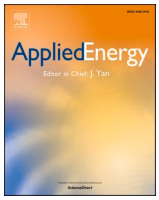
Applied Energy 306 (2022) 118134

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy



Deep neural network battery life and voltage prediction by using data of one cycle only

Chia-Wei Hsu [a], Rui Xiong [b,d], Nan-Yow Chen [c,*], Ju Li [d,*], Nien-Ti Tsou [a,*]

[a] Department of Materials Science and Engineering, National Yang Ming Chiao Tung University, Taiwan
[b] National Engineering Laboratory for Electric Vehicles, School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China
[c] National Center for High-performance Computing, National Applied Research Laboratories, Taiwan
[d] Department of Nuclear Science and Engineering and Department of Materials Science and Engineering, MIT, Cambridge, MA 02139, USA

HIGHLIGHTS

- The first deep neural networks for used battery age, EoL, RUL, and cycle-by-cycle discharge voltage/power prediction.
- Highly accurate predictions based on one cycle only.
- Dimension reduction, last padding technique, and convolutional training strategy enable flexible input and high accuracy of predictions.
- Data-driven features are extracted showing more influence on battery properties than human-picked features.

ARTICLE INFO

Keywords:
Deep neural network
$LiFePO_4$/graphite cells
End-of-life
Remaining useful life
Data-driven features

ABSTRACT

Rechargeable batteries, such as $LiFePO_4$/graphite cells, age differently by variability in manufacturing, charging (energy inflow) policy, temperature, discharging conditions, etc. Great economic and environmental value can be extracted if we can predict how a battery ages and ascertain its current state of health and residual useful life, based on just a few cycles of testing. Here, by developing novel-architecture deep neural networks with a special convolutional training strategy and taking advantage of recently published battery cycling data, we show that one can predict the residual life of a battery to a mean absolute percentage error of 6.46%, using only one cycle of testing. The cycle-by-cycle profiles, such as discharge voltage, capacity, and power curves of any given cycle, of used batteries with unknown age can also be accurately predicted for the first time. Moreover, our models can extract data-driven features from the data which were much more influential on the predicted properties than human-picked features. This work has shown that single cycle data contains a sufficient amount of information to predict essential battery properties with high accuracy. It is expected to provide tremendous economic and environmental benefits since reuse and recycling of batteries can be better planned and less lithium-ion batteries end up in landfills.

1. Introduction

Rechargeable batteries exemplify industrially manufactured devices. Large amounts of energy flow in and out of a battery cell, which has complex chemistry and mechanics inside, but on the outside, manifests as just current–voltage *I-V*, exterior temperature *T*, and other measurables. Battery cycle life depends on (a) the battery chemistry, (b) battery manufacturer and design, (c) variability in manufacturing, (d) history of charging and discharging, e.g. cutoff voltages, charging currents (policy), discharge (use) scenarios, ambient temperature, etc. Highly nontrivial tradeoffs exist between the battery cycle life, the energy output (e.g. by varying the cutoff voltages [1]) and fast-charging [2]. Given a used battery of a certain manufacturer and model, accurately assessing its state of health (SOH) to predict the remaining useful life (RUL), with as little testing as possible, will provide tremendous economic and environmental benefits. It is known that by improving the battery management system (BMS) software, one could potentially double the battery cycle life [3], so less lithium-ion batteries (LIB) end up in landfills, and reuse and recycling of batteries can be better planned. In addition, as these rechargeable batteries, such as $LiFePO_4$ (LFP)

* Corresponding authors.
E-mail addresses: nanyow@narlabs.org.tw (N.-Y. Chen), liju@mit.edu (J. Li), tsounienti@nctu.edu.tw (N.-T. Tsou).

https://doi.org/10.1016/j.apenergy.2021.118134
Received 17 July 2021; Received in revised form 9 September 2021; Accepted 25 October 2021

cells, have been widely used in electric vehicles (EVs) [4], the resale value of used EVs can also be better estimated. Imagine one is given a used rechargeable battery of a known brand, but with an unknown number of pre-existing charge–discharge cycles s. In order to assess its SOH, one can proceed to perform a additional cycles of charge–discharge testing to acquire new $I(t)$-$V(t)$ data. Ideally, if one can take $a = 1$, that is, with just one charge–discharge cycle testing of a used battery of unknown s, and can then predict the future cycle life (RUL) and $I(V)$ characteristics accurately, it will be tremendously helpful for building an intelligent BMS [5]. Using integer end-of-life (EoL) to denote the cycle life of a new battery just out of the factory, then $\underline{\text{EoL}} = \underline{s} + a + \underline{\text{RUL}}$, where underlined quantity denotes the ground truth. But since we do not know the true $\underline{s}$ value, and thereby can only make estimates of s and RUL, there is also $\text{EoL} = s + a + \text{RUL}$, where quantities without underline are one's estimates.

The estimation and control of dynamical systems, traditionally dominated by limited-memory computations such as Kalman filter, is being challenged by deep neural network (DNN) based algorithms. While it can often be computationally intensive to train a DNN, with modern GPU/FPGA/ASIC hardware, a DNN with $\sim 10^7$ weights can be run in real-time in hand-held devices [6–8]. It thus behooves us to see how well state-of-the-art DNNs can perform in RUL-RUL for different acquisition cycling numbers a. In addition to the residual life, we could also task DNN to estimate the current age s, and by applying analysis tools like Deep Taylor Decomposition (DTD) [9], we could task the DNNs to rank the importance of *features* in reducing the root mean square error (RMSE). That is, in the acquired cycling time-sequence data $I(t)$-$V(t)$-$T(t)$, what might be the most important features for making accurate predictions, and how well they agree with human expert intuition.

In their ground-breaking work in 2019, Severson et al. [10] provided the world's largest open-source dataset consisting of 124 $LiFePO_4$/graphite cells being cycled to end-of-life EoL, defined by the cell's discharge capacity Q dropping below 80% of its out-of-factory value. Many different fast-charging policies were used on the 124 cells. EoL was shown to vary widely, ranging from 150 to 2,300 cycles, due to the variabilities (c), (d) in the beginning paragraph, with average $\langle\underline{\text{EoL}}\rangle = 806$. They also proposed human-picked features based regression models to predict EoL [11], using data from the first 100 discharge cycles ($\underline{s} = 0, a = 100$), that is, they provided a solution to the $\text{EoL}(a = 100) \equiv 0 + 100 + \text{RUL}(\underline{s} = 0, a = 100)$ problem. Their relative mean absolute percentage error (MAPE) for EoL was 9.1% for the testing dataset, i.e. RMSE for EoL was about 167 cycles using $a = 100$. The human-picked features consist of the minimum, variance, skewness and kurtosis of the first-100 cycle discharge capacities $\{Q(m = 1\cdots100)\}$, the slope and intercept of the linear fit to the capacity fade curve $Q(m)$ between cycles 2 to 100, the capacity at cycle 2, the difference between the maximum discharge capacity (since many types of batteries have activation) and cycle 2, the average charge time in the first 5 cycles, the integral of temperature over time in cycles 2 to 100, the minimum internal electrical resistance from cycles 2 to 100, and the difference in the internal electrical resistance between cycle 100 and cycle 2. In 2020, Attia et al. [12] utilized the same regression model to optimize closed-loop fast-charging protocols and produced more LIB cells data with even more complex fast-charging policies. Knobloch [13] developed a Convolutional Neural Network (CNN) model, applied it to the same dataset [10] and successfully reduced the required input data amount of consecutive charging cycles down to $a = 20$ while maintaining the accuracy. Tian et al. [14] proposed a DNN to predict complete charging curves by 30 points in given partial charging curves measured within 10 mins. By using the transfer learning technique, pre-trained DNNs can also be adapted to various practical scenarios, such as different ageing states, battery types, and charging strategies [15]. Yang et al. [16] proposed an extreme learning machine-based thermal model to capture the temperature behavior of batteries under external short circuit conditions, achieving better computational efficiency and accuracy than the conventional models. Xiong et al. [17] proposed a multi-stage model fusion algorithm and a proportional–integral–differential observer, which can successfully improve the accuracy of the prediction of state of charge (SOC) and capacity under a complex application environment. Hong et al. [18] adopted the dilated CNN architecture [19] to predict RUL with RMSE of 76 based on 4 cycles ($a = 4$) of acquisition as the input data. Most of the training approaches in the literature are end-to-end, i.e. a set of specific feature correlate to a set of the target value. Thus, in the case of battery EoL prediction, it becomes difficult when only a single cycle is considered. Different from the end-to-end approach, "multi-ends-to-end" is used in the current work, i.e. multiple sets of features correlate to the same set of the target value, by the proposed last padding technique. This will be detailed in the next section.

In the present work, we show that (1) just $a = 1$ is able to provide $\text{RMSE}(\text{RUL}, \underline{\text{RUL}}) < 50$, significantly better than all previous models. Furthermore, the present age s can also be predicted, with $\text{RMSE}(s, \underline{s}) < 35$. (2) The cycle-by-cycle discharge voltage curve $V(Q)$ can be predicted. This offers a great advantage for battery reuse, battery-pack modeling and EV power-system design, as connecting heterogeneous cells are the key and knowing the future $V(Q)_n$ for each cell is essential. Furthermore, the user can check the difference between cycle-by-cycle $V(Q)$ and the predicted $V(Q)$ in the future, offering confidence that the DNN is indeed working or "early-warning system" if it is not, instead of waiting till the EoL event. (3) Our DNN has identified subtle, low-dimensional battery SOH features (LDSOH) that are much sparser and superior than the human-picked features. These low-dimensional features ("latent space" in DNN) may in turn offer continuous gradient-guidelines on future battery manufacturing and usage. That is to say, since the present $\text{RUL/EoL}_{\text{DNN}}$ takes in few-cycle data to make the EoL prediction, we can visualize and "invert" our EoL_{DNN} and RUL_{DNN} efficiently by focusing on LDSOH space (e.g. 2D maps and 3D contours) and ask what the few-cycle data *should* look like in order to have better EoL_{DNN} outcome. This would allow us to ask provocative questions like, given the limited 124 cell samples of EoL ranging from 150 to 2,300 cycles, is it possible to have $LiFePO_4$/graphite cell with the same nameplate weight and performance rating but EoL of 3,000 cycles or even 5,000 cycles under *equally* challenging fast-charging conditions, by tweaking the fast-charging policy or even battery design? For the same battery chemistry, there are still literally hundreds of parameters in battery manufacturing, for example the electrode, separator and current collector thicknesses, calendering pressure, electrolyte, carbon black and binder amounts, the synthesis (e.g. ball-milling, carbon coating and doping of $LiFePO_4$) conditions, electrolyte additives, prelithiation, formation policy, etc. So, is it possible to numerically guide the optimization of these processing and use conditions for long EoL_{DNN} and RUL_{DNN} using cheap-to-calculate and easy-to-visualize LDSOH, instead of the expensive-to-measure EoL? Lastly, is our $\text{EoL/RUL}_{\text{DNN}}$ "explainable"? [20] (4) The architecture and training of our DNN is very flexible and generic, and should be applicable to all industrially manufactured cycling devices like washing machines, supercapacitors and material fatigue experiments. The training outcome is highly self-consistent. Indeed, examination of our RUL_{DNN} has indicated likely slight testing condition anomaly in the Severson et al. [10] dataset between the 55th to 70th testing cycles (before we noticed temperature fluctuations of the environmental chamber in their text), when presumably the testing was at its early/warm-up stage. This kind of automatic detection of the data anomaly for DNN training would allow a robust testing-and-prediction regime to be established for future cycling devices. Note that Table S1 listed n, s, m, a and all the terms used in the manuscript and the corresponding definition is added in supplementary. The illustration of n, s, m, and a are also shown in Fig. S1.

The present work aims to develop DNNs to enable higher accuracy, earlier prediction, greater interpretability and more output features to a wide range of cycling conditions [10]. Three DNNs are trained based on Severson et al.'s 2019 data of 124 $LiFePO_4$/graphite cells [10]. Each of

which has different main objectives, auxiliary learning tasks [21], testing strategy and purpose, summarized and listed in Table 1. The first DNN, referred to as Discharge DNN, allow the prediction of EoL, future charge time at EoL $t_{n=\mathrm{EoL}}^{\mathrm{charge}}$, and future voltage vs. discharge capacity $V(Q)_{n=\mathrm{EoL}}$ curve at EoL for unseen batteries, based on cycle-by-cycle information ($m = 1\cdots a$) including human-picked features F^{human} and the "continuous-t" data of discharge curves without the information about charging. The second DNN referred to as Full DNN, includes information about charging in the training data, and adopted a testing strategy to enable extrapolation for unknown batteries with unknown charging policies. The generality of the second DNN was significantly improved and can be applied to accurately predict EoL, cycle-by-cycle ($n = 1\cdots\mathrm{EoL}$) $V(Q)_n$ of different batteries, even for the other dataset provided by Attia et al. in 2020 [12]. The first two DNNs can take arbitrary a from 1 to $a_{\max} = 100$, but trained to assume $\underline{s} = 0$ and were not tasked to predict s. The third DNN, referred to as Full RUL DNN, predicted RUL, s ($0 \le s$), cycle-by-cycle ($n = 1\cdots\mathrm{EoL}$) $V(Q)_n$, discharge capacity Q_n, and discharge power D_n curves for unknown age of "used" batteries. Note that Full RUL DNN went through a much larger quantity of training data since we take arbitrary starting points in the data. The proposed DNNs can accurately predict EoL/RUL and time-series properties based on an extremely small amount (even one cycle only) of input data, which can start from any cycle in the entire life of a battery. This is achieved due to the special convolutional testing and training strategy and the last padding technique adopted in the current work, giving a great flexibility of the DNNs, and in particular, the ability of detecting data anomaly. The novel deep neural networks provided highly accurate health prognostics with no prior professional knowledge of cell chemistry and the assumption of degradation mechanisms, showing remarkable predictive ability, interpretability and generality.

2. Model overview

The workflow of Discharge DNN, Full DNN, and Full RUL DNN is illustrated in Fig. 1. Note that detailed elements/building blocks/inputs/outputs etc. of each NN pipeline of all the DNNs have been illustrated in Fig. S3 to Fig. S13 in Supplementary. Here, $\underline{s}$ is defined as the number of shifted cycles from cycle number $n = 1$ used as the first cycle of the input data series, and each consisting of a consecutive cycles the testing/training strategy adopted by the DNN, which will be explained in more detail in the next section. We firstly consider the case of $\underline{s} = 0$. The voltage $V(t)$, current $I(t)$, capacity $Q(t) = \int_{t_i}^{t_f} I(\tau)\mathrm{d}\tau$, and can temperature $T(t)$ of the a consecutive cycles during the discharge process provided by Severson et al. [10] were used as input data, where $[t_i, t_f]$ are the start and finish times of the discharge half-cycle. We always discretize these continuous-t data into a vector of dimension 500 by linear interpolation, with time-step $h = (t_f - t_i)/500$, so a $4 \times 500 \times a$ "continuous-t" data block $x_{itm}^{\mathrm{dis}} = \left[V(t)_m^{\mathrm{dis}}, I(t)_m^{\mathrm{dis}}, Q(t)_m^{\mathrm{dis}}, T(t)_m^{\mathrm{dis}}\right]$ for each discharge half-cycle, where superscript " dis " indicates the corresponding data obtained from discharging process only; index i represents each component of the data/feature matrices; index m is defined to describe each cycle used as input and in range of $m = 1\cdots a$. Then, x_{itm}^{dis} was processed by two CNNs, referred to as Dimension reduction 1 and 2, consisting of series of Inception blocks [22], residual bottleneck [23], and Luong-style attention layers [24] followed by global average and maximum pooling layers, i.e. network in network (NiN) [25] architecture. This type of model performed down-sampling by concatenating pairs of adjacent projection inputs [26]. Dimension reduction 1 and 2 gave the first estimation of EoL and charge time, denoted as $\widetilde{\mathrm{EoL}}_m^{\mathrm{dis}}$ and $\left(\tilde{t}_{n=\mathrm{EoL}}^{\mathrm{charge}}\right)_m^{\mathrm{dis}}$, based on every single cycle m during the discharge process, respectively. The two resulting $1 \times 1 \times a$ vectors from Dimension reduction 1 and 2 are the lower-fidelity version of the targets, which have great sensitivity to reflect on the variation of their target of inference. They served as "data-driven features" denoted by $F_{im}^{\mathrm{data,dis}}$ for the networks.

In addition to the "continuous-t" data x_{itm}^{dis} which were all for discharge, we also used six human-picked cycle-by-cycle features ($6 \times 1 \times a$) denoted by F_{im}^{human} provided by Severson et al. [10] that include both charge and discharge information, for Discharge and Full DNNs. These "discrete-m" data include (1) the Charge capacity P_m, (2) Discharge capacity Q_m, (3) Temperature average T_m^{avg}, (4) Temperature minimum $T_m^{\min}$, (5) Temperature maximum $T_m^{\max}$, and (6) Total charge time t_m^{charge} of each charge half-cycle m. Note that, based on our trial runs, the data of internal resistance in the dataset [10] showed the least relevance in deep Taylor decomposition (DTD) [27] analysis compared to the other human-picked features, which is somewhat counter-intuitive. Subsequently, the internal resistance feature was not used in any of our DNNs, to reduce the amount of input and save the cost of the corresponding measurements. These $6 \times 1 \times a$ "discrete-m" features F_{im}^{human} are next concatenated with $2 \times 1 \times a$ data-driven features $F_{im}^{\mathrm{data,dis}} = \left[\widetilde{\mathrm{EoL}}_m^{\mathrm{dis}}, \left(\tilde{t}_{n=\mathrm{EoL}}^{\mathrm{charge}}\right)_m^{\mathrm{dis}}\right]$ described in the previous paragraph, which are also "discrete-m" features, labeled as Discharge features (7) and (8), respectively. Our DNN is flexible with the choice of a, but we deem the practically economic maximum to be $a_{\max} = 100$, which was what Severson et al. [10] used. An $8 \times 1 \times a_{\max}$ data block F_{ik}^{dis} can be obtained from the $8 \times 1 \times a$ data block, by using the last-padding technique, where the last column of the $8 \times 1 \times a$ "discrete-m" data block was copied to fully fill the fixed-width $8 \times 1 \times 100$ data block. This is the key approach to enable a great flexibility of a, and to let the same value of targets (such as EoL) being described by different lengths of input data. Then the input with more cycle information can contribute more adequate gradients for modifying the weights of the DNN and guide the training based on the input with less information, leading to an efficient learning for the DNN. The illustration is shown in Fig. S1 and the effect of this approach is detailed in Fig. S29 and 30 in supplementary.

Next the $8 \times 1 \times 100$ data block was then fed into two CNN subnetworks, referred to as Predictors 1 and 2. Predictor 1 has a structure similar to Dimension reduction with a different set of hyperparameters, and is to predict EoL and $t_{n=\mathrm{EoL}}^{\mathrm{charge}}$, where $t_{n=\mathrm{EoL}}^{\mathrm{charge}}$ prediction was an auxiliary learning task, which assisted the prediction of EoL as a hint. Predictor 2 was used to predict voltage vs. discharge capacity curves at EoL, $V(Q)_{n=\mathrm{EoL}}$, based on the input of $8 \times 1 \times 100$ sequence matrix F_{ik}^{dis}. It adopted an attention-based encoder-decoder structure to enable this sequence-to-sequence prediction.

Full DNN further improved the generality for predicting EoL and $V(Q_j)_n$ curves of cells subjected to different charging policies. This was achieved by taking both charge and discharge curves into account and adopting a new design of Predictor. Similarly, Dimension reduction 3

Table 1
Summary tables for three DNNs in the current work.

Name of DNN	Main objectives	Auxiliary learning	Testing strategy	Purposes of DNN
Discharge DNN	EoL, $V(Q)_{n=\mathrm{EoL}}$	$t_{n=\mathrm{EoL}}^{\mathrm{charge}}$	<1> Unselected random batteries, only trained with 100 cycles.	Predicting for unknown batteries.
Full DNN	EoL, $V(Q)_n$		<2> Unselected random charging policies, only trained with 100 cycles.	Predicting for unknown charging policies.
Full RUL DNN	RUL, Q_n, D_n, $V(Q)_n$	s	<3> Unselected random batteries, trained with 1… EoL-100 cycles.	Predicting for unknown age of unknown used batteries.

