Data science accelerates energy device development

Shuan Cheng,[1,3] Xiao Cui,[2,3] William C Chueh,[2] Shijing Sun[1,*]

[1]Department of Mechanical Engineering, University of Washington, Seattle, WA 98105, USA
[2]Department of Material Science and Engineering, Stanford University, Stanford, CA 94305, USA
[3]These authors contributed equally
*Correspondence: shijing@uw.edu

The Bigger Picture

Rapid development of clean energy technologies is critical for achieving a sustainable and carbon-neutral society. Before rolling out new technology, forecasting future device performance is crucial. Traditionally, this involves extensive testing, a process that is not only time-consuming but also expensive. The integration of data science tools into the development process is now being explored worldwide to expedite testing and diagnostics in a cost-effective manner. This perspective reviews several strategies to augment the data-driven workflows in energy device development, taking solar cells, batteries, and fuel cells as examples. Data science and machine learning (ML) techniques are increasingly used for predicting the longevity and state of health of functional devices. There is also growing interest in understanding why failures occur and in optimizing process parameters through explainable ML. Beyond laboratory testing, the concept of digital manufacturing introduces closed-loop platforms that leverage predictive modeling and ML-guided optimization to improve production processes. While there are challenges in achieving seamless integration of artificial intelligence with traditional human-centric manufacturing workflows, the potential benefits for the advancement of energy devices are significant. Device informatics as a field not only has an impact on clean energy research but is also applicable in many fields in science and engineering, offering accelerated pathways for shortening the time-to-market time of new technologies.

Summary

Data science has become increasingly prevalent in the development of energy devices, offering significant advancements in predicting future behaviors and identifying optimal process parameters in a resource-saving manner. This perspective begins by examining the role of data science and ML in enhancing accelerated aging tests across solar, battery and fuel cells. We present a generalizable data-driven workflow for processing aging test data and predicting the lifespan of different device types. In this perspective, we discuss two strategies to improve our understanding of device failures: integrating physics-based parameters and utilizing interpretable machine learning (ML) techniques. Following a brief review on ML-assisted process optimization, we propose an interpretable closed-loop platform towards digital manufacturing for thin-film solar and Li-ion battery production. Finally, we discuss the current challenges and research gaps in applying data science for accelerated energy device development, aiming to spark further investigation in this field.

Introduction

To address today's pressing environmental and climate challenges, our society is moving towards a sustainable, carbon-neutral future. This crucial transition is underpinned by the widespread electrification of our energy systems. As a main pathway towards decarbonization, electrification spans across energy generation, storage, transmission, and utilization. Over the past decade, we have observed a increasing adoption of renewable energy sources, such as solar, wind, and geothermal energy,[1] paralleled by a technological shift in the transportation sector towards electric and hybrid vehicles.[2] The pursuit of more efficient, reliable, and affordable clean technologies necessitates innovation at every stage, from materials discovery and device testing to system integration and deployment. The advent of high-performance computing, robotics, and artificial intelligence (AI) has heralded digitalization as pivotal accelerators in the development of clean energy technologies, aligning with the Industry 4.0 paradigm of integrating data science tools across research, development (R&D) and manufacturing.[3] Over the past few years, data-driven approaches have emerged as a powerful tool to reduce the lab-to-market time for new materials, as demonstrated from virtual molecular screening to autonomous synthesis of inorganic materials.[4–6] These advances in accelerated materials discovery have led to a substantial increase in candidates poised for potential applications in batteries, hydrogen production, and solar cells. Consequently, there's a need to speed up the development and testing of new devices for both providing timely feedback to materials R&D and for pushing new technologies to deployment. However, few reports have focused on how data-driven approaches can be applied to accelerate testing and diagnostics at the device level. Depending on their functionalities, each device system has its own design principles, material limitations, and aging mechanisms, and therefore demands domain-specific techniques to automate and digitalize existing fabrication workflows and testing protocols. Nonetheless, common challenges exist across device types, including to extend the device's lifetime, to understand the root causes of failures, and to optimize process parameters for high-quality, low-cost, and reproducible device production.[7–9]

In this perspective, we discuss how data science tools, including data-driven workflows and machine learning (ML) techniques, can be employed to assist resolving some of the common challenges in energy device development. We review recent studies that employed informatics tools for tasks such as optimizing performances in a high-dimensional space, reducing the long aging test duration, and drawing insights from the complex and convoluted experimental observations. Using thin-film solar cells, lithium-ion (Li-ion) batteries, and proton exchange membrane (PEM) fuel cells as examples (as illustrated in Fig. 1), we evaluate the concept of generalizable data-driven workflows in laboratory R&D that are expected to be applicable across scientific domains. We also share our perspective on the future of digital manufacturing, where interpretable ML and active learning can be combined forming workhorse workflows for multi-parameter optimization at scale. Finally, we discuss open challenges centered around data availability, scientific understanding, and hardware integration. Overall, we envision data-driven frameworks and models to be increasingly incorporated in future development for functional devices, leading to a paradigm shift in clean energy innovation.

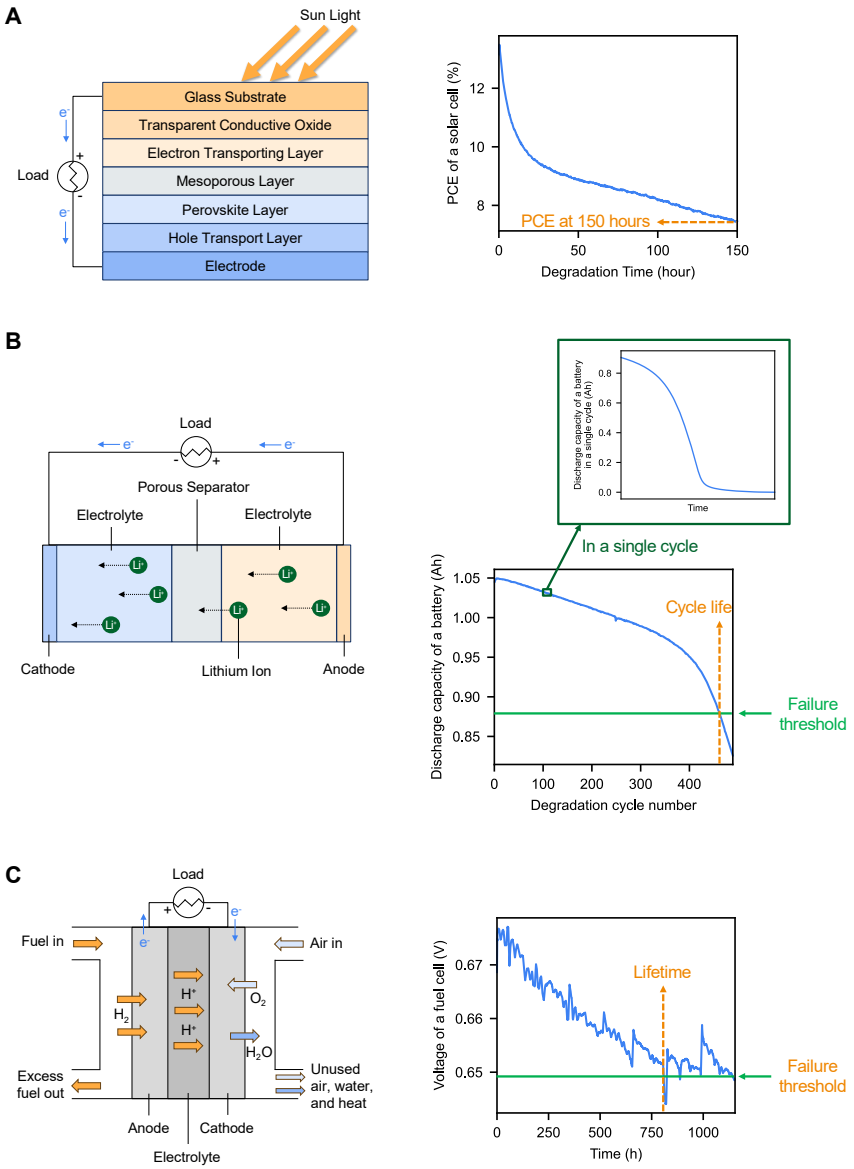

Figure 1. Schematic drawings and degradation curves for different renewable energy devices.
(A) Schematic drawing (left) and degradation curve (right) for a perovskite solar cell, with its power conversion efficiency (PCE) measured to indicate its degradation performance. The evaluation criteria is the PCE at the end.
(B) Schematic drawing (left) and degradation curve (right) for a lithium-ion battery, with its discharge capacity measured to indicate its degradation performance. The failure of the battery is defined as the discharge capacity dropped below a user-specific threshold.
(C) Schematic drawing (left) and degradation curve (right) for a proton exchange membrane fuel cell, with its voltage measured to indicate its degradation performance. The failure of the fuel cell is defined as the voltage dropped below a user-specific threshold.

