Beyond Expert-Level Performance Prediction for Rechargeable Batteries by Unsupervised Machine Learning

Xi Chen, Luhan Ye, Yichao Wang, and Xin Li*

Predicting the performance of rechargeable batteries in real time is of great importance to battery research and industrial production, and hence has been a long pursuit. Previously, sophisticated apparatus is required to measure indicator properties of performance, while machine learning approaches based on feature engineering procedures require a priori expertise that is challenged by the complicated environment of real-world applications. Here, for a more effective real-time prediction of battery life and failure, a novel end-to-end unsupervised machine learning approach is shown; this approach is free from feature engineering and uses only the raw images of the charge–discharge voltage profiles. This model enables unsupervised real-time automatic extraction of latent physical factors that control the performance of Na-ion batteries to classify good or bad cycling performance by using only the voltage profile of the first cycle. This model can also monitor the safety of Li-metal battery systems by giving warnings when the battery is approaching a failure. With the beyond expert-level prediction ability, the abovementioned framework can be a promising prototype to further develop and enable high accuracy predictions of battery performance for real-world applications in the future.

1. Introduction

Rechargeable batteries lie in the center of the clean energy revolution in recent years.\(^1\)\(^-\)\(^5\) The major consumption of Li ion batteries has shifted from portable electronic devices to electric vehicles (EVs) after 2015.\(^6\) It is forecasted that more than 50% of the vehicles will be a certain level of electric-fuel hybrid by 2020.\(^7\) Due to the large capacity and power needed for batteries in EVs and the relatively closed environment of cars, the safety of rechargeable batteries has become a paramount concern, because failure due to short circuit or leakage of chemicals can potentially cause fatal damage to the drivers. Such safety concern has prevented the commercial use of Li metal anodes, although Li metal benefits from an ultra-high capacity (3860 mAh g\(^{-1}\)) compared with the commercial graphite anode (372 mAh g\(^{-1}\)).\(^7\)\(^-\)\(^10\) Also, for a typical battery test in battery research and development, up to thousands of charge–discharge cycles are needed to know the cycling performance. Thus, there is a strong interest across both academic and industrial communities in the possibility of predicting the cycling performance, life, and critical failure event based only on the first cycle, the initial cycles, or the few cycles before battery failure, which holds great potential of accelerating the research and development of battery materials, and improving the safety control procedures of EVs. Earlier theoretical and experimental work attempts to model the cycling performance based on thermodynamic and kinetic principles,\(^11\)\(^-\)\(^13\) or predict the performance based on the specific feature selection and delicate measurement of, for example, the accurate coulombic efficiency or impedance.\(^12\)\(^,\)\(^13\) However, there still lacks an unsupervised prediction capability for battery performance and failure in the absence of any predetermined specific features. It is thus very attractive if all features relevant to performance are extracted automatically in real time from the battery running data for the prediction of battery performance.

In the past decade, Na-ion batteries, as a more cost-friendly counterpart of Li-ion batteries, had attracted an increasing amount of research interest. Due to strong interactions between transition metal (TM) oxides and Na ions in NaTMO\(_2\), as well as possible solid electrolyte interface reactions, the voltage curves of Na-ion batteries often show complicated features because of the interplay with known and unknown phase transitions as a function of Na composition and cycling number.\(^14\)\(^-\)\(^23\) This makes it challenging to predict cycling performance by visual inspection based on human expertise. It is therefore desirable to be able to extract and learn the expertise by linking the feature in voltage profiles with the underlying mechanism for deeper understanding of the latent physical or chemical process during the material degradation. For example, Figure 1 shows the voltage profiles of two Na-ion batteries based on the same cathode material of Na\(_2\)(Fe\(_{0.24}\)Ni\(_{0.76}\))O\(_2\) from the first cycle to the 20th cycle. Although the first or the second cycles of the two batteries are almost identical to the human eye due to the lack of any significant local differences, their 20th cycles show very distinct shape and capacity difference. Thus, a successful prediction method...
should ideally be able to perform unsupervised real-time exploration and identification on the nuanced difference of those nonlocal features, and further understand the roles they play in determining cycling performance.

Recently, with the surge of computational power, machine learning and deep learning have emerged as a powerful method for predictions based on large amount of data.[23–26] For some tasks such as game playing,[27,28] and medical image recognition,[29] where humans were thought to perform really well, machine learning has outperformed human experts. Attempts of using machine learning to predict battery performance have also been made recently,[30–38] using methods ranging from regression to feedforward or recurrent neural on selected features from experiments. From the methodology point of view, these methods still rely on a priori feature selection and engineering of a few features from the field-specific expertise. Explicit machine learning models have been built based on these features, which have achieved certain sound results. However, models built on such fixed features for a specific material system is intrinsically challenging to be migrated for new systems with incomplete knowledge and hence undetermined relevant features.

