenesheli, K.; Wang, R.; Chattopadhyay, A.; Singh, A.; et al. Physics-Informed Machine Learning: Case Studies for Weather and Climate Modelling. *Phil. Trans. R. Soc. A* **2021**, *379*, 20200093. [[CrossRef](#)]
- Ahmad, M.A.; Eckert, C.; Teredesai, A. Interpretable Machine Learning in Healthcare. In Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, Washington, DC, USA, 29 August–1 September 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 559–560. [[CrossRef](#)]
- De Clercq, D.; Wen, Z.; Fei, F.; Caicedo, L.; Yuan, K.; Shang, R. Interpretable Machine Learning for Predicting Biomethane Production in Industrial-Scale Anaerobic Co-Digestion. *Sci. Total Environ.* **2020**, *712*, 134574. [[CrossRef](#)]
- 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](#)]
- Xu, X.; Bi, J.; Moeckel, M.; Wiemer, H.; Ihlenfeldt, S. Quantitative assessment of data volume requirements for reliable machine learning analysis. *IEEE Access* **2025**, *13*, 101545–101557. [[CrossRef](#)]
- Dawson-Elli, N.; Lee, S.B.; Pathak, M.; Mitra, K.; Subramanian, V.R. Data Science Approaches for Electrochemical Engineers: An Introduction through Surrogate Model Development for Lithium-Ion Batteries. *J. Electrochem. Soc.* **2018**, *165*, A1. [[CrossRef](#)]
- Brigato, L.; Iocchi, L. A Close Look at Deep Learning with Small Data. In Proceedings of the 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 2490–2497.
- Westermeier, M. Qualitätsorientierte Analyse Komplexer Prozessketten am Beispiel der Herstellung von Batteriezellen. Ph.D. Thesis, Technische Universität München, München, Germany, 2016.
- Browning, T.R. Applying the design structure matrix to system decomposition and integration problems: A review and new directions. *IEEE Trans. Eng. Manag.* **2001**, *48*, 292–306. [[CrossRef](#)]
- Meng, C.; Griesemer, S.; Cao, D.; Seo, S.; Liu, Y. When physics meets machine learning: A survey of physics-informed machine learning. *Mach. Learn. Comput. Sci. Eng.* **2025**, *1*, 20. [[CrossRef](#)]
- Wang, L.; Zhang, S.; Shi, J.; Jiao, L.; Hassanzadeh, O.; Zou, J.; Wang, C. Schema management for document stores. *Proc. Vldb Endow.* **2015**, *8*, 922–933. [[CrossRef](#)]
- Yang, Z.; Rajagopal, A.; Chueh, C.C.; Jo, S.B.; Liu, B.; Zhao, T.; Jen, A.K. Stable Low-Bandgap Pb-Sn Binary Perovskites for Tandem Solar Cells. *Adv. Mater.* **2016**, *28*, 8990–8997. [[CrossRef](#)]
- Sikos, L.F. *Mastering Structured Data on the Semantic Web: From HTML5 Microdata to Linked Open Data*, 1st ed.; Apress: Berkeley, CA, USA, 2015; pp. xviii–256.
- Lekić, M.; Gardašević, G. IoT sensor integration to Node-RED platform. In Proceedings of the 17th International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia and Herzegovina, 28 February–2 March 2018; pp. 1–5.
- Dean, A.; Voss, D.; Draguljić, D. *Design and Analysis of Experiments*, 2nd ed.; Springer: New York, NY, USA, 2017; p. 180.

27. Xu, X. AutoML based workflow for design of experiments (DOE) selection and benchmarking data acquisition strategies with simulation models. *Sci. Rep.* **2024**, *14*, 32170. [[CrossRef](#)]
28. McKay, M.D.; Beckman, R.J.; Conover, W.J. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics* **1979**, *21*, 239–245. [[CrossRef](#)]
29. Xu, X.; Conrad, F.; Xingyu, X.; Loeprecht, O.; Möckel, M. Comparative Analysis of Small Data Acquisition Strategies in Machine Learning Regression Tasks Addressing Potential Uncertainties. *Int. J. Adv. Softw.* **2023**, *16*, 243–253.
30. Zhao, J.; Gao, Y.; Guo, J.; Chu, L.; Burke, A.F. Cycle life testing of lithium batteries: The effect of load-leveling. *Int. J. Electrochem. Sci.* **2018**, *13*, 1773–1786. [[CrossRef](#)]
31. Spearman, C. The Proof and Measurement of Association between Two Things. *Am. J. Psychol.* **1904**, *15*, 72–101. [[CrossRef](#)]
32. Kaneko, H. Examining variable selection methods for the predictive performance of regression models and the proportion of selected variables and selected random variables. *Heliyon* **2021**, *7*, e07356. [[CrossRef](#)]
33. Smith, P.F.; Ganesh, S.; Liu, P. A comparison of random forest regression and multiple linear regression for prediction in neuroscience. *J. Neurosci. Methods* **2013**, *220*, 85–91. [[CrossRef](#)]
34. Feurer, M.; Eggensperger, K.; Falkner, S.; Lindauer, M.; Hutter, F. Auto-sklearn 2.0: Hands-free AutoML via meta-learning. *J. Mach. Learn. Res.* **2022**, *23*, 261.
35. Hof, A.F.; den Elzen, M.G.J.; Mendoza Beltran, A. The EU 40% greenhouse gas emission reduction target by 2030 in perspective. *Int. Environ. Agreements* **2016**, *16*, 375–392. [[CrossRef](#)]
36. Filz, M.-A.; Gellrich, S.; Turetsky, A.; Wessel, J.; Herrmann, C.; Thiede, S. Virtual Quality Gates in Manufacturing Systems: Framework, Implementation and Potential. *J. Manuf. Mater. Process.* **2020**, *4*, 106. [[CrossRef](#)]
37. Viering, T.; Loog, M. The Shape of Learning Curves: A Review. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 7799–7819. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.