Accelerated Aging Tests and Lifetime Prediction

A key challenge common to the widespread adoption of new solar, battery, and fuel cell

technologies lies in the device durability. For example, despite perovskite solar cells' excellent solar-to-electrical conversion efficiency, their long-term stability lags behind that of silicon solar cells and presents an area of active research.[10] Similarly, Li-ion batteries and fuel cell powertrains are often compared to gasoline cars in the market, the latter last for decades without major degradation of the components.[11] Throughout the R&D phase, testing multi-component devices is essential for quality control and performance evaluation, but can take years. Accelerated aging tests and predictive models that estimate a device's lifespan are therefore necessary to reduce the time from lab to market for new technologies. These predictive models also help plan for warranty services and regular maintenance. Over the past decades, deep learning gained popularity for degradation inference as the technique allows taking the on-board measured current and voltage directly as inputs to the ML models, without additional data featurization step. For example, in a recent study by Ansari et al., a recurrent neural network model was proposed, taking a rolling window of current and voltage time-series inputs to predict the continuous battery capacity fade curves in the near-term and long-term future.[12] Alternatively, when data is scarce, state-of-the-art data-driven models for predicting device lifetime usually follow a standardized workflow that includes structuring data, identifying relevant features, and then using supervised ML models to predict the device's health or its remaining useful life. For example, Paulson et al.[13] used regression models to predict battery cycle life from aging tests of over hundreds of cells. They performed extensive feature engineering, which involved analyzing various parameters including voltage profiles, current rates, and capacity to extract relevant features indicative of battery health and degradation. In the following paragraphs, we select three recent studies of solar cells, batteries, and fuel cells, respectively. For each case study, we review the aging dataset and the data analytics presented and build on the reported work to apply simple ML models for lifetime prediction. The goal of the brief benchmarking in this perspective is to explore the generalizability of data-driven approaches across energy devices with different available data size, types and the objective of the studies.

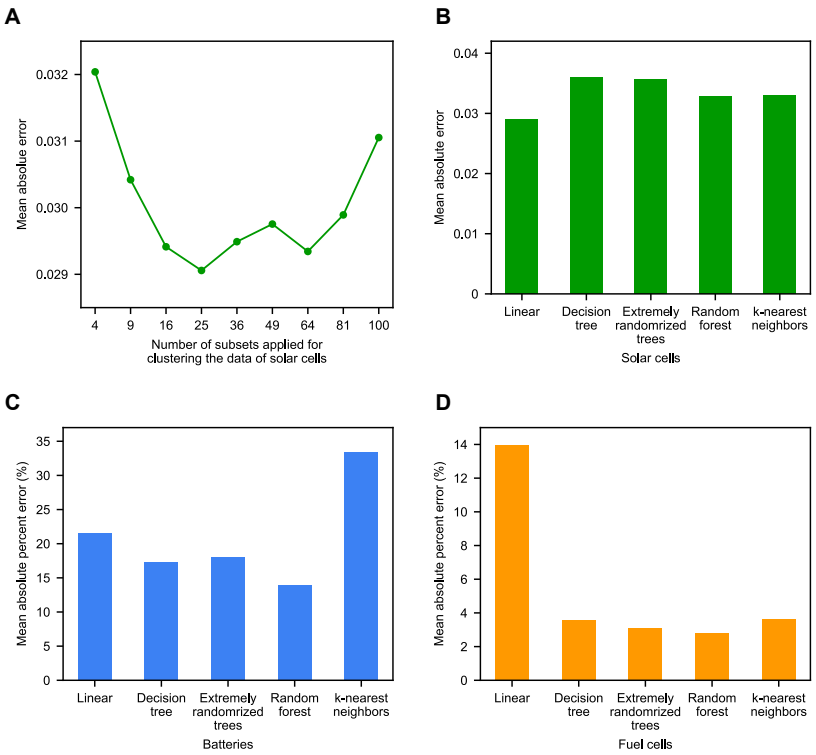

Figure 2. Lifetime prediction performance using regression models. For each case study, 10 times were run with the average taken to ensure the accuracy of the results during the machine learning process.
(A-B) Mean absolute error of the prediction of solar cells' power conversion efficiency (PCE) at 150 hours using (A) different number of subsets applied for clustering and (B) different machine learning regression algorithms, with the self-organizing map parameters set to be sigma = 0.5 and learning rate = 0.1. 2-fold cross validation and grid search were applied on nonlinear machine learning models to tune their hyperparameters.
(C-D) Mean absolute percent error of (C) Lithium-ion batteries' cycle life prediction and (D) fuel cells' lifetime prediction using different machine learning regression algorithms. 5-fold cross validation and grid search were applied on nonlinear machine learning models to tune their hyperparameters.

Case study 1: Perovskite Solar Cells

As shown in Fig. 1(A), perovskite solar cells consist of multiple layers of materials that are applied through printing, coating, or vacuum deposition onto a foundational support layer referred to as the glass substrate,[14] while solar-to-electricity power conversion efficiency (PCE) at 150 hours is a typical indicator of their performance during the aging tests. In their recent study, Hartono et al., applied an unsupervised learning approach to a large dataset of perovskite solar cells aging tests 2245 tests consisting of time-series data of PCEs were clustered into four groups using self-organizing map (SOM) according to the shapes of the PCE curves.[15] SOM was chosen as it efficiently simplified complex, high-dimensional statistical relationships into geometric patterns in lower-dimensional spaces, speeding up the learning process while preserving the data's original structure and relationships.[16]

Lifetime prediction: Building upon the reported study, we combined the results from the SOM approach with a regression model to predict the future PCE values. In our approach, as shown in Fig. 2, clustering was done first to divide the whole dataset into multiple subsets based on the curves from 0 to 100 hours, while for each subset, a simple linear regression model using the PCE value at 100 hours to predict the one at 150 hours was built. It turned out that this prediction method not only ran very fast, but also had a good performance with the least mean absolute error observed if dividing into 25 subsets, as depicted in Fig. 2(A). Notably, in this case study, limiting the model input to single parameters meant that other nonlinear regression models did not offer any advantage over a linear model. In fact, if nonlinear models were applied instead of the linear one, the accuracy would slightly decrease, as illustrated in Fig. 2(B). This case study demonstrates a two-step modeling approach where unsupervised and supervised learning are combined for solar cell life prediction.

Case study 2: Li-ion Batteries

Fig. 1(B) shows an example device structure of Li-ion batteries, which are charged and discharged by lithium ions moving between the negative (anode) and positive (cathode) electrodes. A Lifetime prediction for Li-ion batteries often aims to forecast the number of cycles until discharge capacity reaches 80% of the nominal capacity. In 2019, Severson et al. suggested that the log variance of the change in discharge voltage curves between cycles 10 and 100, denoted as $\mathrm{Var}(\Delta Q_{100\text{-}10}(V))$, exhibited a strong correlation with the battery's cycle life.[17] Whereafter, this variance feature, along with other supplementary features obtained from the initial 100 discharge cycles, were employed as input features for data-driven models to predict battery's cycle life.