Fortunately, unsupervised machine learning technique in principle can extract rich information about a system, with numerous automatically emerged features that may collectively and dynamically determine battery performance. Applying unsupervised learning to a battery dataset may thus also bring new scientific knowledge and insight beyond the current expert level for a better understanding of battery performance and failure processes. In this article, we present an end-to-end unsupervised machine learning platform free from feature engineering, with beyond expert-level performance prediction demonstrated from a proof-of-concept level small dataset. By treating the voltage curves as images, we can straightforwardly make unsupervised automatic extraction and visualization of the key features for parametrizing battery performance, and further build a support vector machine (SVM) to learn and predict the cycling performance and lifetime. SVM as a binary classifier aims to find a maximum margin hyperplane that separates the dataset after mapping them into the kernel space, which is not sensitive to the dataset size and hence suitable for this project with cycling data from limited number of batteries. In addition, we find that convolutional neural network (CNN), which was designed to learn image-like dataset, cannot extract enough effective features to perform classification as accurate as the SVM, due to the known effect of limited dataset size on CNN. However, the transfer learning procedure we designed, which can be considered as a complicated deep CNN model that uses the embedded layer of the Inception network,[39] can effectively comprehend features of the voltage curves and perform equally well as SVM in this small dataset. The advantage of neural networks, especially the Inception, will be magnified by increasing the scale of dataset, because deep neural networks possess a sufficient model capacity that can effectively learn from a large dataset with orders of magnitude of speed advantage over SVM. Therefore, our example also demonstrates a very promising machine learning direction of transferring knowledge from natural images to scientific image processing in general. Using similar insights, we also proposed and implemented an anomaly detector using an autoencoder neural network that can effectively predict the failure of Li-metal batteries. This is a desirable capability for any battery management system of EVs and electronic devices. Note that here we consider the abrupt failure of the battery caused by, for example, a short circuit or malfunction in Li-metal anode batteries, which is a complicated nonlinear phenomenon more closely related to the actual safety problem, compared with previous works that often define failure as the capacity fading beyond a certain level.[44]

2. Results and Discussion

We collected the voltage profiles of 62 Na-ion batteries from 14 different kinds of layered NaTMO$_2$ cathode materials with O3 oxygen stacking, as listed in Experimental Section, with each battery tested for 16 cycles. Note that these batteries were tested during the past several years by different researchers in the laboratory, and therefore, they are not specially prepared for this machine learning project. Intended to learn performance based only on shape information, we mathematically stretch the following cycles to match the horizontal length of the first cycle and prepare each cycling curve as a 2D array. This means that the direct information about capacity fading is largely removed from the raw data, which forces the prediction to focus on critical shapes of voltage curves. We can probe the features of the raw array by principal component analysis (PCA). Figure 2a,b shows the scatter plots of the distribution of raw arrays projected along the two most significant principal components (PC), with each point corresponding to one cycle of a battery. The cyclability

![Figure 1. Cycling profiles of two Na(Fe$_{0.24}$Ni$_{0.76}$)O$_2$ batteries. The battery in a) has significantly lower cyclability than the battery in b). Cyclability (of the first n cycle) is defined as the ratio of the discharge time of the last cycle (the nth cycle) over that of the first cycle. The corresponding 1st, 2nd, 10th, and 20th cycle voltage loops are marked with numbers and different colors. Charge and discharge directions of the voltage loops are marked by the arrows.](image-url)
of the first 16 cycles of a battery, defined as the discharge capacity of the 16th cycle over that of the first cycle, is used for labelling the data. Note that we give the same label to all the 16 data points from each battery based on the category of cycling performance of that battery, as either “high” for batteries with the highest 40% cyclability, “medium” for the middle 20%, or “low” for the lowest 40% among all the 62 batteries in this 2D projection. It can be seen that all data points distribute approximately as an arc, with one end occupied majorly by data points from batteries with “high” performance in green and the other end by “low” performance in blue. We also plot the “medium” ones in orange for the illustration purpose only, which lies around the middle part of the arc. However, in later predictions, we will only use a 50% versus 50% split of the dataset according to the cyclability and label the two parts as “good” or “bad.” For a given material composition, the corresponding batteries could fall into either the good or bad classes. From the two plots in Figure 2a,b, we can identify an interesting evolution trend for the cycling performance of two batteries made by the same material composition. If we call the mostly green corner the “good corner” and the mostly blue corner the “bad corner,” for a particular battery with bad cyclability (Figure 1a and 2a), although the points of its cycles first evolve to the good corner, it rapidly leaps toward the bad corner, as illustrated by those big points, while for a battery with good cyclability (Figure 1b and 2b), its cycling points maintain close to the good corner. This indicates that the battery with bad cyclability has undergone some critical degradation process. The above observation confirms that the cycling performance is strongly correlated with the shape information of voltage curves, as expanded by the distinct separation of data points in the PC space.

