Novel informed deep learning-based prognostics framework for on-board health monitoring of lithium-ion batteries

Article *in* Applied Energy · June 2022

DOI: 10.1016/j.apenergy.2022.119011

CITATIONS
0

READS
37

3 authors, including:

Sung Wook Kim
Pohang University of Science and Technology
6 PUBLICATIONS **21** CITATIONS

SEE PROFILE

Ki-Yong Oh
Chung-Ang University
72 PUBLICATIONS **1,208** CITATIONS

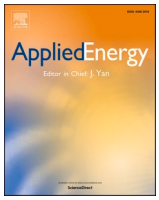
SEE PROFILE

Applied Energy 315 (2022) 119011

Contents lists available at ScienceDirect

Applied Energy

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



Novel informed deep learning-based prognostics framework for on-board health monitoring of lithium-ion batteries

Sung Wook Kim [a], Ki-Yong Oh [b,*], Seungchul Lee [a,c,d,*]

[a] Department of Mechanical Engineering, Pohang University of Science and Technology, 77 Cheongam-ro, Pohang, Republic of Korea
[b] Department of Mechanical Convergence Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul, Republic of Korea
[c] Graduate School of Artificial Intelligence, Pohang University of Science and Technology, 77 Cheongam-ro, Pohang, Republic of Korea
[d] Institute of Convergence Research and Education in Advanced Technology, Yonsei University, 50 Yonsei-ro, Seoul, Republic of Korea

HIGHLIGHTS

- Informed deep learning-based framework is proposed for on-board battery health monitoring.
- Impedance-related features model battery degradation efficiently.
- Layer-wise relevance propagation reveals impedance features' contribution to output.
- Knowledge infusion to a recurrent neural network improves estimation accuracy.
- Monte Carlo dropout secures the model reliability by providing uncertainty measures.

ARTICLE INFO

Keywords:
Lithium-ion battery
On-board health monitoring
State of health estimation
Remaining useful life estimation
Informed deep learning

ABSTRACT

This paper proposes a novel, informed deep-learning-based prognostics framework for on-board state of health and remaining useful life estimations of lithium-ion batteries, which are critical components for strategizing energy and power used in electric vehicles. The framework comprises three phases. First, reliable and online accessible impedance-related features are collected from discharge curves. Second, these features are inputted into the proposed knowledge-infused recurrent neural network, a hybrid model that combines an empirical model with a deep neural network. Third, Monte Carlo dropout, a deep learning method for obtaining a probabilistic prediction of a neural network, is addressed to secure robustness in estimating the state of health and remaining useful life. Layer-wise relevance propagation, a deep learning technique for tracking the evolution of feature importance and offering scientific reasoning of the output, confirms that impedance-related features significantly contribute to the estimation accuracy compared to other features investigated in previous studies. Moreover, the hybrid model improves the estimation accuracy and robustness, whereas Monte Carlo dropout ensures robustness and reliability. Specifically, the estimation results for the public degradation data reveal that the proposed model can output significantly more accurate state of health and remaining useful life estimations than the baseline deep neural networks. The findings of this study provide insight into the explicable and uncertainty-based pipeline of deep neural networks with respect to battery health monitoring, which are highly recommendable features for decision-making and corrective planning of power and energy used in lithium-ion battery cells and packs.

1. Introduction

Lithium-ion batteries have become the most common reversible energy source for electric vehicles, which are increasingly being used on the roads, owing to their excellent characteristics, such as high energy, high power density, low discharge rate, and long service life. Electric vehicles employ several battery packs of which multiple lithium-ion battery cells are stacked inside the casings with the battery management system (BMS). The primary goal of the BMS is the intelligent operation of batteries within pre-defined safe limits by the real-time monitoring of parameters such as the state of health (SOH) and state

* Corresponding authors.
E-mail addresses: kswltd@postech.ac.kr (S.W. Kim), kiyongoh@hanyang.ac.kr (K.-Y. Oh), seunglee@postech.ac.kr (S. Lee).

https://doi.org/10.1016/j.apenergy.2022.119011
Received 19 July 2021; Received in revised form 10 March 2022; Accepted 22 March 2022
Available online 14 April 2022

Nomenclature

Abbreviation

BNN	Bayesian neural network
CCCV	Constant current constant voltage
CDRNN	Cumulative damage recurrent neural network
CDRNN-MC	CDRNN-Monte Carlo
CNN	Convolutional neural network
CT	Charge time
DCNN	Deep convolutional neural network
DNN	Deep neural network
DVS@CC	Discharge voltage slope at a constant current
ECM	Equivalent circuit model
ELBO	Evidence lower bound
FT	Fine-tuning
GPR	Gaussian process regression
HI	Health indicator
IR	Internal resistance
KIRNN	Knowledge-infused recurrent neural network
KIRNN-MC	KIRNN-Monte Carlo
LAM	Loss of active material
LLI	Loss of lithium inventory
LRP	Layer-wise relevance propagation
LSTM	Long short-term memory
MaxE	Maximum error
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MC	Monte Carlo
MLP	Multi-layer perceptron
ODE	Ordinary differential equation
PDE	Partial differential equation
P2D	Pseudo-two-dimensional model
RMSE	Root mean squared error
RMSPE	Root mean squared percentage error
RNN	Recurrent neural network
RUL	Remaining useful life
RVM	Relevance vector machine
SEI	Solid electrolyte interphase
SOC	State of charge
SOH	State of health
SPM	Single-particle model
SVM	Support vector machine
Tmax	Maximum temperature

Symbol

a_t	Accumulated damage at time t
N	Cycle
α	Unknown coefficients of the double exponential model
β	Unknown coefficients of the double exponential model
γ	Unknown coefficients of the double exponential model
λ	Unknown coefficients of the double exponential model
Q_N	Accumulated damage at cycle N, Ah
dQ/dN	Damage increment at cycle N, Ah/cycle
dQ_{KNO}	Damage output by knowledge-infused block, Ah
dQ_{DNN}	Damage output by data-driven block, Ah
ψ	Nonlinear activation function
$W^{(j)}$	Weight of the j th hidden layer
$z^{(j)}$	Bias of the j th hidden layer
M_K	Linear weighting term of dQ_{KNO}
M_D	Linear weighting term of dQ_{DNN}
x_*	Test sample
y_*	Test label
X	Training sample
Y	Training label
w	Model weight
f	Model output
P	Number of stochastic forward passes
p	Probability distribution
μ	Mean of model output
σ	Standard deviation of model output
A	Accelerated dataset
W	Weakly-supervised dataset
T	Fully supervised dataset for testing
x	Input hyper-parameter
x^*	Optimized parameter
Λ	Pre-defined bounded region
y_i^*	Estimated value at the i th test sample
y_i	Target value at i th test sample
$p(w\|X,Y)$	Posterior distribution
$q(w)$	Approximate posterior distribution
η	Weight factor of L2 regularization

of charge (SOC). Precise SOH monitoring is important because the SOH indicates the current battery state, which further indicates the number of charge–discharge cycles remaining before reaching the end-of-life, namely, the remaining useful life (RUL). On-board SOH and RUL estimations have been conventionally performed using three types of methods [1]: model-based, data-driven, and hybrid.

Model-based methods include the electrochemical and empirical models and the equivalent circuit model (ECM). The electrochemical model refers to a single-particle model (SPM) [2,3] or a pseudo-two-dimensional model (P2D) [3,4], both of which account for the complex nonlinear degradation mechanism of batteries over repeated cycles with respect to time. The electrochemical models, particularly P2Ds, represent the internal states of batteries and model the transport and diffusion of charges and ions [4]. Such a model is described by five different governing equations (refer to Appendix A) for charge and mass conservation in solid particles, charge and mass conservation in an electrolyte, and movement of lithium-ion between a solid particle and electrolyte [5]. However, these equations are mostly complex partial differential equations that require to be solved with a high computational cost. Thus, P2D is not a suitable candidate for on-board electric vehicle applications [6]. In contrast, ECM [7,8] is a simplified model aimed at describing the physical and chemical reactions of batteries through circuit analysis to reduce the computational load at the expense of several complex principles. Despite the limitations, it considers the aging mechanism to an extent, including the internal resistance growth. Although ECM is one of the most common models for the online estimation of the SOC because of its reduced computational cost, it is not sufficiently accurate in estimations across a wide range of operating conditions encountered in actual scenarios. Moreover, the lack of physics-based information related to the system states increases the difficulty of estimating the future states, e.g., RUL [6]. The empirical model is a battery degradation model that outputs the SOH based on given input parameters such as the temperature, charging time, SOC, depth of discharge, and the number of cycles. The SOH is generally determined theoretically and heuristically after numerous experiments, and several common models of this type include the exponential [9–11], logarithmic [12], and polynomial [13] models. Empirical models are convenient owing to their simplicity. However, their estimation accuracy using Kalman filter variants [14,15] and particle filter variants [16–18] is highly dependent on the fidelity of the model and does not provide significant physical insight into the battery state [19]. In general, accurate SOH estimations based on the aforementioned model-

based approaches require a high computational cost, and there is no physics-based model for RUL estimations.

Recently, data-driven models have attracted significant attention as an alternative method, as they do not require prior knowledge of batteries, and data collection using these models is significantly more feasible across the industries as compared to model-based approaches. The major advantage of data-driven models is that they allow for the incorporation of historical data and additional domain knowledge (e.g., materials information) into a single model, which has not been considered previously by model-based methods. Thus, the ability for battery state estimation is enhanced [6]. Furthermore, they are suitable for the analysis of a wide range of degradation trends, in addition to operating conditions under which uncommon incidents occur. Data-driven methods include statistical, machine learning, and deep learning methods. In several studies, statistical models such as autoregressive integrated moving average [20,21], the Grey model [22], the Wiener process [23], and entropy analysis [24] have been employed. However, statistical methods are generally not preferable for on-board SOH and RUL estimations because a large number of the models are inappropriate for nonlinear signals and are highly sensitive to the quantity and quality of data. Therefore, they do not demonstrate robustness for long-term predictions. To date, machine learning is the most recognized data-driven method that has a significant focus on the Gaussian process regression (GPR) [25–29], relevance vector machine (RVM) [30–32], support vector machine (SVM) [33–35], and decision tree [36,37]. Despite their widespread use, these machine learning tools are subject to the following drawbacks [1]: high computational cost with large datasets (e.g., GPR and RVM), lack of sparseness (e.g., GPR and SVM), and lack of stability (e.g., RVM). In general, machine learning tools, regardless of the use of parametric or non-parametric models, are ineffective in the case of large datasets, as they require re-training using the entire datasets upon the addition of newly observed data. Moreover, most of the previous studies regarding machine learning-based RUL estimations reported satisfactory performances only under restricted environmental conditions, such as a complete cycle under a constant current, which is not representative of real-world scenarios [38]. Moreover, they require well-handcrafted health indicators as input features, which may vary across different types of batteries, thus requiring expert knowledge of the battery system.

The deep learning approach is more versatile because its model weights can be further updated through fine-tuning with respect to newly observed data, and it extracts meaningful features automatically. Furthermore, it can process big data that reflects numerous operating conditions, thus leading to an improved generalization performance. Due to its versatility, the number of studies that implement deep learning for SOH and RUL estimations is steadily increasing. For example, You et al. [38] implemented an artificial neural network that considered only the raw signals, i.e., voltage, current, and temperature, to estimate the SOH. However, there is a limit to its applications with respect to RUL estimations, as the model cannot effectively capture the temporal characteristics. Conversely, Liu et al. [39] developed an adaptive recurrent neural network to account for the temporal characteristics and the dynamic state during battery degradation. Although the model can adapt to the abrupt state changes as expected, it is not suitable for long-term predictions due to the inherent gradient vanishing problem of RNNs [40]. To address the problem, Zhang et al. [41] presented a detailed discussion on the implementation of long short-term memory (LSTM), which is an RNN variant, for on-board SOH and RUL estimations. The SOH was accurately estimated with or without the offline training data with an acceptable training time. Moreover, a Monte Carlo (MC) simulation was implemented to obtain the probability density function of the estimations, revealing that deep neural networks (DNNs) can provide probabilistic predictions similar to several other data-driven models. Similarly, Shen et al. [19] utilized a deep convolutional neural network (DCNN) for online capacity estimations. The proposed model yielded superior RMSE and MaxE values to those of the RVM for the simulation of the actual case wherein an incomplete discharge process occurred before the subsequent charging cycle. Further recent studies that leverage the power of deep learning algorithms such as DNN [42], CNN [43], LSTM [44,45] prove their promising results in either SOH or RUL estimation. However, regardless of the significant development of such deep learning-based prognostics systems, in most previous studies, the system performances were only validated using a small number of test data collected under limited experimental conditions. Therefore, further investigation on using deep learning models for a much larger dataset is necessary. Moreover, all the essential aspects of the on-board battery prognostics, including probabilistic predictions and training time, were not investigated. With reference to the available literature, the "black-box" models have not been previously investigated and detailed. In parallel with the development of such data-driven models, hybrid models that combine a data-driven approach and model-based approach compose a distinct field of research, as the deficiencies of both model-types are minimized by their integration. Most previous studies were focused on the integration of machine learning with filtering methods. Thus, the potential of the deep learning approach was not considered [1].

In this study, a novel deep learning-based prognostics framework for SOH and RUL estimations is developed with an empirical model-based DNN, namely, a knowledge-infused recurrent neural network with Monte Carlo dropout (KIRNN-MC). The objective of the proposed framework is a high SOH and RUL estimation accuracy with few handcrafted features for on-board applications. The specific contributions of this study are as follows:

1. With reference to the available literature, this is the first study to demonstrate a hybrid methodology wherein an empirical model is leveraged to improve the generalization performance of a DNN. A hybrid model (KIRNN), which combines an empirical model and a DNN improves the mean absolute percentage error (MAPE) and the root mean squared percentage error (RMSPE) by as much as 0.21% and 0.2%, respectively.
2. KIRNN-MC, which applies MC dropout to KIRNN provides a probabilistic prediction for decision-making or corrective action planning and ensures the robustness and reliability of the estimation performance, improving the MAPE and RMSPE by as much as 0.62% and 1.06% compared with KIRNN.
3. Knowledge infusion to a recurrent neural network and MC dropout together secures improved predictive performance and model robustness.
4. Reliable and online accessible impedance-related features are extracted as distinct features for the efficient modeling of the battery capacity. Layer-wise relevance propagation (LRP) is addressed to facilitate the understanding of the feature contribution to modeling for further scientific discovery. This allows for the quantitative interpretation of the deep learning mechanisms in the decision-making of battery health estimations, which was previously challenging due to its "black-box" nature.