Model comparison: Building upon the linear regression algorithm in the study by Severson et al., we further examined nonlinear regression methods for comparison purposes. As shown in Fig. 2(C), the three tree models had less mean absolute percent error than the linear model, and among them, the random forest model achieved superior performance. It is worth noting, however,

that the benefits of using tree models compared to simple linear regression are minimal. This underscores that the ability to identify relevant features effectively for accurate outcome predictions, especially when the data size is limited (as with the 124 battery cells in this case study), is crucial. This case study showcases that with effective data featurization, developing complex models is not always necessary for achieving accurate predictions.

Case study 3: PEM fuel cells

PEM fuel cells use hydrogen as fuel and oxygen from the air as oxidant to convert chemical energy into electricity. Typically, time-resolved voltage curves are utilized for lifetime prediction, with their failure threshold set to be, for example, 95.5% of the initial voltage at the beginning of the test. This standard references the criteria established by the FCLAB Laboratory in 2014, which introduced a challenge featuring a time-series dataset of 10 PEM cells.[18] Fig. 1(C) demonstrates an example PEM cell undergoing aging tests from this dataset. Several papers have performed lifetime predictions using this dataset to train data-driven models. Specifically, deep learning techniques such as artificial neural network, recurrent neural network, or convolutional neural network were employed to predict the succeeding voltage curves over time.[19,20]

Feature engineering: Inspired by the featurization approaches widely employed in the battery community, we constructed features-based lifetime prediction models without using deep learning. According to Tian et al., the output voltage of a fuel cell was negatively correlated with time,[21] making the linear fit of the early time series data of voltage a predictive feature. Moreover, variance features were also included to ensure that the model accounts for the uncertainty present in the voltage measurements in the presence of noise. After applying the same regression methods as for the solar cells and batteries in the previous case study, we observed in Fig. 2(D) that nonlinear ML models exhibited superior performance compared to linear regression, capturing the local nonlinear voltage drops throughout the tests. However, we shall note despite the higher prediction accuracy, training data for these ML models comes from the available aging data of 10 cells. As a result, ML models face overfitting problems with small datasets like in this case study, and the results are not transferable to other datasets. Our case study here exemplifies that sufficient training data is a prerequisite for developing trustworthy ML models.

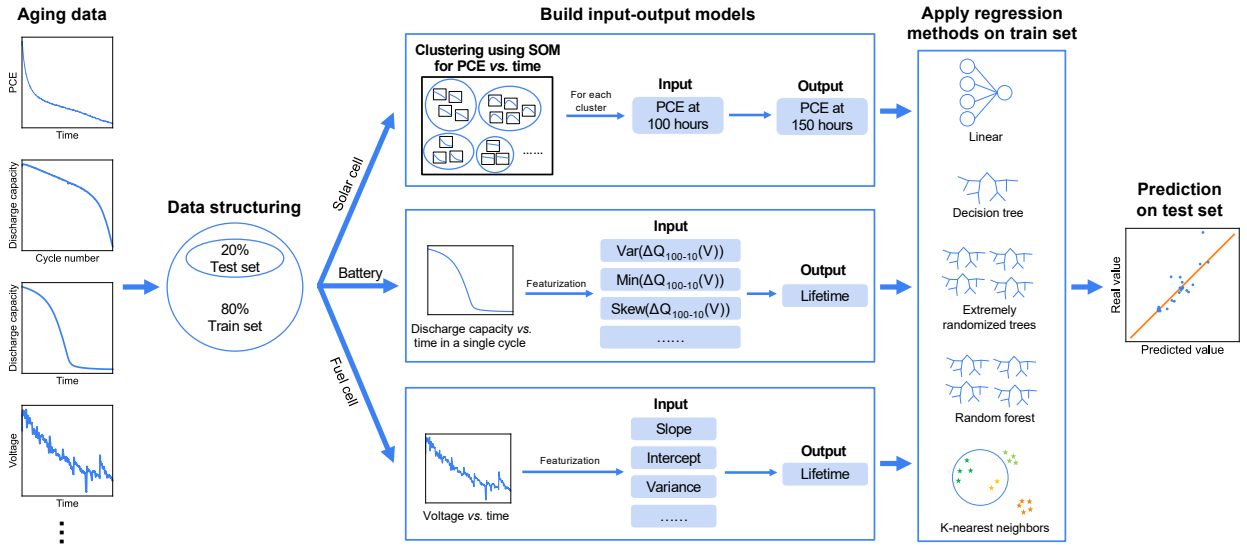

Figure 3. Machine learning process workflow for different renewable energy devices, where SOM stands for self-organizing map, PCE means power conversion efficiency, and Var($\Delta Q_{100\text{-}10}(V)$),

Min($\Delta Q_{100-10}(V)$), Skew($\Delta Q_{100-10}(V)$) represents the variance, minimum, skewness of the change in discharge voltage curves between cycles 10 and 100, respectively.

To sum up, data-driven models have been widely applied for lifetime prediction across solar, battery and fuel cell technologies as a complementary approach to the traditional physics-based models.[22] In Fig. 3, we present a common data-driven workflow employed by researchers. After structuring the data and identifying the representative features for each device, input-output models can be built using the training set with various regression methods applied. Then, predictions are validated with the test dataset. Notably, challenges remain in reducing the high cost of long-term tests for model validation and understanding the complexity of aging mechanisms. Unlike simulations, analytics of real-world testing data suffer from issues such as managing noisy datasets and small available datasets. Instead of large-scale models, identifying predictive features is a powerful approach for small dataset ML modeling and often does not require complex model architectures. Meanwhile, domain knowledge and expert insight play an important role in small dataset modeling. Specifically, collaborating with domain experts leads to effective feature identification and down selections, thus improving the accuracy of prediction and efficiency of model development. For example, in the study conducted by Fathy et al., they built the mathematical model of a PEM fuel cell to help them select the input parameters to be determined before adopting the optimizer, which turned out to be effective in the end.[23] Moreover, if the dataset is large (more than one thousand entries), like our solar cells case study or the batteries study conducted by Deebansok et al.,[24] combining a clustering step with simple linear regression presents another efficient and fairly accurate way for predicting the devices' performance without complex models. Deep learning, on the other hand, can be effective in tackling challenges such as noise reduction and predicting for multiple outputs, given sufficient training dataset. Examples include the use of convolutional neural network[25] or recurrent neural network.[12,26]

Understanding Failures

Identifying the root causes of device failures, whether they are 'expected' due to degradation over a long-term usage or 'unexpected' due to defective systems, is critical for developing safe and reliable energy devices. In the previous section, we discussed various data-driven approaches, however, ML lifetime prediction models usually are black boxes, providing very little physical insights.[27] In this section, we summarize two approaches in recent literature where interpretability can be embedded into the data-driven prediction studies.

The first approach involves directly integrating parameters that have physical meanings as features to the prediction models. Taking battery technology as an example, battery capacity fade are typically attributed to three thermodynamics degradation modes: (1) loss of active materials on the negative electrode side (LAM NE), (2) loss of active materials on the positive electrode side (LAM PE) (3) loss of Li inventory (LLI) , and kinetic degradation.[28] There have been extensive studies in literature to understand those degradation modes through additional materials characterization and device testing, involving reconstructing low-rate full cell performance by manipulating the fresh half cells voltage curves.[29–34] Kinetic degradation can be probed through pulse power tests or electrochemical impedance spectroscopy.[35–38] Those low rate and impedance tests are best known for being non-destructive, and because of this, many introduce low-rate cycles and pulse tests periodically to their testing process to track changes in both thermodynamic and kinetic properties.[39–42] By building data-driven models with fitted physics-based degradation parameters as input, and device performance, such as battery cycle life as

output, we can correlate the device degradation with physic-based parameter change like resistance rise or electrode capacity drop while achieving low prediction errors.[39,40,43]