Our PCA method also automatically extracts the most distinct features that separate batteries via the PC. The three most significant PCs are shown in Figure 2c–e with interpretable features. Figure 2c is a filter that accounts for overall voltage polarization between charge and discharge. Because in this colormap, points that are yellower correspond to more positive value and bluer means more negative value, we can tell that if a voltage curve shows smaller polarization, that is, it lies more in the darker region inside the lighter yellowish region, it gives a more negative coefficient of the first PC (x axis in Figure 2a,b), matching the “good corner” at smaller value in x axis in Figure 2a,b. On the contrary, if the voltage curve has larger polarization and lies in the lighter yellowish region, the coefficient will be more positive, matching the “bad corner” at larger x value. Note that the region outside the yellowish belt, corresponding to extremely large polarization, is also dark. However, this does not mean that larger polarization prefers good cyclability; it simply reflects the fact that our dataset does not include any voltage loop that lies in that region, thus providing no information for PCA to consider. If we include voltage loop data from a battery with such an extremely large polarization, the PCA method will explore this region outside the yellowish belt. The second automatically identified PC in Figure 2d seems to pay more attention to the coulombic efficiency, defined as the ratio between the discharge and charge capacities in each cycle. The lowest part of the bright region ends almost at the origin; thus, a voltage curve with
the final part of discharge overlapping with this bright region, corresponding to a high coulombic efficiency, makes the coefficient of the second PC (y axis in Figure 2a,b) larger that matches the “good corner.” However, some additional features also play an important role here, such as the bright region near the low voltage plateau in discharge and the less bright region near the high voltage step in charge, suggesting that batteries with such clearer features in the voltage profile show better cycling performance. Similarly, the third PC (Figure 2e) accounts for the beginning of low voltage plateau in charge and the beginning of discharge at high voltages. From the second and third PC, we can infer that in general, good cycling performance is positively correlated with the preservation of voltage profile features at high and low voltage regions, while it is less correlated with the ones in the middle voltage region with smooth featureless voltage profiles. It is known that sodium-layered structures often undergo complicated gliding of metal oxide layers to form different stackings during the battery cycling process, with, for example, the P3 type stacking in the middle voltage range and the O3 type stacking at high and low voltages, as illustrated in Figure 2c–e. Thus, the automatically extracted PCA here also suggests that 1) O3 phases might be more important to the battery cyclability than P3 phases, 2) the multiple P3 to O3 phase transitions are also critical, and 3) the P3 to O3 phase transitions at high voltage in charge and at low voltage in discharge might be more important than the O3 to P3 transitions at low voltage in charge and at high voltage in discharge, that is, the direction of phase transition matters and the P3 to O3 direction is more relevant here. This knowledge is more specifically related to the layered NaTMO2 system, which is much less obvious than the general expertise of polarization and coulombic efficiency in governing battery performance. The underlying message, however, might be more complicated and may inspire further investigations, because 1) P3 phases generally possess much higher Na diffusivity than O3 phases at similar interlayer distance, indicating that Na diffusion in some O3 phase might be a limiting factor to cyclability and 2) the P3 to O3 phase transition is a more general phenomenon than the O3 to P3 one at high voltages in layered NaTMO2, for example, the high voltage O3 to P3 transition in discharge can be skipped in the Fe-containing layered Na(Fe,TM)O2 when Fe composition is higher than around 30%, due to the formation of the novel rippling phase toward the end of charge, while the P3 to O3 phase transition in charge still maintains. To summarize, based on the first three most important PCs, we have observed and interpreted that minimizing voltage polarization, maximizing coulombic efficiency, and preserving features related to O3 phases are beneficial for better cyclability.

It is worth noting that although some of these above interpretations may already be known to the field and some are not, the most important aspect here is that all these major PCs are learned in an unsupervised way that surprisingly match the expertise and some of the most advanced knowledge in the field, with rich information to inspire future scientific studies. Apart from the first few PCs, there are also more PCs that contribute to non-negligible variance of the dataset; thus, they are all important to the prediction of cycling performance. We have listed them in Supplementary Information (Figure S1 and S2, Supporting Information) and ranked them as more positively (Figure S1, Supporting Information) or more negatively (Figure S2, Supporting Information) correlated with better cyclability. Some correlations identified by the PC may indicate new physics and are worth further investigations. For example, the formation of some middle to high voltage plateaus are positively correlated to good cyclability (Figure S1b, Supporting Information), and a complete featureless discharge is negatively correlated to good cyclability (Figure S2g, Supporting Information). We expect such analysis to show further importance in the future in learning a larger dataset or analyzing a new material system, where more interpretable PC technically can be extracted for new scientific understandings related to the battery performance.

Although the above 2D projections are indicative enough, they have not fully utilized the classification ability of machine learning algorithms for high-dimensional data. We further feed the coefficients of the PC to an SVM with radial basis function (RBF) kernel, where the optimal prediction ability is achieved using more than 50 PCs simultaneously. We note that although directly interpreting most of these PC features is not easy, their collective contribution is important in predicting battery performance. We find that using only the first cycle of the battery voltage profile can achieve higher than 80% accuracy on average for cyclability prediction, as shown in Figure 3a. Although the performance increases slightly if using later cycles to predict, the result still indicates that the majority of information about cycling performance is already included in the first cycle. Here, training and testing are done by first splitting the 62 batteries into a training set of 50 batteries and a testing set of 12 batteries. All the cycles of the training set batteries are then compared to good cyclability (Figure S1b, Supporting Information). We expect such analysis to show further importance in the future in learning a larger dataset or analyzing a new material system, where more interpretable PC technically can be extracted for new scientific understandings related to the battery performance.
used to train the SVM, while only the data from a particular cycle of the test set batteries are used to make the tests. We note that using all the cycles of the batteries from the training set to train the first cycle can slightly increase the accuracy from the training that uses only the first cycle, because the evolution direction of the data points of the successive cycles nicely follows their category; this fills the gap in the parameter space and thus facilitates the prediction of the kernel SVM. The detailed accuracy by cathode compound type in Figure 3b shows that for most compounds, the accuracy is beyond 80%, with some limited exceptions near 30% Fe composition.

