Lithium-Ion Battery SOH Estimation via Mixture Correntropy Loss-Based PCA and Kolmogorov-Arnold-Enhanced Liquid Neural Networks

Wentao Ma, *Senior Member, IEEE*, Zhuo Li, Peng Guo, Yang Li, *Senior Member, IEEE*, and Badong Chen, *Senior Member, IEEE*

***Abstract*—Accurate state of health (SOH) estimation of lithium-ion batteries is essential for reliable battery management but is particularly challenging under mixed non-Gaussian measurement noise and data scarcity. This article proposes a robust SOH estimation framework that combines noise-resistant feature extraction with data-efficient, interpretable continuous-time dynamic modeling. A mixture correntropy loss-based principal component analysis (MCL-PCA) method is first developed, which employs a hybrid Gaussian–Laplacian correntropy criterion to adaptively suppress mixed non-Gaussian noise and extract stable low-dimensional health features (HFs). These features serve as inputs to a newly designed Kolmogorov-Arnold-enhanced liquid neural network (KLNN), which augments continuous-time liquid dynamics with structured nonlinear mappings to improve stability and nonlinear generalization under limited small-sample conditions. These two components are integrated into a unified framework that yields coherent and physically consistent SOH degradation trajectories. Experimental results on laboratory and public datasets demonstrate that MCL-PCA significantly improves feature robustness, while KLNN achieves superior SOH estimation accuracy in noisy and small-sample scenarios, resulting in notably lower prediction errors than conventional PCA-based methods and consistent advantages over baseline models.**

***Index Terms*—KAN-enhanced liquid neural network (KLNN), lithium-ion batteries, mixture correntropy loss-based principal component analysis (MCL-PCA), small-sample learning, state-of-health (SOH) estimation.**

Received 16 December 2025; revised 24 February 2026; accepted 19 March 2026. This work was supported by the National Natural Science Foundation of China under Grant 62473308. *(Corresponding authors: Wentao Ma; Yang Li.)*

Wentao Ma, Zhuo Li, and Peng Guo are with the School of Electrical Engineering, Xi'an University of Technology, Xi'an 710048, China (e-mail: mawt@xaut.edu.cn; 2241921160@stu.xaut.edu.cn; 2211921110@stu.xaut.edu.cn).

Yang Li is with the School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China (e-mail: yangli@ieee.org).

Badong Chen is with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710048, China (e-mail: chenbd@mail.xjtu.edu.cn).

Digital Object Identifier 10.1109/TIE.2026.3679777

I. Introduction

LITHIUM-ION batteries (LIBs) are widely used in portable electronics, electric vehicles, and grid-scale storage systems due to their high energy density and long lifespan. The safe and reliable operation of LIB systems relies on the battery management systems (BMSs), within which health monitoring plays a critical role. Among various diagnostic indicators, the state of health (SOH) characterizes capacity degradation and reflects the available performance and safety margins of a battery [1]. However, accurate SOH estimation remains challenging in practical applications. Measurement signals are often corrupted by mixed non-Gaussian noise induced by sensor imperfections, impulsive disturbances, abrupt operating fluctuations, and nonstationary operating conditions. Furthermore, the availability of labeled degradation data is typically limited under realistic operating scenarios [2], [3]. These factors jointly complicate robust feature extraction and reliable modeling of battery degradation dynamics.

Approaches for SOH estimation generally fall into two categories: model-based and data-driven methods. Model-based approaches, such as equivalent-circuit models [4] and electrochemical models [5], offer strong physical interpretability but require extensive parameter calibration and are highly sensitive to operating conditions, which limits their suitability for real-time BMS applications [6]. Conversely, data-driven methods learn degradation patterns directly from operational data and can capture complex nonlinear behaviors [7]. However, their performance critically depends on two tightly coupled components: extracting robust health features (HFs) from noisy measurements and building models that maintain generalization and interpretability under limited data.

Accurate SOH estimation depends on extracted HFs that are robust to noise and consistent over time. Early approaches focused on empirical features manually extracted from measured curves under constant current-constant voltage protocols or incremental capacity-differential voltage patterns [8]. Although these features offer interpretability, their high sensitivity to measurement noise and dynamic operating condition variations often perform unreliable

SOH estimation in practical applications [9]. This has motivated a shift toward automated feature extraction, pursued along two primary directions. One line of work leverages deep learning (DL) models, such as convolutional neural networks (CNNs) and attention-based long short-term memory (LSTM) networks, to directly learn latent representations from raw measurement signals [10], [11]. The other line of research focuses on lightweight automated feature extraction based on unsupervised dimensionality-reduction techniques, including independent component analysis (ICA) [12], nonnegative matrix factorization (NMF) [13], and principal component analysis (PCA) [14]. Among these methods, PCA has become the canonical baseline for SOH estimation because it can efficiently compress high-dimensional voltage-capacity measurements into a small set of dominant variance components, thereby reducing noise and multicollinearity in subsequent modeling stages [14]. However, classical PCA relies on a mean square error (MSE) criterion, which makes it highly susceptible to outliers, impulsive disturbances, and heavy-tailed measurement noise [15]. Although robust variants such as maximum correntropy criterion (MCC) based PCA (termed MCC-PCA) [16] and functional PCA [17] alleviate certain noise effects, they remain inadequate under the mixed non-Gaussian noise commonly encountered in battery data. Consequently, achieving reliable feature extraction in such heterogeneous noise environments remains an open and critical challenge.

With stable and accurate HFs obtained, constructing a high-fidelity mapping from these features to SOH becomes the next critical challenge. As battery degradation mechanisms grow increasingly complex, DL models [18] have emerged as the dominant paradigm owing to their strong capability in modeling temporal dependencies. Among them, recurrent neural networks (RNNs) such as LSTM [10] and gated recurrent units (GRUs) [19] networks are widely adopted for capturing temporal dynamics in sequential data. However, their performance typically relies on large volumes of data and tends to suffer from overfitting or underfitting under small-sample conditions, leading to degraded estimation accuracy [20]. In contrast, traditional machine-learning methods such as support vector regression (SVR) [21] and Gaussian process regression (GPR) [22] exhibit better generalization and computational efficiency when data are limited, but they lack the expressive capacity needed to characterize highly nonlinear degradation behaviors. Therefore, developing an SOH estimation framework that is both interpretable and accurate under small-sample constraints continues to be a key challenge. To this end, inspired by biological neural dynamics, Hasani et al. [23] introduced learnable time constants into first-order differential equations to enable adaptive modeling of nonstationary temporal behaviors, thereby proposing a new class of RNNs known as liquid neural networks (LNNs). Compared with conventional RNNs, LNNs leverage an ordinary differential equation (ODE)-based formulation that provides explicit multitimescale dynamics and strong inductive biases (IB), which often leads to improved generalization under limited data conditions [24], [25]. Nevertheless, their nonlinear representation capability remains restricted by fixed activation functions, limiting their ability to emphasize critical and abrupt degradation information. Recently, Kolmogorov-Arnold networks (KANs) have attracted growing attention owing to their superior nonlinear expressiveness and structural transparency [26]. By decomposing multivariate functions into compositions of univariate functions, KANs achieve high-precision nonlinear approximation with fewer parameters while maintaining strong inherent interpretability, providing a more interpretable alternative to traditional neural architectures [27].