The second approach is to integrate interpretable ML into the predictive modeling workflows. Several model-agnostic methods such as function decomposition have been reported[44] for describing the average behavior of a ML model, and local interpretation methods Shapley values are attracting increasing attention for explaining the impact of individual features in a predictive model .[45] SHAP (SHapley Additive exPlanations) proposes interpretation methods based on combinations of Shapley values across the data.[46] In their recent work, Van Vlijmen et al.[39] generated a diverse battery aging dataset by varying 6 cycling parameters, as shown in Fig. 4(A). Given the diverse degradation trajectories resulting from convoluted degradation mechanisms, the traditional method of studying parameter sensitivity by looking at one parameter at a time is no longer feasible. The authors employed pulse measurements and differential voltage analysis to characterize the kinetics and thermodynamics degradation, and applied SHAP analysis to further understand the impact, as illustrated in Fig. 4(B) and (C). The extracted SHAP values can be used to construct a matrix plot where each matrix entry is the impact of a parameter (cycling conditions or physics-based parameters) to the model output. The analysis enables a quick identification of important parameters and degradation diagnosis. For example, the authors found that the most impactful cycling parameters are charge current magnitude and discharge cutoff voltage, and charge current impacts the cycle life through resistance. SHAP and physics-based models enable a quick examination and understanding of parameter impact in a high dimensional parameter space. Similar explainable ML concepts were recently demonstrated in solar cell lifetime prediction studies. Oviedo et al. incorporated interpretable ML to lifetime prediction of thin-film organic solar cells to gain insights into the degradation cause.[47] The authors implemented a black-box neural net model to predict solar cell lifetime from input voltage curves, as seen in Fig. 4(D), and in parallel they calculated the physical parameters from a device model to estimate the contribution of physical parameters to input voltage curves, thus identifying the main degradation driving forces (Fig. 4(E) and (F)).[47] For common energy devices, whether their purpose is to harvest, convert, or store energy, understanding the failure mechanisms is equally, if not more, important than achieving accurate lifetime predictions. Interpretable predictions are powerful as they enable scientific understanding in a high-dimensional space where correlations between inputs can be non-linear or complex. Coupling interpretable ML with accelerated aging tests has the potential to improve operation protocols and even enhance device designs. Methods presented here are highly generalizable and are not only limited to specific energy device applications of battery or solar cells.

Though applied differently, the two strategies discussed in this perspective share a common objective: integrating data-driven techniques with relevant physical parameters, either before or after machine learning models are trained. Importantly, merging these strategies into a unified data-driven workflow can significantly enhance the model's transparency. This integration forms an end-to-end ML pipeline that is more likely to generate actionable knowledge and insights for specific energy systems.

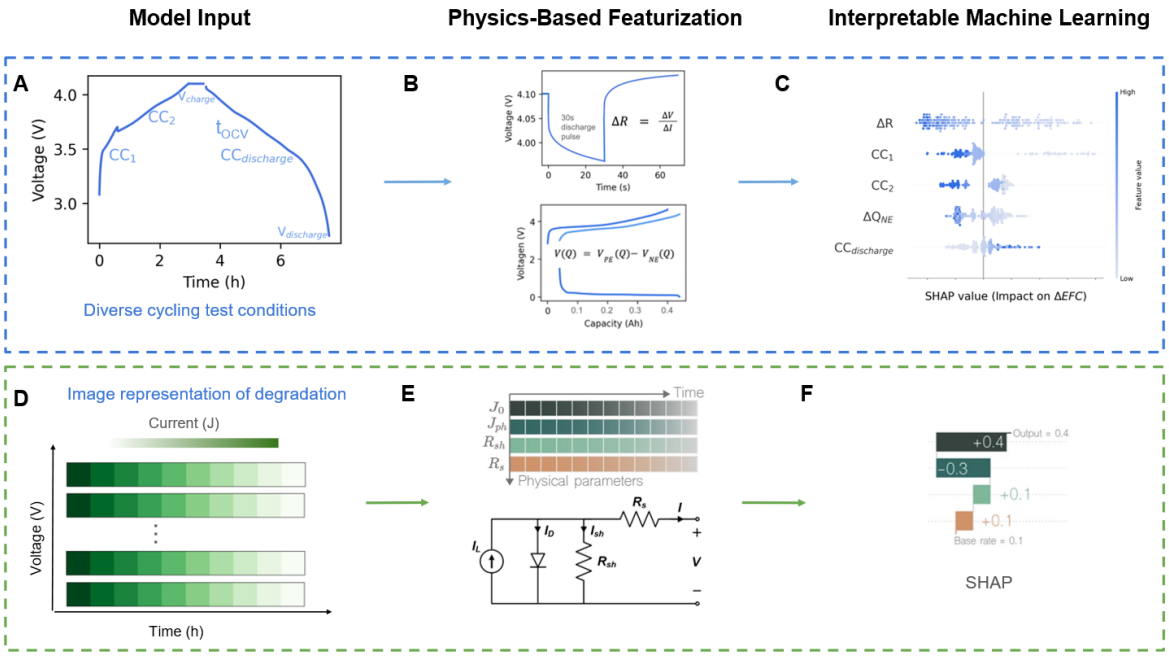

Figure 4. Demonstrations of incorporating interpretable machine learning into the degradation analysis, top row (A)-(C) is an example of using SHAP and physics-based featurization to understand battery cycling parameter impact, and bottom row (D)-(F) is a solar cell example to SHAP and physics-based featurization to understand a black-box prediction model.[47]

(A) A diverse battery aging dataset generated by varying 6 cycling parameters[39]

(B) Physics based measurement (pulse tests for resistance measurements) and model (differential voltage analysis) for parameter estimation of this dataset

(C) SHAP analysis for understanding how cycling conditions and physics-based parameters affect the final battery equivalent full cycles. The extracted SHAP values can be used to construct the matrix plot for a quick visualization of the parameter impact. Darker color in the matrix plot indicates higher parameter impact. This is especially helpful when dealing with a diverse dataset where parameter impact is convoluted.

(D) Time series representation of the initial cell degradation (current density-voltage) as input for neural network prediction.[47]

(E) Time-regularized physical inference. A device model is used to fit the degradation dynamics to the dynamics of various physical parameters on the time axis.

(F) SHAP values are used to quantify the effect of each physical parameter, and this offers insights into the black box prediction model.

Optimizing Processes

Process optimization using ML techniques has the advantages of rapidly assessing and screening a large number of parameters and iteratively converging to optimal decisions.[48,49] A popular approach used in the scientific community research is active learning, a subfield of AI, and have been demonstrated in virtual materials discovery,[50–53] synthesis,[54,55] and electrochemical performance testings.[49] In this section, we review two case studies, where active learning techniques were applied to optimize device fabrication and device operation, respectively.

Bayesian optimization, an example active learning technique, has been employed in a number of studies guiding experimental research and optimizing unknown functions, since the algorithm efficiently balance exploratory experiments with utilizing prior knowledge to identify extrema.[56]

Many researchers use active learning to accelerate materials design and synthesis with targeted properties.[55,57,58] However, there are few demonstrations on the device level. Liu et al. implemented batch Bayesian optimization to identify the optimal conditions for fabricating perovskite solar cells via rapid spray plasma processing, focusing on six process parameters: spray flow rate, plasma nozzle height, plasma gas flow rate, plasma duty cycle, coating speed, and film quality, which altogether present over 40,000 possible conditions.[59] To accelerate the process optimization, the authors developed a data-driven framework that utilizes results from previous batches of experiments as a probabilistic guide to suggest subsequent experiments under a constrained budget. They also combine insights from domain experts as knowledge constraints in their ML models and thereby accelerate the search for process conditions that lead to high-efficiency solar devices. Similar data-driven frameworks have also been reported to optimize device operation rather than fabrication. This approach has been particularly applied in battery management systems, as the state of health of a battery significantly depends on its charging and usage patterns. Attia and Grover et al. designed a closed optimization loop for Li-ion battery fast charging.[49] This optimization loop has two key parts: (1) lifetime prediction using the first 100 aging cycle data to shorten the battery testing time;[17] (2) Bayesian optimization to recommend the next round of parameter values to test. Because of this optimization loop, the authors were able to rapidly identify high-cycle-life charging protocols among 224 candidates in 16 days (compared with over 500 days using exhaustive search without early prediction).