Figure 4 shows the data point distribution of two types of materials with distinctive prediction accuracies, Na(Fe0.8Ni0.2)O2 and Na(Fe0.3Ni0.7)O2, illustrated in the space of the first two PCs. The first cycles of Na(Fe0.3Ni0.7)O2 batteries distribute within the region of low cyclability, which thus can be easily classified by the SVM. The first cycles of Na(Fe0.8Ni0.2)O2 batteries do not lie in the region of their classes, and for quite a few batteries, their first cycles are near the transition between good and bad regions; this causes some difficulty for the SVM to make the correct classification, leading to lower prediction accuracy for this particular compound. In other words, the first cycle of Na(Fe0.3Ni0.7)O2 contains less information about the cyclability than that of other Na(Fe0.8Ni0.2)O2 compounds, suggesting that the lower accuracy of cyclability prediction for Na(Fe0.3Ni0.7)O2 is due to the more complicated electrochemical behavior in the initial cycle of the compound itself rather than the prediction methodology. Interestingly, the prediction accuracy versus Fe composition in Na(Fe0.8Ni0.2)O2 forms a reversed dome shape in Figure 3b with the minimum at around 30% Fe composition. We thus believe that it is most probably related to the complicated collective electronic behavior of FeO6 octahedra near the percolation threshold at around 30% Fe, as discussed specifically in our previous work, where the initial capacity also shows a dome shape versus Fe composition with the maximum at around 30% Fe. The percolation threshold here is the value of Fe composition beyond which the Na diffusion pathway is fully connected by the FeO6 octahedra in the neighboring two sides of the metal oxide layers. Thus, near x ≈ 0.3 in Na(Fe0.8Ni0.2)O2, the system is expected to undergo abrupt changes at high voltages in terms of Na diffusivity and collective electronic behavior, giving novel phases such as the rippling phase.

We further examine this phenomenon by machine learning analysis in the 2D space spanned by the first two PCs in Figure S3 and S4, Supporting Information. Figure S3, Supporting Information plots the first cycle distribution of Na(Fe0.8Ni0.2)O2 compounds near 30% Fe composition in such a 2D space, including Na(Fe0.24Ni0.76)O2, Na(Fe0.27Ni0.73)O2, Na(Fe0.3Ni0.7)O2, and Na(Fe0.33Ni0.67)O2, which show much larger variance than other materials plotted in Figure S4, Supporting Information. The larger variance in PCA here compared with other compounds indicates that near the percolation threshold, the first cycle electrochemical voltage loop of Na(Fe0.8Ni0.2)O2 compounds might be much more sensitive to small variance in the initial value of either the structural parameters of, for example, transition metal clusters formed in the synthesis, or the randomness in battery assembly. The origin of such sensitivity might be closely related to the coupled collective electronic and crystallographic phase transitions near the percolation threshold that is worth further investigations. We also note that if later cycles are used to predict the cycling performance of Na(Fe0.3Ni0.7)O2, the accuracy increases much faster than that of other Na(Fe0.8Ni0.2)O2 compounds near 30% Fe composition due to the more complicated electrochemical behavior in the initial cycle of the compound itself rather than the prediction methodology. Interestingly, the prediction accuracy versus Fe composition in Na(Fe0.8Ni0.2)O2 forms a reversed dome shape in Figure 3b with the minimum at around 30% Fe composition. We thus believe that it is most probably related to the complicated collective electronic behavior of FeO6 octahedra near the percolation threshold at around 30% Fe, as discussed specifically in our previous work, where the initial capacity also shows a dome shape versus Fe composition with the maximum at around 30% Fe. The percolation threshold here is the value of Fe composition beyond which the Na diffusion pathway is fully connected by the FeO6 octahedra in the neighboring two sides of the metal oxide layers. Thus, near x ≈ 0.3 in Na(Fe0.8Ni0.2)O2, the system is expected to undergo abrupt changes at high voltages in terms of Na diffusivity and collective electronic behavior, giving novel phases such as the rippling phase.

![Figure 4](image-url) **Figure 4.** Distribution of the first two principal components for the first cycles of a) Na(Fe0.3Ni0.7)O2 and b) Na(Fe0.8Ni0.2)O2 batteries. Green and blue dots indicate data from other materials. Red and purple circles are the first cycle data of the corresponding material. Here, the separation according to cyclability is 50:50 high and low used in the actual prediction instead of 40:20:40 high, medium, and low separation in Figure 2.
materials, our approach indeed has the ability to identify and learn more generalized features for battery performance, rather than being limited to a specific type of cathode or anode system as in previous literature. We also want to emphasize that a direct comparison of the absolute accuracy of machine learning prediction from different works is not appropriate, as an optimized dataset can easily increase the prediction accuracy beyond 90% even when only based on the “vanilla” machine learning technique such as regression; however, its effectiveness will be challenged in a complicated application environment, where the dataset is more randomly generated from different electrode materials, different batches of electrolytes, and assembled by different researchers, as somewhat represented by our dataset (see Experimental Section).