Motivated by the above challenges, an interpretable SOH estimation framework integrating robust feature extraction with continuous-time neural dynamics is proposed to achieve accurate and reliable SOH estimation under limited and noisy data. The proposed framework advances existing PCA methods, which typically depend on MSE loss and single-kernel correntropy, by extending conventional MCC-PCA through the introduction of a mixture correntropy loss. This enhancement enables effective handling of mixed non-Gaussian noise during the feature extraction stage for SOH estimation. Moreover, unlike KAN based LSTM (KLSTM) [28] type architectures that use KAN modules as static nonlinear replacements for affine transformations in discrete-time recurrent updates, a novel estimation method, called Kolmogorov-Arnold-enhanced LNN (KLNN), is developed by integrating structured KAN mappings into the observation and gating pathways of a continuous-time liquid neural system. This design preserves the inherent ODE-driven dynamics and architectural bias of LNNs while significantly enhancing nonlinear representation capability. The key contributions are given as follows.

1) To address feature distortion in SOH estimation under non-Gaussian measurement noise, this work proposes a mixture correntropy loss-based PCA (MCL-PCA) for feature extraction. By replacing the conventional MSE with a Gaussian–Laplacian mixture correntropy loss (MCL), MCL-PCA adaptively suppresses outliers and heavy-tailed disturbances through kernel-based local similarity weighting, thereby enabling stable low-dimensional health-feature extraction from noisy operational data.
2) To capture the complex nonlinear dynamics in battery degradation, this work develops a KLNN for SOH estimation. The proposed model integrates a KAN-based nonlinear mapping module into the continuous-time dynamics of the LNN via an adaptive gating mechanism (AGM), thereby enhancing its nonlinear representation capacity while preserving the intrinsic ability of LNNs to model temporal evolution.
3) To overcome limitations in decoupled feature-temporal modeling, including inadequate feature fusion and limited end-to-end learning, a cooperative framework combining MCL-PCA and KLNN is proposed that enables end-to-end joint optimization from noisy signals to SOH estimation. Experimental results on both laboratory and public battery datasets demonstrate the framework's superior performance under noisy and small-sample conditions.

The remainder of this article is organized as follows. Section II presents the proposed MCL-PCA method. Section III introduces the KLNN-based SOH estimation model. Section IV describes the experimental setup. Section V discusses the results and analysis. Finally, Section VI concludes the article and outlines future research directions.

II. Robust Feature Extraction Method Based on Mixture Correntropy Loss-Based PCA

A. Mixture Correntropy Loss

Although correntropy provides inherent robustness against outliers through localized similarity evaluation, its effectiveness strongly depends on the choice of a single kernel width, making it unreliable under complex non-Gaussian noise. To address this limitation, this work replaces the conventional MSE objective in PCA with MCL, which adaptively suppresses non-Gaussian noise components and stabilizes the extracted HFs. Given two random variables A and B, the correntropy is defined as follows [29]

$$V(A,B) = E[\kappa(A,B)] = E[\kappa(A-B)] \tag{1}$$

where $E(\cdot)$ denotes the expectation operator, and $\kappa(\cdot)$ is the kernel function. In the traditional MCC. A Gaussian kernel is typically employed and is defined as

$$\kappa_G(A,B) = \exp\left(-\frac{(A-B)^2}{2\sigma_G^2}\right) \tag{2}$$

where σ_G denotes the kernel width.

In fact, the samples are generally finite and of unknown distribution, thus, the sample estimator of correntropy for a given sample set $\{(a_i,b_i)\}_{i=1}^{N}$ can usually be used as

$$\hat{V}(A,B) = \frac{1}{N}\sum_{i=1}^{N}\kappa(a_i,b_i). \tag{3}$$

To further enhance the adaptability and robustness of correntropy when dealing with mixed non-Gaussian noise conditions, this article defines a mixture correntropy (MC) that integrates Gaussian and Laplacian kernels, formulated as

$$\kappa_M(e_i) = \alpha\exp\left(-\frac{e_i^2}{2\sigma_G^2}\right) + (1-\alpha)\exp\left(-\frac{|e_i|}{\sigma_L}\right) \tag{4}$$

where $\alpha \in [0,1]$ is the mixing coefficient, $e_i = a_i - b_i$ denotes the error of the i-th sample, and σ_L is the width parameter of the Laplacian kernel, respectively. The Gaussian kernel emphasizes small residuals due to its rapid exponential decay, whereas the heavier-tailed Laplacian kernel is more resilient to large deviations. Substituting this mixture kernel into the sample correntropy estimator in (3) yields the following:

$$\hat{V}(e_i) = \frac{1}{N}\sum_{i=1}^{N}\kappa_M(e_i). \tag{5}$$

Then, the MCL can be defined as

$$L_{MCL}(e_i) = 1 - \hat{V}(e_i). \tag{6}$$

B. MCL-PCA Method

Conventional PCA with MSE is highly sensitive to outliers and impulsive disturbances, which distort the estimated principal subspace and degrade feature-extraction performance. To enhance robustness, we replace the MSE with the MCL defined in (6), enabling PCA to adaptively down-weight aberrant samples while preserving the intrinsic variance structure. Given the sample matrix $\mathbf{A} = [a_1, a_2, \ldots, a_N] \in \mathbb{R}^{m\times N}$, which is composed of N data samples each of dimension m, the projection matrix $\mathbf{U} \in \mathbb{R}^{m\times k}$ (satisfying $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}$, where $(\cdot)^{\mathrm{T}}$ denotes the transpose and $\mathbf{I}$ is the identity matrix of compatible dimension), and the mean vector μ, the objective function of conventional PCA is formulated as

$$\min_{\mathbf{U},\mu} J_{MSE}(\mathbf{U},\mu) = \sum_{i=1}^{N}\|a_i - \mu - \mathbf{U}\mathbf{U}^{\mathrm{T}}(a_i-\mu)\|_2^2. \tag{7}$$

Here, the objective in (7) is fundamentally reformed by the MCL, leading to the following optimization objective:

$$\min_{\mathbf{U},\mu} J_{MCL}(\mathbf{U},\mu) = \sum_{i=1}^{N} L_{MCL}(\|r_i\|_2) \tag{8}$$

where $r_i = a_i - \mu - \mathbf{U}\mathbf{U}^T(a_i-\mu)$ denotes the reconstruction residual of the i-th sample, and L_{MCL} is defined in (6). Due to the nonlinear and nonconvex nature of (8), we employ the half-quadratic (HQ) optimization framework to convert it into a tractable weighted quadratic form. By using convex-conjugate theory, an auxiliary variable p_i is introduced for each sample, allowing the MCL term to be decomposed into an equivalent convex representation

$$L_{MCL}(\|r_i\|_2) = \inf_{p_i}\left\{\frac{1}{2}p_i\|r_i\|_2^2 + \varphi(p_i)\right\} \tag{9}$$

where $\varphi(\cdot)$ represents the convex conjugate function associated with the mixture kernel. For a fixed residual $\|r_i\|_2$, the optimal solution p_i^* for the above problem with respect to p_i is given by the stationary point condition

$$\frac{\partial}{\partial p_i}\left[\frac{1}{2}p_i\|r_i\|_2^2 + \varphi(p_i)\right] \Rightarrow \frac{1}{2}\|r_i\|_2^2 + \varphi'(p_i^*) = 0. \tag{10}$$