Outlook in Digital Manufacturing

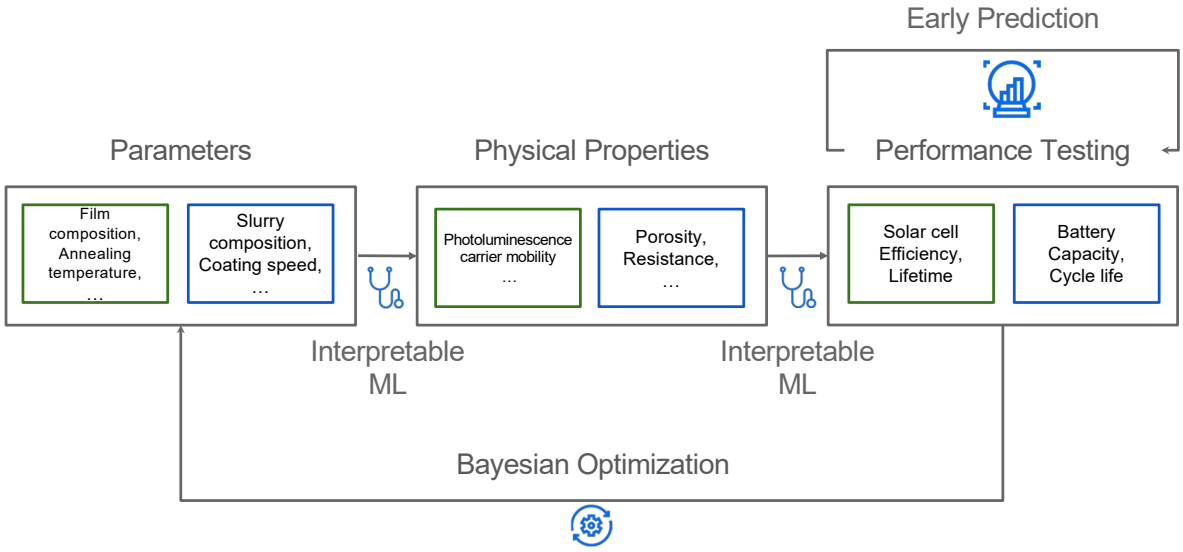

Figure 5. Demonstration of accelerated solar cell and battery R&D via a closed-loop approach with interpretable ML. Blue highlights the parameters for batteries and green highlights the parameters for solar cells. Lifetime prediction here is to shorten the battery/solar cell performance evaluation time, Bayesian optimization is to recommend the next round of parameters to test, and interpretable machine learning helps gain insights into the process-property relations. With parameter space set for individual processes, the same optimization loop can be used for other manufacturing problems not limited to batteries or solar cells.

Building on the data-driven methods reviewed in previous sections, here we propose an general interpretable optimization framework that can be used to accelerate device development at scale,

as shown in the example workflows in Fig. 5. Taking energy storage as an example, battery electrode production, which include slurry preparation and subsequent processing numerous manufacturing steps such as mixing, coating, drying, and calendering,[60] has high dimensional parameters and a long evaluation timeline loop. Data-driven closed-loop optimization can be a promising approach to shorten the optimization period and identify the most important parameters. The first step is to identify parameters to optimize, and listed here (in blue) are some example parameters that can affect the electrode slurry performance such as composition, coating speed, and drying temperature etc. Electrodes are tested in a half/full cell format to further evaluate its performance. Before the cycling test, physical parameters such as porosity, capacity, or resistance can be measured. After this, cells undergo the same cycle life testing. Active learning techniques, such as Bayesian optimization, can be used on the initial dataset to recommend the next round of parameters to test, and lifetime prediction models can be developed to shorten the future cell testing time. Except for the initial dataset, future cycle life prediction can be done during early cycles, and through iterations, we will land with the optimized cell fabrication parameters. By adding interpretable ML between parameters and physical properties, physical properties and cycle life, insights on how parameters affect battery performance could be gained, informing better materials designs in the future. For example, it has previously been shown that though binder can improve mechanical stability, it impedes Li diffusion, leading to higher cell resistance, resulting in shorter cycle life,[60] and such impact can be captured with this interpretable optimization framework. Interpretable ML works efficiently in a high-dimensional space and can identify the important parameters faster.[39] Similar approaches can be applied in the manufacturing of thin-film photovoltaic devices, as exemplified in Fig 5 (highlighted in green).[59,61,62] Such efforts align with the initiatives in digital manufacturing, which emphasis on the automation of production processes with technologies such as AI, internet of things, and cloud computing.[63,64] In the context of the energy industry, digital manufacturing frameworks that integrate a streamlined data workflow that combines interpretable ML with closed-loop active learning into accelerated aging test procedures is promising in accelerating device failure diagnostics. To facilitate this, it's essential to create specific databases that map the relationships between process variables (like thin film composition and annealing temperature in solar cells) and physical properties (such as photoluminescence and carrier mobility of solar materials).[65] Overall, we see great potential in data-driven workflows to enhance equipment efficiency and device reproducibility. These workflows can leverage extensive production data to refine control parameters, thereby optimizing the manufacturing process.

Open Challenges

Despite the great potential of applying data-driven workflows to predict, understand, and optimize the performance of the energy devices, challenges exist to incorporate data science into the industrial R&D and manufacturing workflows beyond proof-of-the concept studies. Here we summarize a few open issues centered around data quality, generalizable insights, and hardware integration.

Data availability: Collecting high-quality data is both expensive and time-consuming. For instance, it is not uncommon to test perovskite solar cells for over 5,000 hours (~208 days), and Li-ion batteries for more than a year in laboratory settings.[39,66] These testing typically are already following accelerated testing strategies, designed to induce device degradation faster than would occur under field operations. Consequently, many academic demonstrations of ML model construction and validation end up using exceedingly non-diverse data sets. To address this, we recommend scrutinizing the generalizability of models. Without this, frequent model re-training

may be necessary when developing new materials formulations or device structures, leading to high upfront costs. Moreover, data collected during manufacturing processes are often proprietary. If more datasets, for example, those concerning previous-generation devices or from exploratory early research stages, could be made available as "challenge datasets" after removing personally identifying information, it would enable community efforts to develop more advanced ML methods and build robust, large-scale models based on real-world data.

Data standardization and model benchmarking: During the preparation of this perspective, we noticed that not all fields are adopting data science at the same pace or converging on standard testing protocols. Compared to batteries, there are fewer sources of fuel cell device aging data available. In the solar cell community, there is an increasing discussion about standard testing and reporting protocols for perovskite photovoltaics. This is exemplified by the 2020 publication of a consensus statement on stability assessment based on ISOS standardized procedures, which was co-authored by 59 researchers.[67] From a model benchmarking angle, among the open-source data available, there often exist different data formats and testing protocols that make the creation of a unified ML-ready dataset challenging. For example, several battery aging datasets were published in the past years,[68] however, the fact each dataset was constructed from testing cells with different electrode chemistry, form factors, and testing conditions confine the broad applicability of the generalizable prediction models and highlight the importance of structured metadata in addition to testing data. To address this, we recommend introducing standard diagnostic tests during regular aging tests to compare the different cycling conditions on a fair basis. Those diagnostic tests can also be used to get more in depth degradation information that can be used in lifetime prediction. Moreover, experimental data can be complemented by synthetic data simulated with different failure modes, and this is a convenient and inexpensive way to expand the training dataset and benchmark the ML models.[69–72]

Scientific understanding and actionable outcomes: There is often a trade-off between model interpretability and its accuracy. In this perspective, we discussed how to incorporate interpretable ML and physics-based parameters to the ML pipeline in order to improve model interpretability. There are also examples of combining physics-based models with data-driven models to achieve desirable accuracy without losing interpretability.[73] While predicting device performance effectively shortens testing durations, relying solely on simple ML models falls short of providing the scientific principles and understanding necessary to guide the direction of next-generation R&D. Multimodal learning that combines both material and device-level characterization and testing data in the model development, along with causal learning that makes predictions based on discovering cause-and-effect relationships, will improve the model's capability in identifying the root causes of device failure. We recommend that researchers focusing on applied ML for energy science to look beyond the immediate benefits of analytical acceleration offered by data science tools. There is significant potential to extract broader, generalizable insights from ML models and to transform these insights into actionable outcomes that can be directly evaluated and tested by device engineers.