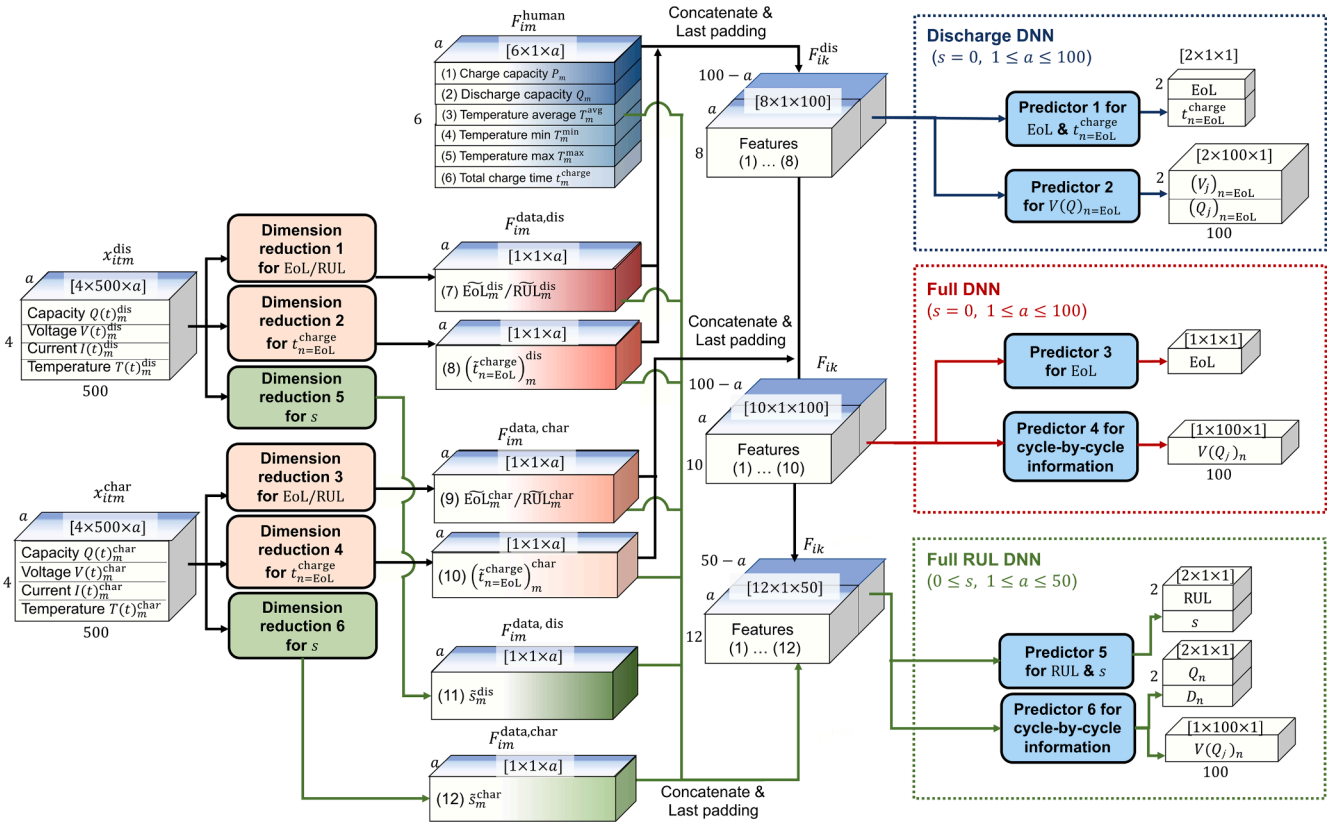


Fig. 1. The workflow of the three DNNs in the current work. Where cubes represent data in tensor form and rounded rectangles represent neural networks. The detailed elements/building blocks/inputs/outputs etc. of each NN pipeline of all the DNNs have been illustrated in Fig. S3 to Fig. S13 in Supplementary.

and 4 were introduced to extract data-driven features $F_{im}^{\text{data,char}} = \left[\widetilde{\text{EoL}}_m^{\text{char}}, \left(\tilde{t}_{n=\text{EoL}}^{\text{charge}}\right)_m^{\text{char}}\right]$, which are labeled as Charge features (9) and (10) respectively, where " char " indicates the data obtained from charging process to avoid the confusion made by the symbols " charge " used by total charge time $t_{n=\text{EoL}}^{\text{charge}}$. The resulting two $1 \times 1 \times a$ vectors were concatenated with the $8 \times 1 \times a$ data block mentioned in the previous paragraphs. Last padding technique was also applied, forming $10 \times 1 \times 100$ data block F_{ik}. Next, Predictor 3 consisted of three CNN sub-networks with deep mutual learning (DML) strategy [28], and the outputted results were ensemble averaged to give the predicted EoL. In which the three sub-networks had different amounts of trainable parameters to preserve the diversity of learning. Predictor 4 provided cycle-by-cycle $V(Q_j)_n$ curves, where n can be specified by users as an integer percentage of EoL, and served as a key for the encoder part of the attention layer. It is noteworthy that Predictor 4 outputted a 1×100 vector of voltage corresponding to pre-defined Q_j ($j = 1 \cdots 100$) with equally-spaced interval of $Q_{\text{nameplate}}/100$. This design reduced the complexity of the prediction of curves enabling a more efficient training/inference of model, and yet, the detail of steep changes of curves may be compromised.

Full RUL DNN has almost no limitation for positive s, i.e. the input sequence does not necessarily begin from the cells first cycle. However, $1 \le a \le 50$ was set to reduce the extremely large amount of data in training process due to the consideration of the wide range of s and a. Therefore, instead of EoL, RUL can be predicted. This also indicates that the networks can adapt and respond correctly based on the features in any cycle as its first input. Full RUL DNN aims at truly well used battery cells, that may have $\underline{s} > \underline{\text{EoL}}/2$ already. Full RUL DNN required two additional Dimension reduction 5 and 6, extracting the data-driven features $\tilde{s}_m^{\text{dis}}$ and $\tilde{s}_m^{\text{char}}$, labeled as (11) and (12), respectively. Thus, in total 12 features of a cycles started from the unknown s^{th} cycle can be obtained.

In addition to RUL, cycle-by-cycle Q_n, D_n, and $V(Q_j)_n$, the value of s was also predicted by Full RUL DNN as an auxiliary learning task. Thus, Full RUL DNN can output the predicted current age s of any well-used batteries, which can be a validation check for any user who doubt the efficacy of the DNN ("ask the Fortune Teller my age, before trusting her prediction of the Doomsday"). A special "convolutional training" strategy was applied to both Full RUL DNN to deal with such complicated tasks where both s and a are varying with little constraint.

2.1. Network setting and training

Our DNNs used the unmodified "raw data" provided by Severson et al. [10] Certain data were, however, omitted due to their unsatisfactory quality specified in the exception notices in the experiment log, such as early stoppage of the tests, temperature recording failure, or data with excess noises. In total, 115 batteries with an acceptable quality of discharge data in the first 100 cycles were considered in our Discharge DNN; out of these, 95 batteries were considered in Full DNN as the charge curves data quality of the omitted 20 batteries were not of satisfaction. Meanwhile, only 81 batteries have acceptable data quality in all the discharge and charge curves in every $0 \le s, m = 1 \cdots a$ cycle, and thus can be considered in Full RUL DNN. Here the constant-current regions of the discharge curves, and the entire curves of charging process of capacity, voltage, current, and temperature were utilized as input datasets.

Three testing strategies were adopted corresponding to the three DNNs proposed in our work. Discharge DNN was designed for the prediction of unlearned battery cells. Thus, its testing strategy, referred to as testing strategy $<1>$, was that the data of entire sequences of certain randomly chosen batteries was omitted from all the considered

batteries. The resulting datasets then trained and forced Discharge DNN to learn the features of overall sequences and to extrapolate for different batteries. Next, Full DNN was designed for batteries with varying charging policies. Testing strategy $< 2>$, where the data of entire sequences of all the batteries with randomly chosen charging policies were deliberately omitted, were used by Full DNN to train the ability of extrapolation for different charging policies. The training of the two DNNs mentioned above was trained by the first 100 cycles ($s = 0$, $1 \le a \le 100$) of each battery only, i.e. the information of the remaining cycles ($a > 100$) was not considered. In contrast, Full RUL DNN adopted a special "convolutional training/testing" (CT) strategy $< 3>$, where the starting cycle with s shifted cycle from $n = 1$ and the number of cycles used a in each set of training data may vary. The kernels with all the possible combinations of s and a then convoluted throughout the entire life cycles of 65 randomly selected batteries (around 80% of the 81 considered batteries) in the training set. This allows Full RUL DNN to possess the ability of multi-vision to obtain full information about the entire life of batteries. Thus, accurately predicting RUL based on an arbitrary s with little information of a cycles can be achieved. The detail of the training strategies $< 1>\ldots<3>$ was illustrated in Fig. S14 in Supplementary.

2.2. Performance of discharge and Full DNNs

The performance of Discharge and Full DNNs is shown in Fig. 2. There are 115 cases for Discharge DNN and 95 cases for Full DNN in the figures, i.e. the total number of batteries considered in the DNNs. The mean absolute percentage error (MAPE) of the predicted values to ground truth of EoL corresponding to the training and testing sets were shown in the legend of each diagram in Fig. 2(a). It is remarkable that very low testing MAPE(EoL, EoL$_-$) = 6.46% can be achieved by Discharge DNN based on $F_{ik}^{\text{discharge}}$, i.e. the discharge curves and the six human-picked features, at the first cycle only ($a = 1$). Full DNN can also give low testing MAPE(EoL, EoL$_-$) = 7.78% for the predicted EoL. These MAPE values based on one cycle only are even lower than 9.1% based on 100 cycles in the literature [10]. Our testing MAPE of Full DNN can be further improved to lower than 3.97% when the data of the first 100 consecutive cycles were used ($a = 100$), but we consider such improvement in quality to be disproportional to the 100 × input measurement work required. In addition to such high overall accuracy of prediction, sparse data points and outlier cases with extremely low or high lifetimes can still be captured by both the DNNs.

Discharge DNN can predict $V(Q)_{n=\text{EoL}}$ curve. Two batteries with the least and the greatest differences on curves at $n = 1$ and EoL, i.e. the best

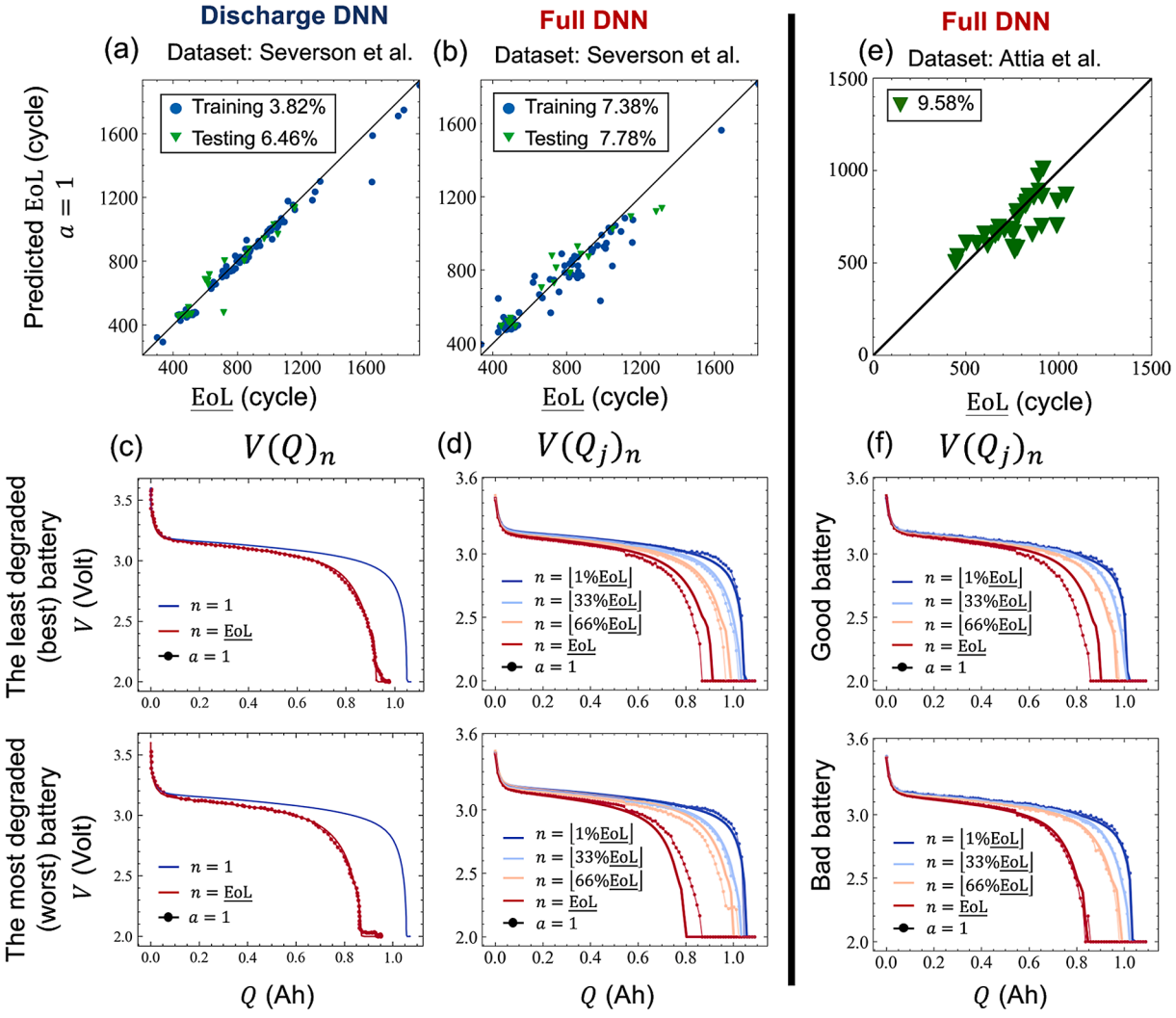


Fig. 2. (a)(b) The predicted values EoL vs. EoL of Discharge and Full DNNs based on the first cycle ($s = 0, a = 1$) in data of Severson et al. Low MAPE (<7.78%) were achieved. (c)(d) The predicted $V(Q)_n$ curves by the two DNNs for the cases of the lease degraded (best) and the most degraded (worst) batteries. In the second part of the figure, Full DNN is applied to predict (e) EoL of 34 cells provided by Attia et al. [12] with totally different charging policies, and (f) $V(Q_j)_n$ curves of two of batteries in (e) are also predicted.

and worst batteries in the dataset, are shown in Fig. 2(c) as examples. Here the difference in $V(Q)_{n=\mathrm{EoL}}$ was defined by the integrated area between the two curves at $n = 1$ (blue solid line) and EoL (red solid line), and can be regarded as an indicator of the level of ageing. The dotted line was the degraded curve predicted based on the first ($a = 1$) cycle, while the red solid line was the ground truth of curves at EoL. It can be observed that the dotted and red solid lines are almost identical, showing a great predictive power of our DNN in both long-lived and short-lived outliers.

We next examine the performance of Full DNN on the prediction of cycle-by-cycle $V(Q_j)_n$ curves, as shown in Fig. 2(d). Where solid lines in blue, light blue, orange, and red representing the ground truth of curves of n = 1%⌊EoL⌋., 33%⌊EoL⌋., 66%⌊EoL⌋., ⌊EoL⌋are chosen for illustration purposes, where $\lfloor x \rfloor$ is floor operation giving the greatest integer less than x. It can be observed that Full DNN accurately predicted these curves and successfully captured the history of the ageing process of cells. Note that our DNN provided a more accurate prediction on the curves for the early-stage usage of cells, since cells in their early stage typically are at a relatively healthy state and with less uncertainty.

The error metrics of properties predicted by Discharge and Full DNNs were shown in Table 2. It is remarkable that, with the average <EoL>=806 cycles and a standard deviation of 377 in the batteries, RMSE of 57 cycles for EoL prediction can be achieved by Discharge DNN based on the first-cycle data only ($s = 0, a = 1$). It is significantly less than the benchmark of RMSE of 122 obtained from Severson et al. based on the first 100 consecutive cycles ($s = 0$, $a = 100$). RMSE was dramatically dropped to 33 when data of the first 100 consecutive cycles were considered in Discharge DNN. In Full DNN, RMSE of 58 and 39 cycles for EoL prediction can be achieved based on data of the first ($a = 1$) and the first 100 cycles ($a = 100$). The ability of extrapolation of Full DNN was excellent and illustrated in Fig. S15 in Supplementary. It is of interest that Full DNN and Discharge DNN produced very similar RMSE. This shows that the auxiliary learning task for predicting future charge time $t^{\mathrm{charge}}_{n=\mathrm{EoL}}$ in Discharge DNN was very helpful in predicting EoL. The effect of this design was almost equivalent to the input data of the entire charge part for Full DNN.

The error metric $\mathscr{M}_n$ for $V(Q)_n$ curves predicted by Discharge DNNs was also shown in Table 2. It is defined by

$$\mathscr{M}_n = \frac{\sqrt{\frac{1}{100}\sum_{j=1}^{100}\left(\left(\Delta V_{jn}/\left(V^{\mathrm{upper}} - V^{\mathrm{lower}}\right)\right)^2 + \left(\Delta Q_{jn}/Q_{\mathrm{nameplate}}\right)^2\right)}}{\int |V'(q)_n| dq}$$

Table 2
Comparison of error metric of properties predicted by the current DNNs and Severson et al. [10] The superscript <·> indicates the specific training strategy adopted by the DNN.

Situation	First cycle $s = 0$ $a = 1$	First 5 cycles $s = 0$ $a = 5$	First 100 cycles s = 0 a = 100	First 100 cycles $s = 0$ $a = 100$
	Discharge DNN			**Discharge Model in [10]**
RMSE(EoL, EoL_) (cycles)	44/57	38/49	15/33	76/122
$RMSE\left(t^{\mathrm{charge}}_{n=\mathrm{EoL}}, \underline{t^{\mathrm{charge}}_{n=\mathrm{EoL}}}\right)$ (mins)	0.37/ 0.6	0.38/ 0.53	0.17/0.45	–
$\mathscr{M}_{n=\mathrm{EoL}}$	32/43	23/36	17/16	–
	Full DNN			**Full Model in [10]**
RMSE(EoL, EoL_) (cycles)	50/58	42/48	35/39	51/167
$\langle \mathscr{M}_n \rangle$	15/26	12/20	10/22	–

where ΔV_{jn} and ΔQ_{jn} are the difference between the j^{th} data point of ground truth and predicted values of voltage and charge for the given cycle n; $[V^{\mathrm{lower}}, V^{\mathrm{upper}}]$ is the discharging and charging voltage cutoffs; $Q_{\mathrm{nameplate}}$ = 1.1 Ah, i.e. the nameplate discharge capacity of these 18,650 $LiFePO_4$/graphite cells; the denominator is the total length of curve. For the prediction of cycle-by-cycle $V(Q_j)_n$ curves by Full DNN, the error metric, $\langle \mathscr{M}_n \rangle$, is determined by averaging $\mathscr{M}_{n=1\%\mathrm{EOL}}$ through $\mathscr{M}_{n=\mathrm{EOL}}$ for all the 100 $V(Q_j)_n$ curves. It is remarkable that even though the complexity of the prediction of cycle-by-cycle curves is significantly higher than it of the prediction of a single-cycle curve at EoL, their error metric values listed in Table 2 are very similar to those of Discharge DNN.

Overall, the error metrics difference between testing and training sets in all items of Table 2 was converged when the first 100 consecutive cycles were considered. This indicated that the feature learning process of the three DNNs were all in good condition and without obvious overfitting phenomenon. The learning progress of the current DNNs was illustrated in detail in Supplementary Fig. S16…18. The reproducibility and reliability of the Discharge and Full DNNs were examined by five-fold cross-validation tests. Five different training sets were used to train the two DNNs. The RMSE of EoL curves show slight fluctuations in the early stage and quickly converge after input cycle $a > 30$ showing high robustness, as shown in Fig. S19 in Supplementary.

With the advantage of highly accurate prediction, our DNNs are particularly suitable for closed-loop optimization for maximizing battery cycle life recently proposed by Attia et al. [12] As RMSE of our $\mathrm{EoL}_{\mathrm{DNN}}$ ($a = 1$) was even less than the benchmark which used $a = 100$ cycles [10], as shown in Table 2, the required measuring number of cycles and optimization time in the protocols can be significantly reduced. Moreover, apart from accurately predicting EoL, our Discharge DNN can predict additional properties with very high accuracy, such as the total charge time. This also enables a useful extension for the protocols to optimize EoL and charge time simultaneously to achieve different objectives defined by the users who may favor the maximum EoL, the shortest charging time, or any preferable combination of the two.

2.3. Application to batteries with different charging policies

The generality of the current work was examined by applying Full DNN based on the year-2019 dataset to predict EoL of batteries included in a completely different dataset which is provided by Attia et al. [12] in 2020, without retraining. The batteries were subjected to very complex charging policies with 5 different constant-current steps. There were 34 batteries chosen from the 2020 dataset [12] to test our Full DNN. Note that 11 batteries in the dataset were omitted since the data condition of their first cycle does not have acceptable quality for the inference of EoL. For example, their current curves are greatly shifted; the patterns of their temperature curves are different from most of the batteries in the dataset. The corresponding data screening process is detailed in the supplementary. Fig. 2(e) shows that Full DNN has MAPE(EoL, EoL_) = 9.58% based on the first cycle only ($a = 1$), which is slightly higher than those when Full DNN predicted EoL in the testing dataset provided by Severson et al. [10] We believe this is because the significant difference in the discharge capacity of batteries used in the Attia et al. [12] (1.051 ± 0.003 Ah) and Severson et al. [10] (1.078 ± 0.002 Ah). Even though, MAPE of 9.58% can still be regarded as sufficiently low for the data and charging policies which were not learned by the DNN. More detail can be found in Fig. S21 and S22 in supplementary.