It is worth noting that our method of preparing voltage curves as images is physically a very natural selection. Although our 2D voltage profiles (Figure 1) contain the same amount of information as the 1D series of raw voltage values, it is observed that if we change the input to the latter, it not only fails to extract physically meaningful PC but also achieves an overall prediction accuracy of only up to 74%. This indicates that although 2D voltage profile does not create more information than the raw data, the form of data organization explicitly emphasizes critical information about temporal correlation of charge and discharge voltage curves that is important to analyze cycling performance, such as the general polarization or specific voltage plateaus between charge and discharge. The 2D voltage profile can also take advantage of the image recognition capability of neural network models. Due to the limited dataset size here, however, we find that SVM can capture more shape features and make better prediction than the CNN, although CNN was designed for learning image-like datasets given sufficient data (Figure 5a).

Also, transfer learning using pre-trained Inception model by Google\textsuperscript{[19]} (See Experimental Section), which is a complicated deep CNN module, outperforms several few-layer CNN design and performs equally well as SVM in this dataset with limited size (Figure 5a). The performance of transfer learning on our Na-ion electrode materials is shown in Figure 5b. Compared with Figure 3b, it shows that transfer learning using Inception treats the features differently than SVM, leading to more uniform performance on different materials, and the model performs better on the 4 batteries outside the Fe/Ni/Co class. The result suggests that various machine learning methods may interplay differently with the same features on the voltage curve as well as the underlying evolution of battery material’s details.

Inspired by the effectiveness of our machine learning model in predicting cyclability, we further consider predicting battery failure. For this purpose, we collected the voltage profiles of 87 Li-metal anode battery tests with cycle life larger than 100 before the failure from our laboratory database. Note that these Li-metal anode battery tests are not specially designed for the machine learning task, which include asymmetric cell tests with the 2D or 3D Cu of various geometric shapes on one side and the 2D Li-metal foil on the other side as current collectors, and symmetric cells with the Li-metal foils on both sides. Here battery failure refers to the abrupt nonlinear change of voltage profile at a certain cycling point rather than the more linear capacity fading as discussed in the previous case. Such failure in Li-metal batteries is most probably caused by the dead Li and detrimental solid electrolyte interface (SEI), the Li dendrite growth-induced internal short circuit, and/or the electrolyte depletion. Figure 6a shows the voltage profiles of the entire cycle life for such four different batteries. Compared with the voltage profiles of Na-ion batteries, the uniqueness of Li-metal batteries here raise the question of whether it is still possible to predict failure. First, the profiles are with much less shape features, and clearly observable change of the profile shape does not exist for every battery. For example, in Figure 6a, only the bottom two profiles change significantly toward the end of cycle life that started tens of cycles before the actual failure, while it seems that no obvious changes can be observed in the top profile. Second, similar to the failure of light bulbs or transistors, it seems the failing process of an Li-metal battery also shares some nature of an exponential stochastic process with memoryless property. For example, in addition to the seemingly sudden failure observed in the top example of Figure 6a, in the third example of Figure 6a, although significant changes occurred at around 40% of its cycle life, it continues to run normally for hundreds of cycles before the actual failure. Third, the number of normal and abnormal cycles are highly unbalanced in a battery test dataset, because the signal of a failing battery, even when observable by visual inspection, often only lies in the immediate cycles before the failure. This impedes the traditional classification schemes that require much more balanced dataset. Apart from these difficulties, the wide distribution of cycle life ranging from 100 to 500 among the 87 batteries shown in Figure 6b, and the fact that unlike cathode materials, the discharge capacities in the raw data here show no general trend of decreasing with cycling, add additional concerns about whether...
there is enough information in the dataset to enable failure prediction here.

We thus designed an anomaly detector using autoencoders\cite{25} to address the aforementioned major difficulties in failure prediction, especially the unbalanced dataset problem. As shown in the schematics of Figure 7a, by feeding the voltage profile pictures to the autoencoder neural network and requiring it to reconstruct the input picture for the output, the autoencoder will extract important features in the middle feature layer with the least number of neurons, also known as the “information bottleneck.” Based on the hypothesis that there are some critical hidden changes in the voltage cycles approaching battery failure, we technically separate the cycles into two categories, that is, the last 30 cycles before the failure (“failing cycles”) and all the cycles before the last 30 cycles (“normal cycles”). By first training an autoencoder using only the normal cycles, we find that the autoencoder produces much larger reconstruction error on a test input from the failing cycles than the normal cycles of all other validation batteries (Figure 7b). We then use this gap of reconstruction error to set up a threshold, beyond which any input cycling profile to the model will receive an anomaly warning. The performance on the test battery dataset is summarized in Figure 7c, where “fail_accu” means the chance of at least detecting one anomaly in the last 30 “failing cycles” and “normal_accu” means the chance of giving no false alarm of anomaly in the “normal cycles.” The figure shows that this anomaly detector is adjustable based on different engineering needs. For example, if the threshold “ERR-recon” is set as 0.04 near the crossover of the two curves, there is around 70% chance that an anomaly is detected in the failing cycles of a battery and also around 70% chance that no false alarm is given. If the threshold is increased to 0.19 instead, then all the predictions are completely free from false alarm in the normal cycles, and there is still around 25% chance that an anomaly will be detected in the failing cycles.