The remainder of this paper is organized as follows. Section 2 describes the overall methodology for on-board battery health monitoring, and Section 3 presents the experimental settings and battery degradation datasets used in this study. Section 4 presents a discussion on the results for different scenarios and combinations of methods and finally, Section 5 summarizes the findings of this study.

2. Methodology

The proposed deep learning-based prognostics framework for on-board SOH estimations is illustrated in Fig. 1. The proposed framework comprises three phases. In Phase A, impedance-related features are extracted to provide sufficient features for model training. The proposed framework then determines whether transfer learning is applied based

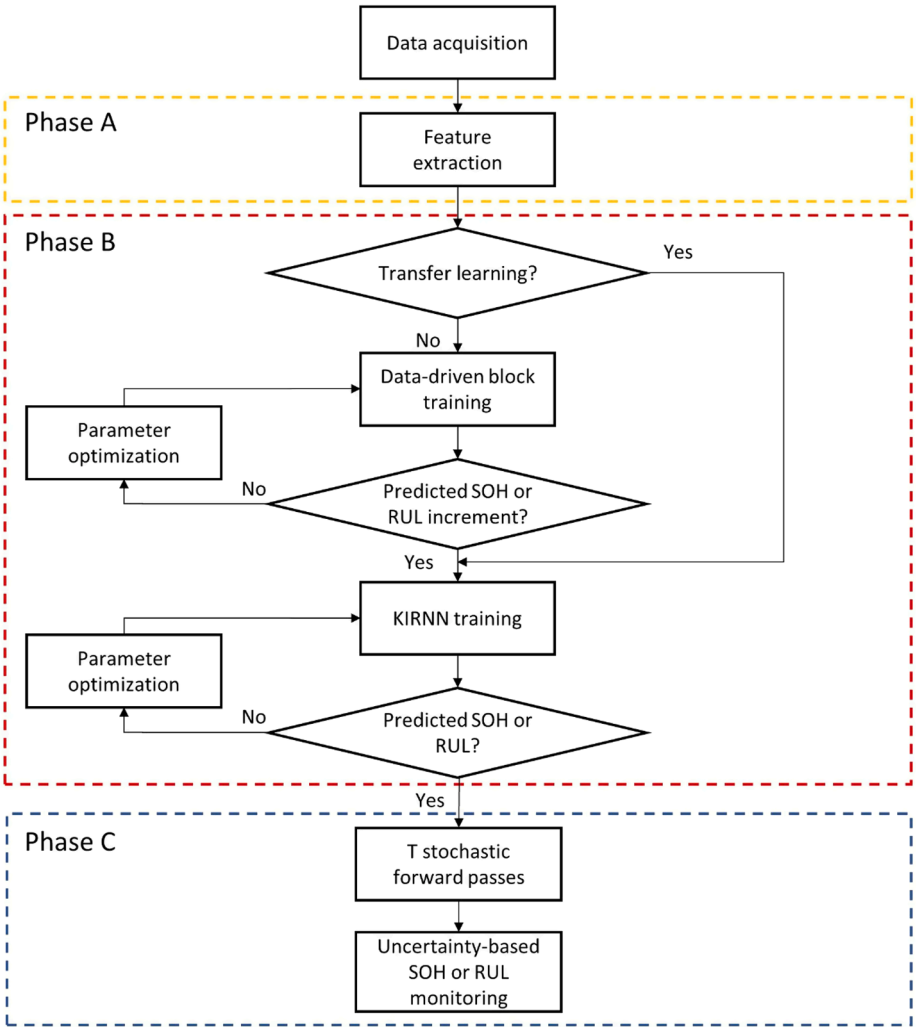


Fig. 1. Flowchart of the proposed framework for on-board SOH and RUL estimations.

on the presence of a pre-trained model in Phase B. In the case wherein a pre-trained model is not present, i.e., there are no initial weights, the data-driven block inside KIRNN should be trained first. The entire model is then trained until an error in a loss function reaches a predefined threshold. It should be noted that the data-driven block is a multi-layer neural network, which should be optimized before training the entire recurrent model to ensure rapid convergence. If there is a pre-trained model, the training of the data-driven block is not conducted, and the KIRNN model is updated. In such a case, the pre-determined trainable parameters are fine-tuned with the newly measured data, which are referred to as online data in the literature [38,41]. The trainable parameters include the four coefficients of the empirical model, the weights and biases from the DNN, and the parameters from the weighted linear combination of the two blocks. In Phase C, the uncertainty-based SOH can be estimated and monitored using stochastic forward passes. The following subsections detail the methods used in each phase.

2.1. Phase A: Impedance-related feature extraction and data processing

Lithium-ion battery degradation is caused mainly by the loss of lithium inventory (LLI), loss of active material (LAM), and conductivity loss [46,47]. The LLI is due to the consumption of available lithium-ions during the solid electrolyte interphase (SEI) film formation and decomposition, electrolyte decomposition, and lithium plating, whereas the LAM can be mainly attributed to graphite exfoliation, electrolyte oxidation, binder decomposition, and crystal structure disorder, thus resulting in a storage capacity loss. The conductivity loss intensifies when the SEI formation consumes electrolyte solvents, lithium plating causes pore-clogging, and particle cracking disrupts the electrical contact between the active particles and the current collector. These complex and diverse aging mechanisms increase the equivalent impedance of a cell during degradation [48], thus implying that the equivalent impedance is highly correlated with degradation and can therefore be considered an effective and sensitive feature for estimating the SOH.

The physical insights into the aging mechanisms allow for impedance-related features to be extracted from the charge or discharge curves. Specifically, ten health indicators (HI1–HI10) were extracted from the discharge curve (the red circle in Fig. 2) at SOCs of 10–100% with intervals of 10%. Hereafter, these HIs are denoted as discharge voltage slopes at a constant current (DVS@CC). These slopes at different SOCs represent impedances, given that $R = \Delta V/\Delta I \cong \Delta V/\Delta t$ in the constant current mode (refer to Appendix B). Because the slopes differ depending on the measurement period, the period length should be set in consideration of the trade-off between practicality and distinction as a good health indicator. The selected time period (Δt) was set as 160 s for all HIs [48]. It should be noted that the HIs were only extracted from the discharge curve in this study, given that the charging protocol was changed several times in the public dataset for accelerated degradation experiments, thus hindering the extraction of unified HIs. Different charge-discharge protocols result in different overpotentials. Hence, different impedances may be extracted for a battery of the same SOH, thus suggesting that the charge-discharge protocol plays a critical role in

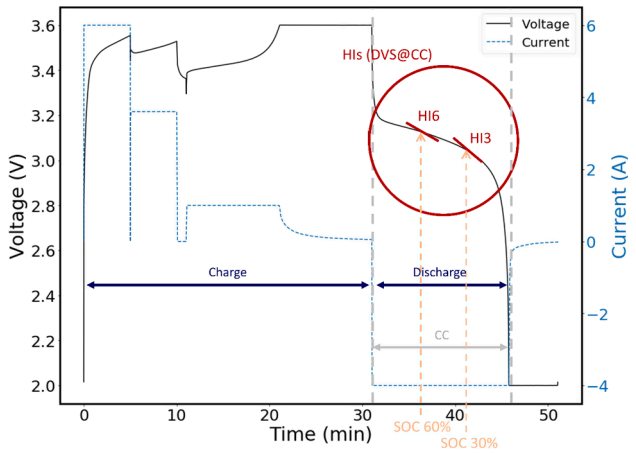


Fig. 2. Graphical illustration of impedance-related feature extraction.

the extraction of reliable features. Moreover, impedance depends on the temperature of a cell, which implies that different C-rates in several charge curves result in different cell temperatures. Hence, extracting impedances at charge curves is ineffective. Leveraging charge curves may be more practical in actual applications because electric vehicles can be charged using the same charge protocol in a garage or a charging station overnight. In particular, in this study, HIs were extracted from the discharge curve due to the limited data availability. Future work includes applying the proposed method to a charge curve, which has the same protocol during degradation experiments.

Before training a DNN, a primary step is to preprocess the HIs. Specifically, the noise reduction of measurements is a common process to ensure the high accuracy of a DNN model. For the public dataset (described in Section 3), raw signals and their extracted features are passed through a Gaussian filter [49] to remove excessive noise. The Gaussian filter works by creating a Gaussian distribution-shaped kernel, and the kernel convolution gives the largest weight to the center point and smaller weights to the nearby points as they get farther away from the center. The only parameter to be determined before its usage is the sigma. This simple parameter selection would be a reason to address this method in this study. The larger the sigma value, the more noise is reduced. The sigma for the Gaussian distribution was set to 4.5 for all data, and the effect of noise reduction is shown in Figure S1 of the Supplementary Material. Their correlations with the discharge capacity or SOH are then estimated to demonstrate the effectiveness of this process. Table 1 clearly reveals that this process increases the correlation between each HI and degradation, thus increasing the estimation accuracy and ensuring robustness.

Three combinations of feature groups are organized for model training because high linear correlations do not ensure an improved SOH estimation, considering the nonlinear degradation phenomenon. Group I consists of the top three relevant HIs (HI8, HI7, and HI6). Note that only the top three HIs are selected for rapid data processing to ensure efficient real-time data collection. Group II consists of the top three relevant features from the given features, including the charge time (CT), maximum temperature (Tmax), and internal resistance (IR). This group is used because its features are effective for estimating the SOH, as suggested in the literature [50]. Group III is a combination of both Group I and II, which comprises the top three correlated feature groups from the extracted features and the given features. The corresponding results are presented in Section 4.

2.2. Phase B: Knowledge-infused recurrent neural network (KIRNN)

In this paper, the KIRNN is proposed to model the capacity degradation over charge–discharge cycles. The recurrent part of the model is in the form of a cumulative damage model [51] that reflects the accumulation of damage over time. The model can be expressed mathematically as $a_t = a_{t-1} + \Delta a_t$, where a_t denotes the accumulated damage at time t. Such a model is applicable to SOH estimations because the parameter can generally be defined as monotonically decreasing despite the small capacity-increasing phenomena of lithium-ion batteries, which can be attributed to the extension of the charge stored in the negative electrode beyond the positive electrode [50,52,53]. Given that the model accounts for the damage increment after each time unit t, it can be modeled using RNNs for the estimation of the battery capacity degradation after each cycle N. This further allows for the solution of empirical models that may be expressed as ordinary differential equations (ODEs) or partial differential equations (PDEs) as a function of the cycle N, which are frequently encountered in engineering problems [54–56].

The inputting of physics information or prior knowledge into a DNN has received significant attention recently because it increases the prediction accuracy and ensures robustness for new test datasets [57–62]. For example, a physics-informed layer modeling an ODE with a data-driven layer can successfully account for the bias term, thus improving the overall prognosis performance [63]. Based on the input of physics information, an empirical model, i.e., a type of prior knowledge, informs the recurrent unit to estimate the capacity degradation, as realized in this study. In particular, an empirical model is fused to a data-driven model to form a hybrid model under the hypothesis that a generalized empirical model may help improve the prediction accuracy of a purely data-driven model. A double-exponential model [64] was employed as the empirical model in this study, which accurately estimated several battery capacity degradation data under different operational conditions [9,11,31,64].

$$Q(N) = \alpha exp(\beta N) + \gamma exp(\lambda N), \tag{1}$$

where Q is the battery capacity over cycle N, and α, β, γ, and λ denote unknown coefficients, which should be determined through training. The Levenberg–Marquardt algorithm [65], which is a nonlinear curve fitting method, is implemented to set initialized coefficients for training the proposed hybrid model as shown in Figure S2 and Table S1 of the Supplementary Material. Hereinafter, the empirical model inside the recurrent unit is referred to as the knowledge-infused block.

Fig. 3 illustrates the details of the proposed neural network, in which knowledge-infused blocks and a data-driven block were combined to form a hybrid recurrent unit for the computation of the damage increment, i.e., the capacity degradation for each charge–discharge cycle with respect to an RNN. Fig. 3 (a) presents the process for pre-training a model, whereas Fig. 3 (b) demonstrates the extension through which

Table 1
Correlation coefficients before and after noise reduction.

Extracted features	HI10	HI9	HI8	HI7	HI6	HI5	HI4	HI3	HI2	HI1
Before	0.38	0.54	0.89	0.80	0.70	0.58	0.46	0.13	−0.02	−0.07
After	0.79	0.88	0.97	0.96	0.95	0.93	0.88	0.29	−0.02	−0.08
Given features	CT	Tavg	Tmin	Tmax	IR					
Before	−0.66	−0.43	−0.09	−0.46	−0.89					
After	−0.73	−0.49	−0.16	−0.50	−0.94					

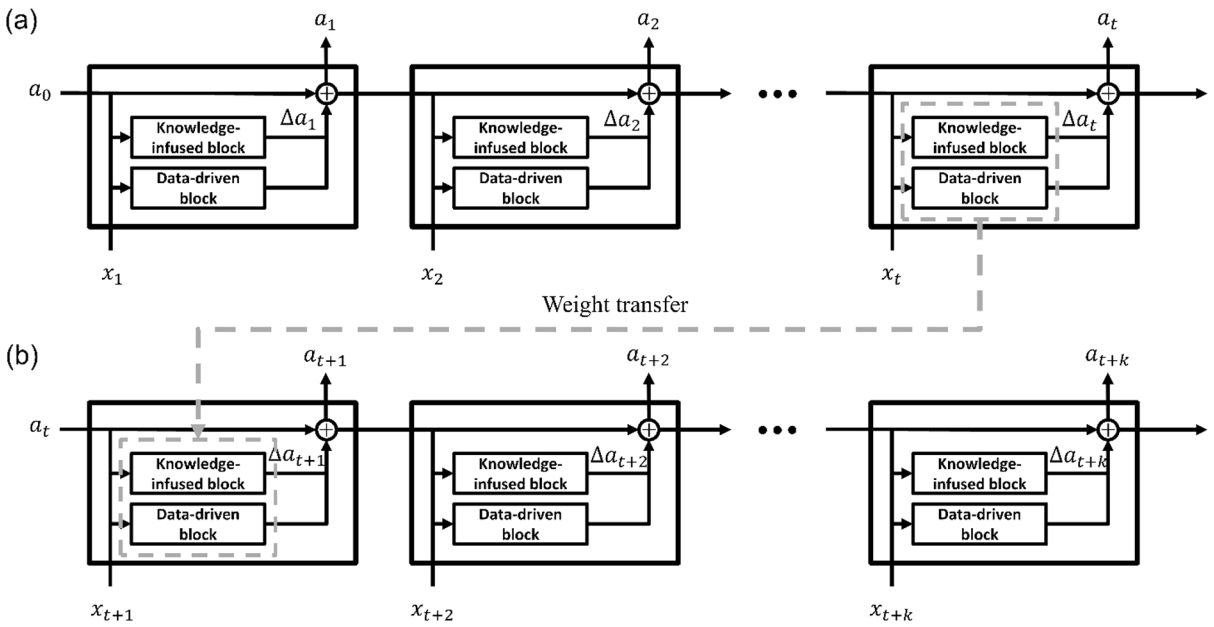


Fig. 3. (a) Illustration of cumulative damage modeling over time using the hybrid recurrent unit and (b) a transfer learning scheme where trainable parameters are transferred for fine-tuning online data.

pre-trained weights from (a) are transferred for updating and fine-tuning the online data of size k.