We next examined the performance of Full DNN in predicting time-series properties $V(Q_j)_n$ of the good and bad batteries. Note that the curves predicted by Full DNN required two automatic pre-processing steps for illustration purposes, as the dataset provided by Attia et al. [12] was not known to the DNN. First, a rigid shift should be performed to make the first point of the predicted curve matches that of the initial

curve of the first cycle (i.e. the voltage at zero discharge capacity). After the rigid shift, the second process was to omit the points in the predicted curve with voltage lower than the minimum applied voltage specified in the dataset (2.0 V in this case). This data pre-processing is completely general and tamper-proof, as all the information required here was included in the input data of the DNN. Fig. 2(f) shows that Full DNN successfully predicted the curves with high accuracy for both cases. It is of interest that, for the case of the bad battery, the ground truth (solid lines) and predicted (dotted lines) curves were almost identical, while there was a notable difference for $n = \underline{\text{EoL}}$ in the case of the good battery. This is because the degradation was insignificant in the good battery, and fewer aging features can be discovered by the DNN in the first cycle. Similar conclusions were made in the work of Zheng [29], where the incremental open circuit voltage test for better estimation of SOC are recommended.

2.4. The most influential features for predicting EoL

The network design enables the interpretation of the current DNNs. The total counts for the human-picked and data-driven features on EoL being the most influential feature in each input cycle in all batteries in the two DNNs can be revealed by Deep Taylor decomposition (DTD), as shown in Fig. 3. There are 115 × 100 and 95 × 100 counts in total in Fig. 3(a) and (b), respectively. In Discharge DNN (Fig. 3(a)), the top 3 influential features were data-driven features (7) $\widetilde{\text{EoL}}_m^{\text{dis}}$, (8) $\left(\tilde{t}_{n=\text{EoL}}^{\text{charge}}\right)_m^{\text{dis}}$, and human-picked feature (5) Temperature maximum $T_m^{\max}$. In Full DNN, the top 3 influential features were data-driven features (10) $\left(\tilde{t}_{n=\text{EoL}}^{\text{charge}}\right)_m^{\text{char}}$, (8) $\left(\tilde{t}_{n=\text{EoL}}^{\text{charge}}\right)_m^{\text{dis}}$, and human-picked feature (5) Temperature maximum $T_m^{\max}$. In contrast, the feature with relatively low relevance in both the DNNs was human-picked feature (1) Charge capacity P_m. Remarkably, the results showed that the most influential features to EoL of a battery are all data-driven features extracted by our Dimension reduction, rather than the human-picked features suggested by the literature, by as much as one order of magnitude. This demonstrated the power of DNNs which is capable of discovering the crucial underlying factors for the performance over and above human intuitions.

We next visualized the most influential feature for each of 95 batteries with different EoL in Full DNN, as shown in Fig. 3(c). It can be observed that data-driven features are more relevant than other features to the batteries with long-range of EoL, showing their great applicability to batteries with different qualities. Furthermore, it is of interest to note that Charge data-driven features (9) $\widetilde{\text{EoL}}_m^{\text{char}}$ and (10) $\left(\tilde{t}_{n=\text{EoL}}^{\text{charge}}\right)_m^{\text{char}}$ are more relevant to the batteries with EoL below the average value of 806 cycles (marked by a dashed line) while Discharge data-driven features (8) $\left(\tilde{t}_{n=\text{EoL}}^{\text{charge}}\right)_m^{\text{dis}}$ are more relevant to those with longer EoL. This result implies that charging policies are the main reason to cause aging and degradation of batteries. It is also worth noting that, batteries with EoL around the average value typically behaved neutrally without significant characteristics. Thus, it is reasonable that both human-picked and data-driven features were possibly the most influential ones. In our Supplementary, more correlations between EoL and input data were determined by layer-wise relevance propagation (LRP) method in Fig. S23, and selected parts of input data suggested by LRP were also visualized and enlarged in Fig. S24.

2.5. The prediction of RUL and the current battery age s

Now we examine the performance of RUL prediction by Full RUL DNN, where $0 \le s, 1 \le a \le 50$. Fig. 4(a) and (b) show the testing RMSE (RUL, $\underline{\text{RUL}}$) and RMSE(s, $\underline{s}$), respectively, where s can be regarded as the unknown age of a used battery. Since this DNN allows accept input from wide range of s (1–1684 cycles), three representative ranges corresponding to "Baby" age (1–100 cycles), "Mid" age (401–500 cycles), and "Old" age (701–800 cycles) were chosen for illustration purpose. RMSE diagrams for RUL and s in the full range are shown in Fig. S25 in Supplementary. It can be observed that RUL was accurately predicted by Full RUL DNN. Most of the regions in Fig. 4(a) are in blue, i.e. RMSE (RUL, $\underline{\text{RUL}}$) < 50.

It is of interest that, in Fig. 4, the regions with relatively high RMSE form 45° sharp lines, i.e. $a + s\_ = c$, where c is a constant. This is because the last padding technique used here repeated the information of the last input cycle, which was equivalent to assign a greater weight to that cycle in the DNNs. In other words, whenever the data of a single cycle with low quality of measurements was the last element in the input data, the prediction was then greatly misled by that particular cycle, resulting in a sudden increase of RMSE. Similarly, RMSE may also drop immediately when the last input element was with an acceptable quality. For example, region A in Fig. 4(a) with relatively high RMSE is a 45° band located around $s\_ = 55$. It is because an anomaly condition (temperature fluctuations of the environmental chamber according to Severson et al. [10]) occurred around those cycles for half of the batteries in the dataset. This indicates that the current DNN successfully reflects the anomaly conditions that occurred in the testing of given batteries. In addition, even the input was with such anomaly conditions, RMSE of the current DNN is still lower than 50 cycles, and it can be further reduced by simply considering more input cycles (greater a). This great flexibility, reliability and robustness are not offered by any other DNN models in the literature which only accept fix amount of input cycles. Another region with relatively high RMSE is $s\_ \le 15, a \le 15$ marked by triangle B. It is reasonable as the predictions were based on small amount of data and batteries were typically at a relatively healthy state with almost no aging features in their early/warm-up stage. It is worth noting that the first column ($s\_ = 0$) in Fig. 4(a) can be compared with the results of Full DNN shown in Table 2. Where their RMSE were both in a similar range between 40 and 80 cycles, even the prediction of RUL is more complex than that of EoL as $\underline{s}$ is unknown and

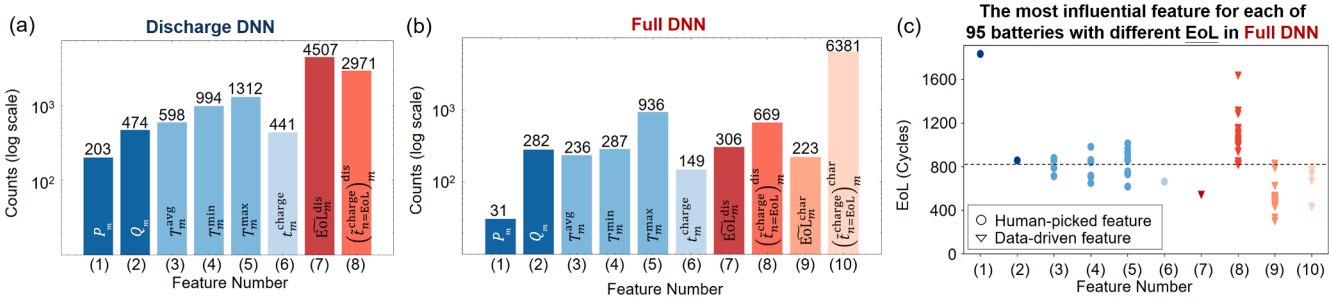


Fig. 3. The total counts for the human-picked and data-driven features on EoL prediction by (a) Discharge and (b) Full DNNs being the most influential feature in each input cycle in all batteries, obtained by Deep Taylor decomposition. (c) The most influential feature for each of the 95 batteries with different EoL in Full DNN.

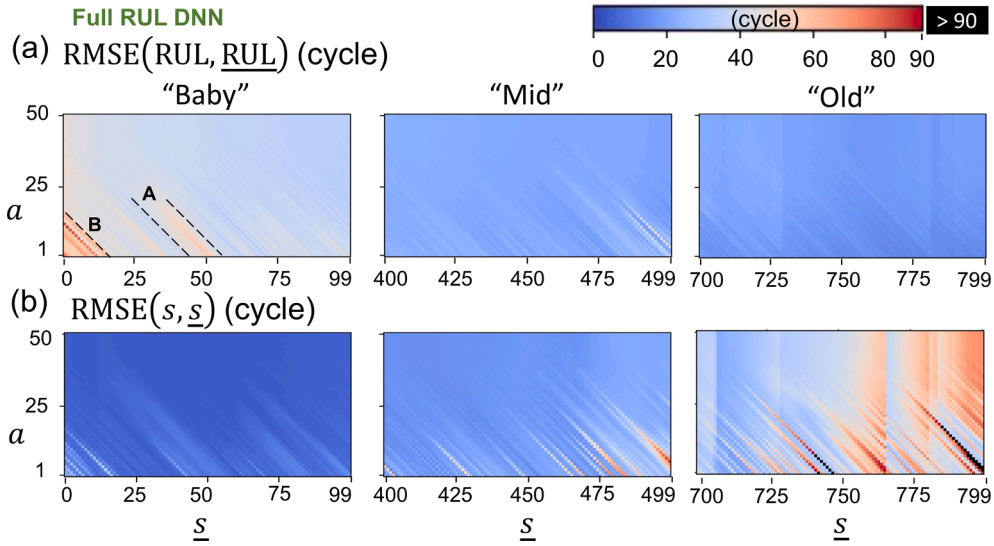


Fig. 4. Testing RMSE of predicted (a) RUL and (b) s, which can be regarded as the current battery age, by Full RUL DNN.

flexible in Full RUL DNN.

Fig. 4(b) shows the performance of age prediction s by Full RUL DNN. It can be observed that the occurrence of 45° sharp edges with relatively high RMSE had good agreement with those in Fig. 4(a), showing a high correlation between RUL and s. Full RUL DNN typically provided a more accurate prediction of s when $\underline{s}$ is small, as shown in Fig. 4(b). In most of the cases, low RMSE(s, $\underline{s}$) (<35 cycles) can be achieved when $\underline{s} < 750$. However, RMSE(s, $\underline{s}$) went up as $\underline{s} > 750$ (see Fig. S25 in Supplementary for more detail). This is because the amount of training data corresponding to the batteries with $\underline{\text{EoL}} > 750$ cycles reduced to 17.3% of the entire dataset (see Fig. S26 in Supplementary), providing insufficient information for Full RUL DNN to learn the corresponding features.

We next examine the performance of Full RUL DNN on cycle-by-cycle information. Fig. 5(a)(b) shows testing MAPE of discharge power D_n and discharge capacity Q_n values at the n^{th} cycle with $0 \le s_{\_}, a = 1$. It is remarkable that our MAPE is typically<4% based on the data at any values of $\underline{s}$, showing excellent cycle-by-cycle predictability. However, D_n and Q_n were relatively difficult to be predicted when n was close to EoL as the batteries became unstable with higher uncertainty, giving a thin black band region at the top of the figures. It is also of interest noting that, for both D_n and Q_n, high accuracy can be achieved especially when the values of the predicting target cycle n and the input data age $\underline{s}$ were close, resulting in a dark blue band from bottom left to top right in Fig. 5 (a)(b). These results were also reasonable as the less cycle difference between n and $\underline{s}$, less uncertainty may be introduced in the batteries. Fig. 5(c) is the error metric $\mathscr{M}_n$ for $V(Q_j)_n$ curves, showing very low values (typically $\mathscr{M}_n < 4\%$). However, a thin black band region at the top of the figures still existed for the same reason mentioned above. A relatively higher value of $\mathscr{M}_n$ close to 4% can be found in the left half of Fig. 5(c). This is because the batteries at their early-stage (less $\underline{s}$) were still healthy with less ageing information. In contrast, age left its traces

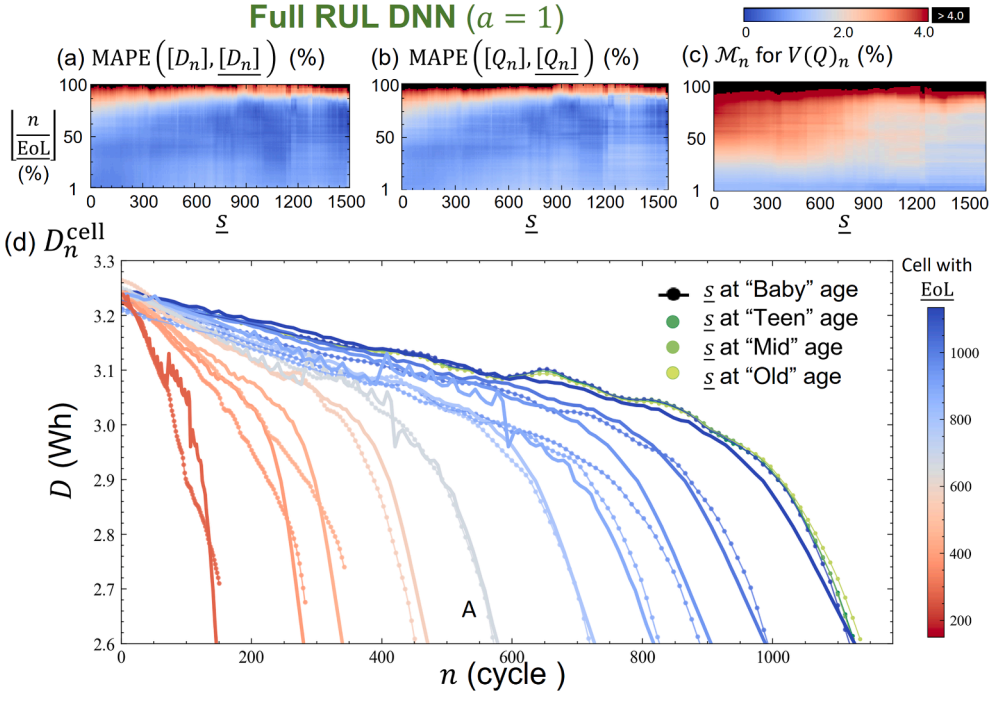


Fig. 5. Testing MAPE of (a) discharge power D_n, (b) discharge capacity Q_n and (c) the error metric $\mathscr{M}_n$ for $V(Q_j)_n$ predicted by Full RUL DNN with $0 \le s_{\_}, a = 1$. (d) The ground truth of discharge power curves D_n^{cell} of ten cells with various $\underline{\text{EoL}}$ provided by Severson et al. [10] and the predicted dotted curves by Full RUL DNN based on one cycle only ($a = 1$) at their "Baby" age ($s_{\_} = 0$). The predicted green dotted curves based on "Teen," "Mid," and "Old" ages, i.e. $\lfloor(\text{EoL}-100)\times 33\%\rfloor$, $\lfloor(\text{EoL}-100)\times 66\%\rfloor$, $\text{EoL}-100$, are also shown for the longest EoL battery.

on the data at the later stage (greater s), providing sufficient information for our DNN to predict $V(Q_j)_n$ curves accurately in a wide range of n.

Fig. 5(d) shows the predicted (dotted lines) and ground truth (solid lines) of discharge power curves D_n^{cell} of ten selected batteries with various EoL provided by Severson et al. [10]. The prediction was performed based on one cycle only ($a = 1$) at the age of "Baby" ($s\_ = 0$). It is remarkable that several important features of the curves of all the ten curves were accurately predicted. For example, the endpoints of most of the predicted curves where $D = 2.6$ Wh matched well with those of ground truth; slope changes of the curves were also accurately predicted; some of the batteries (e.g. battery A marked in Fig. 5(d)) had special fluctuation before the final steepest drop and our DNN can reveal this phenomenon in the predicted curve. We also plot three green dotted predicted D_n^{cell} curves for the batteries with the longest EoL as an example, based on the input cycles of "Teen," "Mid," and "Old" ages (i.e. $\lfloor(\text{EoL} - 100) \times 33\%\rfloor, \lfloor(\text{EoL} - 100) \times 66\%\rfloor, \text{EoL} - 100$), respectively. Note that, the last 100 cycles before EoL were omitted here, as these data were relatively unstable and may have uncommon features of batteries. It can be observed that there was no significant difference between the predicted D_n^{cell} curves based on the four ages of s as input, demonstrating the robustness of the current DNN.

Overall, Full RUL DNN was proven to be a powerful tool to accurately predict RUL, age, and cycle-by-cycle information of used batteries based on input data of few consecutive a cycles started from any cycle s. The cycles under anomaly conditions can also be identified. Most importantly, the results proved that the amount of required data for accurate RUL prediction for any given battery can typically be as small as one cycle ($a = 1$) and including more input cycles ($a > 1$) has limited effect on improving the prediction accuracy. We also show that the data of each cycle of the battery contains sufficient amount of information to be distinguished from each other, provided that the level of precision/resolution of the measurement is sufficiently high in Fig. S27 in the supplementary.

3. Discussion

The proposed DNNs have several advantages compared with the traditional key index measurement methods in practical application. First, a more efficient, convenient, and practicable data storage can be achieved due to the two following reasons. (1) Only a single cycle of data is needed for our DNNs to achieve highly accurate prediction and (2) the stored data of that single cycle can be any cycle before EoL for our Full RUL DNN. This is a significant improvement compared with the conventional machine learning BMSs, which typically required saving the whole sequence of many cycles to predict EoL. Second, take the key index $\Delta Q_{100\text{-}10}(V)$ suggested by Severson et al. [10] as an example. This index shows a very high correlation with EoL of battery, and yet, requires long measurement, i.e. the 10th and 100th cycles. On the other hand, with the help of Dimension Reduction and last padding technique, our DNNs require measurement on any given single cycle and significantly reduce time and cost. Third, the proposed DNNs is suitable to constantly adjust the model to deal with the interference of environmental change and improve prediction accuracy in practical applications. It is because our DNNs are constructed by two parts: Dimension Reduction and Predictor as shown in Fig. 1, offer training flexibility that the users can choose one of the part or both for adjusting the weights.

As for potential applications for lithium-ion batteries, it is possible that our DNNs can be used as the early warning and the safety performance system. For example, since our DNNs can predict the future cycle-by-cycle information (such as $V(Q_j)_n$ and D_n^{cell} curves for the future n-th cycle) based on any previous single cycle of a given battery, these predicted curves can serve as baseline for the future usage of that battery. The users can keep using that battery until the n-th cycle, and compare the actual curves with the predicted base line curves of the n-th cycle. If the actual curves appear worse than predicted baseline curves, it means that the battery has been damaged by unexpected conditions. In this way, the early warning about the health status of that battery can be achieved. Moreover, our DNNs can serve as the predictor in the system of high-throughput battery fabrication, testing, screening and optimization. For example, our DNNs can instantly screen out the just-manufactured batteries with unacceptable predicted EoL; developers can also use our DNNs to predict EoL of batteries subjected to different charging strategies, and thus, charging strategy can be optimized. This approach can also be applied to the optimization of manufacturing parameters.

In this work, we have shown that single cycle data contains a sufficient amount of information to be distinguished from each other and can be used to predict EoL, RUL and other related properties with high accuracy, even for a completely different dataset (Attia et al. 2020 [12]) without retraining. However, we still need to emphasize that the datasets applied in the current work are measured in laboratories. Our DNNs may need more training data to adapt more complex charging strategies, wider engineering/experience of practical aspects, and broader device type/chemistry.

4. Conclusion

This paper first presents two deep neural networks (DNNs) with a novel network design to predict End-of-Life (EoL), charge time t^{charge}, and discharge voltage curve $V(Q)$ curves at EoL of given batteries. To our knowledge, this is the first work to achieve testing root mean square error (RMSE) of EoL prediction as low as 57 and 33 cycles based on inputs of the first cycle ($a = 1$) and the first 100 cycles ($a = 100$) only, respectively (where <EoL> = 806 cycles in the dataset). This was attributed to the introduction of Dimension reduction which extracted data-driven features from the measured profiles of input properties, and the last padding technique. Our results show that these data-driven features played more important roles and were much more influential on the predicted properties in the inference of the DNNs, compared to human-picked features such as the charge/discharge capacity (P_m, Q_m), and the maximum/minimum temperature ($T_m^{\max}$, $T_m^{\min}$). Moreover, our Full DNN was trained to focus on the overall sequences of properties in each cycle with different charging policies, giving very low RMSE of predicted EoL even for batteries with unknown charging polices. The generality of the current DNN was also examined by applying to a completely different dataset, giving the prediction of EoL and cycle-by-cycle information, such as discharge capacity Q_n, discharge power D_n and $V(Q_j)_n$ curves at any given n-th cycle with very high accuracy.

We also proposed a powerful Full RUL DNN which allows the prediction of remaining useful life (RUL) and the current battery age s, and most importantly, allow the flexibility of both s and a. The last padding technique during the training process enables such flexibility, strengthening the connection of the underlying time-series features between sections of data with different length. Also, the performance of the current networks improved as the amount of training data increases, showing a healthy condition of the networks. The anomaly environmental conditions during the tests of batteries can also be detected in the proposed network.