Solving (10) in conjunction with the specific form of the mixture kernels yields the closed-form expression for the theoretical optimal solution p_i^*

$$p_i^* = \frac{\alpha}{2\sigma_G^2}\exp\left(-\frac{\|r_i\|_2^2}{2\sigma_G^2}\right) + \frac{1-\alpha}{2\sigma_L\|r_i\|_2}\exp\left(-\frac{\|r_i\|_2}{\sigma_L}\right). \tag{11}$$

The second term on the right-hand side of (11) contains an inverse-residual factor $1/\|r_i\|_2$ which diverges as the residual approaches zero, rendering direct use of p_i^* numerically unstable. To address this numerical issue and obtain a more interpretable weighting scheme, we define a stable adaptive weight

$$w_i = \frac{\alpha}{\sigma_G^2}\exp\left(-\frac{\|r_i\|_2^2}{2\sigma_G^2}\right) + \frac{(1-\alpha)}{\sigma_L}\exp\left(-\frac{\|r_i\|_2}{\sigma_L}\right) \tag{12}$$

where w_i decreases monotonically with the residual magnitude, providing natural suppression of outliers. Substituting w_i into the HQ-augmented objective shows that the term associated with the original MCL potential depends only on the weights computed from the previous iteration and is therefore constant with respect to the optimization variables. Removing this constant simplifies the objective, reducing the MCL term to a weighted residual minimization

$$\min_{\mathbf{U},\mu} \sum_{i=1}^{N} w_i \|r_i\|_2^2. \tag{13}$$

Thus, each HQ iteration reduces the nonconvex MCL-PCA objective to a weighted PCA update, in which kernel weights decay with residual magnitude, yielding robustness to outliers while preserving the principal subspace. In implementation, the Gaussian and Laplacian kernel widths, together with the mixing coefficient α, are determined via Bayesian optimization with cross-validation prior to HQ optimization and then fixed during the subsequent iterative updates, while the auxiliary sample weights are iteratively updated based on the residuals during the HQ alternating procedure.

Based on (12), the weighted mean μ and the weighted covariance matrix C_W are computed as

$$\mu = \frac{\sum_{i}^{N} w_i a_i}{\sum_{i}^{N} w_i}, C_w = \frac{\sum_{i=1}^{N} w_i (a_i - \mu)(a_i - \mu)^{\mathrm{T}}}{\sum_{i}^{N} w_i}. \tag{14}$$

Subsequently, eigenvalue decomposition is performed on C_w, and the projection matrix $\mathbf{U}$ is updated by selecting the eigenvectors corresponding to the top k largest eigenvalues

$$C_w \mathbf{U} = \mathbf{U}\mathbf{\Lambda} \tag{15}$$

where $\mathbf{\Lambda}$ is a diagonal matrix with diagonal elements equal to the eigenvalues, typically arranged in descending order.

At the end of each iteration, the value of the original MC objective J_{MCL} is recomputed according to (8). Convergence is achieved when the relative change between two consecutive objective values falls below a predefined tolerance ε, set to 10^{-6} in our implementation to ensure stable and efficient optimization. The stopping criterion is formally given by

$$\frac{|J_{MCL}^{(t+1)} - J_{MCL}^{(t)}|}{|J_{MCL}^{(t+1)}|} < \varepsilon \tag{16}$$

where $J_{MCL}^{(t)}$ denotes the objective value computed at the i-th iteration. This adaptive criterion terminates the optimization once the solution stabilizes, providing a favorable balance between accuracy and computational cost.

To justify this stopping rule, note that the alternating minimization scheme yields a monotonically decreasing objective sequence, ensuring stable convergence. The final output of the MCL-PCA algorithm is shown as follows:

$$\mathbf{Z} = (\mathbf{A} - \mu)\mathbf{U}, \hat{\mathbf{A}} = \mathbf{Z}\mathbf{U}^{\mathrm{T}} + \mu, \mathbf{E} = \mathbf{A} - \hat{\mathbf{A}} \tag{17}$$

where $\mathbf{Z}$ denotes the low-dimensional principal component representation, which serves as the robust HFs and is fed directly into the downstream KLNN model for SOH estimation. The matrices $\hat{\mathbf{A}}$ and $\mathbf{E}$ represent the reconstructed samples and the residual matrix, respectively. The norm of $\mathbf{E}$ provides a quantitative measure of robustness, reflecting the algorithm's ability to suppress mixed noise.

Algorithm 1: MCL-PCA.

Input: Data matrix $\mathbf{A}$, kernel parameters σ_G, σ_L mixing coefficient α, convergence threshold ε, target Dimension d.
Output: Projection matrix $\mathbf{U}$, weighted mean μ, low- dimensional representation $\mathbf{Z}$.
Preprocessing: Determine kernel parameters σ_G and σ_L using Bayesian optimization on the validation dataset.
1: Initialize $\mathbf{U}$, initialize $\mu = \text{mean}(\mathbf{A})$ via standard PCA.
2: **repeat**
3: Compute reconstruction residuals r_i for all samples.
4: Update w_i for all samples according to (12).
5: Update μ according to (14).
6: Update $\mathbf{U}$ by solving the eigenvalue problem in (15).
7: **if** the difference of the MCL in (8) is smaller than ε, **then**
8: Break loop.
9: **end if**
10: **until** convergence
11: Compute low-dimensional representation $\mathbf{Z} = (\mathbf{A} - \mu)\mathbf{U}$.

III. Proposed KLNN Model for SOH Estimation

A. Liquid Neural Network

The degradation of LIBs is a continuous-time process, combining long-term capacity fade with short-term fluctuations. Discrete-time models, operating on discrete observations, often struggle to accurately capture these cross-timescale dynamics, resulting in suboptimal SOH estimation. LNN addresses this limitation by modeling system evolution through ODEs with learnable time constants, yielding a physically interpretable continuous-time framework. Central to this framework is the liquid time-constant (LTC) neuron, with state dynamics described as

$$\begin{aligned}\dot{\mathbf{h}}(t) &= \frac{d\mathbf{h}(t)}{dt} = \frac{-\mathbf{h}(\mathbf{t})}{\boldsymbol{\tau}_i} + f_{lnn}(\mathbf{h}(t), \mathbf{x}(t))(\mathbf{E}_i - \mathbf{h}(t)) \\ &= -\frac{\mathbf{h}(t)}{\boldsymbol{\tau}_i} + f_{lnn}(\mathbf{W}_{in}\mathbf{x}(t) + \mathbf{W}_{rec}\mathbf{h}(t) + \mathbf{b}_t)(\mathbf{E}_i - \mathbf{h}(t))\end{aligned} \tag{18}$$

where $\mathbf{x}(t)$ denotes the input feature vector at time t, $\mathbf{h}(t)$ represents the hidden state vector, $\mathbf{W}_{in}$ and $\mathbf{W}_{rec}$ are the input and recurrent weight matrices respectively, $\mathbf{b}_i$ is the bias vector, and the key interpretability parameter τ_i is the time constant vector governing the decay or response rate of i-th neuronal state. The nonlinear activation function $f(\cdot)$ is typically chosen as $\tanh(\cdot)$, and E_i indicates the reversal potential constraining the state range. The first term on the right-hand side embodies the system's natural decay mechanism, while the second term captures input–state coupling over time.