The specific components of the hybrid recurrent unit are illustrated in Fig. 4. Specifically, the green boxes constitute the knowledge-infused blocks, whereas the red box and the blue box correspond to the data-driven block and the linear combination block, respectively. The data-driven block includes dropout layers, which is described in the next subsection in detail. The knowledge-infused blocks are designed based on Equation (2a), which is a derivative of Equation (1), and its output is denoted as dQ/dN, where N is the cycle number. The data-driven block is a DNN with an output of dQ_{DNN}, as expressed by Equation (2b), where ψ, $W^{(3)}$, and $z^{(3)}$ correspond to the activation function, weights in the third layer, and the given input at the third layer, respectively. The specific implementation of Group I is exemplified in Fig. 4, where the inputs to the data-driven block are the cycle number, HI8, HI7, and HI6. The linear combination block was positioned such that the outputs from the knowledge-infused blocks and the data-driven block are weighted respectively by M_K and M_D, in accordance with Equation (2c), to calculate the final damage increment, ΔQ_N. Finally, ΔQ_N is added to Q_{N-1} to obtain the accumulated damage Q_N for the period of cycle N (Equation (2d)).

$$dQ/dN = dQ_{KNO} = \alpha\beta exp(\beta N) + \gamma\lambda exp(\lambda N) \tag{2a}$$

$$dQ_{DNN} = \psi(W^{(3)} z^{(3)}) \tag{2b}$$

$$\Delta Q_N = M_K(dQ_{KNO}) + M_D(dQ_{DNN}) \tag{2c}$$

$$Q_N = Q_{N-1} + \Delta Q_N \tag{2d}$$

2.3. Phase C: Monte Carlo dropout

For several applications where a misclassified decision leads to fatal incidents, e.g., self-driving, medical judgments, and manufacturing, it is preferable that the system indicates a high level of uncertainty in the case of unknown test data rather than executing a random decision. This allows for the final decision to be handled by the domain experts. However, a large number of deep learning models are "deterministic," thus providing a mere point estimate per test sample when making decisions, e.g., a regression problem. Although Bayesian neural networks

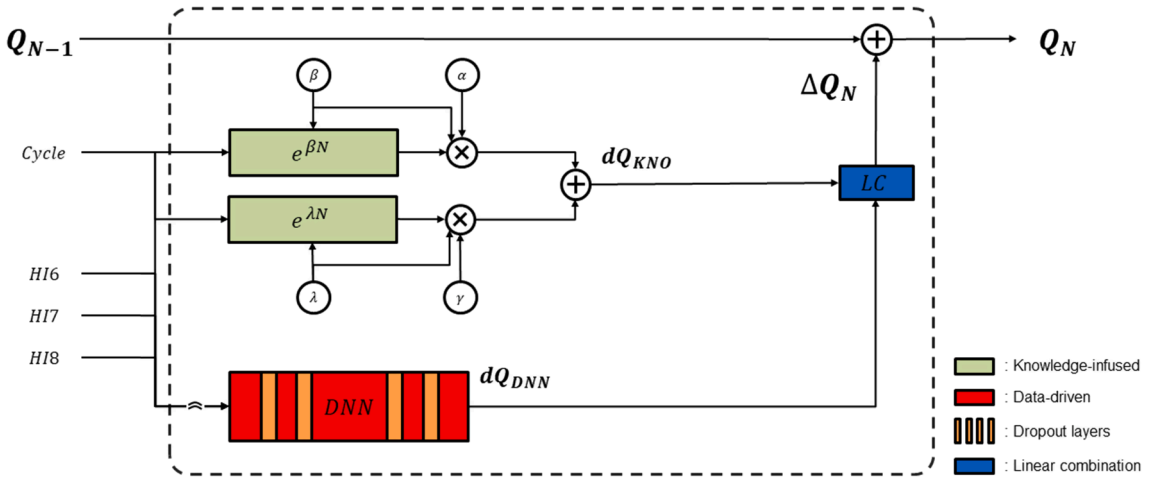


Fig. 4. Schematic diagram of the proposed hybrid recurrent unit.

(BNNs) solve this problem by considering model parameters as a distribution or density rather than a point, their accuracy may be insufficient. This is because the integration for predictive inference is typically intractable and therefore replaced by an approximated value by minimizing the Kullbeck–Leibler divergence [66–69]. Moreover, the training process requires excessive computational time.

To address this issue, Gal et al. [70] proposed the MC dropout technique, which demonstrated an improved uncertainty estimation when compared with BNNs. Moreover, its implementation is significantly simpler with a low computational cost. The underlying principle of the MC dropout technique is as follows. The traditional dropout scheme, which was previously applied to DNNs during training for regularization, has the same effect as that of BNNs if it is combined with an MC simulation in the test phase. In particular, applying dropout for training generates an ensemble of the model, from which the output distribution is obtained after the test data is evaluated for P stochastic times. The following equation is its simplified mathematical formulation of predictive inference for obtaining the probability distribution of test label y_*, given test sample x_*, training samples X, and its corresponding labels Y.

$$p(y_*|x_*,X,Y) = \int p(y_*|x_*,w)p(w|x_*,X,Y)dw \approx \frac{1}{P}\sum_{n=1}^{P} p(y_*|x_*,w_n) = \frac{1}{P}\sum_{n=1}^{P} f(x_*,w_n), \tag{3}$$

where w and $f(x_*,w_n)$ denote the model weights and the output of a neural network. Fig. 5 depicts the stochasticity of each forward pass, which is achieved by randomly nullifying the given number of neurons, i.e., red circles, in each layer. The second line of Equation (83) is a direct result of an MC integration to obtain an average result for P stochastic forward passes. The standard deviation of the results corresponds to the predictive uncertainty. A detailed explanation of the equivalence between the application of the MC dropout and the solution of BNNs using a variational inference technique is given in [70].

In this study, the advantages of the MC dropout technique, namely, the ease of use and high computation speed, were leveraged to achieve the predictive uncertainty for SOH and RUL estimations. In several previous studies, other methods such as MC simulations [41] and deep ensembles [71] were investigated to add predictive uncertainty to the estimations. However, the former method requires a high computational cost, whereas the latter requires the modification of the network structure. Furthermore, the MC dropout improves the prediction results due to its classical role as a regularizer, as detailed in Section 4.

Several dropout layers were added inside the data-driven block of the hybrid recurrent unit to utilize the MC dropout technique, as illustrated by the orange vertical stripes shown in Fig. 4. With all of the aforementioned components described in subsection 2.2 combined with the MC dropout technique, the proposed neural network was designated as a KIRNN-MC.

3. Experimental procedure

3.1. Capacity fade data 1

Capacity fade data [50] collected from 124 LFP/graphite A123 APR18650M1A cells cycled under 72 fast charging conditions are used to train, validate, and test the proposed framework in this study. It is the largest public dataset of its type, with cycle lives ranging from 150 to 2300. Thus, it is a highly suitable candidate for battery health prognosis tasks. The dataset is split into training, validation, and test sets as 41, 43, and 40 cells, respectively, according to [50]. It is worth noting that all models addressed hereafter are first trained using the training set and then fine-tuned with the early 100 cycles of the validation and test sets when evaluating. What is optimized with the training set is referred to as the pre-trained model, and this procedure starts off with training the

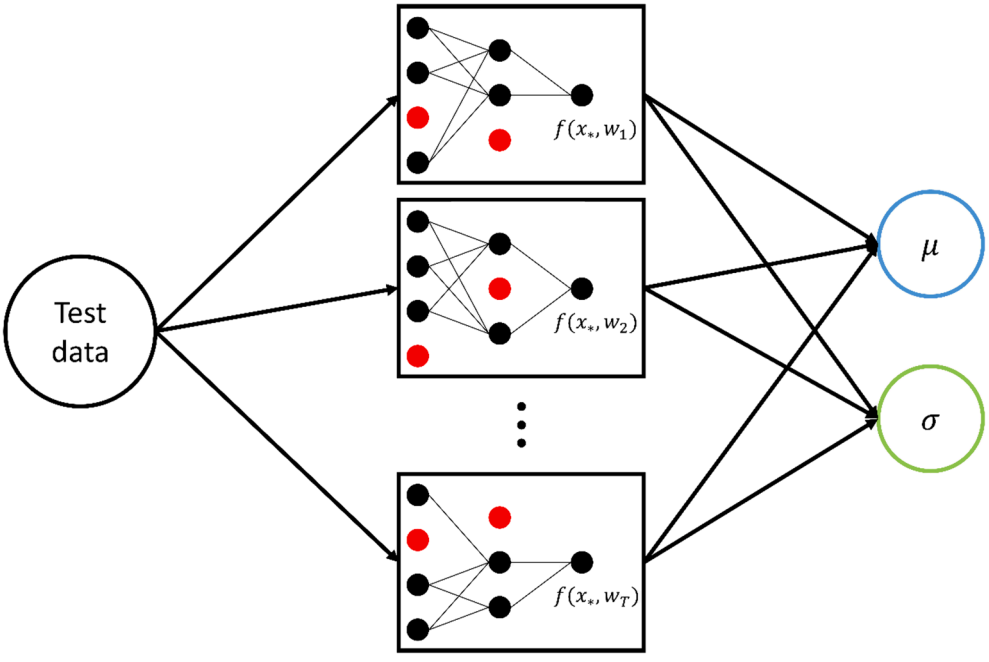


Fig. 5. P stochastic forward passes of a neural network on test data to compute the mean (μ) and the standard deviation (σ) of the estimations.

Table 2
Detailed cycle profiles for the capacity fade data 1. The charging policies consist of 72 different combinations of current steps ranging from a state of charge of 0 to 80%.

	Policy	SOC level		
		0%	50%	80%
Charging	One-step	Apply 3.6C (C-rate)		
	Two-step	Apply 6C	Apply 4C	Apply 1C (CCCV) to 3.6 V with current cutoff of C/50
Discharging	One-step	Apply 4C to 2.0 V with current cutoff of C/50		

data-driven block, as illustrated in Fig. 1. The fine-tuning starts off at the KIRNN training in the flow chart. The early 100 cycles can be considered as the online data that are collected in real-time, and they are depicted in the grey shaded region of Fig. 7. The overall data distributions and their statistics are displayed in Figure S3. The cells were placed inside an environmental chamber to maintain a temperature of 30 °C. The specific profiles for the degradation experiments are summarized in Table 2.

3.2. Capacity fade data 2

Another battery aging dataset [72], made publicly open as of August 2021 is used in the study for further evaluation. The data is collected from three types of commercial batteries, i.e., FST-3350 mAh, ME-2600 mAh, and SY-2150 mAh, under 15 different cycle profiles. These batteries are made of lithium nickel manganese cobalt oxides as a cathode material, whereas capacity fade data 1 from subsection 3.1 use lithium iron phosphate as a cathode material. Hence, applying the proposed method to these data is an effective way to validate the robustness of the proposed method in that this study applies the proposed method to different shapes of a battery with different active materials. Table 3 describes the collected datasets under the cycle profiles. In the table, '*A*' denotes datasets obtained under accelerated charging conditions, thus the relatively low aging cycles compared to those of other datasets. '*W*' stands for weakly-supervised datasets for which only a few capacity labels are given. Lastly, '*T*' refers to fully supervised datasets with the entire capacity labels given. In this study, all '*A*' and '*W*' datasets are used for training (SY-2150 W_1 is used for validation), whereas all '*T*' datasets except for those of SY-2150 are used for testing. These SY-2150 datasets are excluded for testing because of the two successive cycling processes, which cause inconsistent feature extraction due to different C-rates during the aging process. Specifically, the impedance-related features proposed in Section 2.1 are only valid for use under the constant C-rates because they depend on C-rate due to overpotential. In addition, it should be noted that '*W*' datasets are trained with the entire reference capacity labels because weakly-supervised learning is out of scope of this study. Please refer to [72] for further details of the data configuration.

Table 3
Detailed cycle profiles for the capacity fade data 2. The first C-rate stands for the constant-current charge rate and the second C-rate is the constant-current discharge rate. Then comes the number of cycles until the end-of-life. '+' denotes two successive cycling processes.

	FST-3350	ME-2600	SY-2150
A	1.0C-1.0C × 50 cycles	1.25C-1.25C × 70 cycles	1.0C-1.0C × 525 cycles
W_1	0.3C-0.3C × 370 cycles	0.48C-0.48C × 1094 cycles	0.5C-1.0C × 844 cycles
W_2	0.5C-0.5C × 260 cycles	0.67C-0.67C × 543 cycles	0.7C-1.0C × 525 cycles
T_1	0.3C-0.3C × 395 cycles	0.29C-0.29C × 695 cycles	0.7C-1.0C × 300 + 1.0C-1.0C × 316 cycles
T_2	0.4C-0.4C × 430 cycles	0.67C-0.67C × 444 cycles	1.0C-1.0C × 210 + 0.5C-1.0C × 653 cycles
T_3	0.5C-0.5C × 235 cycles	0.77C-0.77C × 179 cycles	1.5C-2.0C × 90 + 0.5C-1.0C × 719 cycles

Table 4
Proposed KIRNN-MC architecture.