With the help of the proposed special convolutional training strategy throughout the entire life cycles, the DNN enables a more accurate prediction (RMSE(RUL, RUL) < 50 cycles in the most of the range of cycles), a more accurate cycle-by-cycle information prediction (error metrics < 4% for the predicted properties), a much smaller amount of required cycles of input data (typically only 1 cycle is needed), and more flexible length of input data (s and a can be arbitrary) than the state-of-the-art algorithm. Thus, practically, our key contributions, including high accuracy of prediction, high efficiency on saving battery history and excellent ability of extrapolation for unknown charging policies and unknown age of used batteries, make our DNNs suitable to be an

important component of modern quality control systems in battery manufacturing and battery management systems. The current work demonstrated the power of machine learning, which sheds light on solving the long-lasting problems of health estimation and lifetime prediction of lithium-ion batteries, providing great economic and environmental benefits.

4.1. Methods

In the current work, all the data sets were standardized, such that

$$Z \equiv \frac{x-\mu}{\sigma}$$

where Z is standardized value; x is the original value from the raw data; μ and σ represent the mean and standard deviation of the raw data, respectively. These standardized data served as the training and testing data sets for the all DNNs. The performance of the DNNs was evaluated by the root mean square error (RMSE), such that

$$\mathrm{RMSE} \equiv \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2}$$

where n is the total amount of sample; y_i and $\widehat{y}_i$ are the ground truth and predicted values, respectively. All the DNNs achieved relatively lower RMSE than other competitors, showing that key features in the profiles of the properties were successfully captured and correlated with EoL/RUL of batteries. Mean absolute percentage error (MAPE) also served as an error metric for cycle-by-cycle information, e.g. Q_n and D_n, such that

$$\mathrm{MAPE} \equiv \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i-\widehat{y}_i}{y_i}\right|$$

We also used Huber loss [30] during the training process for predicting $V(Q_j)_n$ curves, Q_n and D_n, such that

$$\mathrm{Huberloss} \equiv \begin{cases} \frac{1}{2}(y_i-\widehat{y}_i)^2, & \text{if } |y_i-\widehat{y}_i|\langle\delta \\ \delta|y_i-\widehat{y}_i|-\frac{1}{2}\delta^2 \end{cases}$$

where δ is a hyperparameter which makes the behavior of loss function similar to MSE or MAE.

We also applied Deep Taylor decomposition (DTD) [9] to the DNNs to enable the feature explanation. It redistributed the relevance layer by layer from the predicted results back to the input data. This analysis provided the most influential key feature on target, showing that the data-driven features obtained by Dimension reduction were, in general, more important than human-picked features.

Moreover, in order to obtain the attribution along with the profile of input data, relevance score R_j for the j-th neuron in each layer based on the k-th neuron in the next layer was calculated by using either LRP- $\alpha\beta$ or LRP- ε methods according to the type of the present layer (convolution or dense layers, respectively). These relevance scores were then reweighted by the flat method to have their summation as a unity. The formulation of LRP- $\alpha\beta$ and LRP- ε method can be expressed as

$$R_j = \sum_k \left(\alpha\frac{(a_j\omega_{jk})^+}{\sum_{0,j}(a_j\omega_{jk})^+} - \beta\frac{(a_j\omega_{jk})^-}{\sum_{0,j}(a_j\omega_{jk})^-}\right)R_k$$

$$R_j = \sum_k \left(\frac{a_j\omega_{jk}}{\epsilon+\sum_{0,j}a_j\omega_{jk}}\right)R_k$$

where a and ω are the output and weight of the current layer; the notation $(x)^+$ and $(x)^-$ are with the value of 0 or $\times$ whichever is the greater or the fewer ones, respectively; α, β, and ϵ are coefficients with values of 1, 0, and 0.1. In this way, a relevance score unit vector $\widehat{\mathbf{R}}_{\mathrm{cell}}$ corresponding to the curve $\mathbf{Y}_{\mathrm{cell}}$ can be obtained, where $\mathbf{Y}_{\mathrm{cell}}$ can be the curves of voltage, current, capacity, and temperature of a given cell. The current DNNs can then be back-propagated layer by layer, turning a black-box into an explainable model.

All training process was based on python 3.7 and implemented in Keras (v2.3.1) with Tensorflow GPU backend (v2.1.0). Bayesian optimization with tree parzen estimator [31] technique has applied to all DNNs to optimize the hyper-parameters, such as the number of filters, kernel size, pooling type, and the suitable weight for the regularizer of layers. The process included 80 trials and each of which consisted of 80 epochs. All DNNs were trained by Amsgrad [32] optimizer with the default setting. We performed 5-fold cross-validation where 20% of the dataset was chosen as testing sets and 80% as training sets. During the training of the networks, spatial dropout [33] and early stopping [34] methods were adopted to prevent overfitting and to terminate the training. The specification of the current DNNs was detailed in Tables in Supplementary.

4.2. Data and codes

- The codes for prediction and the two experiments in supplementary are released in GitHub: https://github.com/acctouhou/Prediction_of_battery
- The entire pre-trained models and dataset of this work are available as follow: https://drive.google.com/drive/folders/1Aq-wfoQ8ltDqziyHUcka7oUncQ7NgSP8?usp=sharing

CRediT authorship contribution statement

Chia-Wei Hsu: Writing – original draft, Conceptualization, Methodology, Software, Validation, Visualization. **Rui Xiong:** Conceptualization, Validation. **Nan-Yow Chen:** Conceptualization, Methodology, Validation. **Ju Li:** Conceptualization, Methodology, Writing – review & editing, Project administration, Validation. **Nien-Ti Tsou:** Writing – original draft, Conceptualization, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

RX acknowledges support by National Science Foundation for Excellent Young Scholars of China (Grant No. 51922006). NYC acknowledges support by 109-2221-E-492-015 in MOST, Taiwan. JL acknowledges support by NSF CBET-2034902, United States. NTT acknowledges support by 109-2221-E-009 -054 in MOST, Taiwan. We thank National Center for High-Performance Computing of Taiwan for providing computational and storage resources.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apenergy.2021.118134.

References

[1] Zhu Z, Yu D, Shi Z, Gao R, Xiao X, Waluyo I, et al. Gradient-Morph LiCoO2single Crystals with Stabilized Energy Density above 3400 W h L-1. Energy Environ Sci 2020;13(6):1865–78.
[2] Tian Yu, Lin C, Li H, Du J, Xiong R. Detecting Undesired Lithium Plating on Anodes for Lithium-Ion Batteries – A Review on the in-Situ Methods. Appl Energy 2021; 300:117386. https://doi.org/10.1016/j.apenergy.2021.117386.
[3] Remaining the Battery Experience - Qnovo. (2020). Available at Https://Qnovo.Com/.

[4] Panchal S, Mcgrory J, Kong J, Fraser R, Fowler M, Dincer I, et al. Cycling Degradation Testing and Analysis of a LiFePO4 Battery at Actual Conditions. Int J Energy Res 2017;41(15):2565–75.
[5] Tian J, Xiong R, Shen W. A Review on State of Health Estimation for Lithium Ion Batteries in Photovoltaic Systems. eTransportation 2019;2:100028. https://doi.org/10.1016/j.etran.2019.100028.
[6] Xu X, Ding Y, Hu SX, Niemier M, Cong J, Hu Y, et al. Scaling for Edge Inference of Deep Neural Networks. Nat Electron 2018;1(4):216–22.
[7] Sehgal A, Kehtarnavaz N. Guidelines and Benchmarks for Deployment of Deep Learning Models on Smartphones as Real-Time Apps. arXiv 2019;1(1):450–65.
[8] Tan M, Pang R, Le QV. EfficientDet: Scalable and Efficient Object Detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2020.
[9] Montavon G, Samek W, Müller K. Methods for Interpreting and Understanding Deep Neural Networks. Digit Sig Process 2018;73:1–15.
[10] Severson KA, Attia PM, Jin N, Perkins N, Jiang B, Yang Z, et al. Data-Driven Prediction of Battery Cycle Life before Capacity Degradation. Nat Energy 2019;4(5):383–91.
[11] Farmann A, Waag W, Marongiu A, Sauer DU. Critical Review of On-Board Capacity Estimation Techniques for Lithium-Ion Batteries in Electric and Hybrid Electric Vehicles. J Power Sources 2015;281:114–30.
[12] Attia PM, Grover A, Jin N, Severson KA, Markov TM, Liao YH, et al. Closed-Loop Optimization of Fast-Charging Protocols for Batteries with Machine Learning. Nature 2020;578(7795):397–402.
[13] Hannes Knobloch. Predicting Battery Lifetime with CNNs.(2019). Available at Https://Towardsdatascience.Com/Predicting-Battery-Lifetime-with-Cnns-C5e1faeecc8f.
[14] Tian J, Xiong R, Shen W, Lu J, Yang XG. Deep Neural Network Battery Charging Curve Prediction Using 30 Points Collected in 10 Min. Joule 2021;5(6).
[15] Tian J, Xiong R, Shen W, Lu J. State-of-Charge Estimation of LiFePO4 Batteries in Electric Vehicles: A Deep-Learning Enabled Approach. Appl Energy 2021;291:116812. https://doi.org/10.1016/j.apenergy.2021.116812.
[16] Yang R, Xiong R, Shen W, Lin X. Extreme Learning Machine-Based Thermal Model for Lithium-Ion Batteries of Electric Vehicles under External Short Circuit. Engineering 2021;7(3):395–405.
[17] Xiong R, Wang Ju, Shen W, Tian J, Mu H. Co-Estimation of State of Charge and Capacity for Lithium-Ion Batteries with Multi-Stage Model Fusion Method. Engineering 2021. https://doi.org/10.1016/j.eng.2020.10.022.
[18] Hong J, Lee D, Jeong ER, Yi Y. Towards the Swift Prediction of the Remaining Useful Life of Lithium-Ion Batteries with End-to-End Deep Learning. Appl Energy 2020;278:115646.
[19] Lea C, Flynn MD, Vidal R, Reiter A, Hager GD. Temporal Convolutional Networks for Action Segmentation and Detection. In: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR *2017*; 2017; Vol. 2017-January.
[20] Castelvecchi D. Can We Open the Black Box of AI? Nature 2016;538(7623):20–3.
[21] Ruder S. An Overview of Multi-Task Learning for Deep Learning. Sebastian Ruder 2017.
[22] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2016; Vol. 2016-December.
[23] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2016; Vol. 2016-December.
[24] Luong MT, Pham H, Manning CD. Effective Approaches to Attention-Based Neural Machine Translation. In: Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing; 2015.
[25] Lin M, Chen Q, Yan S. Network in Network. In 2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings; 2014.
[26] Tang X, Zou C, Yao K, Lu J, Xia Y, Gao F. Aging Trajectory Prediction for Lithium-Ion Batteries via Model Migration and Bayesian Monte Carlo Method. Appl Energy 2019;254:113591.
[27] Montavon G, Lapuschkin S, Binder A, Samek W, Müller K-R. Explaining Nonlinear Classification Decisions with Deep Taylor Decomposition. Pattern Recogn 2017;65:211–22.
[28] Zhang Y, Xiang T, Hospedales TM, Lu H. Deep Mutual Learning. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2018.
[29] Zheng F, Xing Y, Jiang J, Sun B, Kim J, Pecht M. Influence of Different Open Circuit Voltage Tests on State of Charge Online Estimation for Lithium-Ion Batteries. Appl Energy 2016;183:513–25.
[30] Huber PJ. Robust Estimation of a Location Parameter. Ann Math Stat 1964;35(1):73–101.
[31] Bergstra J, Yamins D, Cox DD. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. In: 30th International Conference on Machine Learning, ICML *2013*; 2013.
[32] Reddi SJ, Kale S, Kumar S. On the Convergence of Adam and Beyond. In: 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings; 2018.
[33] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Dropout SR. A Simple Way to Prevent Neural Networks from Overfitting. J Mach Learn Res 2014;15.
[34] Caruana R, Lawrence S, Giles L. Overfitting in Neural Nets: Backpropagation, Conjugate Gradient, and Early Stopping. In: Advances in Neural Information Processing Systems; 2001.

Supplementary

Table S 1: The definition of the symbols used in the current work.

Symbol	Component	Definition	Symbol	Component	Definition
n		Cycle number in the dataset, $n = 1 \ldots \mathrm{EoL}$	F_{im}^{human}		Human-picked features in cycle *m*
s		Number of shifted cycles from $n = 1$ used as the first cycle of input to the DNNs.	F_{1m}^{human}	P_m	Charge capacity in cycle *m*
m		Cycle number in the input dataset to the DNNs, where $m = 1 \ldots a$.	F_{2m}^{human}	Q_m	Discharge capacity in cycle *m*
a		Total number of cycles used as input to the DNNs, i.e. $a = \max(m)$.	F_{3m}^{human}	T_m^{avg}	Average temperature in cycle *m*
t		Time step	F_{4m}^{human}	T_m^{min}	Temperature minimum in cycle *m*
EoL		End-of-life	F_{5m}^{human}	T_m^{max}	Temperature maximum in cycle *m*
RUL		Residual useful life	F_{6m}^{human}	t_m^{charge}	Total charge time in cycle *m*
D_n		Discharge power of cycle number *n*			
x_{itm}^{dis}		Continuous-*t* data in discharge half-cycle *m*	$F_{im}^{\mathrm{data,\,dis}}$		Data-driven features in discharge half-cycle *m*
x_{1tm}^{dis}	$Q(t)_m^{\mathrm{dis}}$	Capacity curve in discharge half-cycle *m*	$F_{1m}^{\mathrm{data,\,dis}}$	$\widetilde{\mathrm{EoL}}_m^{\mathrm{dis}}$ / $\widetilde{\mathrm{RUL}}_m^{\mathrm{dis}}$	Data-driven feature for end-of-life/ residual useful life in discharge half-cycle *m*
x_{2tm}^{dis}	$V(t)_m^{\mathrm{dis}}$	Voltage curve in discharge half-cycle *m*	$F_{2m}^{\mathrm{data,\,dis}}$	$\left(\tilde{t}_{n=\mathrm{EoL}}^{\mathrm{charge}}\right)_m^{\mathrm{dis}}$	Data-driven feature for total charge time in discharge half-cycle *m*
x_{3tm}^{dis}	$I(t)_m^{\mathrm{dis}}$	Current curve in discharge half-cycle *m*	$F_{3m}^{\mathrm{data,\,dis}}$	$\tilde{s}_m^{\mathrm{dis}}$	Data-driven feature for age in discharge half-cycle *m*
x_{4tm}^{dis}	$T(t)_m^{\mathrm{dis}}$	Temperature curve in discharge half-cycle *m*			
x_{itm}^{char}		Continuous-*t* data in charge half-cycle *m*	$F_{im}^{\mathrm{data,\,char}}$		Data-driven features in charge half-cycle *m*
x_{1tm}^{char}	$Q(t)_m^{\mathrm{char}}$	Capacity curve in charge half-cycle *m*	$F_{1m}^{\mathrm{data,\,char}}$	$\widetilde{\mathrm{EoL}}_m^{\mathrm{char}}$ / $\widetilde{\mathrm{RUL}}_m^{\mathrm{char}}$	Data-driven feature for end-of-life/ residual useful life in charge half-cycle *m*
x_{2tm}^{char}	$V(t)_m^{\mathrm{char}}$	Voltage curve in charge half-cycle *m*	$F_{2m}^{\mathrm{data,\,char}}$	$\left(\tilde{t}_{n=\mathrm{EoL}}^{\mathrm{charge}}\right)_m^{\mathrm{char}}$	Data-driven feature for total charge time in charge half-cycle *m*
x_{3tm}^{char}	$I(t)_m^{\mathrm{char}}$	Current curve in charge half-cycle *m*	$F_{3m}^{\mathrm{data,\,char}}$	$\tilde{s}_m^{\mathrm{char}}$	Data-driven feature for age in charge half-cycle *m*
x_{4tm}^{char}	$T(t)_m^{\mathrm{char}}$	Temperature curve in charge half-cycle *m*			

The illustration of *n*, *s*, *m*, and *a*

Since we have enabled the flexibility of the input data for the inference, four symbols related to each cycle, i.e. *n*, *s*, *m*, and *a* are defined, and each of which has different definition. Symbol *n* represents cycle number in the dataset, and thus, the range of *n* for a given battery is $n = 1 \ldots \text{EoL}$. Symbols *s*, *m*, and *a* are defined for sampling a segment of cycles from the entire life time of the given battery. Symbol *s* is defined as the number of shifted cycles from cycle number $n = 1$ used as the first cycle of input for either inference or training. Symbol *a* is defined as the total number of cycles used as input. Symbol *m* is defined to describe each cycle used as input, and thus, the range of *m* for a given input is $m = 1 \ldots a$.

Fig. S 1 shows two examples. The first case of *s*=0, *a*=5 indicates that the first five cycles of the battery (cycle number: 1…5 in the dataset) are extracted and with the features in the fifth cycle padding to form the input data for the inference of EoL, t_m^{charge}… etc. by Discharge DNN. The second case of *s*=5, *a*=48 indicates that 48 cycles of the battery from the cycle number 6 (cycle number: 6…53 in the dataset) are extracted and with the features in the 48th cycle padding twice to form the input data for the inference of RUL, *s*… etc. by Full RUL DNN.

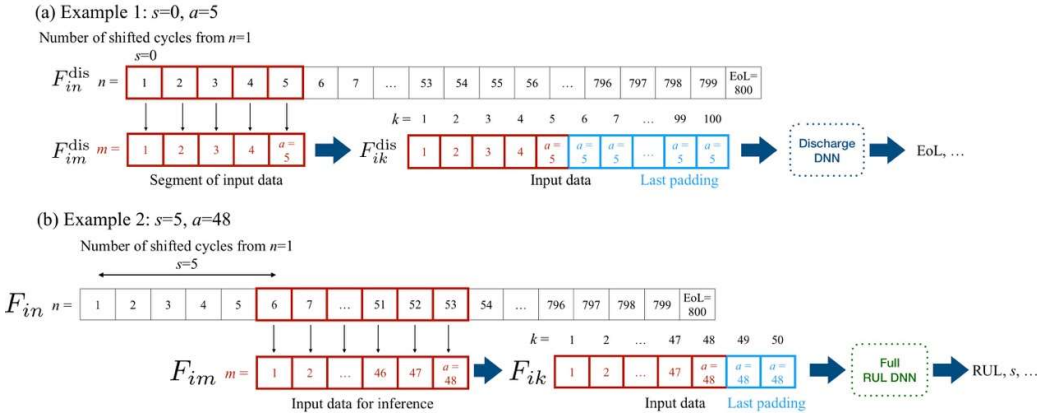


Fig. S 1: The illustration of two examples. (a) *s*=0, *a*=5 and (b) *s*=5, *a*=48 for the inference by Discharge DNN and Full RUL DNN, respectively.

The detail of each NN pipeline

The workflow of the three DNNs in the current work, as shown in Fig. S 2. Where cubes represent data in tensor form and rounded rectangles represent neural networks. The elements/building blocks/inputs/outputs etc. of each NN pipeline have been illustrated in Fig. S 3 to Fig. S 13, showing all the detail of the current DNNs. Note that, due to the limitation of the article length, Q_n and D_n predicted by Predictor 4 for cycle-by-cycle information are not mentioned in manuscript and shown in supplementary only. The codes for prediction and the two experiments in supplementary are released in GitHub:

https://github.com/acctouhou/Prediction_of_battery

The entire pre-trained models and dataset of this work are available as follow:

https://drive.google.com/drive/folders/1Aq-

wfoQ8ltDqziyHUcka7oUncQ7NgSP8?usp=sharing

Note that this link is shared by author's personal Google Drive and only available for the current review process.

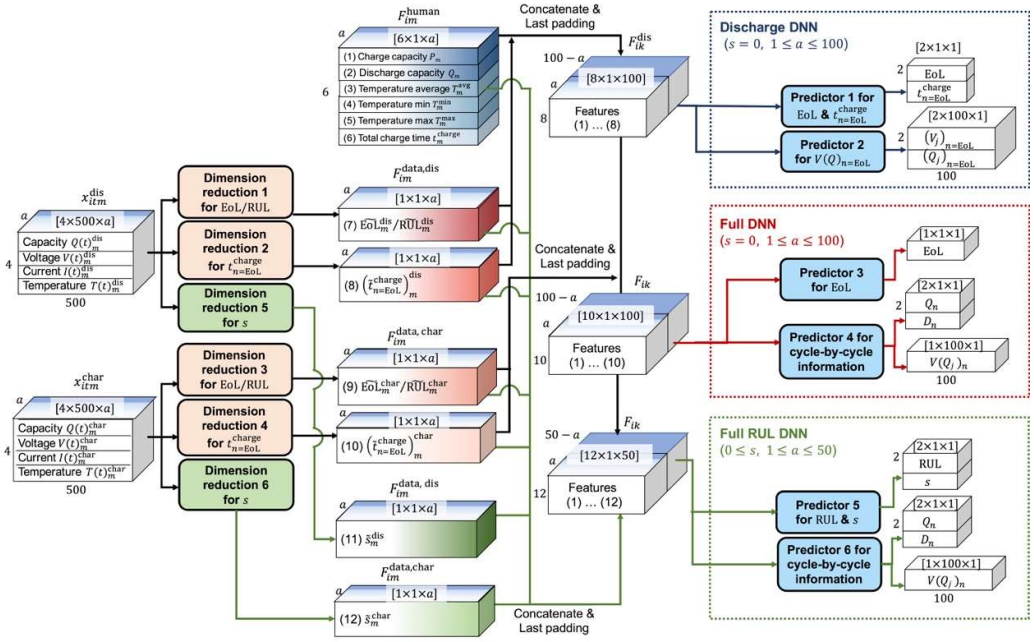


Fig. S 2: The workflow of the three DNNs in the current work. Where cubes represent data in tensor form and rounded rectangles represent neural networks.