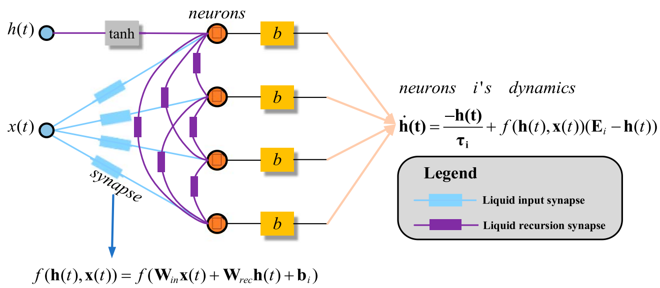

Fig. 1. Structure of the liquid time-constant (LTC) network.

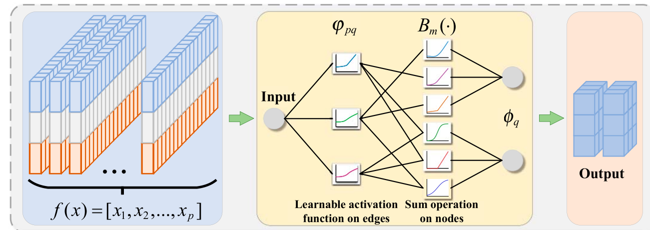

Fig. 2. Architecture of the Kolmogorov–Arnold network (KAN).

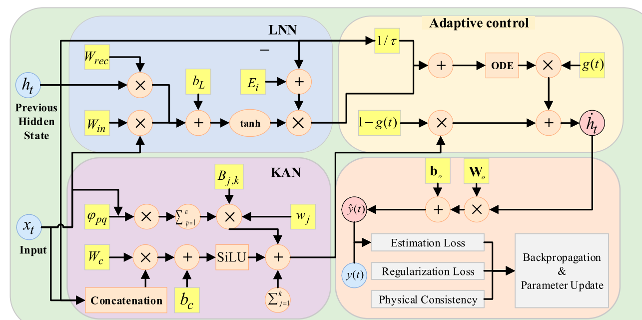

Fig. 3. Architecture of the proposed KLNN.

As illustrated in Fig. 1, adaptive time constants allow hidden states evolve continuously, enabling concurrent modeling of long-term and short-term degradation dynamics while retaining interpretability. However, the fixed nonlinear form limits expressive flexibility for abrupt or highly nonlinear degradation, motivating architectures with enhanced nonlinear representation.

B. Kolmogorov-Arnold Network (KAN)

To overcome the limited nonlinear representation capability of LNN, we introduce the KAN architecture to characterize the intricate degradation patterns. Different from the conventional multilayer perceptron (MLP) that relies on fixed activation functions with learnable weights, KAN employs learnable univariate functional modules, thereby providing substantially greater flexibility in approximating nonlinear mappings.

The expressive power of KAN is theoretically supported by the Kolmogorov–Arnold representation theorem, which states that any multivariate continuous function $f(x_1, x_2, \ldots, x_n)$ can be represented as a finite composition of univariate functions

$$f_{kan}(x) = \sum_{q=1}^{Q} \Phi_q \left(\sum_{p=1}^{d} \phi_{pq}(x_p) \right) \tag{19}$$

where Φ_q performs aggregation of transformed inputs, while each $\phi_{pq}(x_p)$ individually transforms the p-th input feature. The indices p and q denote transformation stages, d represents the input dimension, and Q indicates the decomposition rank. This formulation allows KAN to construct high-dimensional mappings through combinations of low-dimensional nonlinear functions, thereby enhancing both approximation capability and interpretability.

As illustrated in Fig. 2, the KAN module first constructs independent univariate functions $\phi_{pq}(\cdot)$ for each dimension of the input vector $x = [x_1, \ldots, x_d]$, implemented using B-spline basis expansions. These transformed features are subsequently aggregated to achieve an efficient representation of multivariate functions. At the implementation level, each univariate function is parameterized as

$$\phi_{pq}(x_p) = \sum_{m} c_{pq,m} \cdot B_m(x_p) \tag{20}$$

where $B_m(\cdot)$ denotes the B-spline basis functions and $c_{pq,m}$ represents the trainable coefficients. In (19), each univariate function such as $\phi_{q,p}$ and Φ_q has its own independent set of coefficients $c_{pq,m}$. Therefore, the complete KAN module is parameterized by a large collection of all such coefficient sets across the network, which will be denoted as θ_{KAN}.

C. The Proposed KLNN Model Architecture

To integrate the continuous-time dynamics of LNN with enhanced nonlinear modeling capability, a novel KLNN framework is proposed in this subsection. While KAN provides strong nonlinear representation capability, prior studies have mainly integrated it into recurrent architectures via straightforward activation replacement, as seen in KLSTM [28]. Such substitution is well-suitable for discrete-time networks, where activations serve as static nonlinearities. In contrast, within continuous-time LNNs, these nonlinear functions are embedded in the state evolution operator; directly replacing them therefore alters the governing dynamics and may compromise the structural consistency of the underlying ODE system. To overcome this, we introduce an AGM that couples LNN and KAN through parallel pathways, retaining dynamical consistency while substantially improving nonlinear expressiveness. Notably, the KAN branch operates as a static nonlinear mapping of the input features without modifying the intrinsic ODE structure of LNN state evolution. The cubic B-spline basis functions defining this mapping are fixed in form, with only their spline coefficients treated as trainable parameters during optimization. Consequently, the continuous-time stability properties inherited to the LNN backbone are structurally preserved. The overall network architecture is shown in Fig. 3.