Knowledge-infused block		
Layer	Type	Dimensions
Input	Cycle	1
Hidden 1	Exponential terms	2
Output	Capacity (dQ_{KNO})	1
Data-driven block		
Layer	Type	Dimensions
Input	Cycle + HIs	1 + 3
Hidden 1	Dense layer	50
	ReLU activation	–
	Dropout	–
Hidden 2	Dense layer	10
	ReLU activation	–
	Dropout	–
Hidden 3	Dense layer	10
	ReLU activation	–
	Dropout	–
Hidden 4	Dense layer	10
	ReLU activation	–
	Dropout	–
Hidden 5	Dense layer	5
	ReLU activation	–
	Dropout	–
Output	Dense layer (dQ_{DNN})	1
Linear combination block		
Layer	Type	Dimensions
Input	$dQ_{KNO} + dQ_{DNN}$	2
Output	ΔQ	1

3.3. Construction of deep neural networks

The proposed architecture of the KIRNN-MC was determined heuristically, as summarized in Table 4. Furthermore, several baseline DNNs, including LSTM, CNNs, and multi-layer perceptron (MLP) were tested and evaluated in parallel to validate the superiority of the proposed framework. It should be noted that the structures of these models were well-established through hyperparameter optimization to ensure a fair comparison. This is because there are no benchmark models of the same types for the particular dataset used in this study. Therefore, the Bayesian optimization method was implemented to optimize the hyperparameters of the model structures.

In particular, Bayesian optimization functions by the iteration of posterior distribution construction due to the typical fitting of the Gaussian process on given data points, followed by the selection of a new set of hyperparameter space that is likely to be used in the next iteration [73–75]. The selection of hyperparameters at each iteration is determined by the posterior distribution combined with an exploratory strategy, such as expected improvement, which determines the local maxima of the acquisition function. In contrast with the conventional methods, such as the naïve grid search and the random search, which use random combinations of hyperparameters, Bayesian optimization utilizes the prior knowledge based on the fact that similar inputs generate similar outputs. Consequently, its convergence is significantly more rapid with an improved result. Specifically, the objective of the Bayesian optimization process is to solve the following objective function:

$$x^* = \underset{x\in\Lambda}{argmax} f(x), \tag{4}$$

where x is the set of input hyperparameters, and Λ represents the predefined bounded regions of the entire hyperparameter space. Through the iterative Bayesian optimization process, the parameter x^* that best predicts the battery capacity is adopted.

The LSTM, CNN, and MLP models were optimized through Bayesian optimization, given the searching space of the hyperparameters, as listed

Table 5
Training process details of the baseline deep learning models optimized using Bayesian optimization.

LSTM		CNN		MLP	
Num. of input nodes	10	Num. of conv. layers	6	Num. of input nodes	27
Num. of hidden layers	1	Num. of filters	16	Num. of hidden layers	4
Num. of hidden nodes	10	Batch size	256	Num. of hidden nodes	46
Step size	5	Kernel size	2	Batch size	979
Batch size	502	Stride	1	Early stopping	50
Early stopping	50	Pool size	2	Loss function	MSE
Loss function	MSE	Num. of dense layers	1	Learning rate	2.3e-3
Learning rate	7e-3	Num. of dense nodes	64	Optimizer	Adam
Optimizer	Adam	Early stopping	50	Weight decay	5e-3
Weight decay	1.5e-3	Loss function	MSE		
		Learning rate	1e-4		
		Optimizer	Adam		
		Weight decay	1e-2		

in Table S2. The optimized architectures of each model, as well as the training process details, are summarized in Table 5. The input dimensions of the data-driven block may change depending on the type of dataset used, e.g., Group I. The provided architectures may not be the optimal models due to the lack of resource and time. However, Bayesian optimization generally guarantees more accurate results than manual optimization with limited resource and time, as claimed by numerous authors and practitioners [74,75].

The programming was conducted using Python 3.5.2, Tensorflow 1.14.0, and Keras 2.2.5. The computation was carried out using a GeForce RTX 2080 Ti GPU. The mean absolute percent error (MAPE) and root mean squared percent error (RMSPE) [71] were computed for the quantitative evaluation of the models for SOH monitoring as follows.

$$MAPE = \frac{1}{P}\sum_{i=1}^{P}\frac{|y_i^* - y_i|}{y_i} \tag{5}$$

$$RMSPE = \sqrt{\frac{1}{P}\sum_{i=1}^{P}\left(\frac{y_i^* - y_i}{y_i}\right)^2} \tag{6}$$

where y_i^* and y_i denote the estimated value and the target value at the i th occurrence from a total of P test samples. The MAE and RMSE were used for RUL estimations to provide a more intuitive understanding of the results represented in terms of the cycle number. They are equivalent to the MAPE and RMSPE, except for the denominators.

4. Results and discussion

The effectiveness of the proposed method was first validated, as detailed in subsections 4.1–4.3 (from Phase A to Phase C) and 4.4 (overall discussion), using the validation set consisting of 43 cells. The established models were then applied to the test set consisting of 40 cells to demonstrate its generalization capability and robustness on data that exhibit different capacity fading behaviors. The practical on-board battery health monitoring was examined, as presented in subsection 4.5, including practical scenarios of the proposed framework for SOH estimations and the feasibility of applying RUL estimations.

4.1. Contribution of impedance-related features

Impedance-related features were extracted and used as distinct features for training the proposed model in this study, whereas other features were used in the literature [50]. The quantitative contribution of these features to the estimation accuracy and their effectiveness as health indicators were compared with those of other features, as presented below. Specifically, LRP was addressed to evaluate the contribution of each feature to the output. LRP is a technique based on the Taylor series, which is suitable for explaining the decision of a neural network through decomposition [76]. This method re-distributes the relevance score in a top-down manner from the output node toward the input nodes. A detailed LRP formula is presented in Appendix C, and Fig. C1 illustrates the backward propagation of the relevance score. In summary, LRP implements saved weights at every epoch when training the KIRNN-MC and ranks the input features with respect to their contributions to the prediction output.

The LRP was evaluated using Group III, i.e. all the features, as the analysis clearly compared the relative contribution of the impedance-related features to that of other features employed in previous studies. Fig. 6 (a) presents the evolution of the relative feature importance during training, which was computed by normalizing the relevance scores at each epoch using a min–max normalization method. Hence, the highest-ranked feature had a relative importance of one, and the lowest-ranked feature had a relative importance of zero. The relative importance first fluctuated because the neural network was not sufficiently trained during the period of these epochs. In contrast, the relative importance of each feature gradually converged to certain values with an increase in the number of epochs. Specifically, HI7 demonstrated the highest relative importance after 1400 epoch, whereas Tmax contributed slightly to the SOH estimation, given that its relative importance was low. The relative importance of HI8, HI6, and CT gradually increased after 4000 epochs, whereas that of other features decreased with an increase in the number of epochs. As the model was sufficiently optimized, the difference was more significant, and the overall rank converged. Hence, the relative importance was finalized after 7500 epochs, after which the optimal prediction accuracy for the validation set was achieved. Fig. 6 (b) reveals that the HIs exhibited higher relevance scores than the given features throughout the training epochs. Specifically, three HIs, which exhibited the highest linear correlations, significantly contributed to the SOH estimation. Moreover, the CT and IR contributed to an extent to the SOH estimation, whereas Tmax contributed minimally. The results confirm that impedance-related HIs are superior health indicators to the other features proposed in the literature. The results correspond with the estimation accuracy listed in Table 6. In the validation set, Group I (HI8, HI7, and HI6) demonstrated the highest estimation accuracy with respect to the MAPE, whereas Group II (the given features) demonstrated the highest estimation accuracy with respect to the RMSPE. Moreover, the errors denoted by the MAPE and RMSPE exhibited similar orders of magnitude. Hence, it is difficult to select the optimal HIs, although LRP suggests Group I to be effective HIs. In contrast, the analysis of the estimation accuracy for the test set clearly revealed that Group I demonstrated the highest estimation accuracy with respect to both metrics, thus suggesting that impedance-related features were highly-correlated health indicators for battery health monitoring. Conversely, Group II and Group III, which contained less correlated features compared with those of Group I, exhibited lower estimation accuracies. These results indicate that the presence of irrelevant features (e.g., Tmax) negatively influences the estimation performance, given that the non-zero weights assigned to these features play a role in noise addition with respect to estimating outcomes [77]. The contribution of these HIs to the accuracy and robustness was more clearly demonstrated in the test set. Specifically, the ratios of the MAPE and RMSPE for Group II and Group III in comparison with that of Group I were larger in the test set, as shown in the parentheses in Table 6. This observation suggests that the validation set was used for the optimization of hyperparameters. Therefore, overfitting may occur for Group II and Group III. Hence, the neural network trained with these groups exhibited a low generalization capability and robustness, thus resulting in a low estimation accuracy for the test set.

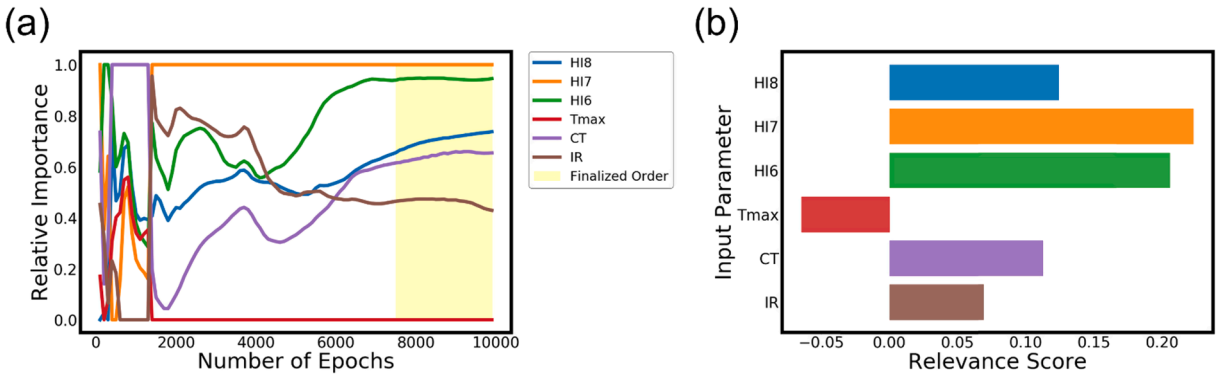


Fig. 6. Layer-wise relevance propagation results show the evolution of the relative feature importance over (a) all the training epochs and (b) the finalized relevance score at Epoch 7500.

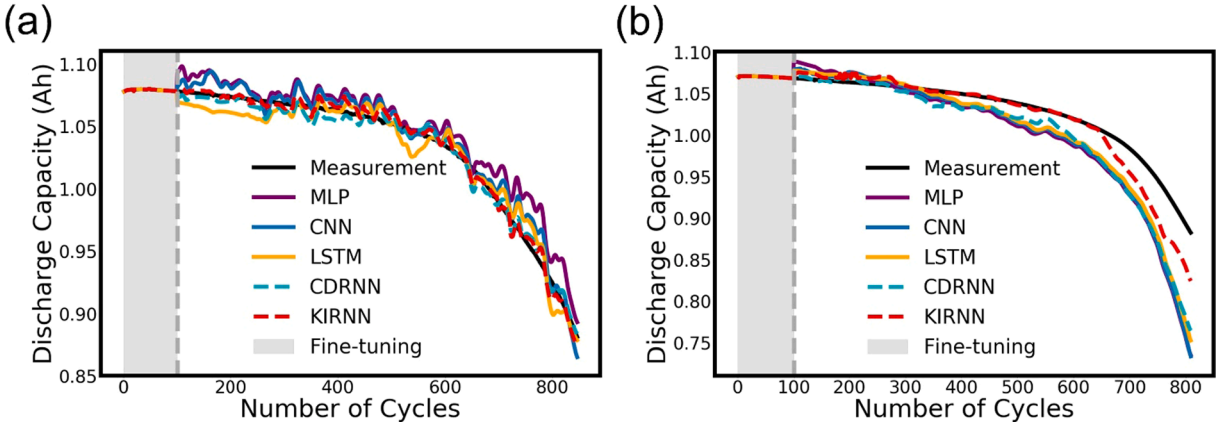


Fig. 7. State of health estimation of (a) Cell #25 from the validation set and (b) Cell #11 from the test set.

Table 6
Performance evaluation of pre-defined feature groups based on KIRNN-MC.

		Group I	Group II	Group III
Validation set	MAPE (%)	**1.49**	1.50 (1.01)	1.51 (1.01)
	RMSPE (%)	2.36	**2.24** (0.95)	2.33 (0.99)
Test set	MAPE (%)	**1.0**	1.58 (1.58)	1.38 (1.38)
	RMSPE (%)	**1.36**	2.31 (1.70)	1.76 (1.29)

4.2. KIRNN implementation

The effect of leveraging prior knowledge in model construction was analyzed, as presented in this subsection. Specifically, the KIRNN was compared with the cumulative damage RNN (CDRNN). In particular, the KIRNN is the cumulative damage framework embedded with an empirical model (a double exponential model), whereas the CDRNN only has the architecture of a cumulative damage model in the absence of all the knowledge-infused blocks. Both models do not include the dropout layers in the data-driven block. The contribution of the layers is discussed in the following subsection. In addition to the two models, the commonly adopted baseline DNNs, including LSTM, CNN, and MLP whose architectures were optimized using Bayesian optimization were evaluated to demonstrate the superiority of the proposed method. The results of the SOH estimation of a sample cell (Cell #25 from the validation set) using all the models are exemplified in Fig. 7 (a). It should be noted that a small portion of online data, which corresponds to a set of data collected from the first 100 cycles (below the gray dashed horizontal line), was used at the start to tune the pre-trained models in advance, as illustrated in Fig. 3. The KIRNN and CDRNN were determined to be more suitable than the baseline DNNs for the estimation using the validation set. However, no significant difference was observed between the two models. Moreover, the estimation accuracies

Table 7
Summary of state of health estimation using several deep learning models with and without the fine-tuning process. FT refers to fine-tuning. Optimal scores are shown in bold font.