Integration with existing human-centric workflows: The development of cyber-physical platforms that merges the software and hardware solutions into intelligent systems provides avenues towards autonomous operation. However, most of the workflows currently in use in research labs and factories are designed to be human-centric, leading to barriers to adopt data-driven approaches. In the energy storage field, data hubs like *Battery Archive* are good examples of open sourcing and standardizing data from different cyclers and testing methods to make data-driven models more accessible.[74] Intuitive analytics platforms allow users to visualize, evaluate and predict device degradation without deep technical expertise, as exemplified by *BEEP* analytics platform for batteries.[75] Furthermore, embedding transparency and explainability in data

models demystifies prediction and optimization strategies, fostering trust and adoption. Today's automated workflows in R&D still rely on human feedback and supervision to ensure model integrity. Therefore, it is essential to reduce the barriers for domain experts to use data science tools by offering software interfaces that allow them to interact with the model and provide timely feedback. With the support of enhanced data infrastructure, education of the workforce in data science, and progress in software for human-machine interfaces, we can expect data-driven strategies to unlock new possibilities towards autonomous device research, development, and manufacturing, leading to increased efficiency and reduced costs.

Acknowledgments

XC is supported by Stanford Data Science Scholars Fellowship, SS thanks support from the Automated Robotics for Energy Storage Laboratory, Energy Storage Materials Initiative at the Pacific Northwest National Laboratory, and from Department of Mechanical Engineering, University of Washington.

Declaration of Interests

The authors declare no competing interests.

References

1. Gielen, D., Boshell, F., Saygin, D., Bazilian, M.D., Wagner, N., and Gorini, R. (2019). The role of renewable energy in the global energy transformation. Energy Strategy Rev. *24*, 38–50. 10.1016/j.esr.2019.01.006.

2. Rizvi, S.A.A., Xin, A., Masood, A., Iqbal, S., Jan, M.U., and Rehman, H. (2018). Electric Vehicles and their Impacts on Integration into Power Grid: A Review. In 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2) (IEEE), pp. 1–6. 10.1109/EI2.2018.8582069.

3. Dalenogare, L.S., Benitez, G.B., Ayala, N.F., and Frank, A.G. (2018). The expected contribution of Industry 4.0 technologies for industrial performance. Int. J. Prod. Econ. *204*, 383–394. 10.1016/j.ijpe.2018.08.019.

4. Szymanski, N.J., Rendy, B., Fei, Y., Kumar, R.E., He, T., Milsted, D., McDermott, M.J., Gallant, M., Cubuk, E.D., Merchant, A., et al. (2023). An autonomous laboratory for the accelerated synthesis of novel materials. Nature *624*, 86–91. 10.1038/s41586-023-06734-w.

5. Gomes, C.P., Selman, B., and Gregoire, J.M. (2019). Artificial intelligence for materials discovery. MRS Bull. *44*, 538–544. 10.1557/mrs.2019.158.

6. Merchant, A., Batzner, S., Schoenholz, S.S., Aykol, M., Cheon, G., and Cubuk, E.D. (2023). Scaling deep learning for materials discovery. Nature *624*, 80–85. 10.1038/s41586-023-06735-9.

7. Stevens, M.B., Anand, M., Kreider, M.E., Price, E.K., Zeledón, J.Z., Wang, L., Peng, J., Li, H., Gregoire, J.M., Hummelshøj, J., et al. (2022). New challenges in oxygen reduction catalysis: a consortium retrospective to inform future research. Energy Environ. Sci. *15*, 3775–3794. 10.1039/D2EE01333A.

8. Čulík, P., Brooks, K., Momblona, C., Adams, M., Kinge, S., Maréchal, F., Dyson, P.J., and Nazeeruddin, M.K. (2022). Design and Cost Analysis of 100 MW Perovskite Solar Panel Manufacturing Process in Different Locations. ACS Energy Lett. *7*, 3039–3044. 10.1021/acsenergylett.2c01728.

9. Liu, Y., Zhang, R., Wang, J., and Wang, Y. (2021). Current and future lithium-ion battery manufacturing. iScience *24*, 102332. 10.1016/j.isci.2021.102332.

10. Olaleru, S.A., Kirui, J.K., Wamwangi, D., Roro, K.T., and Mwakikunga, B. (2020). Perovskite solar cells: The new epoch in photovoltaics. Sol. Energy *196*, 295–309. 10.1016/j.solener.2019.12.025.

11. Schloter, L. (2022). Empirical analysis of the depreciation of electric vehicles compared to gasoline vehicles. Transp. Policy *126*, 268–279. 10.1016/j.tranpol.2022.07.021.

12. Ansari, M., Torrisi, S.B., Trewartha, A., and Sun, S. (2024). History-agnostic battery degradation inference. J. Energy Storage *81*, 110279. 10.1016/j.est.2023.110279.

13. Paulson, N.H., Kubal, J., Ward, L., Saxena, S., Lu, W., and Babinec, S.J. (2022). Feature engineering for machine learning enabled early prediction of battery lifetime. J. Power Sources *527*, 231127. 10.1016/j.jpowsour.2022.231127.

14. Ebrahimi, M. (2023). Solar power plants. In Power Generation Technologies (Elsevier), pp. 419–461. 10.1016/B978-0-323-95370-2.00010-7.

15. Hartono, N.T.P., Köbler, H., Graniero, P., Khenkin, M., Schlatmann, R., Ulbrich, C., and Abate, A. (2023). Stability follows efficiency based on the analysis of a large perovskite solar cells ageing dataset. Nat. Commun. *14*, 4869. 10.1038/s41467-023-40585-3.

16. Kohonen, T. (2001). Self-Organizing Maps (Springer Berlin Heidelberg) 10.1007/978-3-642-56927-2.

17. Severson, K.A., Attia, P.M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M.H., Aykol, M., Herring, P.K., Fraggedakis, D., et al. (2019). Data-driven prediction of battery cycle life before capacity degradation. Nat. Energy *4*, 383–391. 10.1038/s41560-019-0356-8.

18. FCLAB Research (2014). IEEE 2014 Data Challenge Data. https://repository.lboro.ac.uk/articles/dataset/IEEE_2014_Data_Challenge_Data/3518141.

19. Zhou, D., Al-Durra, A., Zhang, K., Ravey, A., and Gao, F. (2018). Online remaining useful lifetime prediction of proton exchange membrane fuel cells using a novel robust methodology. J. Power Sources *399*, 314–328. 10.1016/j.jpowsour.2018.06.098.

20. Zhang, L., Lin, J., Liu, B., Zhang, Z., Yan, X., and Wei, M. (2019). A Review on Deep Learning Applications in Prognostics and Health Management. IEEE Access *7*, 162415–162438. 10.1109/ACCESS.2019.2950985.

21. Tian, Q., Chen, H., Ding, S., Shu, L., Wang, L., and Huang, J. (2023). Remaining Useful Life Prediction Method of PEM Fuel Cells Based on a Hybrid Model. Electronics *12*, 3883. 10.3390/electronics12183883.

22. Finegan, D.P., Zhu, J., Feng, X., Keyser, M., Ulmefors, M., Li, W., Bazant, M.Z., and Cooper, S.J. (2021). The Application of Data-Driven Methods and Physics-Based Learning for Improving Battery Safety. Joule *5*, 316–329. 10.1016/j.joule.2020.11.018.