Specifically, the state evolution equation of KLNN is reformulated as follows:

$$\begin{aligned}\dot{\mathbf{h}}(t) &= \frac{d\mathbf{h}(t)}{dt} = F(\mathbf{h}(t), \mathbf{x}(t), t, \Theta)\\ &= \mathbf{g}(t) f_{lnn}(\mathbf{h}(t), \mathbf{x}(t)) + (1-\mathbf{g}(t)) f_{kan}(\mathbf{h}(t), \mathbf{x}(t))\\ &= \mathbf{g}(t)\Bigg(\frac{-\mathbf{h}(t)}{\tau_i} + f(\mathbf{W}_{in}\mathbf{x}(t) + \mathbf{W}_{rec}\mathbf{h}(t) + \mathbf{b}_i)\\ &\qquad (\mathbf{E}_i - \mathbf{h}(t))\Bigg)\\ &\quad + (1-\mathbf{g}(t))\Phi(\mathbf{h}(t), \mathbf{x}(t), \theta_{KAN})\end{aligned} \tag{21}$$

where $f_{lnn}(\cdot)$ and $f_{kan}(\cdot)$ are the nonlinear mappings of LNN and KAN branches, respectively, and $\mathbf{g}(t)$ is an AGM that dynamically balances the contribution of the two branches.

$$\mathbf{g}(t) = \text{sigmoid}(\mathbf{W}_g[\mathbf{h}(t), \mathbf{x}(t)] + \mathbf{b}_g) \tag{22}$$

where $\mathbf{W}_g$ and $\mathbf{b}_g$ denote the learnable weight matrix and bias vector of the gating unit, respectively. The AGM allocates the contributions of the LNN and KAN branches according to the input feature patterns. It favors the LNN pathway under smoothly evolving health indicators to preserve the continuous-time dynamical structure of LNN, while increasing the KAN contribution when the features exhibit rapid variations or nonlinear transitions associated with abrupt capacity changes. This input-dependent weighting mechanism preserves bounded state interactions while enabling the model to capture complex degradation behaviors. From a dynamical-system perspective, the AGM performs a bounded convex-like combination between two nonlinear mappings, thereby ensuring that the overall state update remains governed by the continuous-time dynamics of the LNN branch. The SOH estimate at a given time is obtained through a linear mapping of the hidden state as

$$\hat{y}(t) = \mathbf{W}_o\dot{\mathbf{h}}(t) + \mathbf{b}_o \tag{23}$$

where $\hat{y}(t)$ is the estimated SOH value, $\mathbf{W}_o$ and $\mathbf{b}_o$ are the weight matrix and bias of the output layer, respectively.

Based on this output formulation, the complete set of model parameters $\Theta = \{\tau_i, \mathbf{W}_{in}, \mathbf{W}_{rec}, \mathbf{b}_i, \theta_{KAN}, \mathbf{W}_g, \mathbf{b}_g, \mathbf{W}_o, \mathbf{b}_o\}$ is trained by minimizing the loss function given in (24). All model parameters, including the gating mechanism, are optimized jointly under this unified objective without auxiliary supervision. As a result, the gating weights are thus learned implicitly through the overall estimation task

$$\begin{aligned}\mathbf{L} &= \frac{1}{N}\sum_{j=1}^{N}(\hat{y}(t_j) - y(t_j))^2 + \lambda_1\|\tau_i\|^2\\ &\quad + \lambda_2\sum_{j=1}^{N}\max(0, -\Delta\hat{y}) + \lambda_3\|\Theta\|_1.\end{aligned} \tag{24}$$

The formulation employs $y(t_j)$ and $\hat{y}(t_j)$ denote the true and estimated SOH at the j-th observation time step, while λ_1, λ_2, and λ_3 are the corresponding regularization coefficients. The Heaviside function penalizes a negative difference $\Delta\hat{y} = \hat{y}(t_{j-1}) - \hat{y}(t_j)$ between consecutive observation steps thereby discourages nonmonotonic degradation in SOH estimation. The second term constrains the time constants to stabilize the underlying dynamical behavior, and the fourth term promotes sparsity in the model parameters. Although (24) is explicitly written as four terms, from an optimization perspective they can be grouped into three functional components: the primary SOH estimation loss, the continuous-time dynamical regularization, and the Heaviside-weighted gated structural consistency constraint, which simplifies hyperparameter tuning and improves training stability.

To preserve the continuous-time formulation during training, the forward state trajectory in (21) is obtained by numerically integrating the underlying ODE. Since gradients cannot be directly computed through the solver, the adjoint sensitivity method is employed, in which an adjoint state $\mathbf{a}(t) = \partial\mathbf{L}/\partial\mathbf{h}(t)$ is introduced and propagated backward along the same continuous-time dynamics to compute parameter gradients. It should be noted that the adjoint computation is required only during offline training for gradient evaluation and does not participate in the inference stage. Both forward and backward integrations are implemented using MATLAB's ode45 solver with adaptive step-size control. The forward pass adopts stricter relative and absolute tolerances (10^{-3} and 10^{-5}) to prevent numerical drift in long-term SOH evolution, whereas slightly relaxed tolerances (10^{-2} and 10^{-4}) are used in the adjoint computation to reduce computational overhead without compromising convergence stability. These tolerance settings strike a practical balance between numerical accuracy and computational efficiency for continuous-time SOH modeling.

The adjoint state evolves according to a backward-time ODE

$$\dot{\mathbf{a}}(t) = \frac{d\mathbf{a}(t)}{dt} = -\mathbf{a}(t)^T\frac{\partial F(\mathbf{h}(t), t, \Theta)}{\partial\mathbf{h}(t)}. \tag{25}$$

By solving this adjoint equation backward in time from t_1 to t_0, the gradients of the loss with respect to all parameters can be computed. For instance, the gradients for the time constant vector τ_i and the gating weight matrix $\mathbf{W}_g$ are derived as

$$\frac{\partial\mathbf{L}}{\partial\tau_i} = \int_{t_0}^{t_1}\mathbf{a}(t)^{\mathrm{T}}\left(\mathbf{g}(t)\frac{\mathbf{h}(t)}{\boldsymbol{\tau}_i^2}\right)dt + 2\lambda_1\boldsymbol{\tau}_i + \lambda_2\cdot sign(\tau_i) \tag{26}$$

$$\frac{\partial\mathbf{L}}{\partial\mathbf{W}_g} = \int_{t_0}^{t_1}\mathbf{a}(t)^{\mathrm{T}}\frac{\partial F(\mathbf{h}(t), t, \Theta)}{\partial\mathbf{W}_g}dt + \lambda_2\cdot sign(\mathbf{W}_g) \tag{27}$$

where $sign(\cdot)$ denotes the direction of the residual. The gradient in (26) determines the temporal response characteristics of the network, while the gradient in (27) controls the adaptation and fusion strength between two branches. Gradients for the remaining parameters (e.g., $\mathbf{W}_{in}$, $\mathbf{W}_{rec}$, and θ_{KAN})are computed analogously via the chain rule within this adjoint framework.