			KIRNN-MC	KIRNN	CDRNN-MC	CDRNN	LSTM	CNN	MLP
Validation set	MAPE (%)	w/ FT	**1.49**	1.61	1.51	1.65	1.77	2.00	2.20
		w/o FT	**1.56**	1.70	1.61	1.77	1.88	2.09	2.34
	RMSPE (%)	w/ FT	**2.36**	2.62	2.42	2.59	2.65	3.13	3.24
		w/o FT	**2.47**	2.74	2.52	2.74	2.79	3.26	3.37
Test set	MAPE (%)	w/ FT	**1.00**	1.62	1.94	2.2	1.83	2.73	3.03
		w/o FT	**1.12**	1.71	2.06	2.30	1.96	2.84	3.19
	RMSPE (%)	w/ FT	**1.36**	2.42	2.69	3.4	2.62	4.3	4.44
		w/o FT	**1.51**	2.57	2.78	3.56	2.80	4.44	4.59
Time required for FT of new online data at transfer learning (~100 cycles, s)			339	304	347	263	5064	65	**36**

of all the models were of similar orders of magnitude. It is difficult to evaluate whether the improved accuracies of the KIRNN and CDRNN can be attributed to network superiority or hyperparameter optimization. In contrast, the estimated SOH of another sample cell, i.e., Cell #11 from the test set, revealed that the KIRNN outperformed the CDRNN and other DNNs, as it was more in accordance with the ground truth line (Fig. 7 (b)). The comparison results confirmed that the KIRNN ensures a high estimation accuracy and robustness on a new dataset. Specifically, the MAPE and RMSPE decreased by 0.58% and 0.98% (Table 7), respectively, by knowledge infusion into the network when the models were evaluated on the test set. This implies that the empirical model, which supervises the degradation behavior of lithium-ion batteries, plays a critical role in a purely data-driven deep learning model. Moreover, this phenomenon is more significantly observed for data from a different probability distribution. In conclusion, the supervision of physical knowledge increases the estimation accuracy and improves the robustness of the neural network.

4.3. MC dropout implementation

The contribution of the MC dropout technique to the estimation accuracy was analyzed. The effect of the MC dropout implementation was first validated on a sample cell from the validation set (Cell #5), followed by a sample cell from the test set (Cell #20). The results for other cells were similar, although they were not included in this manuscript for conciseness. For all cases, 50 stochastic passes were executed to compute the mean and standard deviation of the estimations in the test phase. Fig. 8 illustrates the SOH estimation results for the cases. In particular, Fig. 8 (a) and (c) compare the CDRNN and CDRNN-MC with Cell #5 and Cell #20, respectively, whereas Fig. 8 (b) and (d) compare the KIRNN and KIRNN-MC with Cell #5 and Cell #20, respectively. The solid black line represents the ground truth. The blue dashed line denotes the point estimation without using the MC dropout technique, whereas the red dashed line and surrounding yellow region represent the estimated mean and three-sigma confidence bounds when using the MC dropout technique.

With respect to the validation data (Cell #5), using the MC dropout slightly improved the estimation performance. The difference was almost negligible, although the overall MAPE and RMSPE were improved by 0.14% and 0.17%, respectively, for the CDRNN-MC and 0.12% and 0.26%, respectively, for the KIRNN-MC (Table 7). No significant changes were observed between the CDRNN-MC and KIRNN-MC. First, the difference between the CDRNN-MC and CDRNN was only observed at the small end portion of the estimation points toward the end-of-life. Second, the mean estimation points exhibited significant fluctuations throughout the cycles, thus providing relatively low confidence for the estimation. Therefore, it can be concluded that the MC dropout technique did not contribute significantly to the SOH estimation, despite the slight increase in performance. The only significant difference was the relatively higher estimation confidence intervals for the KIRNN-MC when compared with the CDRNN-MC. This observation supports the hypothesis that a generalized empirical model helps improve the estimation accuracy of a purely data-driven model.

The effect of the MC dropout was more significant for the test data (Cell #20), as shown in Fig. 8 (c) and (d). Specifically, the overall MAPE and RMSPE were improved by 0.26% and 0.71%, respectively, for the CDRNN-MC and 0.62% and 1.06%, respectively, for the KIRNN-MC (Table 7). Different from the previous case in which the effect of the MC dropout was similar for the CDRNN-MC and KIRNN-MC, the effect was more evident for the KIRNN-MC when applied to the test set. This suggests that the MC dropout contributes significantly to the SOH estimation when prior knowledge is infused. The mean estimation of the KIRNN-MC was closer to the ground truth, and the difference between the models increased in accordance with the rapid downward trend of

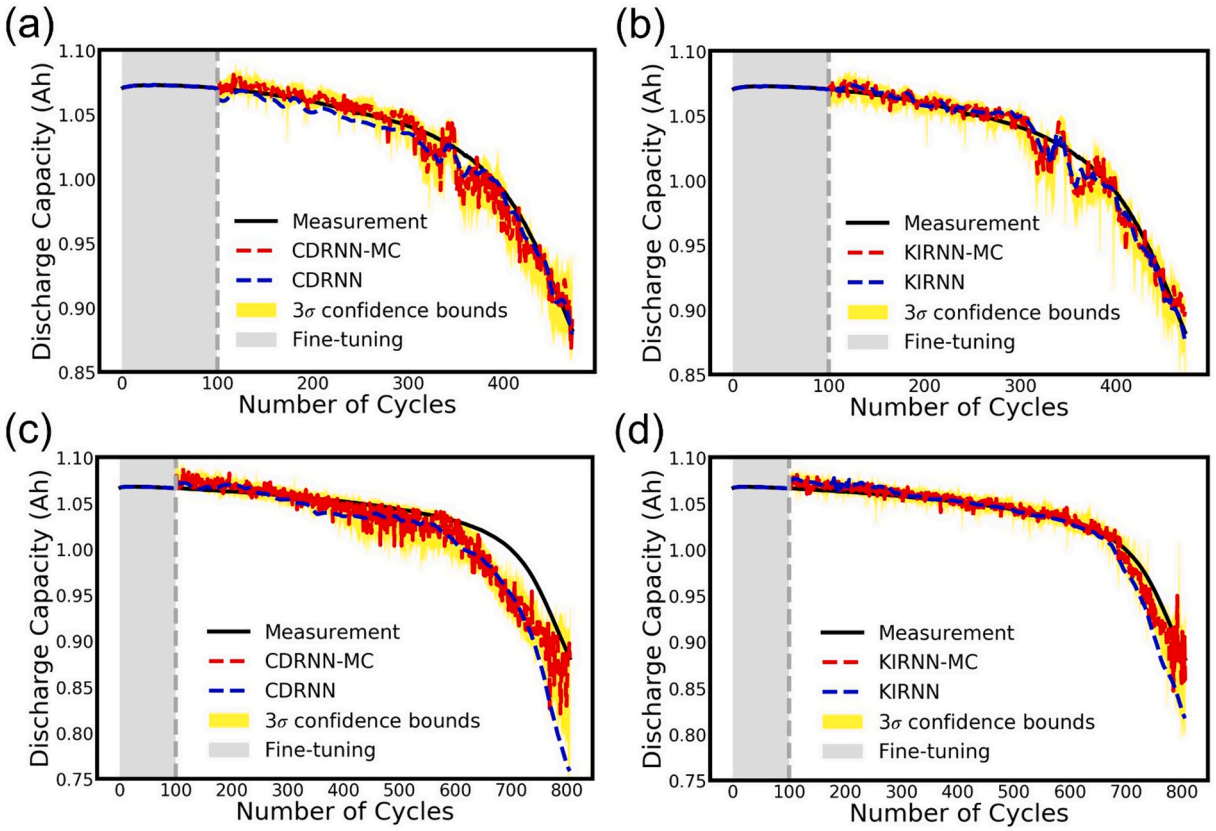


Fig. 8. Contribution of MC dropout technique to state of health estimation with (a) the CDRNN and (b) KIRNN for Cell #5 from the validation set, and the (c) CDRNN and (d) KIRNN for Cell #20 from the test set.

the curve. In particular, the difference was first observed at an earlier stage. The KIRNN output was almost in parallel with the lower bound of the KIRNN-MC output, thus implying the predictive generalization effect of the MC dropout technique when applied in conjunction with knowledge infusion. It should be noted that there were significantly fewer fluctuations in the estimation with the narrower upper and lower bounds of the confidence interval. Consequently, the MC dropout technique is the most effective when used simultaneously with the KIRNN (KIRNN-MC), as confirmed by the results of the overall test set (Table 7). This phenomenon can be attributed to the inherent nature of the MC dropout, as it mainly accounts for the model uncertainty, i.e., epistemic uncertainty, which can be reduced by optimizing the model and conducting the training process using a sufficient amount of data [78]. It is highly probable that infusing empirical knowledge into model construction optimizes the model and, therefore, reduces the related uncertainty. The effect appears more clearly around points towards the end-of-life where the uncertainty is large due to a sudden change in the aging trajectory. Moreover, the observation is in accordance with the qualitative analysis of MC dropout, in which predictive performance is improved when MC dropout is applied to a standard neural network [78]. This particular regularization term originates from the Bayesian approximation of dropout in which the Kullbeck-Leibler divergence between the approximate posterior $q(w)$ and the posterior, $p(w|\mathrm{X}, \mathrm{Y})$ is minimized to ultimately satisfy the Bayes rule and provide probabilistic predictions on unseen data points. Dropout on neural network adds stochasticity or randomness to the output of neural network, securing robustness. Specifically, it can be represented as endowing Bernoulli distribution upon the weights and biases of the neural network. Therefore, the randomness (Kullbeck-Leibler divergence) inherent in the reparametrized form of evidence lower bound (ELBO), which is a trick to compute the posterior of a Bayesian neural network, is qualitatively equivalent to the same terms (Bernoulli distributed weights and biases). Therefore, adding dropout to neural network mathematically equates to adding an L2 regularization term (stochastic) weighted by the factor η to a data-fit term (deterministic), commonly observed objective terms for deep learning models. The regularization term minimizes the weight values of neural network, which will lead to reduced effect by the local noise of training data. This prevents overfitting of the network and thus generalizes the network output. Further investigation will be conducted in future work.

4.4. Discussions on the proposed method

The effectiveness of the proposed models was demonstrated by quantitative comparison with the commonly adopted baseline DNNs using the entire capacity fade datasets described in subsection 3.1.

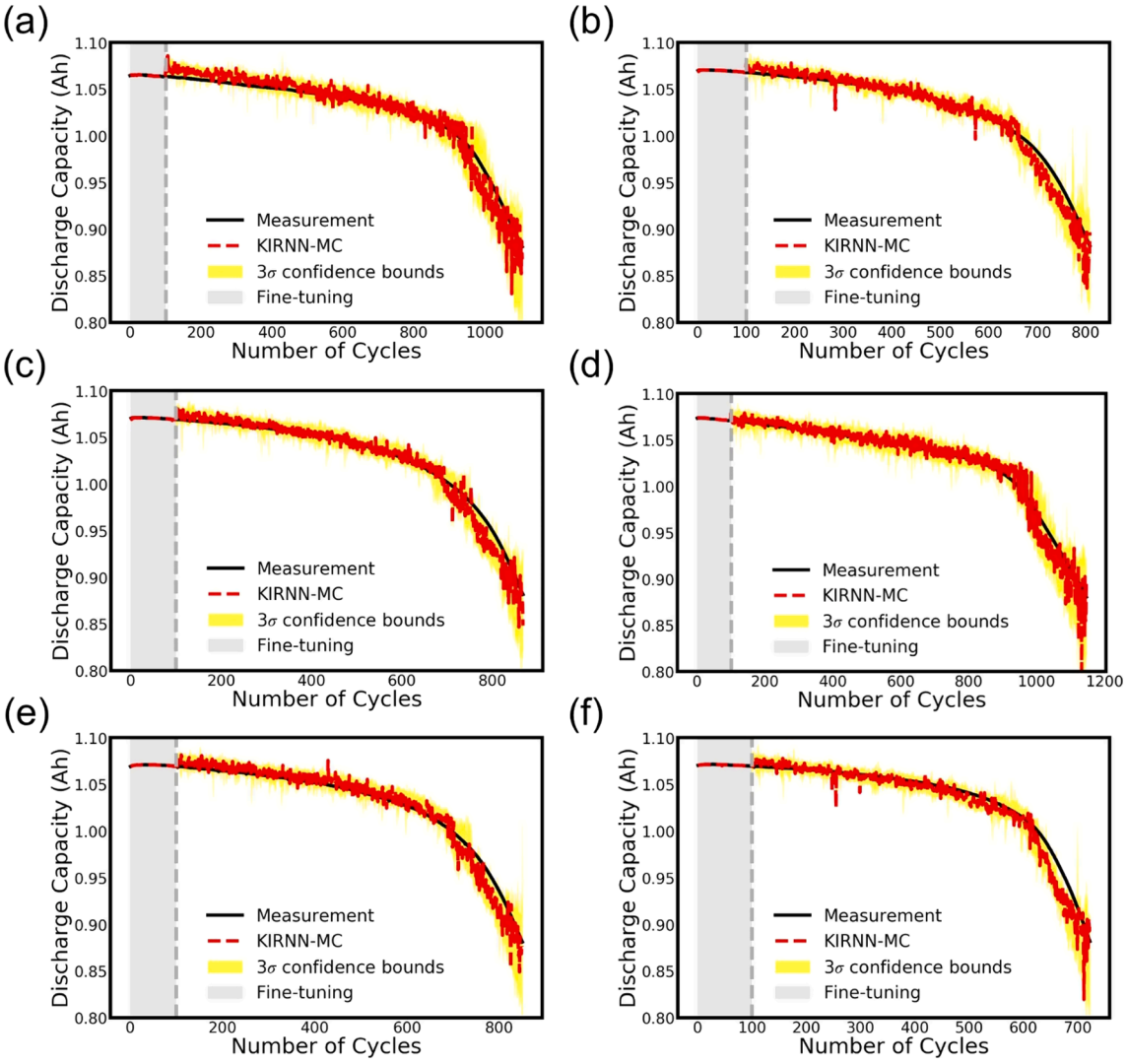


Fig. 9. The state of health estimation of the test cells using the proposed KIRNN-MC. The test cells are (a) Cell #3, (b) Cell #11, (c) Cell #15, (d) Cell #19, (e) Cell #29, and (f) Cell #31.