Note that in the algorithm, we can also use the last 20 or 40 cycles instead of 30 cycles to define the failing cycles. However, our numerical experiment in Figure 7c–e corresponding to the choice of the last 30, 20, and 40 cycles as failing cycles, respectively, show that the choice of 30 cycles gives the largest separation between the reconstruction errors of normal cycles and failing cycles and thus the best prediction performance. This finding also indicates that anomalies related to battery failure in our voltage curve dataset majorly lie in the last 30 cycles in these Li-metal batteries, regardless of the material or geometry of the Cu current collectors, the electrolytes, and the additives in our experiments. This new fact about the importance of the last 30 cycles found by our machine learning analysis may suggest that the critical issue that triggers the failing process takes around 30 cycles on average to evolve to the actual failure in these Li-metal batteries. The exact Li cycling dynamics in the last 30 cycles, including dendrite growth, electrolyte depletion or dead Li, and SEI deposition mechanisms, is thus worth a further exploration.

We then compare the performance of this autoencoder anomaly detector with linear methods such as PCA. A similar anomaly detector based on the reconstruction using PC is shown in Figure S5, Supporting Information. Due to the abundance of the normal cycles for training, the autoencoder, which benefits from the nonlinear operation of neural networks, shows better performance than PCA. The entire curve of failure prediction (“fail_accu”) given by PCA in Figure S5, Supporting Information is lower than that by autoencoder in Figure 7c. Also, the PCA method cannot detect any anomaly if we require no false alarm (normal_accu = 1). Although the result of PCA is inferior to that of nonlinear neural network autoencoders, PCA as a linear model is interpretable; thus, its PC can provide understanding about the features used for detecting anomaly. The autoencoder model may also use similar features, but its superior prediction power relies on the nonlinear operation that is difficult to interpret. Figure 7f shows the first two PC of the normal cycles and the last 30 failing cycles. Note that we process the charge and discharge curves to 2D images in a similar way to the previous case of Na-ion battery. Remember that in the Na-ion battery case, to focus on shape information, the absolute value of capacity in each cycle was removed, while the voltage information of the curves was retained, which was learned to give the PC about overall polarization in Figure 2c. This procedure was justified because those Na-ion batteries were based on the same half-cell configuration, with different cathode voltages versus the same anode material of Na metal. However, in our Li-metal battery dataset, this is not the case due to the mixed configurations of

**Figure 6.** Voltage profiles and cycle life distribution of Li-metal battery tests. a) Voltage profiles of the entire cycle life of four batteries. b) The distribution of the cycle life of the 87 batteries used in this study.
symmetric and asymmetric cells (see Experimental Section), thus making a direct comparison of battery voltage problematic. We therefore take each charge–discharge voltage profile as a complete 2D image here with the absolute voltage value on the $y$ axis.

Figure 7. Anomaly detector for battery failure prediction. a) Schematic illustration of the neural network autoencoder used for building the anomaly detector. b) Reconstruction error of the cycles from the testing battery. c) Prediction accuracy versus the threshold of the reconstruction error ERR-recon as anomaly indicator, where normal_accu corresponds to the accuracy of detecting no anomaly in the entire “normal cycle” life before the last 30 “failing cycles,” and fail_accu corresponds to the accuracy of detecting at least one anomaly in the failing cycles. d,e) Prediction accuracy corresponding to labeling the last 20 and 40 cycles as “failing cycles”. f) The first and second principal components of “normal cycles” and “failing cycles.” Brighter and darker colors indicate more positive and negative values, respectively, similar to Figure 2. g,h) Examples of original curves of “normal cycles.” i,j) Examples of original curves of “failing cycles.”
removed, that is, the 2D images do not have the same voltage scale, and our learning here focuses completely on the shape information.

We show that these PCs again may contain some physically meaningful information. Compared with the PC of normal cycles, we can see in Figure 7f the feature of sharper peaks in the failing cycles defined by the yellow region toward the end of charge on the right side of the first and second PCs as well as the feature of larger slope in voltage curves toward the end of discharge on the left side of the first PC. Some exemplar cycling curves from the normal cycles (Figure 7g,h) and the failing cycles (Figure 7i,j) are plotted, where both features can be observed. The sharp peak (Figure 7h,i) that is observed only in asymmetric cells means a quick increase of polarization and resistance toward the end of charge, corresponding to the end processes of Li stripping from the Cu side and plating to the Li side. The first feature of the sharper peak in failing cycles hence suggests that the resistance associated with either of the two end processes increases faster when approaching the battery failure. This seems to be more directly related to electrolyte depletion or increased thickness of dead Li or SEI layer, rather than to short circuit by Li dendrite penetration. However, catastrophic dendrite growth can obviously contribute significantly to these related processes. The second feature of larger slope at the end of discharge in the failing cycles (Figure 7j) can usually be found after several hundred normal cycles in either symmetric (Figure 7g) or asymmetric cells (Figure 7h), which was previously related to the dissolution of moss-like Li2O2 generated through cycling. More careful examination of those batteries with observable changes upon cycling reveals that most of the symmetric batteries approach failure with similar voltage curves like those shown in Figure 7f, while asymmetric cells may fail in either way of Figure 7i or 7j.