23. Fathy, A., Rezk, H., Alharbi, A.G., and Yousri, D. (2023). Proton exchange membrane fuel cell model parameters identification using Chaotically based-bonobo optimizer. Energy *268*, 126705. 10.1016/j.energy.2023.126705.

24. Deebansok, S., Deng, J., Le Calvez, E., Zhu, Y., Crosnier, O., Brousse, T., and Fontaine, O. (2024). Capacitive tendency concept alongside supervised machine-learning toward classifying electrochemical behavior of battery and pseudocapacitor materials. Nat. Commun. *15*, 1133. 10.1038/s41467-024-45394-w.

25. Rosato, A., Araneo, R., Andreotti, A., Succetti, F., and Panella, M. (2021). 2-D Convolutional Deep Neural Network for the Multivariate Prediction of Photovoltaic Time Series. Energies *14*, 2392. 10.3390/en14092392.

26. Succetti, F., Rosato, A., Araneo, R., and Panella, M. (2020). Deep Neural Networks for Multivariate Prediction of Photovoltaic Power Time Series. IEEE Access *8*, 211490–211505. 10.1109/ACCESS.2020.3039733.

27. Hong, J., Lee, D., Jeong, E.-R., and Yi, Y. (2020). Towards the swift prediction of the remaining useful life of lithium-ion batteries with end-to-end deep learning. Appl. Energy *278*, 115646. 10.1016/j.apenergy.2020.115646.

28. Edge, J.S., O'Kane, S., Prosser, R., Kirkaldy, N.D., Patel, A.N., Hales, A., Ghosh, A., Ai, W., Chen, J., Yang, J., et al. (2021). Lithium ion battery degradation: what you need to know. Phys. Chem. Chem. Phys. *23*, 8200–8221. 10.1039/D1CP00359C.

29. Dahn, H.M., Smith, A.J., Burns, J.C., Stevens, D.A., and Dahn, J.R. (2012). User-Friendly Differential Voltage Analysis Freeware for the Analysis of Degradation Mechanisms in Li-Ion Batteries. J. Electrochem. Soc. *159*, A1405–A1409. 10.1149/2.013209jes.

30. Birkl, C.R., Roberts, M.R., McTurk, E., Bruce, P.G., and Howey, D.A. (2017). Degradation diagnostics for lithium ion cells. J. Power Sources *341*, 373–386. 10.1016/j.jpowsour.2016.12.011.

31. Weng, A., Siegel, J.B., and Stefanopoulou, A. (2023). Differential voltage analysis for battery manufacturing process control. Front. Energy Res. *11*, 1087269. 10.3389/fenrg.2023.1087269.

32. Dubarry, M., Truchot, C., and Liaw, B.Y. (2012). Synthesize battery degradation modes via a diagnostic and prognostic model. J. Power Sources *219*, 204–216. 10.1016/j.jpowsour.2012.07.016.

33. Dubarry, M., Svoboda, V., Hwu, R., and Yann Liaw, B. (2006). Incremental Capacity Analysis and Close-to-Equilibrium OCV Measurements to Quantify Capacity Fade in

Commercial Rechargeable Lithium Batteries. Electrochem. Solid-State Lett. *9*, A454. 10.1149/1.2221767.

34. Bloom, I., Jansen, A.N., Abraham, D.P., Knuth, J., Jones, S.A., Battaglia, V.S., and Henriksen, G.L. (2005). Differential voltage analyses of high-power, lithium-ion cells. J. Power Sources *139*, 295–303. 10.1016/j.jpowsour.2004.07.021.

35. Barai, A., Uddin, K., Widanage, W.D., McGordon, A., and Jennings, P. (2018). A study of the influence of measurement timescale on internal resistance characterisation methodologies for lithium-ion cells. Sci. Rep. *8*, 21. 10.1038/s41598-017-18424-5.

36. Barai, A., Uddin, K., Dubarry, M., Somerville, L., McGordon, A., Jennings, P., and Bloom, I. (2019). A comparison of methodologies for the non-invasive characterisation of commercial Li-ion cells. Prog. Energy Combust. Sci. *72*, 1–31. 10.1016/j.pecs.2019.01.001.

37. Mc Carthy, K., Gullapalli, H., Ryan, K.M., and Kennedy, T. (2021). Review—Use of Impedance Spectroscopy for the Estimation of Li-ion Battery State of Charge, State of Health and Internal Temperature. J. Electrochem. Soc. *168*, 080517. 10.1149/1945-7111/ac1a85.

38. Xie, Y., Kim, J., Lee, M., Park, J., and Na, W. (2021). An Impedance Behavior Study of Commercial NCA Cylindrical Battery Cells at Different SOCs. J. Electrochem. Soc. *168*, 090548. 10.1149/1945-7111/ac2706.

39. Vlijmen, B.V., Lam, V., Asinger, P.A., Cui, X., Ganapathi, D., Sun, S., Herring, P.K., Gopal, C.B., Geise, N., Deng, H.D., et al. (2023). Interpretable Data-Driven Modeling Reveals Complexity of Battery Aging (Chemistry) 10.26434/chemrxiv-2023-zdl2n-v2.

40. Weng, A., Mohtat, P., Attia, P.M., Sulzer, V., Lee, S., Less, G., and Stefanopoulou, A. (2021). Predicting the impact of formation protocols on battery lifetime immediately after manufacturing. Joule *5*, 2971–2992. 10.1016/j.joule.2021.09.015.

41. Harlow, J.E., Ma, X., Li, J., Logan, E., Liu, Y., Zhang, N., Ma, L., Glazier, S.L., Cormier, M.M.E., Genovese, M., et al. (2019). A Wide Range of Testing Results on an Excellent Lithium-Ion Cell Chemistry to be used as Benchmarks for New Battery Technologies. J. Electrochem. Soc. *166*, A3031–A3044. 10.1149/2.0981913jes.

42. Ma, X., Harlow, J.E., Li, J., Ma, L., Hall, D.S., Buteau, S., Genovese, M., Cormier, M., and Dahn, J.R. (2019). Editors' Choice—Hindering Rollover Failure of Li[Ni $_{0.5}$ Mn $_{0.3}$ Co $_{0.2}$]O $_2$ /Graphite Pouch Cells during Long-Term Cycling. J. Electrochem. Soc. *166*, A711–A724. 10.1149/2.0801904jes.

43. Jones, P.K., Stimming, U., and Lee, A.A. (2022). Impedance-based forecasting of lithium-ion battery performance amid uneven usage. Nat. Commun. *13*, 4806. 10.1038/s41467-022-32422-w.

44. Huang, J.Z. (1998). Functional ANOVA Models for Generalized Regression. J. Multivar. Anal. *67*, 49–71. 10.1006/jmva.1998.1753.

45. Song, E., Nelson, B.L., and Staum, J. (2016). Shapley Effects for Global Sensitivity Analysis: Theory and Computation. SIAMASA J. Uncertain. Quantif. *4*, 1060–1083. 10.1137/15M1048070.

46. Lundberg, S., and Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In 2017 31st Conference on Neural Information Processing Systems (NIPS 2017) (Curran Associates Inc.), pp. 4768–4777.

47. Oviedo, F., Hayden, D.S., Heumeuller, T., Wortmann, J., Perea, J.D., Naik, R., Xue, H., Lavista, J., Fisher Iii, J., Brabec, C.J., et al. (2023). DeepDeg: Forecasting and explaining degradation in novel photovoltaics (Chemistry) 10.26434/chemrxiv-2023-j2pvp.

48. Sliwoski, G., Kothiwale, S., Meiler, J., and Lowe, E.W. (2014). Computational Methods in Drug Discovery. Pharmacol. Rev. *66*, 334–395. 10.1124/pr.112.007336.

49. Attia, P.M., Grover, A., Jin, N., Severson, K.A., Markov, T.M., Liao, Y.-H., Chen, M.H., Cheong, B., Perkins, N., Yang, Z., et al. (2020). Closed-loop optimization of fast-charging protocols for batteries with machine learning. Nature *578*, 397–402. 10.1038/s41586-020-1994-5.