Table 7 summarizes the quantitative results for the validation set and the test set using all the models, including the variants of the proposed model, to confirm the overall generalization effect of the knowledge infusion and the MC dropout implementation. Furthermore, the training time of the online data (~100 cycles) for the model parameter update was computed for each model until a low error was achieved, as it is an important metric for implementing the proposed method in an on-board BMS. Fig. 9 presents the results of the SOH monitoring for six test cells (Cell #3, Cell #11, Cell #15, Cell #19, Cell #29, and Cell #31). It should be noted that the estimations were conducted by the KIRNN-MC, which demonstrated the most superior performance overall, and they were in accordance with the ground truth in most cells, as depicted in Fig. 9.

It can be seen from Table 7 that the proposed KIRNN-MC exhibited the most superior performance in terms of the MAPE and RMSPE for the validation set and test set. Given that the models were optimized based on the validation set, their estimation performances with respect to the test set were thoroughly investigated. In the absence of the knowledge-infused blocks and MC dropout, i.e., the CDRNN, the model outperformed the CNN and MLP. However, it exhibited poorer performance than LSTM, thus implying that the cumulative damage modeling is not sufficient to accurately estimate the SOH. Furthermore, the implementation of the MC dropout technique (CDRNN-MC) decreased the metric scores by 0.26% (MAPE) and 0.71% (RMSPE). However, the result was slightly less accurate than that of LSTM. Embedding an empirical model within the hybrid recurrent unit (KIRNN) would be beneficial, as described in subsection 4.2. The model outperformed LSTM. An excellent performance was achieved when the model was integrated with an empirical model under the application of the MC dropout technique (KIRNN-MC), as it exhibited a more significant decrease in the MAPE (0.94%) and RMSPE (1.33%). Compared with LSTM, the KIRNN-MC exhibited a 0.83% decrease in the MAPE and 1.26% decrease in the RMSPE. Thus, the effectiveness of the proposed model in comparison with the existing deep learning models was validated.

An important concern regarding the fine-tuning process is whether or not the true SOH label of the first 100 cycles can be collected in advance during the periodic maintenance of batteries. To alleviate such concerns, the same results without the fine-tuning process (w/o FT) are also listed in Table 7. The results reveal that although the performance slightly degrades overall for all models, the severity of degradation is not significant enough to change the conclusion. Therefore, it can be concluded that the fine-tuning process is conducted to ensure a better fit to a new test cell.

The training times of the online data provide significant insight into the trade-off between the accuracy and training time of models. The KIRNN-MC required approximately 339 s for the fine-tuning of the newly measured data. This is not comparable with the training times of the CNN and MLP. However, it is practical for on-board applications because the interval time between cycles is significantly longer than the training time. In contrast, LSTM required over 5000 s, which is multiple times larger than that of the proposed model. Hence, this study was focused on the CDRNN instead of LSTM. In the case wherein the estimation accuracy is more important than the training time of online data, LSTM can be supervised using physical information.

The proposed algorithm is further validated on another capacity fade dataset introduced in Section 3.2. For this particular case, only the KIRNN-MC is compared against the baseline deep learning models as other variants, including KIRNN, CDRNN-MC, and CDRNN, have proven to be less efficient. As mentioned in Section 3.2, '*A*' and '*W*' datasets are utilized for training the models (SY-2150 W_1 is used for validation) and '*T*' datasets (FST-3350 and ME-2600 only) for testing afterward. Fig. 10 (a) and (b) illustrate the predicted SOH of FST-3350 T_3 and ME-2600 T_1, respectively. The figure reveals that KIRNN-MC indicated by the red dashed line shows the most accurate estimation than the other models. In addition, no fine-tuning has been conducted for this testing to validate the proposed methodology under the circumstance in which no reference SOH label is available for the fine-tuning. Testing results of the remaining datasets are provided in Figure S4 of the Supplementary Material for the sake of brevity.

The quantitative results are provided in Table 8. In contrast to the results of Table 7 (capacity fade data 1), KIRNN-MC demonstrates relatively poor performance for the validation set. This phenomenon

Table 8
Summary of state of health estimation using several deep learning models. Optimal scores are shown in bold font.

			KIRNN-MC	LSTM	CNN	MLP
Validation set	SY-2150 W_1	MAPE (%)	0.96	0.71	**0.55**	0.75
		RMSPE (%)	1.18	0.83	**0.70**	0.90
Test set	FST-3350 T_1	MAPE (%)	**0.80**	0.96	2.04	2.71
		RMSPE (%)	**1.25**	1.46	2.28	2.93
	FST-3350 T_2	MAPE (%)	**0.81**	0.84	2.05	2.88
		RMSPE (%)	1.06	**0.93**	2.43	3.30
	FST-3350 T_3	MAPE (%)	**0.86**	0.99	1.64	2.70
		RMSPE (%)	**1.16**	1.33	1.80	2.96
	ME-2600 T_1	MAPE (%)	**0.49**	0.51	0.56	0.70
		RMSPE (%)	**0.63**	0.64	0.67	0.79
	ME-2600 T_2	MAPE (%)	**1.09**	1.13	1.51	2.14
		RMSPE (%)	**1.46**	1.47	1.64	2.32
	ME-2600 T_3	MAPE (%)	0.62	**0.56**	1.27	2.13
		RMSPE (%)	0.84	**0.72**	1.49	2.37

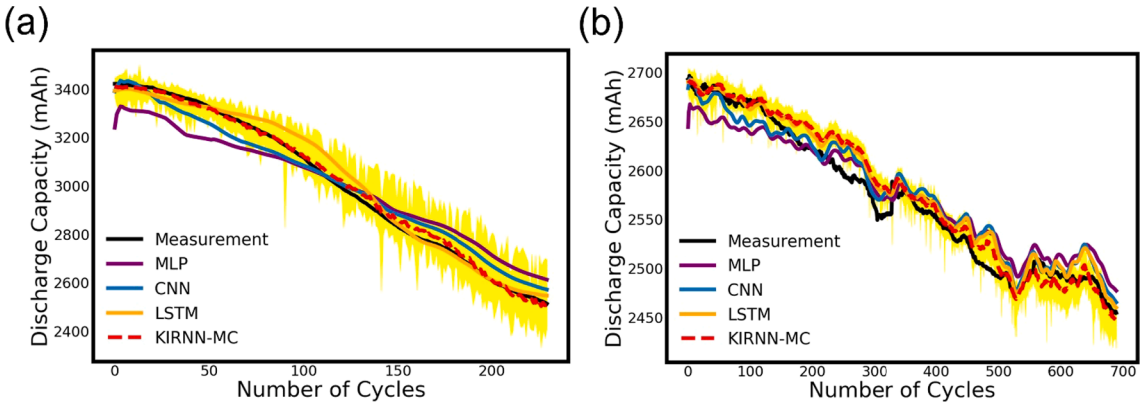


Fig. 10. The estimated SOH of (a) FST-3350 T_3 and (b) ME-2600 T_1 using KIRNN-MC and the baseline deep learning models.

attributes to KIRNN-MC having stronger regularization terms due to the empirical model and MC dropout, which prevent overfitting to the trained data. On the other hand, it performs much better on the test set as expected. KIRNN-MC outperforms LSTM by as much as 0.05% MAPE and 0.02% RMSPE on average, thanks to its better generalization performance. The exceptional result of ME-2600 T_3 may have been caused by its close similarity to one of the training datasets, e.g., ME-2600 W_2 with respect to the extracted impedance-related features. The practical on-board application scenario is further discussed in subsection 4.5.1.

4.5. Electric vehicle applications

4.5.1. Scenario for on-board SOH monitoring

The SOH estimation was conducted upon the collection of new input features from the charge–discharge cycles in real-time. The suggested input features in the proposed framework were the impedance-related features in Group I (HI3, HI4, and HI5), which were extracted from the readily accessible voltage curve during the discharging process of a constant current protocol due to the high estimation accuracy of this feature combination. However, the discharging process of a constant current protocol until a battery is fully discharged is not feasible or applicable with respect to the usage behavior of batteries in electric vehicles. Typically, most batteries are exposed to stochastic discharging processes depending on the driving conditions [79]. The extraction of such features from the voltage curve during the charge may be an alternative method, given that batteries are typically charged using the same CCCV protocol overnight in a garage or a charging station. It should be noted that this scenario is not applicable to the public dataset employed in this study, given that the charging protocol was changed several times for the acceleration of the degradation experiments, thus limiting the extraction of unified HIs in a charge curve. Therefore, the suggested solution to address the issue is SOH estimation at scheduled maintenance, e.g., twice annually or per a specified number of cycles [80]. In scheduled maintenance, the same discharge protocol used for the dataset can be applied, and impedance-related features can be extracted from voltage curves under similar conditions to ensure the robustness of the proposed model.

4.5.2. On-board RUL monitoring

The KIRNN-MC and the other compared models were employed to estimate RULs using the test set, given that the RUL is a critical parameter that requires estimation in addition to the SOH in electric vehicle applications. It should be noted that RULs have been monitored by training structurally equivalent additional DNNs with the target SOH values replaced by the target RUL values. Moreover, the training should be conducted separately for SOH and RUL estimations because the

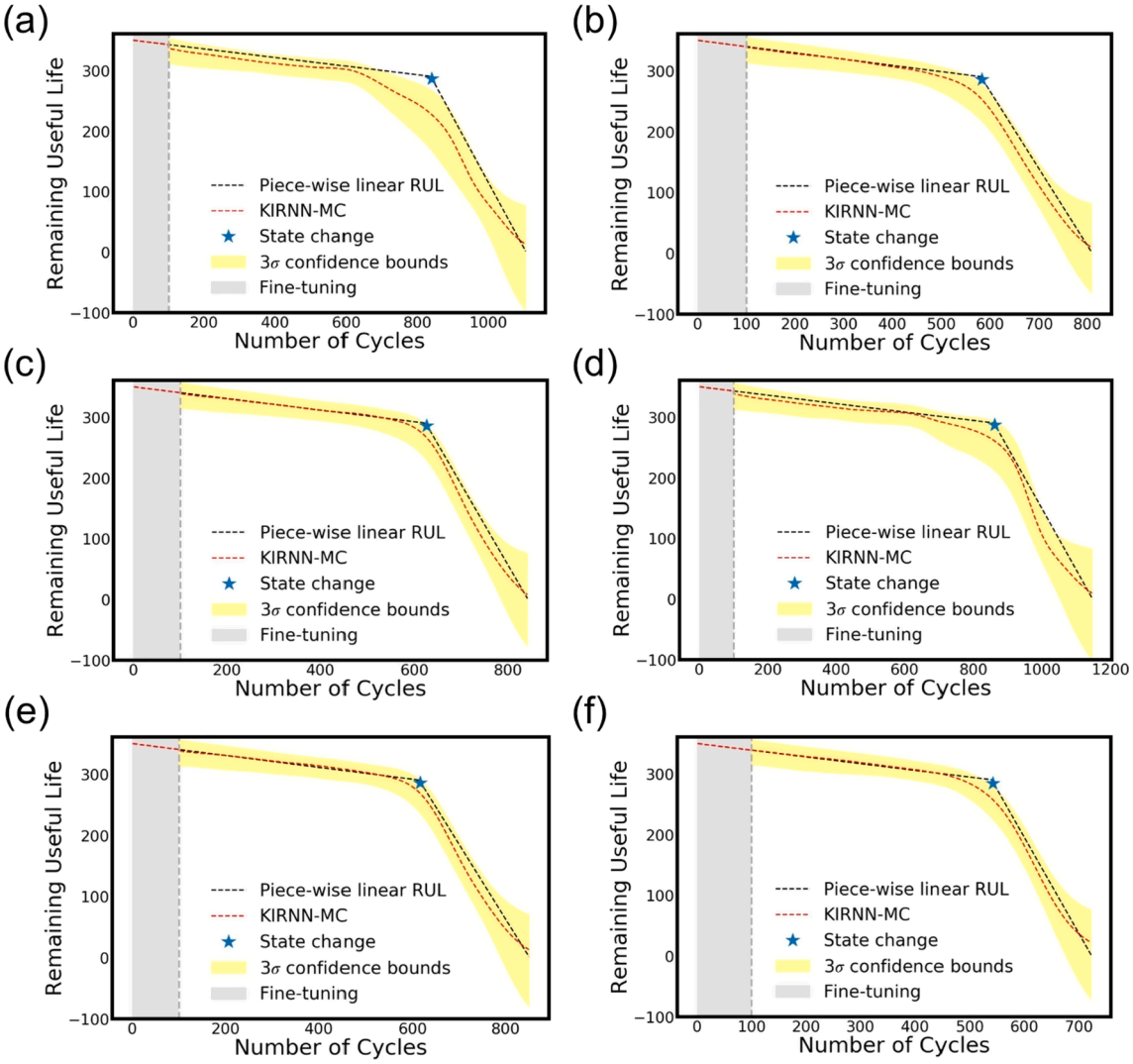


Fig. 11. RUL estimation of test cells using the proposed KIRNN-MC. The test cells are (a) Cell #3, (b) Cell #11, (c) Cell #15, (d) Cell #19, (e) Cell #29, and (f) Cell #31.

parameters are qualitatively different. For future SOH and RUL estimations, only the cycle steps and the extracted health indicators (HIs) are provided. In most previous studies, the RUL was represented as a quantity that decreased linearly and inversely with respect to the cycle of an actual system. However, for various applications, the RUL is generally more difficult to directly predict than the SOH due to the different target RULs of each system. This is different from the SOH, which starts at 100% and degrades thereafter. In particular, RULs cannot be normalized effectively. This characteristic destabilizes the training of numerous existing deep learning models. To address this issue, the target RUL linear degradation model was modified to a piece-wise linear degradation model [81–83], such that the starting points of the RULs, i.e., the initial RULs, were equivalent and reflected the actual system degradation phenomenon, which is negligible until a given instant. In this study, the piece-wise linear degradation model was modified further to more accurately reflect the battery degradation mechanism. The early phase of the model, which was previously considered negligible, was changed to a linear model with a different gradient. Thus, a different type of initial RUL was required. In particular, a second linear model was used to reflect the early degradation phase, during which the RUL decreased at a significantly lower rate than during the later phase. Therefore, the piece-wise linear RUL model consists of two linear lines, the first of which represents slow aging in the first half, and the other representing accelerated aging in the second half. The two linear models and initial RULs are shown in Fig. 11, as indicated by the dashed lines.