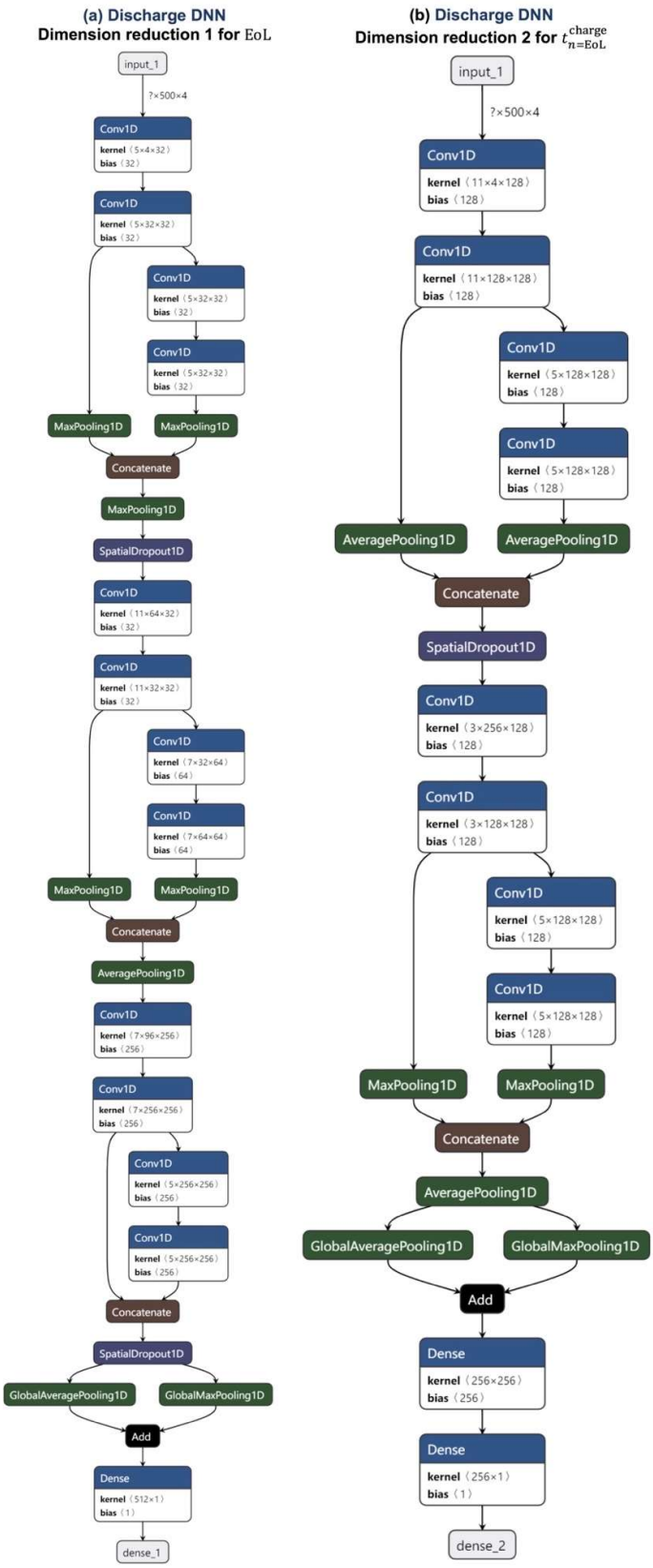


Fig. S 3: The architecture of Dimension reduction (a) 1 and (b) 2 in Discharge DNN.

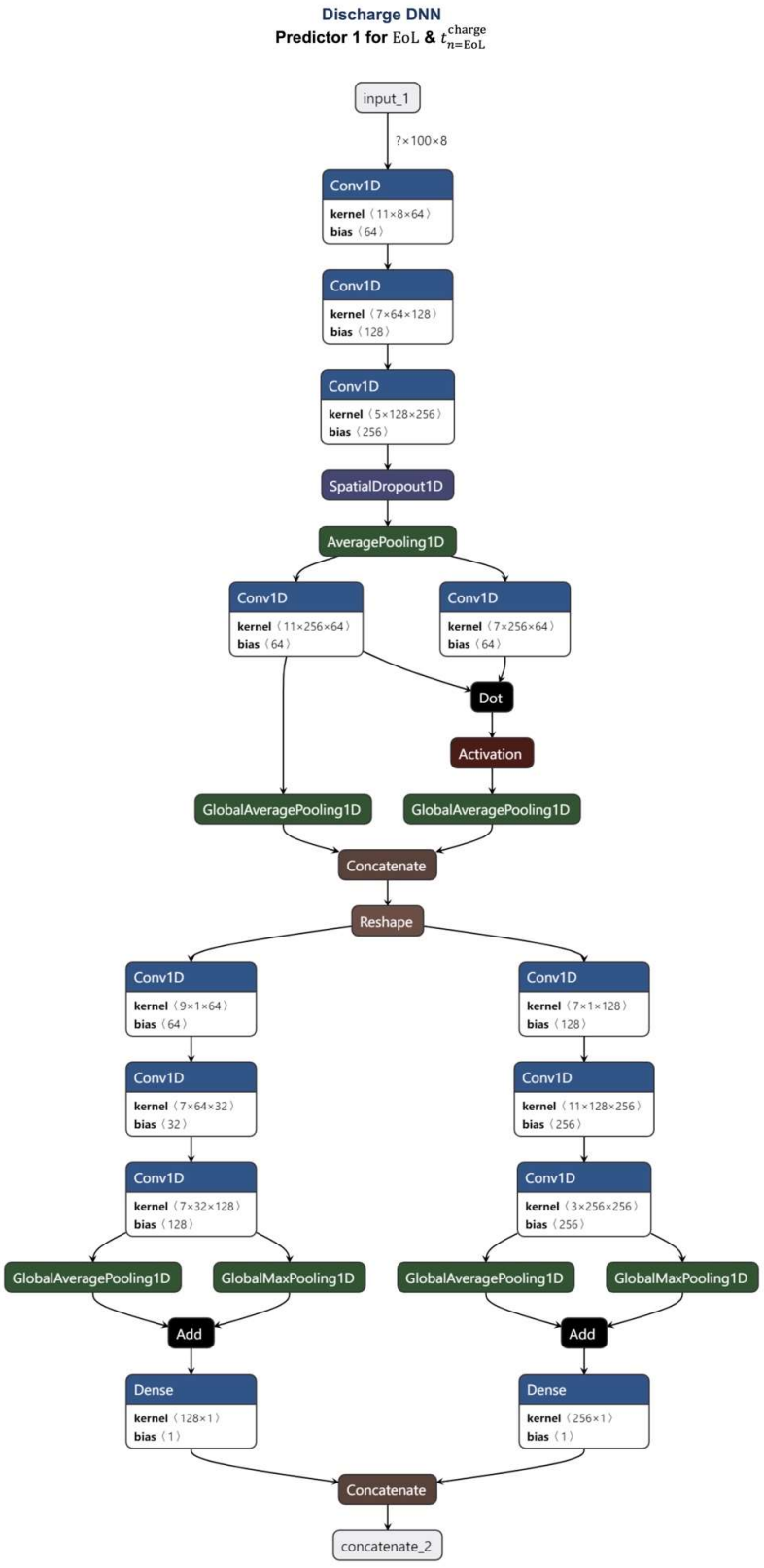


Fig. S 4: The architecture of Predictor 1 in Discharge DNN.

.

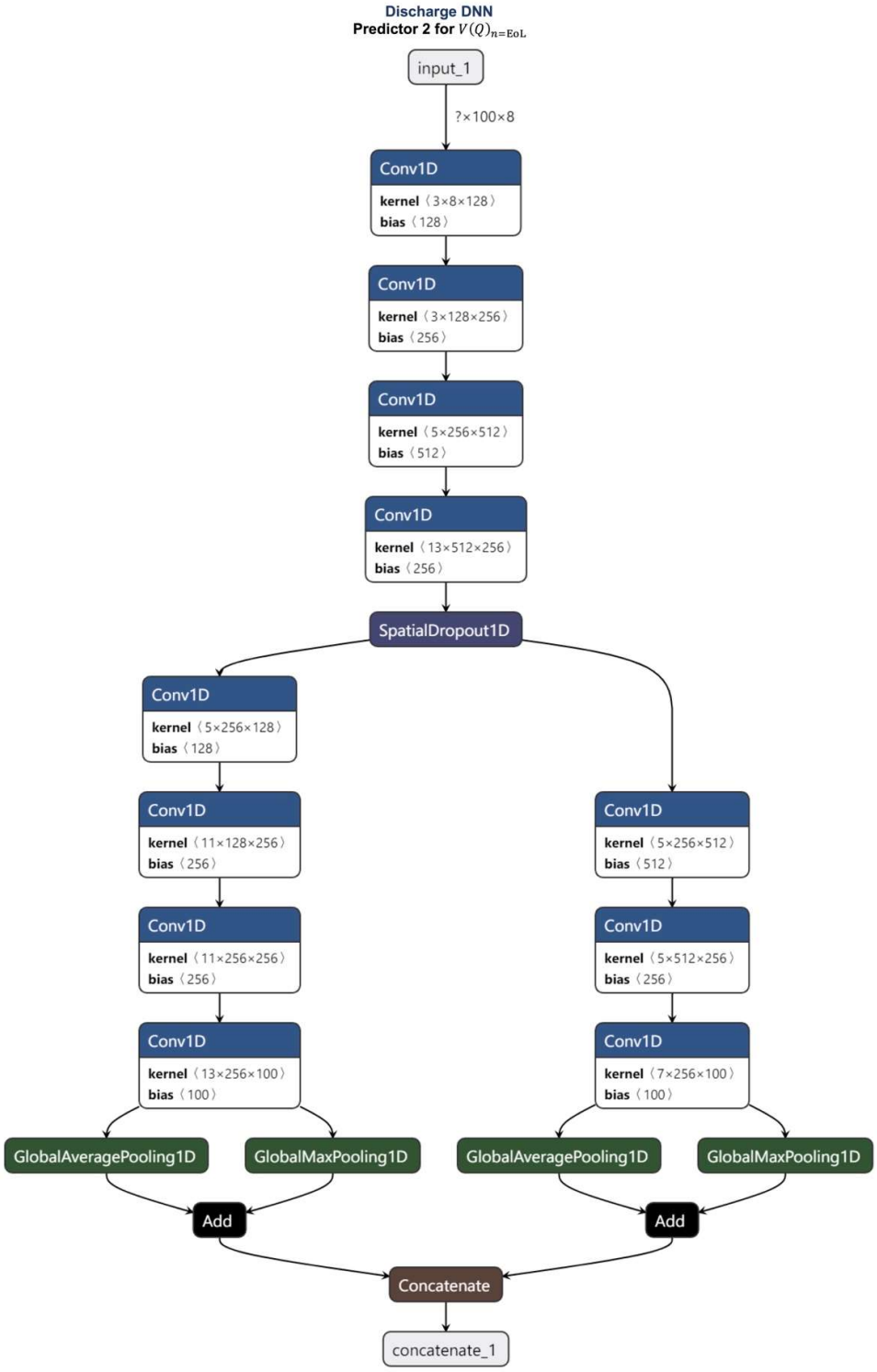


Fig. S 5: The architecture of predictor 2 in Discharge DNN.

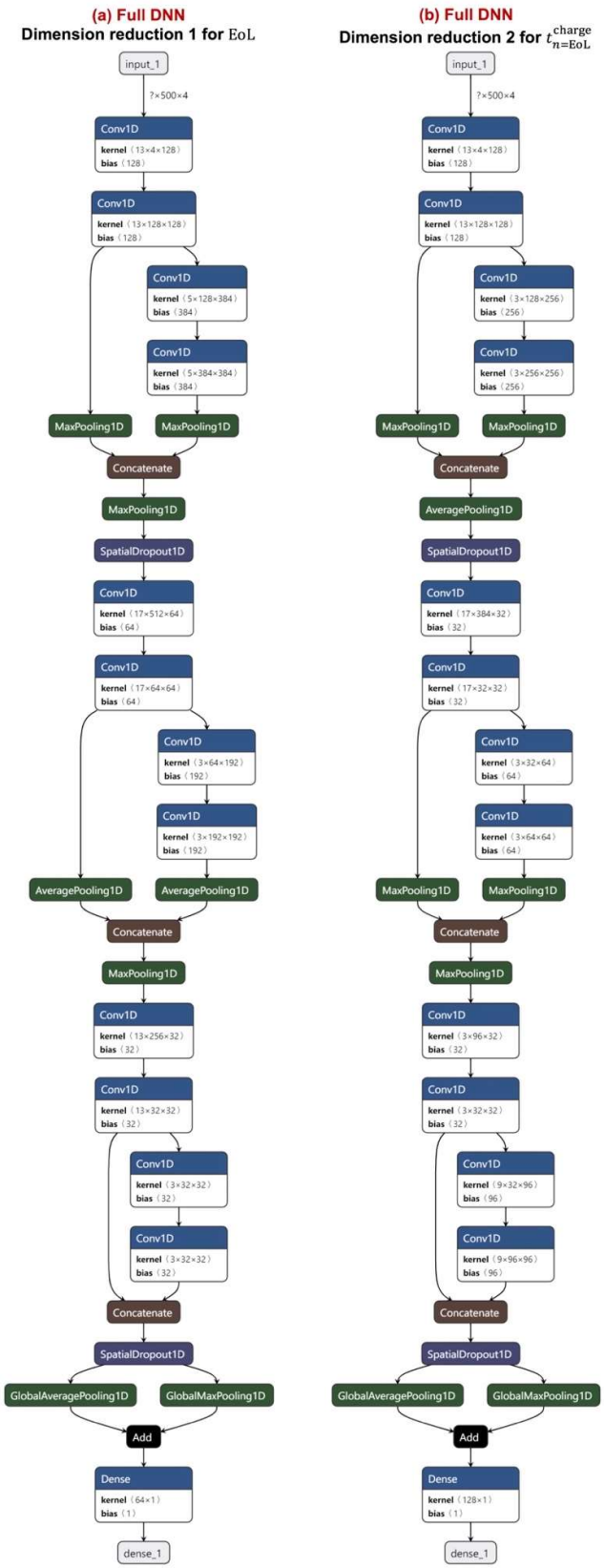


Fig. S 6: The architecture of Dimension reduction (a) 1 and (b) 2 in Full DNN.

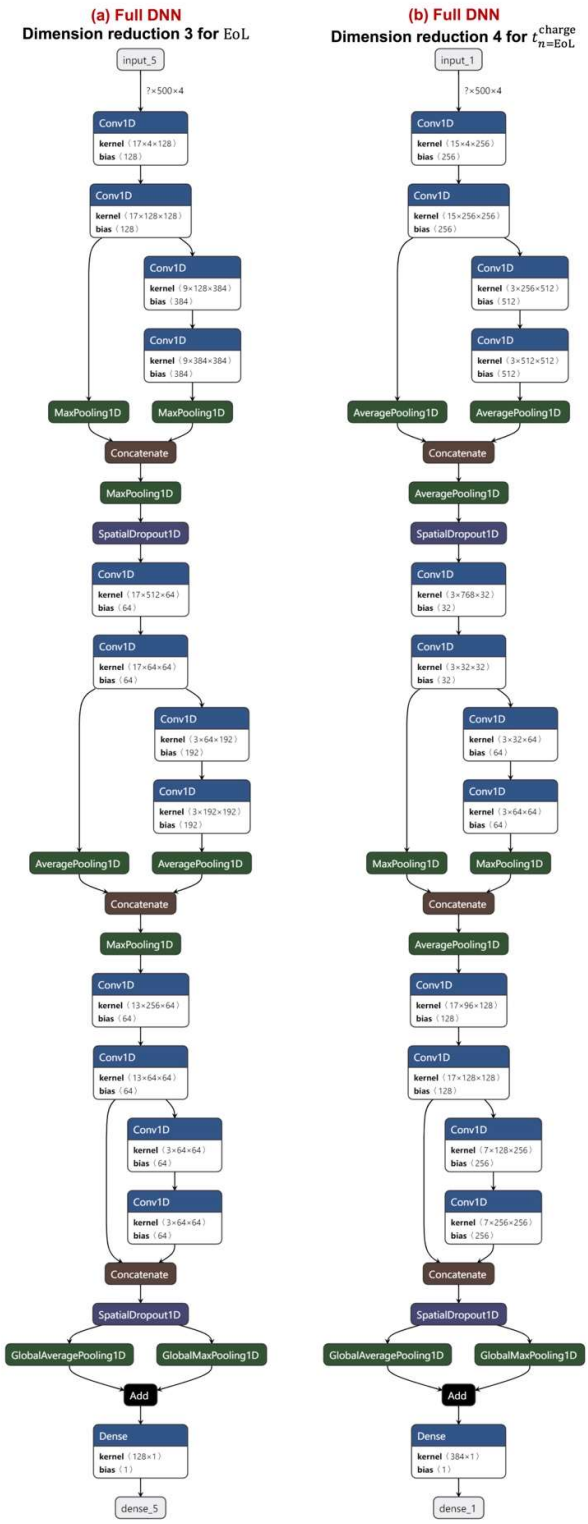


Fig. S 7: The architecture of Dimension reduction (a) 3 and (b) 4 in Full DNN.

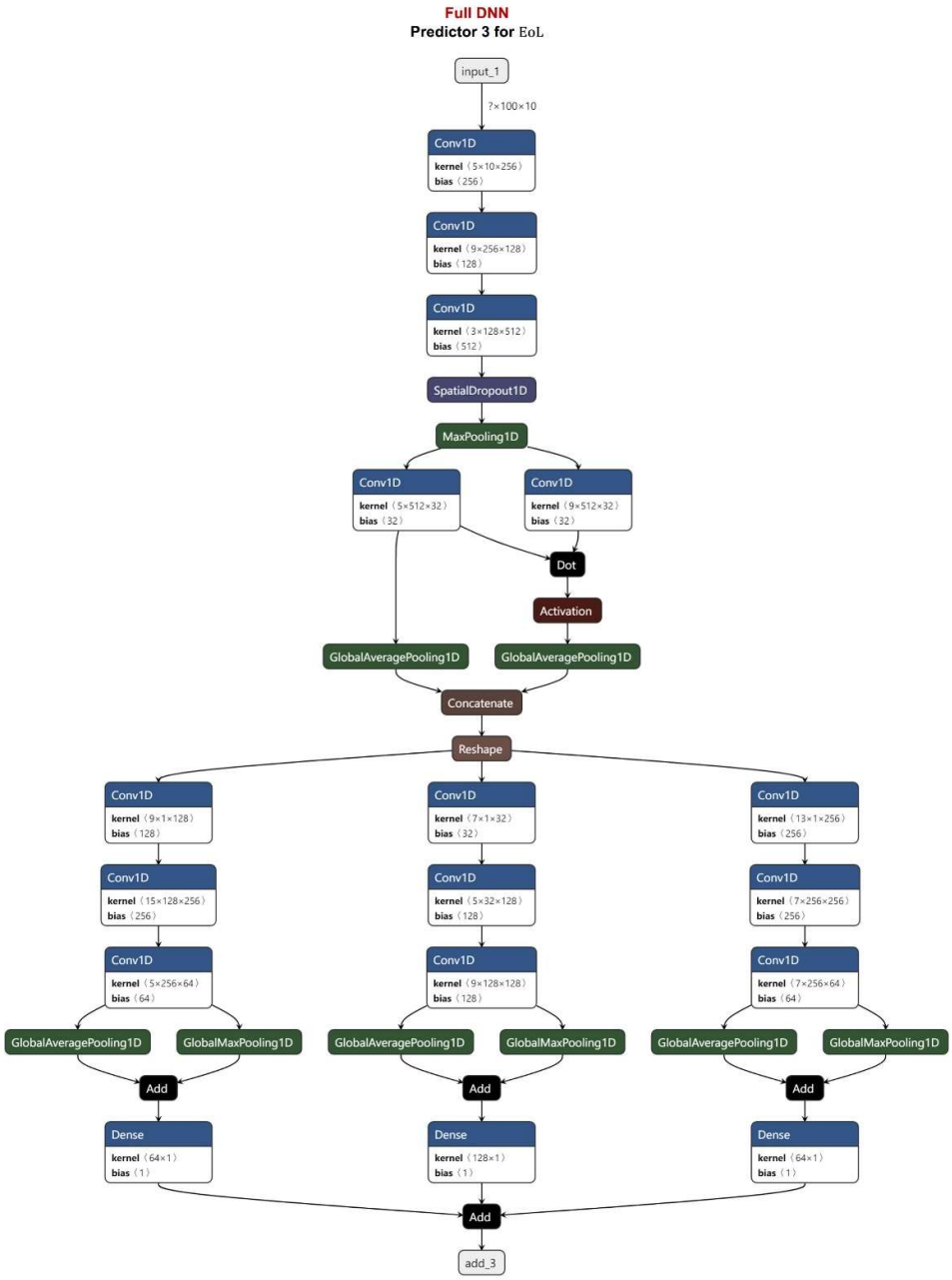


Fig. S 8: The architecture of Predictor 3 in Full DNN.