In summary, our machine learning platform has discovered two most important failing signals based on analyzing the 2D images of voltage curves. The first is a gradually sharpening voltage spike at the end of charge (Figure 7i) for asymmetric cells. The second is an increasing voltage slope at the end of discharge (Figure 7j) for both symmetric and asymmetric cells. Both signals in the failing cycles are not directly related to Li dendrite penetration, but they may be generated by a catastrophic dendrite growth. More importantly, we demonstrate for the first time that unsupervised machine learning is able to recognize, with such numerical confidence, these reasonable and meaningful shape features in the voltage curve as the most critical PC that correlate with the electrochemical failure in Li-metal batteries. Note that similar to the case of Na-ion batteries, here the features learned are also general across batteries with various kinds of current collector designs, symmetric and asymmetric cell configurations, battery assemblies, and electrolyte systems that were not optimized for this machine learning project.

We emphasize that a huge advantage of our method is that it is powerful when new relevant features emerge through battery cycling, which may not be relevant at the early stage of battery running or outside the field-specific knowledge. Any such new feature will trigger a large reconstruction error of the autoencoder and thus raises an anomaly warning. This capability is especially critical in monitoring battery failure, as demonstrated in the example related to Figure 6 and 7 here, because abrupt catastrophic changes in the battery electrochemical environment is often recipient to the actual failure, where case-dependent new features often emerge collectively from the chaos, to fail any method that relies on the engineering of a fixed set of universal battery features a priori known by battery experts. Our method thus can largely improve the safety of the battery system in EVs. Technically, our platform goes beyond the elementary regression-based machine learning and completely abandoned feature engineering, thereby exhibiting for the first time the ability to dynamically learn from the battery running data. We believe that our method here is closer to the essence of machine learning and also a leap toward the practical application, where real-time battery failures often follow complicated mechanisms and patterns.

3. Conclusion

In this work, we have implemented an end-to-end, feature engineering-free unsupervised approach to learn and predict battery cycling performance, including cyclability, lifetime, and failure. Compared to previous machine learning works on battery performance prediction, our work is free of feature engineering and can make predictions of cyclability by using only the first cycle or first few cycles for a dataset consisting of various kinds of cathode materials. This suggests that our method will likely be more effective in complicated real-world applications with the ability of learning every relevant correlation emerged from the data that is often beyond a fixed input of expert knowledge, a critical fact that restricts the performance of a predetermined feature ensemble by a priori expertise.

By normalizing the range of the voltage loops and keeping only the shape information, applying PCA on this dataset reveals critical features to battery performance. An SVM with RBF kernel classifies good or bad cycling performance with above 80% accuracy on average by using only the voltage loop of the first cycle. Due to the limited amount of experimental data, the CNN does not possess equal accuracy. However, we find that large-scale, deep CNN models pre-trained on a vast number of natural pictures such as the Inception network, show great potential of transfer learning from natural images to scientific images. The Inception network outperforms various few-layer CNN designs, with a performance equal to that of SVM on this sparse dataset. We believe that with more experimental data, it is possible to make transfer learning outperform traditional multilayer classifiers.

We also implemented an anomaly detector using autoencoder for predicting the battery failure, which shares the same spirit as predicting the cycling performance. Our method gives anomaly warnings close to battery failure with superior performance to linear models. It also shows the engineering flexibility to balance the failure prediction accuracy and the false positive alarm, thus showing great potential for real-world applications. Although the current dataset is only at a proof-of-concept size, our platform already shows its powerfulness. In application to large-scale dataset, recurrent neural networks (RNN)[25] might be a good candidate to explore more information encoded in the successive changes of the voltage loops.
4. Experimental Section

Experiment: For predicting the cyclability of Na-ion batteries, the dataset has 62 batteries made of 14 different kinds of materials, including 3 NaNiO2, 7 Na(Fe0.1Ni0.9)O2, 3 Na(Fe0.2Ni0.8)O2, 8 Na(Fe0.24Ni0.76)O2, 2 Na(Fe0.3Ni0.7)O2, 3 Na(Fe0.4Ni0.6)O2, 2 Na(Fe0.5Ni0.5)O2, 2 Na(Fe0.27Co0.73)O2, 5 Na(Fe0.27Ni0.73)O2, 11 Na(Fe0.3Ni0.7)O2, 8 Na(Fe0.33Ni0.67)O2, 2 Na(Fe0.4Co0.6)O2, 3 Na(Fe0.4Ni0.6)O2, 3 Na(Fe0.5Ni0.5)O2, and 4 batteries with various Cu-metal anode batteries, the dataset consists of 87 batteries made with 62 different researchers in the laboratory, that is, these data were not specially prepared for the machine learning project. For predicting the failure of Li-metal anode batteries, the dataset consists of 87 batteries made with various Cu fiber substrates similar to those reported in previous publications,[45,46] with voltage cutoff of 4.5 V and current rate of 0.1 C. These batteries were assembled and tested in a time span from November 2015 to September 2016 by different researchers in the laboratory, that is, these data were not specially prepared for the machine learning project. The failure of Li-metal anode batteries, the dataset consists of 87 batteries made with various Cu fiber substrates similar to those reported in previous publications,[45,46] with voltage cutoff of 4.5 V and current rate of 0.1 C. These batteries were assembled and tested in a time span from November 2015 to September 2016 by different researchers in the laboratory, that is, these data were not specially prepared for the machine learning project.