All parameters are updated using an adaptive optimization algorithm. In this work, we employ the AdaBelief optimizer [24], which adapts the step size based on the belief in the gradient directions. The update rule for a generic parameter $\theta \in \Theta$ at step k is as follows:

$$\mathbf{m}_k = \gamma_1\mathbf{m}_{k-1} + (1-\gamma_1)\mathbf{g}_k \tag{28}$$

$$\mathbf{s}_k = \gamma_2\mathbf{s}_{k-1} + (1-\gamma_2)(\mathbf{g}_k - \mathbf{m}_k)^2 + \delta \tag{29}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

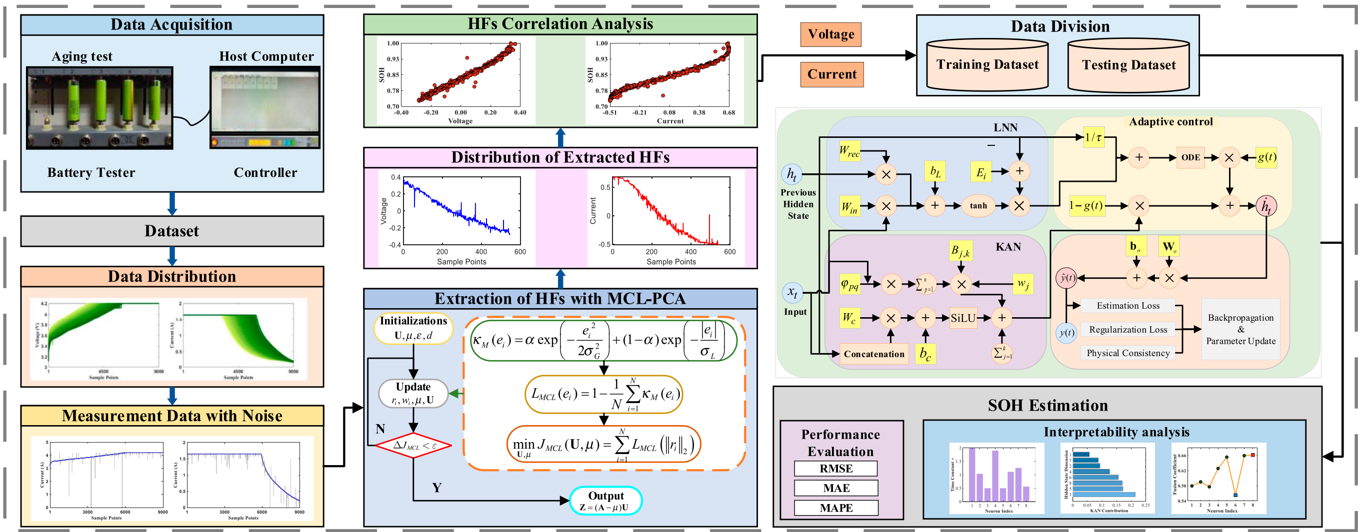

Fig. 4. Schematic diagram of the overall SOH estimation framework.

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} - \eta \frac{\mathbf{m}_k}{\sqrt{\mathbf{s}_k} + \delta} \tag{30}$$

where $\mathbf{g}_k = \partial \mathbf{L} / \partial \boldsymbol{\theta}$ is the gradient at step, $\mathbf{m}_k$ and $\mathbf{s}_k$ are the first and second moment estimates of the gradients, respectively, γ_1 and γ_2 are exponential decay rates for the moment estimates, η is the learning rate, and δ is a smoothing term used to prevent division by zero.

This optimization strategy, coupled with adjoint-based gradient computation, preserves consistency between parameter updates and the underlying continuous-time dynamics, thereby enhancing numerical stability during training. Moreover, since the KAN module augments only the nonlinear observation pathway without modifying the fundamental ODE formulation, the structural form of the LNN state evolution remains unchanged. The spline-based nonlinear mappings retain the regularity properties inherent to the original dynamical systems. From a computational perspective, the integration of the KAN module does not increase the dimensionality or order of the underlying ODE system. The spline-based mappings operate on low-dimensional latent states and introduce only a limited number of basis function evaluations per integration step. Consequently, computational overhead grows modestly and scales linearly with the state dimension, while the dominant runtime complexity continues to be governed by the ODE integration process.

IV. Experimental Setup

To comprehensively evaluate the estimation accuracy and robustness of the proposed MCL-PCA and KLNN framework, extensive experiments are conducted on both laboratory and public battery datasets. The overall validation workflow, illustrated in Fig. 4, covers data acquisition, noise simulation, feature extraction, and SOH estimation.

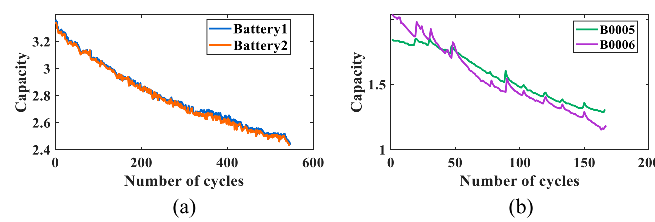

Fig. 5. Capacity degradation curves. (a) Dataset A. (b) Dataset B.

TABLE I
Specifications of the Lithium-Ion Battery Used in Experiments

Parameter	Value
Rated capacity	3.30 Ah
Nominal voltage	3.60 V
Maximum voltage	4.25 V
Discharge cutoff voltage	2.50 V
Maximum discharge current	3.00 C

A. Dataset Description

To comprehensively evaluate algorithm performance, three battery datasets with distinct noise characteristics and aging patterns are employed.

Dataset A: This dataset was collected using a battery testing platform assembled for this study, as shown in Fig. 4. The setup consists of a battery tester, a microcontroller for data control and buffering, and a host computer for supervisory operation. Two NCR18650GA lithium-ion cells (3.3 Ah) were cycled under CC-CV charge/discharge protocols at different ambient temperatures with a sampling interval of 1 s. Key specifications are summarized in Table I, and the corresponding nonlinear capacity degradation trajectories are shown in Fig. 5(a).

Dataset B: The second dataset uses the publicly available NASA battery data for cells B0005 and B0006. Measurements

TABLE II
ITEMS OF BATTERY CHARGING DATA

Items	Unit	Resolution
Time	yyyy-mm-dd hh:mm:ss	-
Current	A	
Pack voltage	V	0.1 V
SOC	-	0.1
Maximum cell voltage	V	0.001
Minimum cell voltage	V	0.001
Maximum cell temperature	°C	1 °C
Minimum cell temperature	°C	1 °C

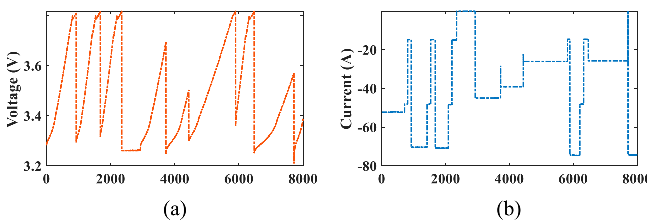

Fig. 6. Charging curves of a vehicle.

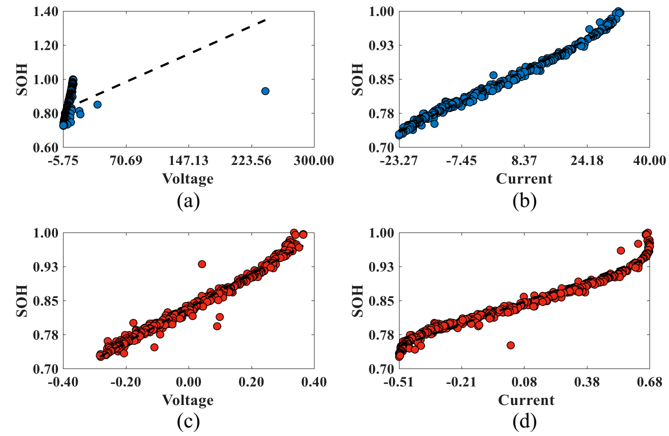

Fig. 7. HFs extraction results. (a) and (b) PCA. (c) and (d) MCL-PCA.

were acquired in a controlled laboratory environment at 24 °C following standardized cycling procedures. The corresponding degradation trajectories are presented in Fig. 5(b).