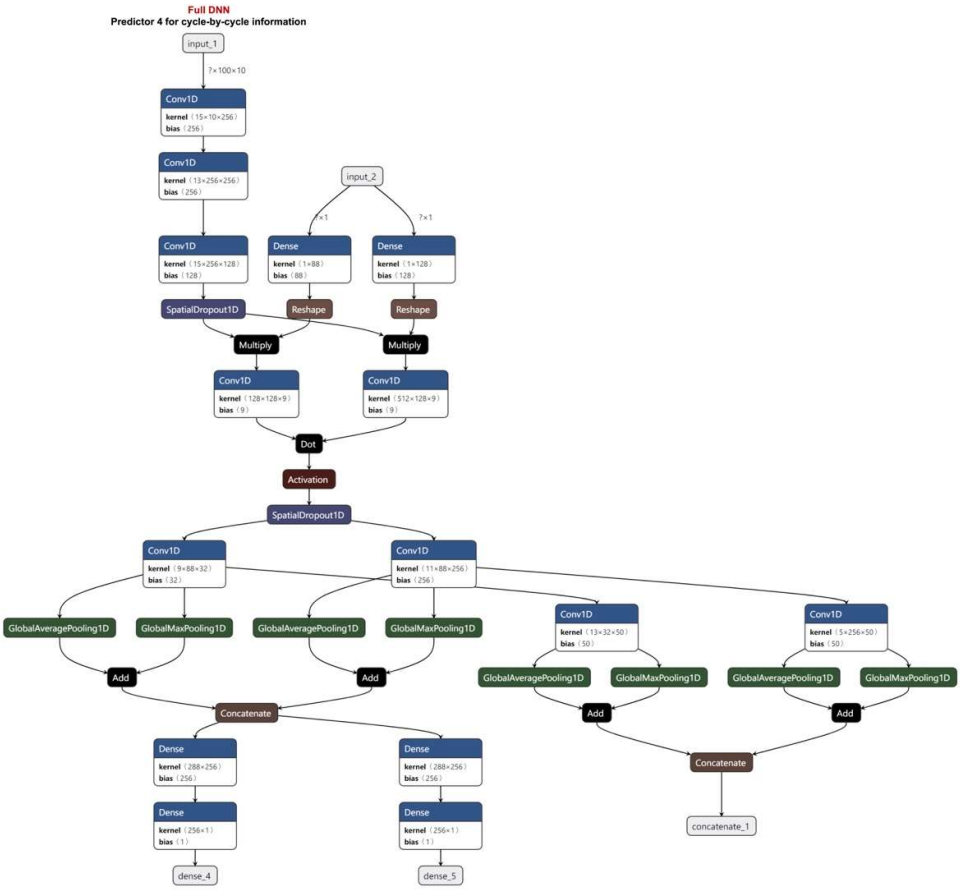


Fig. S 9: The architecture of Predictor 4 in Full DNN.

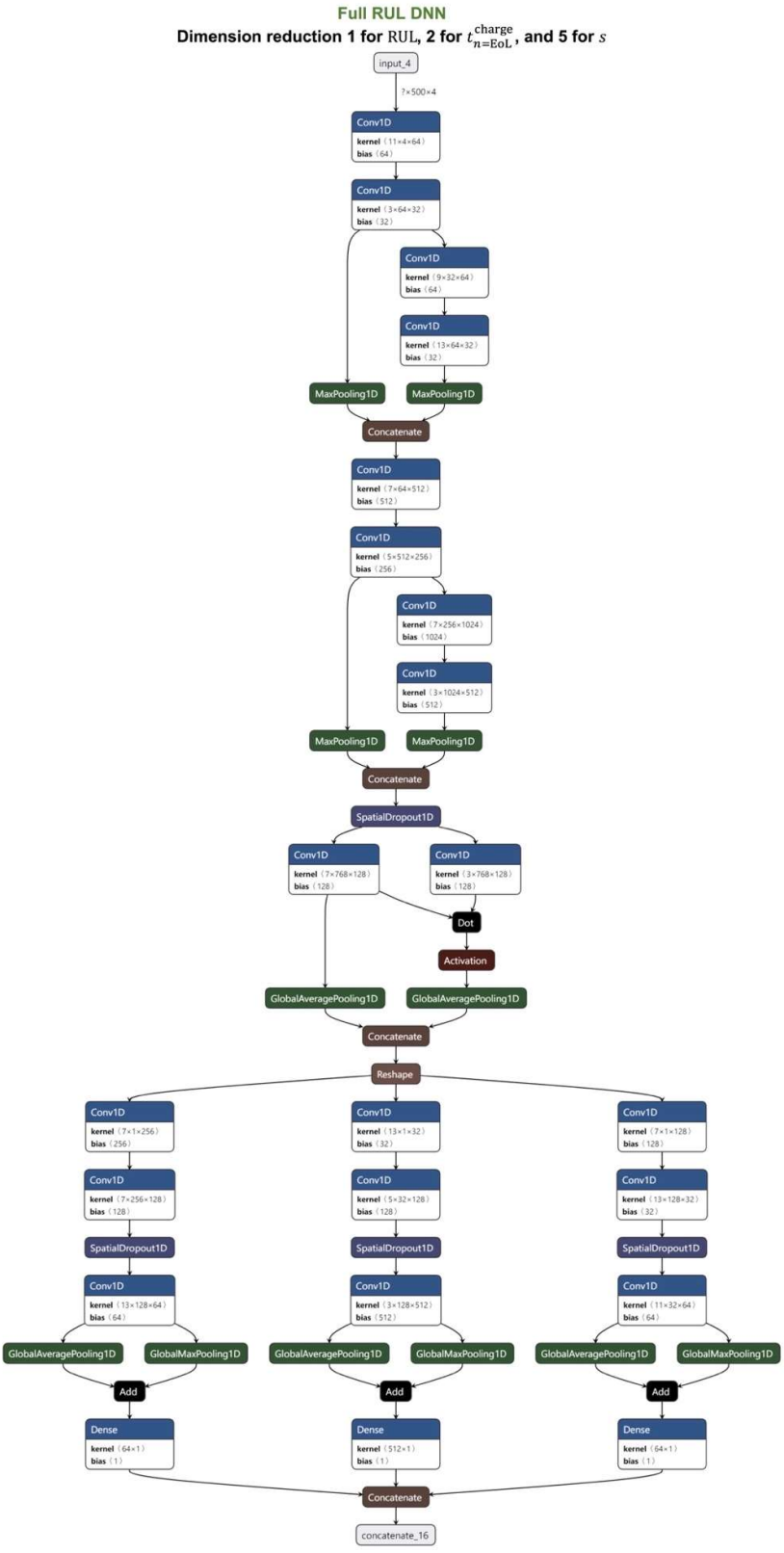


Fig. S 10: The architecture of Dimension reduction 1, 2, and 5 in Full RUL DNN.

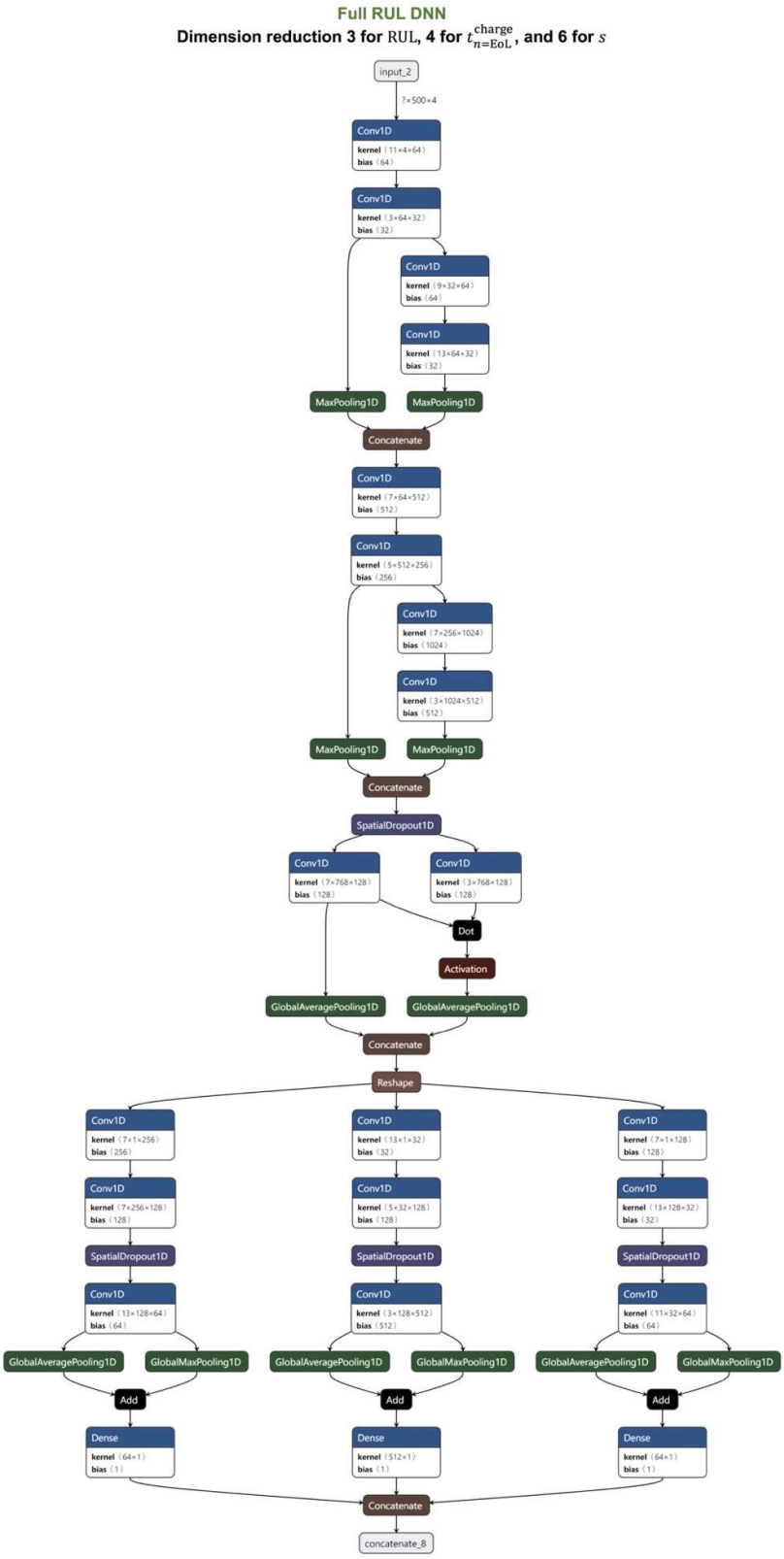


Fig. S 11: The architecture of Dimension reduction 3, 4, and 6 in Full RUL DNN.

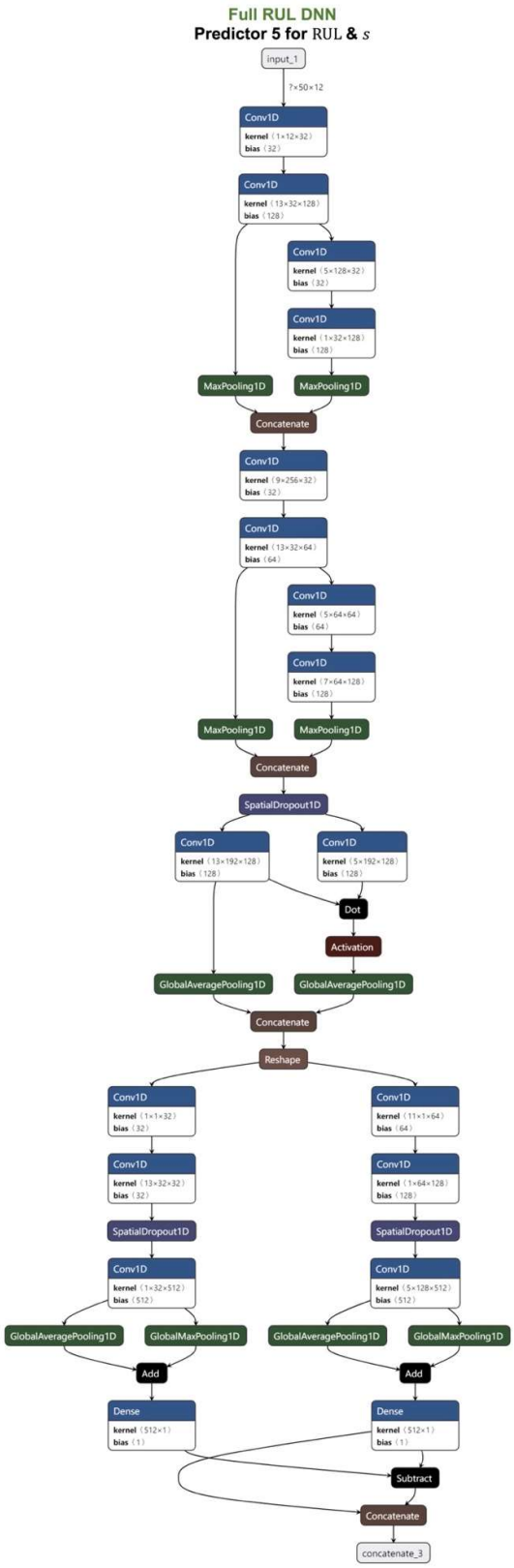


Fig. S 12: The architecture of Predictor 5 in Full RUL DNN.

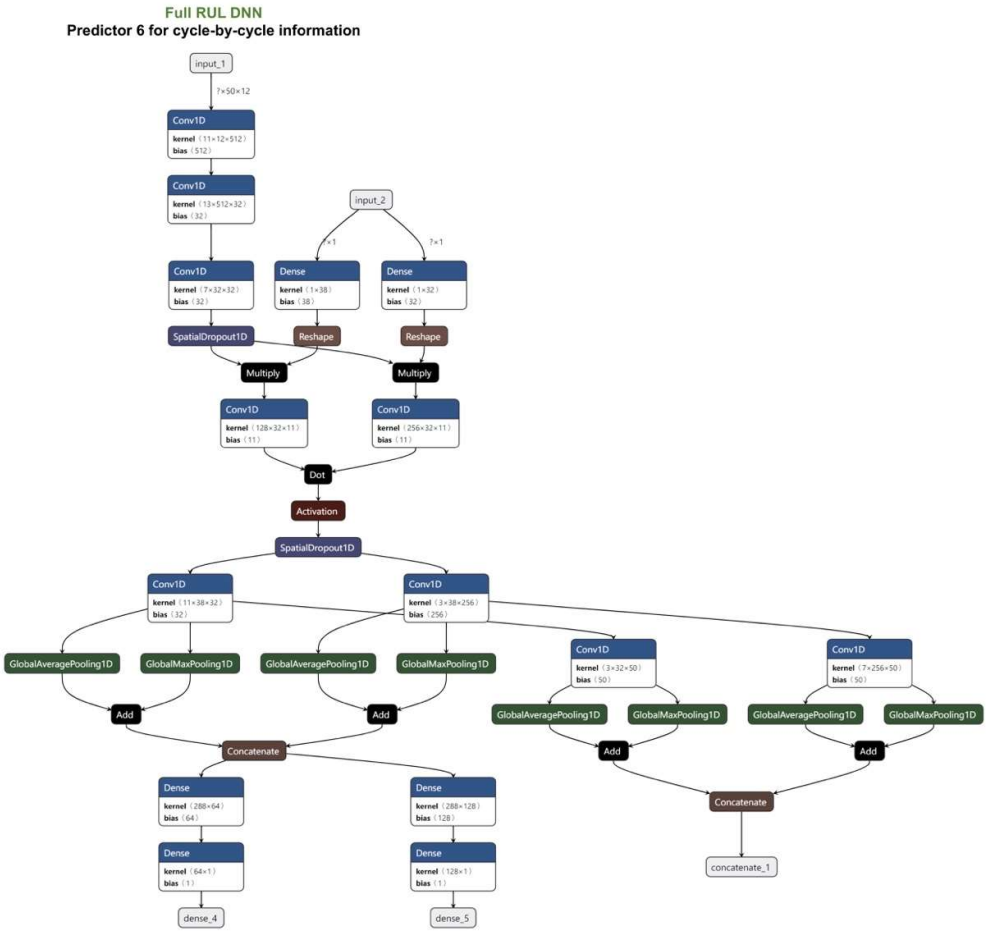


Fig. S 13: The architecture of Predictor 6 in Full RUL DNN.

The illustration of the testing strategies used in the current DNNs

As mentioned in the main texts, three different testing strategies were adopted in the DNNs. The illustration of the three selection methods for training and testing datasets was shown in Fig. S 14. Where each grid represents a single cycle, and each column represents the case of a battery. Fig. S 14 (a) shows that testing strategy <1> unselected the entire sequences of randomly chosen batteries; Fig. S 14(b) shows that testing strategy <2> unselected the entire sequences of all the batteries with randomly chosen charging policies. Sampling started from the first cycle ($s = 0$) and the length of each sampling may vary from $n = 1 \ldots 100$. Fig. S 14 (c) shows that testing strategy <3> was similar to strategy <2>. However, convolutional sampling was adopted here where s may vary and $a = 1 \ldots 50$, corresponding to Full RUL DNN. Note that the convolution sampling for each selected battery continued until the cycle of $\mathrm{EOL} - 100$ was reached, in order to avoid the unstable properties and behavior of the battery near its EoL. These testing strategies were designed to force DNNs to achieve their own objectives.

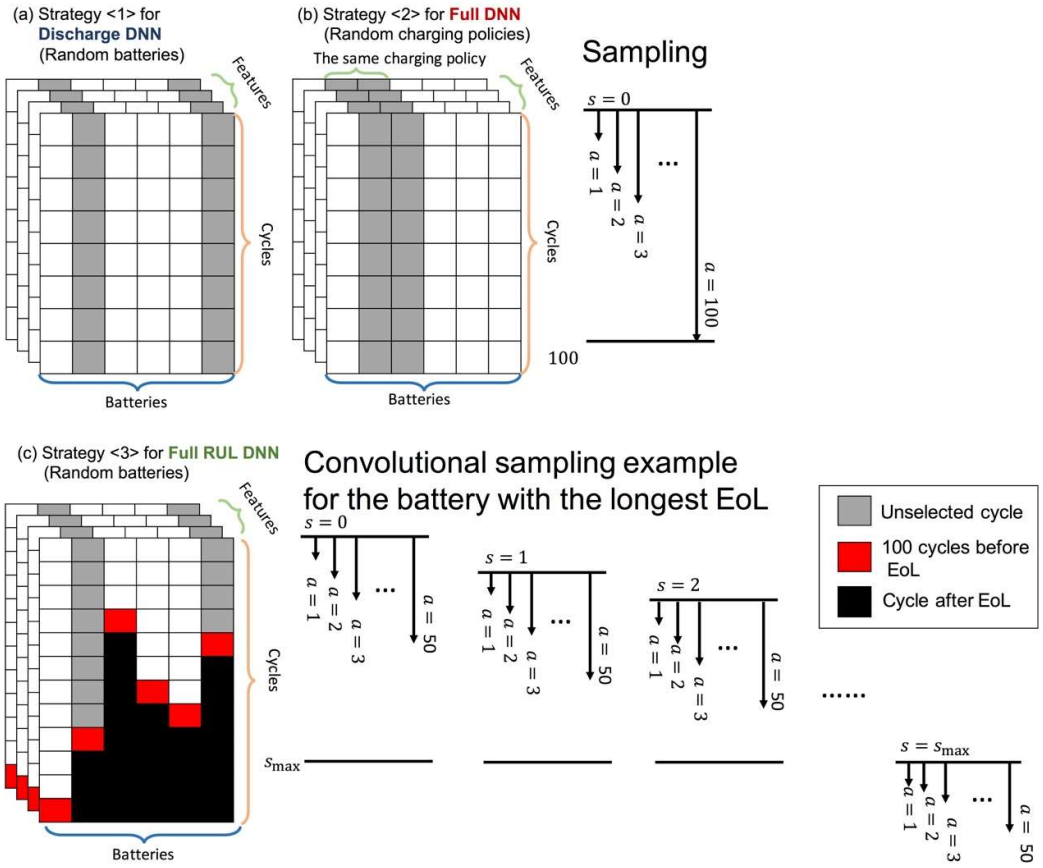


Fig. S 14: Three testing strategies. The grids colored in gray will be used as testing sets; the white grids will be used as training sets; the red grids are the 100 cycles before EoL; the black grids are the cycles after EoL of that battery.