Given that RUL exhibits a negligible degradation up to a given point, followed by a linear decrease, the RUL estimation models should detect the instant of the abrupt change of state, so-called 'knee point' [84–86], to reflect the gradual decrease until zero RUL is reached. The knee point is generally caused by a phase shift in the underlying degradation mechanism. The first phase of degradation is mainly dominated by LAM. The second phase of degradation, which induces an abrupt decline in the battery capacity, is dominated by LAM and LLI. The knee points on SOH curves were detected by the "Kneedle algorithm," which determines the knee point of a concave curve [87]. Moreover, their locations are indicated by the blue stars in Fig. 11.

Fig. 11 illustrates the RUL estimation results of the proposed KIRNN-MC for the same test cells ((a)–(f)) as those used for the SOH monitoring. As can be seen from the figure, the model captured the decreasing RUL trend with a high confidence level until the state changed at the knee point. The confidence interval widened after the point. The results are satisfactory because the early phase of degradation was the same for all the training data until the state changed, followed by variations throughout the cells. Moreover, in general, the knee points of the predicted curves and actual state changes were in good agreement, thus implying that the model successfully detected the start of the later degradation phase. The estimations after the state changed were sufficiently accurate to utilize the proposed model for on-board applications. In general, the KIRNN-MC exhibited a significantly superior predictive performance with respect to RUL estimations than the other deep learning models. The effects of knowledge infusion and the MC dropout implementation were significant. The averaged results for all the test cells are summarized in Table 9. Similar to the other results, the proposed method outperformed the other DCNN, especially in the test set, thus suggesting that the proposed method is effective for SOH and RUL estimations with respect to accuracy and robustness.

5. Conclusions

A novel deep learning-based prognostics framework was proposed, and its applicability for the on-board state of health and remaining useful life estimations of lithium-ion batteries was demonstrated. The proposed method features three characteristics as the key contributions. First, reliable and online accessible impedance-related features or health indicators extracted from discharge curves can be leveraged to accurately model the battery capacity degradation over charge–discharge cycles. Moreover, the layer-wise relevance propagation analysis quantitatively validates the effectiveness of health indicators as distinct features in describing the complex and nonlinear degradation phenomena of batteries. Second, the estimation performance is significantly

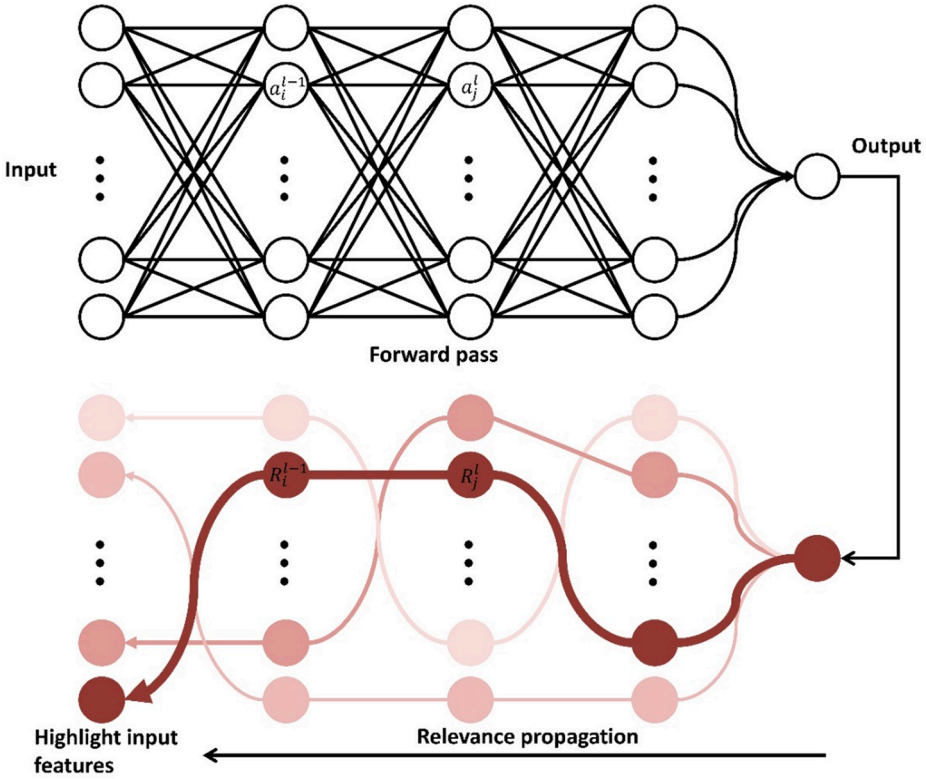


Fig. C1. Relevance score propagation from the output node toward the input nodes.

Table 9
Summary of remaining useful life estimation using several deep learning models. Optimal scores are shown in bold font.

		KIRNN-MC	KIRNN	CDRNN-MC	CDRNN	LSTM	CNN	MLP
Validation set	MAE (cycle)	24.8	**24.1**	24.4	33.7	26.8	25.0	28.0
	RMSE (cycle)	**42.9**	47.0	45.4	60.3	44.1	47.3	45.5
Test set	MAE (cycle)	**19.9**	20.6	36.8	46.6	35.2	40.5	49.8
	RMSE (cycle)	**38.6**	39.9	53.6	68.6	53.5	60.9	70.7

improved through knowledge infusion to the recurrent unit, which generalizes its estimation capacity with respect to newly added datasets. Third, the robustness of the proposed model and uncertainty-based estimations are realized by addressing the Monte Carlo dropout technique. This endows the deep learning-based decisions with higher reliability. Specifically, the proposed model outperforms several baseline deep learning models by the maximum of 2.03% and 3.08% with respect to the mean absolute percent error and root mean squared percent error, respectively, in state of health monitoring, and 29.9 cycles and 32.1 cycles with respect to the mean absolute error and root mean squared error, respectively, in remaining useful life monitoring in the test phase. A discussion is presented on the applicability of the proposed framework to on-board systems, and the scheduled maintenance is suggested as a solution. Future research will include the application of the proposed method to a charge curve using the same protocol employed during degradation experiments to confirm its feasibility with respect to actual applications. It is possible to infuse empirical knowledge into model construction to optimize the model and, therefore, reduce the related uncertainty. This aspect should be investigated further. Moreover, the focus will be directed toward validating the proposed framework using a more accurate dataset that reflects the dynamic charge–discharge process, despite the long period of time required for data acquisition.

CRediT authorship contribution statement

Sung Wook Kim: Methodology, Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. **Ki-Yong Oh:** Conceptualization, Formal analysis, Writing – review & editing, Funding acquisition. **Seungchul Lee:** Writing – review & editing, Funding acquisition, Supervision.

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.

Acknowledgement

This research was financially supported by the Institute of Civil Military Technology Cooperation funded by the Defense Acquisition Program Administration and Ministry of Trade, Industry and Energy of Korean government under grant No. 19-CM-GU-01, and the Research and Development on Fire Safety Technology for ESS Hydrogen Facilities, 20011568, Development of Automatic Extinguishing System for ESS Fire, funded by the National Fire Agency (NFA, Korea).

Appendix A. Battery modeling governing equations

The following four governing equations model how a lithium-ion battery works.

$$\sigma \frac{\partial^2 \varnothing}{\partial x^2} = a_s j, \tag{A1}$$

$$\rho C_p \frac{\partial T}{\partial t} + \frac{\partial}{\partial t}\left(-\lambda \frac{\partial T}{\partial x}\right) = a_s j(\varnothing_s - \varnothing_e - U) + a_s j T \frac{\partial U}{\partial T} + \frac{j^2 R_f}{a_s}, \tag{A2}$$

$$j = j_0[exp\left(\frac{\alpha_a nF}{RT}\eta\right) - exp(\frac{-\alpha_c nF}{RT}\eta)], \tag{A3}$$

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}(D\frac{\partial C}{\partial x}), \tag{A4}$$

Equation (A1) is the Ohm's law, where σ, ∅, a_s, and j denote electric conductivity, potential, specific surface area, and current density, respectively. This law describes the amount of charge in electrodes and electrolyte. Equation (A2) is the energy balance equation, where ρ, C_p, T, λ, U, and R_f denote the density of solid particle, heat capacity, temperature, heat conductivity, equilibrium potential, and contact resistance. This equation describes the energy balance in the battery system. Equation (A3) is Butler-Volmer equation, where j_0, α_a, n, F, R, and η denote exchange current density, transfer coefficient, the number of electrons, Faraday constant, gas constant, and overpotential. It describes the relationship between the potential and current. Equation (A4) is Fick's law, where C and D denote concentration and diffusivity coefficient. This law describes how ions diffuse across different materials.

Appendix B. Proof of concept

Assume a constant current mode,

$$\Delta SOC = \int_{t_1}^{t_2} Idt = I(t_2 - t_1) \cong \Delta t \tag{B1}$$

$$R = \frac{\Delta V}{\Delta I} \cong \frac{\Delta V}{\Delta SOC} = \frac{\Delta V}{\int_{t_1}^{t_2} Idt} = \frac{\Delta V}{\Delta t} \tag{B2}$$

$$\therefore R = \frac{\Delta V}{\Delta t} \approx \frac{dV}{dt}(Slope) \tag{B3}$$

In the above formulae, $\Delta SOC, R, V, I, and\ t$ denote the change of battery capacity, resistance, voltage, current, and time, respectively. To elaborate, according to Ohm's law, $R = \Delta V/\Delta I$. With regards to ΔI, it is directly proportional to ΔSOC because SOC or capacity equals to the area under the current over time according to the Coulomb counting method. It should be noted that ΔI can just be represented as I or Δt (for the above case) since it is operating under a constant current mode.

Appendix C. Layer-wise relevance propagation

A simple LRP assumption is that every input neuron contributes to the output, and such relevance to the output is preserved for every layer. Mathematically, the contribution amount of each input to the output $f(x)$ can be computed as the partial derivative of the output with respect to the input, that is, x_p: $\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_p}$. Given that a neural network is a universal function approximator, the output can be represented by the first-order Taylor series, as follows:

$$f(x) = f(a) + \sum_{p=1}^{d} \frac{\partial f}{\partial x_p}|_{x=a}(x-a) + \epsilon, \tag{C1}$$

where a is a real value and ∈ is an error term for the higher-order polynomial terms of the Taylor series. The second term on the right-hand side of the equation indicates the change in $f(x)$ according to the variations in x_p. Equation (12C1) enables the decomposition of the output into the relevance score. However, terms a and ∈ should be resolved. To overcome this challenge, Montavon et al. [88] proved that $\epsilon = 0$ when ReLU activation functions are used for neural networks and that numerous a values satisfy $f(a) = 0$. Therefore, Equation (C1) can be reformulated in the context of a neural network, as follows:

$$\begin{aligned} f(x) &= f(a) + \sum_{p=1}^{d} \frac{\partial f}{\partial x_p}|_{x=a}(x-a) + \epsilon \\ &= \sum_{p=1}^{d} \frac{\partial f}{\partial x_p}|_{x=a}(x-a) \\ &= \sum_{p=1}^{d} w_p x_p + b \\ &= \sum_{p=1}^{d} R_p, \end{aligned} \tag{C2}$$

where w_p and b are the weights and bias, and R_p is the relevance score. According to the above equation, the following condition is true for any two consecutive layers in a DNN: $\sum_i R_i = \sum_j R_j = f(x)$. This is equivalent to a conservation property, where the data received by a neuron is redistributed to the lower layer in an equal amount. In a neural network, the relevance score propagates backward based on the trained weights and bias, and it is calculated as follows:

$$R_i^{(l)} = \sum_j \frac{a_i^{(l)} w_{ij}^{(l)}}{\sum_i a_i^{(l)} w_{ij}^{(l)}} R_j^{(l+1)}, \tag{C3}$$

where $a_i^{(l)}$ and $w_{ij}^{(l)}$ are neuron activations and weights connecting Neuron i to Neuron j, respectively, at the l-th layer. The objective of the weighted sum in the denominator is to ensure that the values are normalized to establish the conservation property. In this study, the LRP was applied to the data-driven block, shown in Fig. 4, for every training epoch to track the evolution of the feature importance with respect to the number of cycles.

Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apenergy.2022.119011.

References

[1] Hu X, Xu L, Lin X, Pecht M. Battery lifetime prognostics. Joule 2020;4(2):310–46.
[2] Guo M, Sikha G, White RE. Single-particle model for a lithium-ion cell: Thermal behavior. J Electrochem Soc 2010;158(2):A122.
[3] Jokar A, Rajabloo B, Désilets M, Lacroix M. Review of simplified Pseudo-two-Dimensional models of lithium-ion batteries. J Power Sources 2016;327:44–55.
[4] Chun H, et al. Adaptive exploration harmony search for effective parameter estimation in an electrochemical lithium-ion battery model. IEEE Access 2019;7:131501–11.
[5] Torchio M, Magni L, Gopaluni RB, Braatz RD, Raimondo DM. Lionsimba: a matlab framework based on a finite volume model suitable for li-ion battery design, simulation, and control. J Electrochem Soc 2016;163(7):A1192–205.
[6] Ng M-F, Zhao J, Yan Q, Conduit GJ, Seh ZW. Predicting the state of charge and health of batteries using data-driven machine learning. Nature Machine Intelligence 2020;2(3):161–70.
[7] Goebel K, Saha B, Saxena A, Celaya JR, Christophersen JP. Prognostics in battery health management. IEEE Instrum Meas Mag 2008;11(4):33–40.
[8] B. Saha, S. Poll, K. Goebel, and J. Christophersen, "An integrated approach to battery health monitoring using Bayesian regression and state estimation," in *2007 IEEE Autotestcon*, 2007: Ieee, pp. 646-653.
[9] Y. H. Lui *et al.*, "Physics-based prognostics of implantable-grade lithium-ion battery for remaining useful life prediction," *Journal of Power Sources,* vol. 485, p. 229327, 2021.
[10] Zhang L, Mu Z, Sun C. Remaining useful life prediction for lithium-ion batteries based on exponential model and particle filter. IEEE Access 2018;6:17729–40.
[11] Xu B, Oudalov A, Ulbig A, Andersson G, Kirschen DS. Modeling of lithium-ion battery degradation for cell life assessment. IEEE Trans Smart Grid 2016;9(2):1131–40.
[12] Yang F, Wang D, Xing Y, Tsui K-L. Prognostics of Li (NiMnCo) O2-based lithium-ion batteries using a novel battery degradation model. Microelectron Reliab 2017;70:70–8.