Dataset C: The third dataset comprises real-world operational data from a commercial electric vehicle fleet [30]. It captures long-term battery degradation dynamics under actual driving and charging conditions, with records spanning approximately 29 months from 20 identical electric taxis operating in the same urban region. During charging, BMS data were acquired through the controller area network at an 8-s sampling interval. The measured variables-essential for evaluating pack-level consistency and thermal behavior-include pack current and voltage, state of charge (SOC), and extreme cell voltages and temperatures, as detailed in Table II. The charging curves of a representative vehicle are shown in Fig. 6, which concatenates multiple charging cycles within a certain period.

To evaluate the robustness of the proposed framework under mixed non-Gaussian noise conditions, we inject a simulated hybrid noise, composed of four distinct types, into the original voltage and current signals. The composite noise ζ_n is formulated as: $\zeta_n = \zeta_n^{Gauss} + \zeta_n^{Cauhy} + \zeta_n^{Laplace} + \zeta_n^{Non-stat}$, where ζ_n^{Gauss} represents Gaussian background noise with variance approximately 0.1% of the original signal amplitude, simulating small-amplitude measurement fluctuations; ζ_n^{Caushy} is generated from a Cauchy (t-distribution with $\nu = 1$) component, where approximately 5% of the samples are replaced with heavy-tailed values to emulate abrupt spike disturbances; $\zeta_n^{Laplace}$ denotes zero-mean Laplacian noise simulating moderate jumps and drifts, with peak amplitude around 3% of the signal; Finally, $\zeta_n^{Non-stat}$ corresponds to nonstationary perturbations generated by superimposing low-amplitude sine waves, with maximum amplitude around 0.1% of the signal peak, emulating time-varying fluctuations. These injected components are intended to represent diverse disturbance characteristics typically encountered in battery monitoring systems, such as stochastic measurement noise, impulsive outliers, and time-varying perturbations. Rather than attempting to fully replicate actual sensor noise, the objective of this design is to examine whether the proposed SOH estimation framework can sustain stable performance under heterogeneous and compounded disturbance conditions.

In addition to experiments with simulated hybrid noise experiments, Dataset C, comprising real-world EV operational data is evaluated without any artificial noise injection. This allows the robustness of the system to be assessed under unregulated field noise conditions.

B. Extraction and Evaluation of HFs

In general, the original measurement data may be affected by complex and heterogeneous noise in practical operating conditions. This study applies the MCL-PCA method to perform unsupervised extraction of low-dimensional HFs from voltage, and current signals. This feature extraction stage is conducted offline prior to model training and does not introduce additional computational burden during online SOH inference. The results of HF extraction are shown in Fig. 7. To quantitatively evaluate the quality of the extracted features under heterogeneous noise conditions, both the Spearman rank correlation coefficient and the maximal information coefficient (MIC) are adopted. Spearman measures the monotonic relevance between the features and SOH, whereas MIC captures their nonlinear dependency.

As shown in Table III, the features extracted by MCL-PCA exhibit markedly stronger correlations with SOH, with Spearman coefficients remaining above 0.98, whereas conventional PCA fluctuates substantially and drops to 0.64 in certain samples. The magnitude of correlation improvement, however, differs across signal channels. Voltage-derived features are more susceptible to mixed non-Gaussian disturbances due to nonlinear voltage behavior and relaxation dynamics during charge-discharge cycles, which can degrade the effectiveness of conventional PCA. In contrast, the adaptive residual weighting

TABLE III
CORRELATION COEFFICIENTS BETWEEN FEATURES AND SOH

Features	Evaluating	Model	Battery 1	Battery 2
Voltage	Spearman	PCA	0.9569	0.9688
		MCL-PCA	0.9913	0.9912
	MIC	PCA	0.8999	0.8741
		MCL-PCA	0.9225	0.9229
Current	Spearman	PCA	0.9965	0.9933
		MCL-PCA	0.9927	0.9925
	MIC	PCA	0.9453	0.9436
		MCL-PCA	0.9384	0.9474

mechanism in MCL-PCA suppresses such heterogeneous perturbations, thereby yielding more stable voltage-derived features. For current-derived features, particularly under CC-CV protocols, the signal profile is comparatively smoother and less noise-sensitive. In these cases, conventional PCA is often sufficient to capture dominant SOH-related variations, achieving correlation levels comparable to those from MCL-PCA. These results demonstrate that the MCL-PCA improves the stability (defined here as robustness to mixed noise perturbations under similar SOH conditions) of feature extraction without necessarily improving correlation in already stable feature domains.

Regarding nonlinear dependency, MCL-PCA yields higher MIC values, with the score increasing from 0.57 for PCA to 0.78. This improvement indicates that the resulting feature subspace preserves the nonlinear relationships associated with SOH more effectively. Overall, the results indicate that MCL-PCA improves the robustness of SOH-related feature extraction under mixed-noise conditions and provides more reliable representations of nonlinear degradation behavior.

C. Experimental Scenarios and Evaluation Metrics

To evaluate the proposed framework under limited data availability, three experimental scenarios are designed as

1) Scenario 1: Dataset A is limited in size as shown in Fig. 5(a), representing a typical small-sample scenario, and is split 7:3 into training and testing subsets to evaluate baseline estimation accuracy.
2) Scenario 2: Extreme Small-Sample Validation (Dataset B). A more stringent experiment is conducted using the NASA dataset seen in Fig. 5(b) to examine the model's performance under severely limited data availability. The model is trained using only the first 50% of lifecycle data and tested on the remaining unseen 50%, evaluating its robustness and ability to capture intrinsic degradation dynamics.
3) Scenario 3: Validation on Real-World EV Operating Data (Dataset C). This scenario is designed to assess the model's practicality and generalization under severe data constraints typical of real-world deployments. To evaluate cross-vehicle generalization, a rigorous validation scheme is implemented utilizing data from two vehicles in Dataset C for training, while reserving a third, unseen vehicle for testing. This setup simulates a scenario where only limited historical data from a small fleet is available to predict the health of a new vehicle. Given the absence of complete charge-discharge profiles, handcrafted monthly statistical features are used instead of MCL-PCA-extracted ones.