50. Xiang, X.-D., Sun, X., Briceño, G., Lou, Y., Wang, K.-A., Chang, H., Wallace-Freedman, W.G., Chen, S.-W., and Schultz, P.G. (1995). A Combinatorial Approach to Materials Discovery. Science *268*, 1738–1740. 10.1126/science.268.5218.1738.

51. Montoya, J.H., Winther, K.T., Flores, R.A., Bligaard, T., Hummelshøj, J.S., and Aykol, M. (2020). Autonomous intelligent agents for accelerated materials discovery. Chem. Sci. *11*, 8517–8532. 10.1039/D0SC01101K.

52. Aykol, M., Hegde, V.I., Hung, L., Suram, S., Herring, P., Wolverton, C., and Hummelshøj, J.S. (2019). Network analysis of synthesizable materials discovery. Nat. Commun. *10*, 2018. 10.1038/s41467-019-10030-5.

53. Diwale, S., Eisner, M.K., Carpenter, C., Sun, W., Rutledge, G.C., and Braatz, R.D. (2022). Bayesian optimization for material discovery processes with noise. Mol. Syst. Des. Eng. *7*, 622–636. 10.1039/D1ME00154J.

54. He, J., Bahr, J., Chisholm, B.J., Li, J., Chen, Z., Balbyshev, S.N., Bonitz, V., and Bierwagen, G.P. (2008). Combinatorial Materials Research Applied to the Development of New Surface Coatings X: A High-Throughput Electrochemical Impedance Spectroscopy Method for Screening Organic Coatings for Corrosion Inhibition. J. Comb. Chem. *10*, 704–713. 10.1021/cc8000458.

55. Wahl, C.B., Aykol, M., Swisher, J.H., Montoya, J.H., Suram, S.K., and Mirkin, C.A. (2021). Machine learning–accelerated design and synthesis of polyelemental heterostructures. Sci. Adv. *7*, eabj5505. 10.1126/sciadv.abj5505.

56. Kusne, A.G., Yu, H., Wu, C., Zhang, H., Hattrick-Simpers, J., DeCost, B., Sarker, S., Oses, C., Toher, C., Curtarolo, S., et al. (2020). On-the-fly closed-loop materials discovery via Bayesian active learning. Nat. Commun. *11*, 5966. 10.1038/s41467-020-19597-w.

57. Deshwal, A., Simon, C.M., and Doppa, J.R. (2021). Bayesian optimization of nanoporous materials. Mol. Syst. Des. Eng. *6*, 1066–1086. 10.1039/D1ME00093D.

58. Bassman Oftelie, L., Rajak, P., Kalia, R.K., Nakano, A., Sha, F., Sun, J., Singh, D.J., Aykol, M., Huck, P., Persson, K., et al. (2018). Active learning for accelerated design of layered materials. Npj Comput. Mater. *4*, 74. 10.1038/s41524-018-0129-0.

59. Liu, Z., Rolston, N., Flick, A.C., Colburn, T.W., Ren, Z., Dauskardt, R.H., and Buonassisi, T. (2022). Machine learning with knowledge constraints for process optimization of open-air perovskite solar cell manufacturing. Joule *6*, 834–849. 10.1016/j.joule.2022.03.003.

60. Drakopoulos, S.X., Gholamipour-Shirazi, A., MacDonald, P., Parini, R.C., Reynolds, C.D., Burnett, D.L., Pye, B., O'Regan, K.B., Wang, G., Whitehead, T.M., et al. (2021). Formulation and manufacturing optimization of lithium-ion graphite-based electrodes via machine learning. Cell Rep. Phys. Sci. *2*, 100683. 10.1016/j.xcrp.2021.100683.

61. Sun, S., Tiihonen, A., Oviedo, F., Liu, Z., Thapa, J., Zhao, Y., Hartono, N.T.P., Goyal, A., Heumueller, T., Batali, C., et al. (2021). A data fusion approach to optimize compositional stability of halide perovskites. Matter *4*, 1305–1322. 10.1016/j.matt.2021.01.008.

62. Xu, W., Liu, Z., Piper, R.T., and Hsu, J.W.P. (2023). Bayesian Optimization of photonic curing process for flexible perovskite photovoltaic devices. Sol. Energy Mater. Sol. Cells *249*, 112055. 10.1016/j.solmat.2022.112055.

63. Ayerbe, E., Berecibar, M., Clark, S., Franco, A.A., and Ruhland, J. (2022). Digitalization of Battery Manufacturing: Current Status, Challenges, and Opportunities. Adv. Energy Mater. *12*, 2102696. 10.1002/aenm.202102696.

64. Pandey, V., Sircar, A., Bist, N., Solanki, K., and Yadav, K. (2023). Accelerating the renewable energy sector through Industry 4.0: Optimization opportunities in the digital revolution. Int. J. Innov. Stud. *7*, 171–188. 10.1016/j.ijis.2023.03.003.

65. Hsu, J.W.P., and Xu, W. (2022). Accelerate process optimization in perovskite solar cell manufacturing with machine learning. Matter *5*, 1334–1336. 10.1016/j.matt.2022.04.004.

66. Jiang, Q., Tirawat, R., Kerner, R.A., Gaulding, E.A., Xian, Y., Wang, X., Newkirk, J.M., Yan, Y., Berry, J.J., and Zhu, K. (2023). Towards linking lab and field lifetimes of perovskite solar cells. Nature *623*, 313–318. 10.1038/s41586-023-06610-7.

67. Khenkin, M.V., Katz, E.A., Abate, A., Bardizza, G., Berry, J.J., Brabec, C., Brunetti, F., Bulović, V., Burlingame, Q., Di Carlo, A., et al. (2020). Consensus statement for stability assessment and reporting for perovskite photovoltaics based on ISOS procedures. Nat. Energy *5*, 35–49. 10.1038/s41560-019-0529-5.

68. Dos Reis, G., Strange, C., Yadav, M., and Li, S. (2021). Lithium-ion battery data and where to find it. Energy AI *5*, 100081. 10.1016/j.egyai.2021.100081.

69. Dubarry, M., and Beck, D. (2020). Benchmark Synthetic Training Data for Artificial Intelligence-based Li-ion Diagnosis and Prognosis (MATHEMATICS & COMPUTER SCIENCE) 10.20944/preprints202005.0106.v1.

70. Dubarry, M., Costa, N., and Matthews, D. (2023). Data-driven direct diagnosis of Li-ion batteries connected to photovoltaics. Nat. Commun. *14*, 3138. 10.1038/s41467-023-38895-7.

71. Ward, L., Babinec, S., Dufek, E.J., Howey, D.A., Viswanathan, V., Aykol, M., Beck, D.A.C., Blaiszik, B., Chen, B.-R., Crabtree, G., et al. (2022). Principles of the Battery Data Genome. Joule *6*, 2253–2271. 10.1016/j.joule.2022.08.008.

72. Che, Y., Hu, X., and Teodorescu, R. (2023). Opportunities for battery aging mode diagnosis of renewable energy storage. Joule *7*, 1405–1407. 10.1016/j.joule.2023.06.014.

73. Aykol, M., Gopal, C.B., Anapolsky, A., Herring, P.K., Van Vlijmen, B., Berliner, M.D., Bazant, M.Z., Braatz, R.D., Chueh, W.C., and Storey, B.D. (2021). Perspective—Combining Physics and Machine Learning to Predict Battery Lifetime. J. Electrochem. Soc. *168*, 030525. 10.1149/1945-7111/abec55.

74. De Angelis, V., and Preger, Y. (2020). Sandia National Laboratories Study Overview. https://www.batteryarchive.org/snl_study.html.

75. Herring, P., Balaji Gopal, C., Aykol, M., Montoya, J.H., Anapolsky, A., Attia, P.M., Gent, W., Hummelshøj, J.S., Hung, L., Kwon, H.-K., et al. (2020). BEEP: A Python library for Battery Evaluation and Early Prediction. SoftwareX *11*, 100506. 10.1016/j.softx.2020.100506.