The ability of extrapolation of Full DNN for unlearned charging policies

There were 68 different charging policies in the datasets provided by Severson et al. Full DNN adopted special network structure and considered the data of charge process, enabling accurate prediction of EoL even charging policies that have not been learned. This remarkable extrapolation ability for EoL prediction can be illustrated by Fig. S 15. The three axes of the figures ($C1$, $C2$, and $Q1$) represent the first and second applied discharge currents, and the state of charge current (%) where the current switched, respectively (the illustration for these symbols is given in Fig. S 15(c)). This indicates that each point specifies a charging policy. Fig. S 15(a) shows the RMSE of predicted EoL for each charging policy, based on the first cycle only, when all the cells with that charging policy were used as testing sets, i.e. they were not included in training sets. It can be observed that most of the points are in dark blue color (low RMSE) in Fig. S 15(a). Points with colors other than dark blue typically accumulated in the region of $C2 < 4$. This implied that lower currents applied at the second stage of charge process may be less harmful to cells, leading to less aging features in the first few cycles for Full DNN to inference. Next, Fig. S 15(b) shows the reduction of RMSE for each charging policy when the first 100 cycles were considered. It is interesting that Fig. S 15(a) and (b) look almost the same, indicating that cells with certain charging policies which were not well captured by

Full DNN with fewer input cycles, had significant improvement as long as large enough number of cycle used as input.

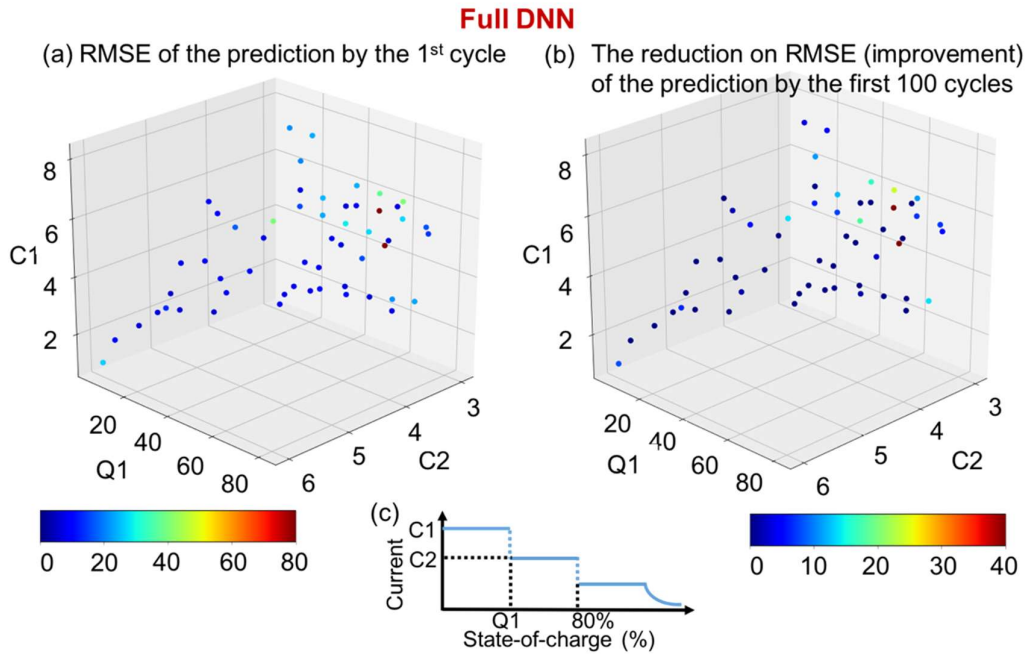


Fig. S 15: (a) RMSE of predicted EoL for each charging policy, based on the first cycle only, when all the cells with that charging policy were used as testing sets. (b) The reduction of RMSE for each charging policy when the first 100 cycles were considered. (c) The illustration of the charging policy adopted in Severson et al.

The learning progress of the current DNNs

Here we take Discharge DNN as an example to illustrate the learning progress of the DNNs on the current battery problems. The loss (RMSE) of Discharge DNN corresponding to training epochs and first *a* cycles as input data was visualized as a 3-dimensional landscape in Fig. S 16. At the early stage of the training, the loss for the cases with more input cycles decreased more significantly compared to the cases with fewer input cycles. After 50 training epochs, the loss for cases with more than the first 80 input cycles converged at the minimum. This indicated that cases with more input cycles dominated the weights and biases of the DNN. Then, the optimizers of the DNN were able to learn about the cases with fewer input cycles more efficiently due to the assistance of those existed parameters predetermined by the cases with more input cycles.

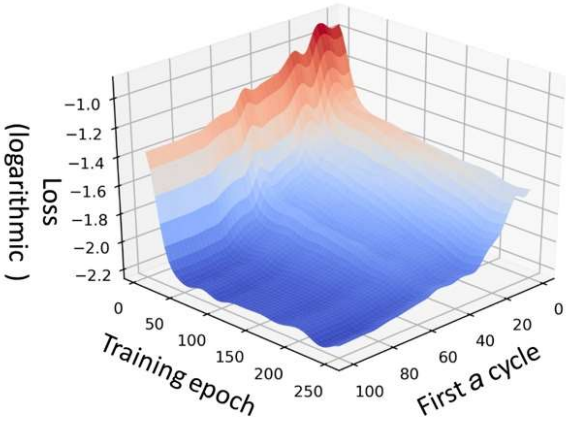


Fig. S 16: The 3-dimensional landscape of the loss of Discharge DNN corresponding to training epochs and first *a* cycles as input data.

We also used the relevance score of each feature corresponding to the first *a* input cycles obtained by Deep Taylor decomposition (DTD) to visualize the learning progress of the two DNNs about each feature, as shown in Fig. S 17. For the illustration purpose, features of similar types, such as features related to capacity and temperature, were averaged and merged into one curve. Note that the area under these curves represented the information about the corresponding features learned by the DNN. It can be observed that, for both the two DNNs, a significant amount of information has been learned before the first 40 input cycles and the data after the first 60 or 80 input cycles were less influential to the predicted results.

It is of interest noting that the curve for Discharge DNN indicated that the predicted results were still greatly affected by the data around the 80th cycle. This is because of the uncertainty introduced by the unlearned charging policies, and apparently, the features of these policies were very difficult to be extracted by Discharge DNN with the discharge input data only. Full DNN resolved this problem by considering both the charging and discharge input data. The sufficient amount of data assisted Full DNN to extract features.

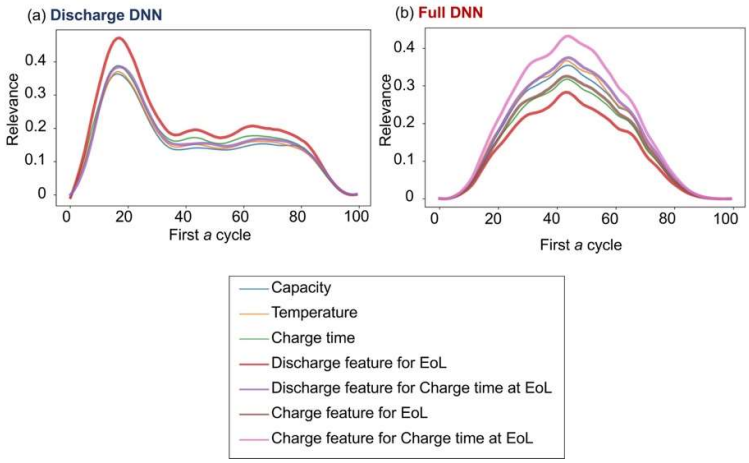


Fig. S 17: The DTD relevance score of each feature corresponding to the first a input cycles for (a) Discharge and (b) Full DNN. A significant amount of information has been learned by all the two DNNs before the first 40 input cycles and the data after the first 60 or 80 input cycles were less influential to the predicted results.

The DTD relevance score of each feature was used again to demonstrate the flexibility and the robustness of EoL prediction of our DNNs. Since network in network (NiN) structure was adopted in Predictors in all the current DNNs, the sequence length of the input data can be varied by using the last padding technique. Here, input data with a length of 60, 80, and 100 (the present setting described in the main texts) cycles for the previous version of Discharge DNN was taken as an example here. The bounded DTD approach was applied to interpret the behavior of the DNN and the corresponding relevance scores of human-picked and data-driven features with different transparency are shown in Fig. S 18. It can be observed that the shapes of the curves are all similar, showing great robustness of the learning progress of the current DNN even with fewer input cycles. More importantly, the peaks of all the three cases are in the range of the 20^{th} and 40^{th} cycles. This demonstrates that the power of the current DNN in terms of serving as an early predictor of EoL of batteries.

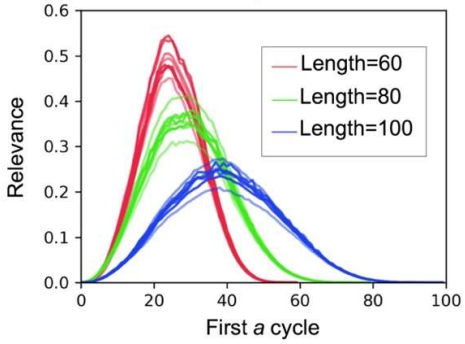


Fig. S 18: The DTD relevance score of the previous version of Discharge DNN with an input data length of 60, 80, and 100 cycles, showing flexibility and the robustness of EoL prediction.

Five-fold cross-validation tests of the current DNNs

The reproducibility and reliability of the two DNNs were examined by five-fold cross-validation tests. Five different training sets were used to train the three DNNs. Fig. S 19 shows the range of RMSE of EoL vs. the number of consecutive input cycles for all the five cases in blue; the average of the curves was marked by a dashed line. It can be observed that the fluctuation of RMSE occurred in both DNNs in the early stage. This was because batteries were typically at a relatively healthy state initially, showing almost no aging features in the curves. Thus, it is reasonable that DNNs may produce relatively inaccurate results based on fewer cycles in the early stage. Another reason was that the quality of measurement in each cycle is more influential to the performance of prediction when fewer input cycles were considered. These issues can be solved by either feeding more data of cycles to the DNNs or providing data of the batteries with early occurrence of aging where extreme charging/discharge policies were applied.

As the number of input cycle *a* becomes greater than around 30, all DNNs reached their converged values of RMSE. Wherein, the converged RMSE value of Discharge DNN was less than it of Full DNN. This is reasonable due to the different nature of the two DNNs. As mentioned above, the training datasets of Discharge DNN missed the data of some randomly selected batteries (testing strategy <1>); those of Full DNN missed the data of certain batteries with some randomly selected charging policies

(testing strategy <2>). Thus, Discharge DNN and Full DNN were forced to have the ability to extrapolate for unknown batteries and charging policies, respectively. Thus, the prediction of Full DNN encountered a much higher complexity than Discharge DNN. Fortunately, Fig. S 19 shows that both of DNNs achieved high robustness as the input cycle number increased.

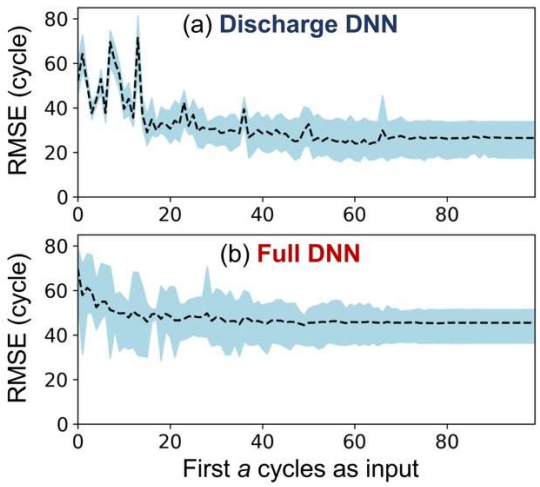


Fig. S 19: The range of RMSE of EoL vs. the first *a* cycles as input data for the five-fold cross-validation tests colored in blue. Where the average of the five curves marked with the dashed line.

Data screening process for the dataset provided by Attia et al.

We have implemented a program and screen out the unusable data of batteries both manually and automatically in the dataset provided by Attia et al., based on four types of reason. First, the experiment log provided in the dataset package stated that batteries in channel 5, 12, 45, 46, 48 are problematic. Second, after the data was visualized, we found that curves of some batteries are greatly shifted between different cycles. For example, current curves of batteries in channel 9, 17, 31, 34 and voltage curves of batteries in channel 9, 17, 30, 34, 39 are greatly shifted, where one of the cases were shown in Fig. S 20(a)(b). In addition, Fig. S 20 (b) also shows there was an unusual voltage drop at the early stage of the measurement step in one of the cycles, marked by a green circle. Third, the patterns of the measured curves of certain batteries are weird and different from the most of the batteries in the dataset. For example, temperature curves of batteries in channel 14, 15, 16, 17, 30, 34, 44 are messy, where one of the cases were shown in Fig. S 20(c). Forth, an unreasonable negative value recorded in time log at a random measurement step in the 31st to 35th cycle during the charging or discharging process in all the batteries. This systematic problem was not recoded in their experiment log. Thus, we did not inference these batteries as they did not pass our data screening program.

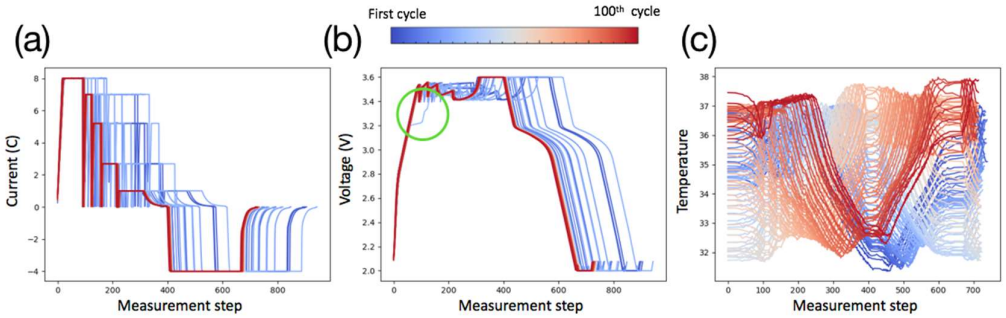


Fig. S 20: One of the examples that (a)(b) the measured curves are greatly shifted between different cycles, and (c) the patterns of the curves in the battery are different from the most of the batteries in the dataset.

We also have applied Full DNN to inference EoL of 34 batteries in Attia et al. These 34 batteries were chosen because their data of the first cycle only can pass our data screening program, while the quality of the remaining cycles may not have acceptable quality for the inference of EoL. The predicted values EoL vs. EoL is shown in Fig. S 21(a). It can be observed that our Full DNN successfully predicted EoL of the 34 batteries with very low mean-absolute-percentage-error of 9.6%, based on the data of the first cycle ($s = 0, a = 1$) only.

We further colored these points based on the categories defined in Attia et al. in Fig. S 21(b). It can be observed that the EoL prediction of the batteries in the category of CLO TOP 3 was with relatively higher error. We believe this is because, during the charging process of batteries of CLO TOP 3, the current curves had deep gaps between steps in the given charging policies (marked by green circles in Fig. S 22(a)), and the gaps were not severe (Fig. S 22(b)) in the other batteries marked by blue and green dots in Fig. S 21(b).

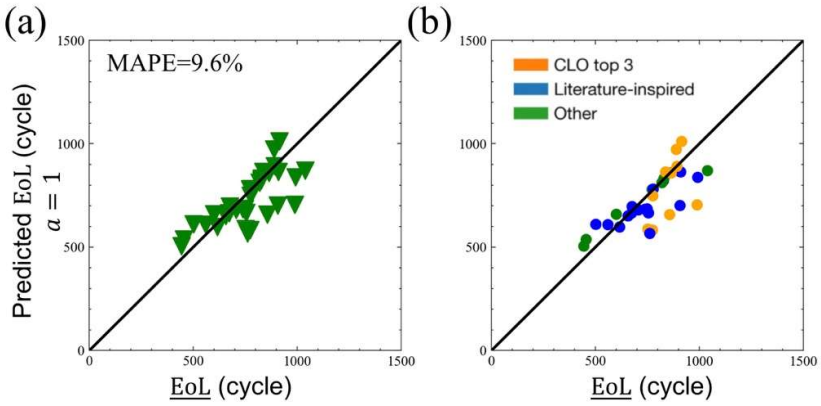


Fig. S 21: The predicted values EoL vs. EoL plot with MAPE of about 9.6%. Where

Full DNN predicted 34 batteries provided by Attia et al. based on the data of the first cycle only ($s = 0, a = 1$). (b) Data points are colored based on the charging protocol specified in Attia et al., where the EoL prediction of the batteries in the category of CLO TOP 3 was with relatively higher error.

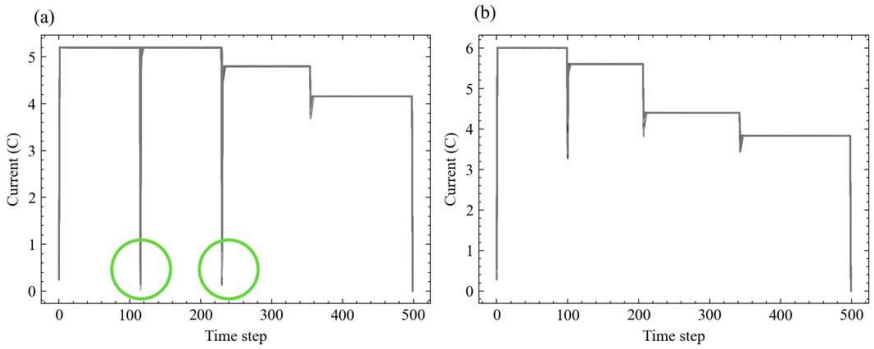


Fig. S 22: Typical examples of current curves corresponding to the charging policies (a) "CLO TOP3" and (b) "Literature inspired" and "Other" provided by Attia et al.

Attribution of the current DNNs

A backward propagation method, layer-wise relevance propagation (LRP), is used to analyze Full DNN and determine the relevance between EoL (the output layer) and all data points in the discharge curves of voltage in each cycle (the input layer). We took four batteries with the fourth and second shortest and longest EoL as examples without losing generality. Their EoL were 335, 429, 1284, and 1638 cycles. Relevance scores determined by LRP were plotted on the curves from the 1st to the 100th cycles, as shown in Fig. S 23. It can be observed that features which lead a short EoL for batteries (colored in red) can be discovered in many regions in the two bad batteries by our DNN. Similarly, features which lead a long EoL for batteries (colored in green) can also be discovered in the two good batteries. This indicates that our DNN can capture the key features to the prediction of EoL.

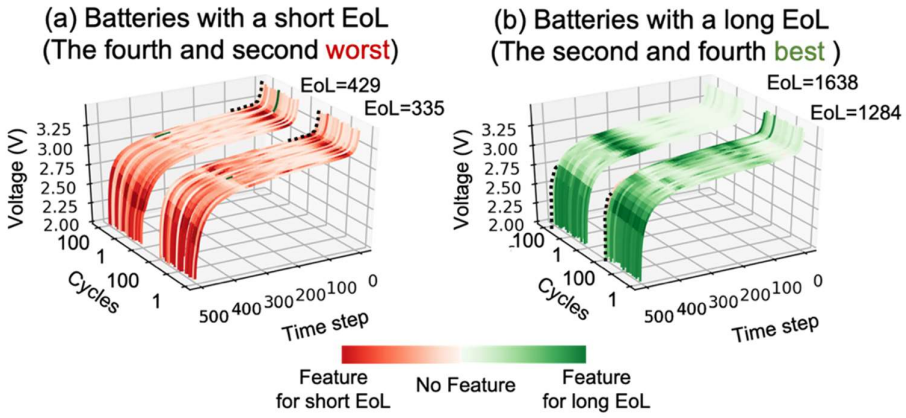


Fig. S 23: Relevance scores determined by LRP plotted on voltage curves in discharging process of batteries with a (a) short and (b) long EoL. Where green and red colors indicate features for leading to a long and short EoL, respectively, while white indicates no feature for EoL were found. Segments marked by dashed lines will be shown in Fig. S 24.

It is of interest to correlate between the value of voltage and EoL. Two segments of the voltage curves were chosen marked by dashed lines in Fig. S 23. Then, the voltage curves at that segment of all the batteries in the dataset and their corresponding EoL were plotted in Fig. S 24. It can be observed that the results show that the higher voltage at these segments has strong correlation with longer EoL[1]. This section demonstrated DNN and LRP can provide in-depth insight to design batteries or discharge conditions.