Machine Learning Models:

1) Principal component analysis. PCA is a dimensional reduction method. Given a dataset, it tries to find an orthogonal subspace of the original sample space with maximized covariance. For a dataset written as a p by n matrix $X$, where $p$ is the number of samples and $n$ is the length of each sample, the $i$th PC is the eigenvector $\eta_i$ of the covariance matrix $C = (X - \bar{X})(X - \bar{X})^T$ corresponding to the $i$th largest eigenvalue, where $\bar{X}$ represents matrix transposition and $X$ is the mean of $X$. Because the eigenvectors $\{\eta_i\}$ span an orthogonal space, using the first $P$ PCs, each data point can be approximated as

$$X_j = \bar{X} + \sum_{i=1}^{P} c_j \eta_i$$

where the coefficients $c_i = (X_i - \bar{X}) \eta_i$. For the example of Na-ion battery prediction in this article, the input vector $X_j$ is a length of 400*250 vector corresponding to a voltage profile image in the first part of the Na-ion battery prediction, and is a length of 500 vector of successive change in the shape of the voltage profile in the second part of the Li-metal anode prediction. The PCA is implemented with the sklearnn[46] package. The amount of variance of the data that is explained by each $\eta_i$ is decreasing very fast, namely, the spread of the original data can be usually captured by the first few or first tens of $\eta_i$, depending on dataset complexity.

2) Support vector machine. SVM with kernel transformation is a binary classifier that aims to find a maximum margin hyperplane that separates the dataset after mapping them into the kernel space.[44,45,46] We used SVM with soft margin and RBF kernel. The mathematical formulation of the optimization problem[46] given the training data in the form of vectors $x_i \in \mathbb{R}^n, i = 1, \ldots, n$ and the labels $y_i \in \{-1, 1\}^n$, is

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha$$

subject to $y_i \alpha = 0$

$$0 \leq \alpha_i \leq C, \quad i = 1, \ldots, n$$

where $e$ is the vector of all ones and $C > 0$ is a parameter that penalizes misclassification. $Q$ is an $n$ by $n$ positive semidefinite matrix, $Q_{ii} \equiv y_i y_j K(x_i, x_j)$, where

$$K(x_i, x_j) = e^{-\gamma ||x_i - x_j||^2}$$

is the RBF kernel function. Similar to the PCA, for the actual implementation of SVM in this work, $x$ are the 400*250 vector of voltage profile images and $y_i$ are the labels of “good”/“bad” cyclability. All above equations are combined into a decision function for the test data $x$:

$$\text{sgn} \left( \sum_{i=1}^{n} y_i \alpha_i K(x_i, x) + \beta \right)$$

where $\beta$ is an intercept related to the dataset. The SVM in this article is implemented with the sklearnn package.

3) Neural networks. In this work, we used the Inception-v3[39] model provided by TensorFlow Hub and the self-constructed CNNs.[25] CNN is a type of deep neural network modified from the fully connected multi-layer perceptrons (MLP), especially suitable for data with dimension higher than two, such as images or movie clips that can be viewed as an image array. Instead of mapping all information from every layer to the successive layer by a nonlinear activation function on top of the linear transformation in MLP, CNN has dedicated convolutional layers, where many convolutional filters with variable sizes scan through different local areas of the high-dimensional data and collect the output of convolution. This enables CNN to identify and emphasize the local correlations in the data. The Inception model was a large-scale, state-of-art neural network image classifier with CNN as the central building block.[39] It was trained on millions of natural images from 1000 different classes on the ImageNet. In this work, we use the Inception to apply a transfer learning technique by collecting the output of the Inception network from a layer (“feature layer”) before the classification layer. The structure from the input layer to this feature layer acts as a general feature extractor, and it can be used to extract features in our battery voltage profiles. With all the parameters in this feature extractor, the Inception model freezes, that is, sets itself as un-trainable; we then attach a new classification layer for voltage profiles to it and use the data to re-train this classification layer. The resulting model from this transfer learning technique shows much better performance than the vanilla CNN model and is as good as the SVM on our data. The neural network parts are implemented with the TensorFlow[47] package. The autoencoder[25] neural network we constructed is a three-layer MLP with 32 neurons as the middle feature layer. The input and output layer equals to the number of pixels in the picture, which is 4096 in our case. An autoencoder neural network constructed based on CNN achieves similar results.

In this article, each cycle of voltage loop is prepared as a 400 × 250 2D array, mathematically stretched by interpolating 400 points in the horizontal time axis and 250 points in the vertical voltage axis. For predicting the cyclability of Na-ion batteries, we used a train/test split of 50:12 to ensure that the model is trained sufficiently. For predicting the failure of Li-metal anodes, we put more samples into the test set because of better data availability, which makes a train/test split of 60:27. Variation in the train/test split will not qualitatively affect the learning result.

Supporting Information
Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements
The authors thank helpful discussions with Michael Brenner.

Conflict of Interest
The authors declare no conflict of interest.
Keywords
battery failure prediction, battery performance prediction, machine learning