[13] Xing Y, Ma EW, Tsui K-L, Pecht M. "A case study on battery life prediction using particle filtering," in Proceedings of the IEEE 2012 Prognostics and System Health Management Conference (PHM-2012 Beijing). In: IEEE; 2012. p. 1–6.

[14] Walker E, Rayman S, White RE. Comparison of a particle filter and other state estimation methods for prognostics of lithium-ion batteries. J Power Sources 2015; 287:1–12.

[15] A. P. Lall, B. H. Zhang, and C. R. Lall, "PHM of state-of-charge for flexible power sources in wearable electronics with EKF," in *2018 IEEE International Reliability Physics Symposium (IRPS)*, 2018: IEEE, pp. P-SR. 2-1-P-SR. 2-6.

[16] Zhang X, Miao Q, Liu Z. Remaining useful life prediction of lithium-ion battery using an improved UPF method based on MCMC. Microelectron Reliab 2017;75: 288–95.

[17] Zhang H, Miao Q, Zhang X, Liu Z. An improved unscented particle filter approach for lithium-ion battery remaining useful life prediction. Microelectron Reliab 2018; 81:288–98.

[18] Liu J, Wang W, Ma F. A regularized auxiliary particle filtering approach for system state estimation and battery life prediction. Smart Mater Struct 2011;20(7): 075021.

[19] Shen S, Sadoughi M, Chen X, Hong M, Hu C. A deep learning method for online capacity estimation of lithium-ion batteries. J Storage Mater 2019;25:100817.

[20] Saha B, Goebel K, Christophersen J. Comparison of prognostic algorithms for estimating remaining useful life of batteries. Trans Inst Meas Control 2009;31 (3–4):293–308.

[21] Long B, Xian W, Jiang L, Liu Z. An improved autoregressive model by particle swarm optimization for prognostics of lithium-ion batteries. Microelectron Reliab 2013;53(6):821–31.

[22] Zhou Z, et al. "Lithium-ion battery remaining useful life prediction under grey theory framework," in 2014 Prognostics and System Health Management Conference (PHM-2014 Hunan). In: IEEE; 2014. p. 297–300.

[23] Zhang Z, Si X, Hu C, Lei Y. Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods. Eur J Oper Res 2018;271 (3):775–96.

[24] Y. Chen, J. Bao, Z. Xiang, and W. Jian, "Predictability analysis of lithium-ion battery remaining useful life with multiscale entropy," in *2013 International Conference on Computational and Information Sciences*, 2013: IEEE, pp. 1052-1055.

[25] Richardson RR, Birkl CR, Osborne MA, Howey DA. Gaussian process regression for in situ capacity estimation of lithium-ion batteries. IEEE Trans Ind Inf 2018;15(1): 127–38.

[26] Liu D, Pang J, Zhou J, Peng Y, Pecht M. Prognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression. Microelectron Reliab 2013;53(6):832–9.

[27] D. Liu, J. Pang, J. Zhou, and Y. Peng, "Data-driven prognostics for lithium-ion battery based on Gaussian Process Regression," in *Proceedings of the IEEE 2012 Prognostics and System Health Management Conference (PHM-2012 Beijing)*, 2012: IEEE, pp. 1-5.

[28] Richardson RR, Osborne MA, Howey DA. Gaussian process regression for forecasting battery state of health. J Power Sources 2017;357:209–19.

[29] L. Li, P. Wang, K.-H. Chao, Y. Zhou, and Y. Xie, "Remaining useful life prediction for lithium-ion batteries based on Gaussian processes mixture," *PloS one,* vol. 11, no. 9, p. e0163004, 2016.

[30] Liu D, Zhou J, Pan D, Peng Y, Peng X. Lithium-ion battery remaining useful life estimation with an optimized relevance vector machine algorithm with incremental learning. Measurement 2015;63:143–51.

[31] Wang D, Miao Q, Pecht M. Prognostics of lithium-ion batteries based on relevance vectors and a conditional three-parameter capacity degradation model. J Power Sources 2013;239:253–64.

[32] Zhou Y, Huang M, Chen Y, Tao Y. A novel health indicator for on-line lithium-ion batteries remaining useful life prediction. J Power Sources 2016;321:1–10.

[33] Tao T, Zhao W. "A support vector regression-based prognostic method for li-ion batteries working in variable operating states," in 2016 Prognostics and System Health Management Conference (PHM-Chengdu). In: IEEE; 2016. p. 1–5.

[34] J. Zhou, D. Liu, Y. Peng, and X. Peng, "Dynamic battery remaining useful life estimation: An on-line data-driven approach," in *2012 IEEE international instrumentation and measurement technology conference proceedings*, 2012: IEEE, pp. 2196-2199.

[35] Wang S, Zhao L, Su X, Ma P. Prognostics of lithium-ion batteries based on battery performance analysis and flexible support vector regression. Energies 2014;7(10): 6492–508.

[36] Mansouri SS, Karvelis P, Georgoulas G, Nikolakopoulos G. Remaining useful battery life prediction for UAVs based on machine learning. IFAC-PapersOnLine 2017;50(1):4727–32.

[37] Donato TH, Quiles MG. Machine learning systems based on xgBoost and MLP neural network applied in satellite lithium-ion battery sets impedance estimation. Adv Computational Intelligence: An Int J (ACII) 2018;5:1–20.

[38] You G-W, Park S, Oh D. Real-time state-of-health estimation for electric vehicle batteries: A data-driven approach. Appl Energy 2016;176:92–103.

[39] Liu J, Saxena A, Goebel K, Saha B, Wang W. An adaptive recurrent neural network for remaining useful life prediction of lithium-ion batteries. National Aeronautics And Space Administration Moffett Field CA Ames Research 2010.

[40] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Networks 1994;5(2):157–66.

[41] Zhang Y, Xiong R, He H, Pecht MG. Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. IEEE Trans Veh Technol 2018;67(7):5695–705.

[42] Hsu C-W, Xiong R, Chen N-Y, Li J, Tsou N-T. Deep neural network battery life and voltage prediction by using data of one cycle only. Appl Energy 2022;306:118134.

[43] Li H, Zhao W, Zhang Y, Zio E. Remaining useful life prediction using multi-scale deep convolutional neural network. Appl Soft Comput 2020;89:106113.

[44] Li P, et al. State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network. J Power Sources 2020;459:228069.

[45] Cheng G, Wang X, He Y. Remaining useful life and state of health prediction for lithium batteries based on empirical mode decomposition and a long and short memory neural network. Energy 2021;232:121022.

[46] Pastor-Fernández C, Uddin K, Chouchelamane GH, Widanage WD, Marco J. A comparison between electrochemical impedance spectroscopy and incremental capacity-differential voltage as Li-ion diagnostic techniques to identify and quantify the effects of degradation modes within battery management systems. J Power Sources 2017;360:301–18.

[47] Pop V, Bergveld HJ, Regtien PPL, Op het Veld JHG, Danilov D, Notten PHL. Battery aging and its influence on the electromotive force. J Electrochem Soc 2007;154(8): A744.

[48] Kwak E, Jeong S, Kim J-H, Oh K-Y. Prediction of compression force evolution over degradation for a lithium-ion battery. J Power Sources 2021;483:229079.

[49] H. J. Blinchikoff and A. I. Zverev, "Filtering in the time and frequency domains," 1976.

[50] Severson KA, Attia PM, Jin N, Perkins N, Jiang B, Yang Zi, et al. Data-driven prediction of battery cycle life before capacity degradation. Nat Energy 2019;4(5): 383–91.

[51] Fatemi A, Yang L. Cumulative fatigue damage and life prediction theories: a survey of the state of the art for homogeneous materials. Int J Fatigue 1998;20(1):9–34.

[52] Gyenes B, Stevens DA, Chevrier VL, Dahn JR. Understanding anomalous behavior in coulombic efficiency measurements on Li-ion batteries. J Electrochem Soc 2015; 162(3):A278–83.

[53] Lewerenz M, Münnix J, Schmalstieg J, Käbitz S, Knips M, Sauer DU. Systematic aging of commercial LiFePO4| Graphite cylindrical cells including a theory explaining rise of capacity during aging. J Power Sources 2017;345:254–63.

[54] A. Dourado and F. A. Viana, "Physics-informed neural networks for corrosion-fatigue prognosis," in *Proceedings of the Annual Conference of the PHM Society*, 2019, vol. 11, no. 1.

[55] Nascimento RG, Viana FA. Cumulative damage modeling with recurrent neural networks. AIAA Journal 2020;58(12):5459–71.

[56] Nascimento RG, Viana FA. Fleet prognosis with physics-informed recurrent neural networks. arXiv preprint arXiv:1901.05512. 2019.

[57] M. A. Chao, C. Kulkarni, K. Goebel, and O. Fink, "Fusing physics-based and deep learning models for prognostics," *arXiv preprint arXiv:2003.00732,* 2020.

[58] Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J Comput Phys 2019;378:686–707.

[59] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations," *arXiv preprint arXiv:1711.10561,* 2017.

[60] Yang L, Zhang D, Karniadakis GE. Physics-informed generative adversarial networks for stochastic differential equations. arXiv preprint arXiv:1811.02033. 2018.

[61] Kim SW, Kim I, Lee J, Lee S. Knowledge Integration into deep learning in dynamical systems: an overview and taxonomy. J Mech Sci Technol 2021:1–12.

[62] M. Diligenti, S. Roychowdhury, and M. Gori, "Integrating prior knowledge into deep learning," in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2017: IEEE, pp. 920-923.

[63] Dourado AD, Viana F. Physics-informed neural networks for bias compensation in corrosion-fatigue. In: in *AIAA Scitech 2020 Forum*; 2020. p. 1149.

[64] He W, Williard N, Osterman M, Pecht M. Prognostics of lithium-ion batteries based on Dempster-Shafer theory and the Bayesian Monte Carlo method. J Power Sources 2011;196(23):10314–21.

[65] Laadissi E, Anas E, Zazi M, Jaouad K. Parameter identification of a lithium-ion battery model using Levenberg-Marquardt algorithm. J Eng Appl Sci 2019;14: 1267–73.

[66] Neal RM. Bayesian learning for neural networks. Springer Science & Business Media; 2012.

[67] A. Graves, "Practical variational inference for neural networks," in *Advances in neural information processing systems*, 2011: Citeseer, pp. 2348-2356.

[68] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114,* 2013.

[69] Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature 2015;521(7553):452–9.

[70] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*, 2016: PMLR, pp. 1050-1059.

[71] Roman D, Saxena S, Robu V, Pecht M, Flynn D. "Machine learning pipeline for battery state-of-health estimation," *Nature*. Machine Intelligence 2021;3(5): 447–56.

[72] Tang X, Liu K, Li K, Widanage WD, Kendrick E, Gao F. Recovering large-scale battery aging dataset with machine learning. Patterns 2021;2(8):100302.

[73] MacKay DJ. A practical Bayesian framework for backpropagation networks. Neural Comput 1992;4(3):448–72.

[74] Snoek J, Larochelle H, Adams RP. Practical bayesian optimization of machine learning algorithms. arXiv preprint arXiv:1206.2944. 2012.

[75] P. I. Frazier, "A tutorial on Bayesian optimization," *arXiv preprint arXiv: 1807.02811,* 2018.

[76] Binder A, Bach S, Montavon G, Müller K-R, Samek W. "Layer-wise relevance propagation for deep neural network architectures," in *Information science and applications (ICISA)*. Springer 2016;2016:913–22.
[77] May RJ, Maier HR, Dandy GC, Fernando TG. Non-linear variable selection for artificial neural networks using partial mutual information. Environ Modell Software 2008;23(10–11):1312–26.
[78] R. Seoh, "Qualitative analysis of monte carlo dropout," *arXiv preprint arXiv: 2007.01720,* 2020.
[79] Zhang J, Wang Z, Liu P, Zhang Z. Energy consumption analysis and prediction of electric vehicles based on real-world driving data. Appl Energy 2020;275:115408.
[80] Li Yi, Zou C, Berecibar M, Nanini-Maury E, Chan J-W, van den Bossche P, et al. Random forest regression for online capacity estimation of lithium-ion batteries. Appl Energy 2018;232:197–210.
[81] F. O. Heimes, "Recurrent neural networks for remaining useful life estimation," in *2008 international conference on prognostics and health management*, 2008: IEEE, pp. 1-6.
[82] S. Zheng, K. Ristovski, A. Farahat, and C. Gupta, "Long short-term memory network for remaining useful life estimation," in *2017 IEEE international conference on prognostics and health management (ICPHM)*, 2017: IEEE, pp. 88-95.
[83] Q. Wang, S. Zheng, A. Farahat, S. Serita, and C. Gupta, "Remaining useful life estimation using functional data analysis," in *2019 ieee international conference on prognostics and health management (icphm)*, 2019: IEEE, pp. 1-8.
[84] 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.
[85] Tang X, Liu K, Wang X, Gao F, Macro J, Widanage WD. Model migration neural network for predicting battery aging trajectories. IEEE Trans Transp Electrif 2020;6 (2):363–74.
[86] Liu K, Tang X, Teodorescu R, Gao F, Meng J. Future ageing trajectory prediction for lithium-ion battery considering the knee point effect. IEEE Trans Energy Convers 2021.
[87] Satopaa V, Albrecht J, Irwin D, Raghavan B. In: IEEE; 2011. p. 166–71.
[88] 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.