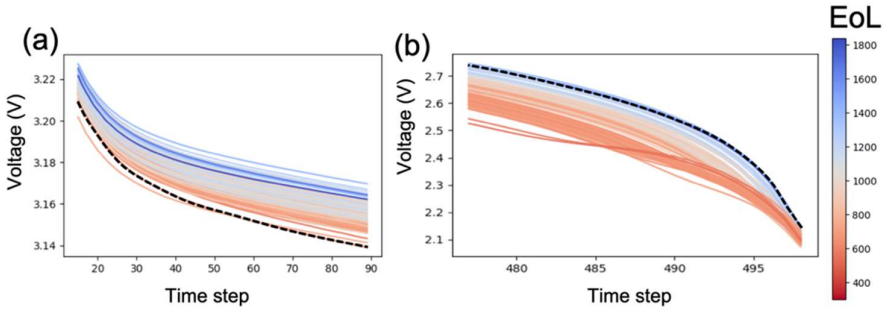


Fig. S 24: EoL and selected segments of voltage curves in discharging process for all batteries in the dataset at the 100th cycle, showing that higher voltage within these segments having longer EoL. Here, dash lines in (a)(b) are corresponding to those shown in Fig. S 23(a)(b), respectively.

Full range of RMSE RUL and *s* and the amount of sampling in training set

Testing RMSE diagrams for RUL and s in the full range by Full RUL DNN are shown in Fig. S 25. It can be observed that most of regions of RMSE of RUL were less than 40. It is also true for RMSE of s when $\underline{s} < 750$. However, RMSE of s dramatically increased around the 800th, 900th, and 1000th—1500th cycles. This is because the significant reduction of amount of sampling of testing strategy <3> corresponding to the batteries with $\mathrm{EoL} > 750$ cycles in training dataset. Fig. S 26 shows that the number of batteries can be used for the given value of $\underline{s}$. The amount of convolutional sampling throughout the entire dataset is in total was 53027, while that with $\underline{s} > 750$ was 9166, i.e. 17.3% of the entire dataset. Thus, insufficient information can be learned by Full RUL DNN.

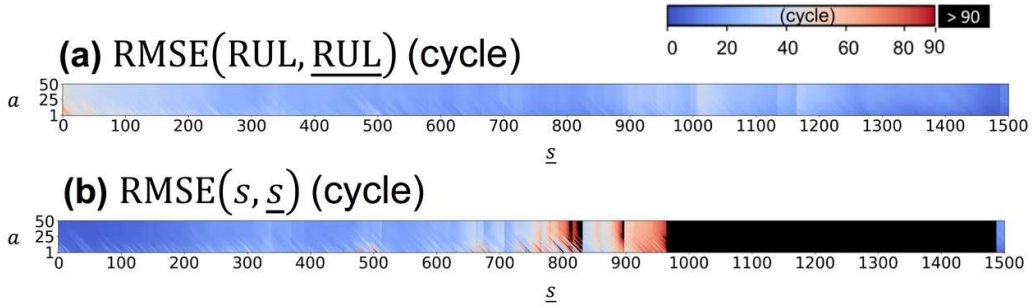


Fig. S 25: Testing RMSE of predicted (a) RUL and (b) *s*, which can be regarded as the current battery age, in full range from cycle 1--1500 by Full RUL DNN.

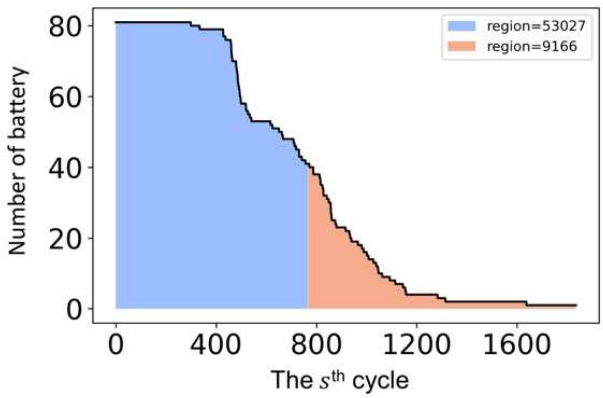


Fig. S 26: The number of batteries can be used for the given value of s, where the area represents the number of training sample.

The specification of the current DNNs

The specifications of the Discharge DNN, Full DNN and RUL DNN were shown in Table S 2…Table S 4, respectively. Where the DNNs with an input data length of 60, 80, and 100 cycles for testing purposes mentioned in Fig. S 18 were also listed. The numbers were estimated by TensorFlow (v2.1.0) and Keras (v2.3.1). It can be observed that the number of parameters in Discharge DNN were around a million for the prediction of single value properties and 4 million for time-series curves. The computation requirements, referred to as the number of floating-point operation (FLOPs), were less than 40 million. Feature selectors (i.e. Dimension reduction in the manuscript) in Full DNN required more the number of parameters (around 3 million) and the three "student networks" in DML required less than a million parameters. It can be concluded that the current DNNs are very efficient and effective. Our DNNs can be run in real-time in typical hand-held devices

Table S 2: The specification of the Discharge DNN.

Component	Objective	Parameters (M)	FLOPs (M)
Feature selector	EoL/RUL (Discharge)	1.3	4.1
	Charge time at EoL (Discharge)	0.7	6.3
Predictor	EoL & Charge time at EoL (60 cycle)	1.3	12.6
	EoL & Charge time at EoL (80 cycle)	1.3	25.2
	EoL & Charge time at EoL (100 cycle)	1.3	37.8
	$V(Q)_{n=\mathrm{EoL}}$	4.2	32.3

Table S 3: The specification of the Full DNN.

Component	Objective		Parameters(M)	FLOPs(M)
Feature selector	EoL (Discharge)		3.1	11.7
	Charge time at EoL (Discharge)		3.4	20.6
	EoL (Charge)		2.1	29.5
	Charge time at EoL (Charge)		0.8	32.9
Predictor	Integrated DML Predictor	Student 1	0.1	
		Student 2	0.2	
		Student 3	0.4	
		Total	0.8	6
	$V(Q)_{n=\mathrm{EoL}}$		1.3	21.2

Table S 4: The specification of the Full RUL DNN.

Component	Objective	Parameters(M)	FLOPs(M)
Feature selector	Discharging part (1&2&5)	5.9	33.5
	Charging part (3&4&6)	5.9	23.8
Predictor	RUL & s $(0 \leq \mathrm{s}, a \leq 50)$	2.7	37.9

The data of each cycle of the battery contains sufficient amount of information and the power of proposed last padding technique

In our work, the proposed DNNs determined and take the advantage of lower-fidelity version of the target, i.e. our data-driven features, to show higher sensitivity to the variation of the principal components/axes of the reduced latent feature space of APR18650M1A batteries than human-picked features. This reduces the complexity degree of space, and thus, we can extract high influence factors from one single cycle.

Here we proposed two statements to show the first cycle can provide sufficient information for the prediction of EoL. Firstly, we found that sufficiently high precision (resolution) of measurement can provide many degrees of freedom to describe battery status comprehensively. It allows the "battery genome" embedded in time-series properties in a single cycle. Secondly, our data augmentation (last padding technique) is the key approach to enable a great flexibility of a, and to let the same value of targets (such as EoL) being described by different lengths of meaningful input data during the

training process, leading to an efficient learning for the DNN. Each of the statements will be demonstrated by a simple toy testing model in the follows.

For the first statement, we adopted the model of natural language process to demonstrate how the battery behavior can be encoded in the curves with sufficiently high precision of measurement. The algorithm is shown in Algorithm R 1. Here the K-means[2] Clustering method is used to group each timestep k in standardized capacity, voltage, current, and temperature curves, $[C_k\ V_k\ I_k\ T_k]_{ij}$, of battery i during the j-th discharge half-cycle by the given number of cluster N into a sequence S_{ijk}. Where N serves as the number of available vocabularies of the current version of battery language, and can also be regarded as the level of precision (resolution) of the data measurement. Since there is no physical meaning between the clusters in S_{ijk}, we next applied positional encoding method[3][4], a typical nature language process (NLP) used with transformers to tokenize group ID for each time step based on its frequency and position of occurrence, giving S'_{ijk}. These tokenized sequences, i.e. "battery language," were then fed into ALBERT[5], which is the complex unsupervised neural network consisting of 223 million parameters trained by dataset of multi-languages.

Algorithm R 1: Testing model 1 to obtain the results in Fig. S 27(a).

Require: Battery $i = 1 \ldots n$; cycle number $j = 1 \ldots 100$; time step $k = 1 \ldots 500$.
Require: $[C_k\ V_k\ I_k\ T_k]_{ij}$: Standardized Capacity, Voltage, Current, Temperature.
Require: N: Vocabulary size.
$S_{ijk} \leftarrow \mathrm{K-mean\ clustering}\left([C_k\ V_k\ I_k\ T_k]_{ij}\right)$ with vocabulary size N.
$S'_{ijk} \leftarrow \mathrm{Positional\ Encoding}(S_{ijk})$
$Y_{i1k} \leftarrow \mathrm{ALBERT}(S'_{i1k})$
$Y_{i100k} \leftarrow \mathrm{ALBERT}(S'_{i100k})$
$D = \sum_{i=1}^{n} \sum_{k=1}^{500} |Y_{i1k} - Y_{i100k}|$
Return $\mathrm{Normalize}(D/n)$

Next, Manhattan distance is used to measure the difference between the activation values outputted by ALBERT based on different N corresponding to the first and the 100th cycles. **Error! Reference source not found.**(a) shows that, the resulting Manhattan distance (normalized with the reference of the case with N=100) increased with the vocabulary size N. This indicated that it is possible that the tokenized sequences of the two cycles possessing the features of aging effect of the considered battery can be successfully distinguished by ALBERT with sufficiently high level of precision of the data measurement.

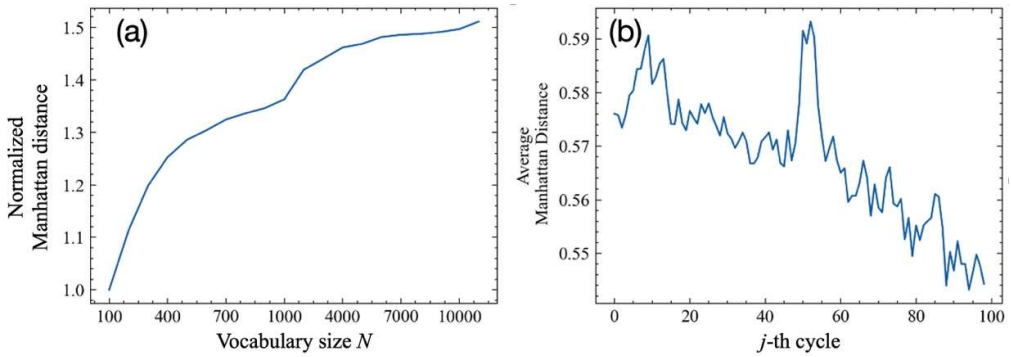


Fig. S 27: (a) Normalized Manhattan distance vs. vocabulary size *N*. (b) The averaged Manhattan distance throughout different tested number of vocabulary size *N* between the tokenized sequences of the *j*-th ($j = 1 \ldots 99$) and the 100th cycles.

Next, we illustrate the averaged Manhattan distance throughout different tested number of vocabulary size *N* between the tokenized sequences of the *j*-th ($j = 1 \ldots 99$) and the 100th cycles by

$$\langle D_j \rangle_N = \langle \frac{\sum_{i=1}^{n} \sum_{k=1}^{500} \left| Y_{ijk}^N - Y_{i100k}^N \right|}{n \times 500 \times 784} \rangle_N$$

where $\langle \cdot \rangle$ is the symbol of average. The results are shown in Fig. S 27(b). It can be observed that ALBERT seems to find a one-to-one mapping value corresponding to the distinct cycle number, and the relationship between each cycle can also be revealed. Thus, it implies that it is possible to accurately predict the future properties of batteries based on single cycle by a properly designed DNN. It is also interesting that there is an unusual peak around the 50th to 55th cycles, where anomaly temperature fluctuation occurred during the measurement as mentioned in Severson et al., showing that the mapping values can successfully describe the behavior of the batteries.

For the second statement, we will demonstrate that our last padding technique is the key approach to make the inference based on single cycle possible. In the literature, most of the training approaches is end-to-end, i.e. a set of specific feature correlate to a set of target value. Thus, in the case of battery EoL prediction, it becomes extremely difficult when only single cycle is considered. Different from the end-to-end approach, "multi-ends-to-end" is used in the current work, i.e. multiple sets of feature correlate to the same set of target value, where the multiple sets can be augmentable.

We trained the Discharge DNN with *a*=1, *a*=2, *a*=3, … *a*=100 all together for each selected battery. Where the cases of *a*<100 adopt the last-padding technique to repeat the information of the last cycle to fully fill the fixed-width of 100 data block. For example, the training input of the case of *a*=1 contains the features of cycle 1 repeated 100 times; the training input of the case of *a*=2 contains the features of cycle 1 and those of cycle 2 repeated 99 times. Note that these 100 input blocks correlate with the same EoL of that selected battery, i.e. multi-ends-to-end approach.

It is proved by Severson et al. that the input block of *a*=100 contains sufficient information to accurately predict EoL, and of course, this task is easier than it of *a*=99. Thus, input of *a*=100 can contribute more adequate gradients for modifying the weights of the DNN during the training process. Based on the guidance provided by the input of *a*=100, the DNN then become more experienced to find the correlation between the input of *a*=99 and the same EoL, and so on so forth. With this properly designed training strategy, the DNN can gradually learn how to predict EoL by the input of *a*=1.

The above-mentioned learning process is very similar to it of U-Net. During the training, the DNN temps to start its update of weights based on the input with the highest correlation to the targets, then based on those with lower correlation (i.e. the more difficult task with the less cycle information). Thus, when the DNN performed the inference for the difficult tasks, it can either be based on the parameters obtained from the learning process of the easier tasks, or be based on its own independent gateway to make the prediction for the target.

To compare the effect between the traditional end-to-end and our multi-ends-to-end approaches, we applied both approaches to replace the regression procedure done by gradient descent method on a problem of linear regression. We first generated 1000 lines in x, y space which are in the form of $y = wx + b$, where w and b are uniformly randomly chosen in the range $0 \leq w, b \leq 10$. Each of these lines was described by five points where $x = [1, 2, 3, 4, 5]$ and the corresponding y. Then, gradient descent method was applied to fit each line with the same initial guess of $w^0 = 0, b^0 = 0$ based on the five points. We recorded (w^j, b^j), where $j = 1 \ldots 10$, and (w^{10}, b^{10}) is expected to be the result closest to the answer (w^*, b^*). Fig. S 28 shows the error of w and b are dramatically altered between the iterations, showing that this task is not as easy as we expected if gradient descent method is used. After the data preparation, we adopted three models, (1) ANN with end-to-end approach, (2) ANN with our approach, (3) CNN with our approach, to learn how to predict (w^*, b^*) based on (w^1, b^1). Where the three models have similar complexity of network with the number of parameters around 10^5.

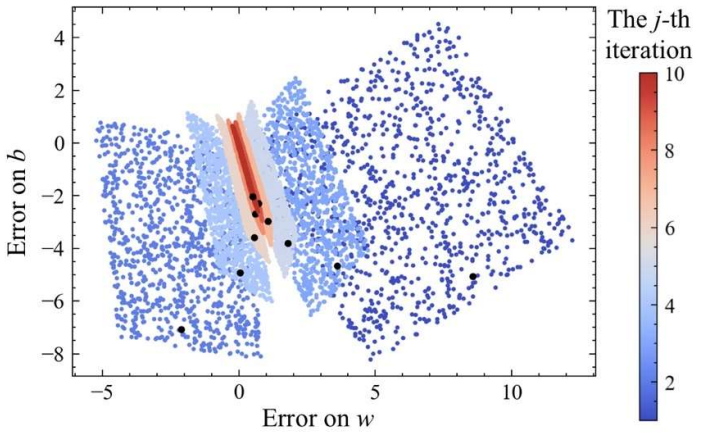


Fig. S 28: The error of w and b of 1000 lines in the j-th iteration. Where the black dots show the path of the convergence of one of the lines, showing the error dramatically altered between the iterations

The first model directly used all the sets of (w^1, b^1) and (w^*, b^*) as input and output to train the network. Where the structure of [2,128,128,2] unit of dense lays was used, where batch normalization[6] was applied to each layer. The second model adopted the structure of [20,128,128,2] and was trained by the proposed last padding technique. Where 20 are 10 for w and 10 for b. For example, if only the first step (w^1, b^1) is used, then they will be repeated for another 9 times to fill the input matrix; if nine steps are used, and then (w^9, b^9) will be repeated once. Note that although we used all the ten steps to train our second model, its inference of (w^*, b^*) during the testing is still based on the first step (w^1, b^1). The third model was very similar to the second one and adopted convolution layers (CNN). The performance of the three models on the testing sets is shown in Fig. S 29. It can be observed our approach, Model (2) and (3), enables excellent convergence on both the progress and results. Also, the relationship between each time step in the entire time series can be well captured by convolution layers in Model (3).

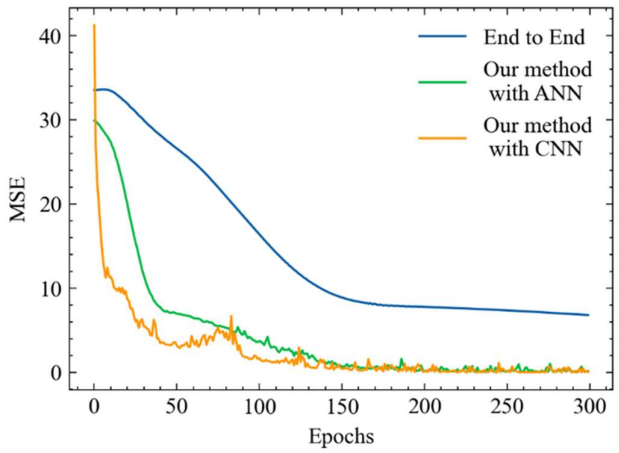


Fig. S 29: The performance of the three models on the testing sets. Note that performance of models evaluated by (w^1, b^1).

In order to demonstrate the behavior of learning, we plot the gradient summation contributed by the input matrix containing the information of $(w^1, b^1) \ldots (w^j, b^j)$ in each training batch, as shown in Fig. S 30. It can be observed that the input matrices containing more meaningful information (i.e. greater j) contributed more and those containing less meaningful information (i.e. less j) in the early stage of training (before around the 100th training batch). However, in the later stage of the training, the model has completely learned the features provided by the input matrices with greater j, and started to focus on the learning about the input matrices with less j. With the two demos mentioned above, we show that (1) the data of each cycle of the battery contains sufficient amount of information to be distinguished, provided that the level of precision/resolution of the measurement is sufficiently high; (2) the proposed last padding technique enables an efficient and effective learning and make the most of the dataset compared with the traditional end-to-end approach.

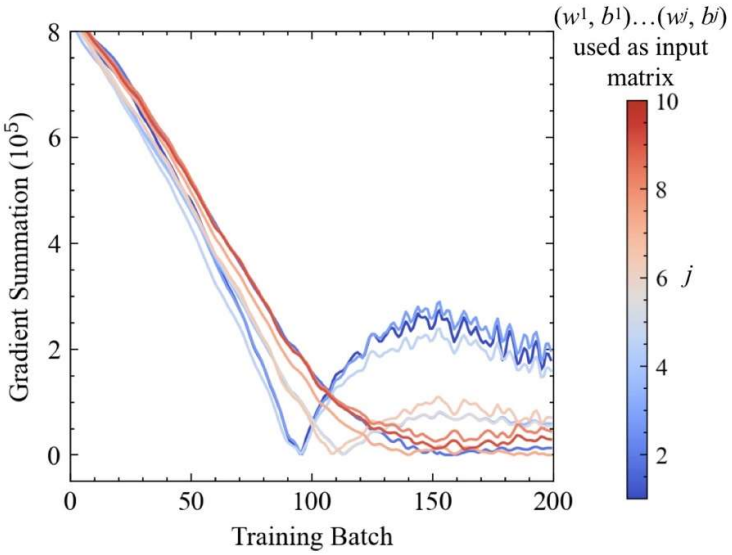


Fig. S 30: The gradient summation contributed by the j-th input matrix in each training batch.

Reference in Supplementary

[1] XiongR. Battery management algorithm for electric vehicles. 2019. https://doi.org/10.1007/978-981-15-0248-4.

[2] ArthurD, VassilvitskiiS. K-means++: The advantages of careful seeding. Proc. Annu. ACM-SIAM Symp. Discret. Algorithms, vol. 07-09-January-2007, 2007.

[3] VaswaniA, ShazeerN, ParmarN, UszkoreitJ, JonesL, GomezAN, et al. Attention is all you need. Adv. Neural Inf. Process. Syst., vol. 2017- December, 2017.

[4] DevlinJ, ChangMW, LeeK, ToutanovaK. BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf., vol. 1, 2019.

[5] LanZ, ChenM, GoodmanS, GimpelK, SharmaP, SoricutR. Albert: A lite bert for self-supervised learning of language representations. ArXiv 2019.

[6] IoffeS, SzegedyC. Batch normalization: Accelerating deep network training by reducing internal covariate shift. 32nd Int. Conf. Mach. Learn. ICML 2015, vol. 1, 2015.