To ensure a consistent and fair comparison across different models, baseline configurations were established through an unified optimization procedure. The kernel parameters in the MCL formulation were tuned via Bayesian optimization on the training dataset, where the Gaussian and Laplacian kernel widths were over a logarithmic scale within [1, 100], while the mixing coefficient was selected from the interval [0.001, 0.999]. The optimization objective was to maximized the statistical consistency between the extracted health features and the SOH labels, as quantified by the average absolute correlation coefficient. The resulting parameters set was kept fixed and applied consistently across all subsequent experiments. The hyperparameters of SVR, LSTM, and GRU were also determined through Bayesian optimization. The optimized SVR configuration used an RBF kernel with a regularization coefficient of 497.54, a kernel width of 19.80, and an insensitive-tube threshold of 0.0747. The LSTM and GRU models adopted three stacked hidden layers with 235 and 233 units per layer, respectively, and the optimizer identified learning rates of 0.000422 for LSTM and 0.000343 for GRU. In contrast, the LNN and KLNN models exhibited robust performance across reasonable hyperparameter ranges during parameter selection due to their strong structural inductive biases. Therefore, they were configured via principled manual tuning, unlike the baseline models that required systematic Bayesian optimization, which maintains a consistent comparison protocol while avoiding unnecessary computational search. Specifically, the LNN employs a learning rate of 0.001, a hidden dimension of four, and time constants within the interval [0.5, 2.0]. Using this backbone, the KLNN introduces a single-rank (Q = 1) KAN component, constructed with 7 cubic B-spline basis functions to perform independent nonlinear transformations on each input dimension, along with an adaptive gating parameter initialized at 0.5.

With the model configurations finalized, performance was assessed using the mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage Error (MAPE), defined in (31). All experiments were implemented in MATLAB R2024a on a workstation equipped with an AMD Ryzen 7 3750H CPU operating at 2.30 GHz.

$$\begin{cases} \mathrm{MAE} = \dfrac{1}{N}\sum_{i-1}^{N} |y_i - \hat{y}_i| \times 100\% \\ \mathrm{RMSE} = \sqrt{\dfrac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \times 100\% \\ \mathrm{MAPE} = \dfrac{1}{N}\sum_{i-1}^{N} \left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \end{cases} \tag{31}$$

where y_i represents the true SOH value, $\hat{y}_i$ denotes the model-estimated value and N is the number of samples.

TABLE IV
SOH ESTIMATION PERFORMANCE USING HFS FROM PCA AND MCL-PCA

Evaluating	Model	Battery 1	Battery 2
	PCA	1.82	2.12
RMSE (%)	MCC-PCA	0.62	0.68
	MCL-PCA	0.45	0.52
	PCA	1.52	1.70
MAE (%)	MCC-PCA	0.46	0.51
	MCL-PCA	0.33	0.40
	PCA	2.03	2.27
MAPE (%)	MCC-PCA	0.60	0.68
	MCL-PCA	0.43	0.52

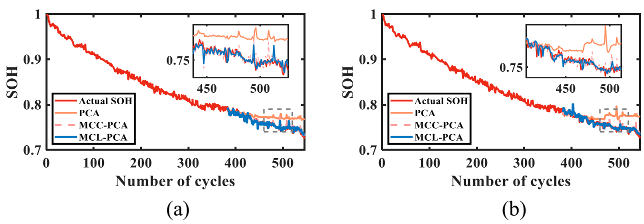

Fig. 8. SOH estimation results obtained using HFs extracted by MCL-PCA. (a) Battery 1. (b) Battery 2.

V. EXPERIMENTAL VERIFICATION AND ANALYSIS

A. Robustness Verification of Feature Extraction

To evaluate the robustness of feature extraction under noisy conditions, SOH estimation performance was compared using features obtained from conventional PCA, MCC-PCA, and the proposed MCL-PCA within the same KLNN model. As summarized in Table IV and illustrated in Fig. 8, MCL-PCA consistently achieves the highest estimation accuracy across both test batteries. Compared with standard PCA, yields markedly lower RMSE and MAE on both Battery 1 and Battery 2, demonstrating its robustness under mixed non-Gaussian noise. Although MCC-PCA significantly outperforms conventional PCA by suppressing severe outliers, its single-kernel correntropy formulation may retain moderate-scale residual effects, which is reflected in larger local fluctuations in the estimated SOH trajectories, as highlighted in the enlarged views. In contrast, MCL-PCA employs a Gaussian-Laplacian mixture kernel to impose multiscale residual penalization, which can more effectively suppresses heterogeneous noise. This yields smoother health features with reduced local variability, leading to more accurate and stable SOH estimation when integrated with KLNN.

B. Performance Comparison of the Proposed KLNN Model

To evaluate the performance of the KLNN model under small-sample conditions, we conducted a systematic comparison using experimental Dataset A (Scenario 1). The trajectories in Fig. 9 show that KLNN stays closely aligned with the true SOH over the entire lifecycle. Its predictions form a noticeably tighter error envelope, and the peak deviations in strongly

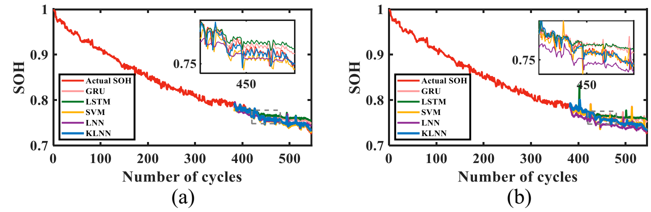

Fig. 9. Comparison of SOH estimation results by different models. (a) Battery 1. (b) Battery 2.

TABLE V
SOH ESTIMATION PERFORMANCE OF DIFFERENT MODELS

Battery	Evaluating	GRU	LSTM	SVR	LNN	KLNN
	RMSE (%)	0.83	1.10	0.65	0.78	0.45
Battery 1	MAE (%)	0.67	0.90	0.53	0.63	0.33
	MAPE (%)	0.89	1.20	0.69	0.83	0.43
	RMSE (%)	1.11	1.24	0.73	1.28	0.52
Battery 2	MAE (%)	0.89	0.97	0.53	1.14	0.40
	MAPE (%)	1.19	1.29	0.69	1.49	0.52

nonlinear regions, including the capacity knee point, are significantly smaller than those of SVR, LSTM, and GRU. The agreement between quantitative metrics and trajectory behavior confirms the robustness of KLNN in modeling complex degradation patterns.

As shown in Table V, KLNN attains RMSE values of 0.45 percent and 0.52 percent on the two test cells, yielding an average improvement of about 31.5 percent over the best conventional baseline. Compared with the LNN, KLNN further reduces MAE by 47.6 percent and 64.9 percent, highlighting the enhanced nonlinear representation offered by the KAN module. Collectively, these results demonstrate that KLNN delivers accurate SOH estimation under limited sample availability and effectively captures the multiscale nonlinear degradation characteristics inherent to lithium-ion battery aging.

C. Ablation Study and Interpretability Analysis

We conducted systematic ablation studies to dissect KLNN's performance gains and quantify component contributions.

1) Ablation Study: All experiments were conducted on battery 2 from Dataset A using MCL-PCA-extracted features to ensure consistency. We compared the following models: the LNN backbone (with KAN and gating removed), a KLNN variant with fixed gating (KLNN-no-adaptive, gating set to 0.5), and the full KLNN model.

As shown in Fig. 10, the results highlight the complementary and essential roles of each module. Ablating the KAN module causes the RMSE to surge from 0.52% to 1.28%, revealing the inadequacy of linear temporal dynamics alone in capturing complex nonlinear degradation. Fixing the gating mechanism further increases the RMSE to 1.53%, highlighting the critical importance of adaptive gating in dynamically integrating time-varying features. The complete KLNN model achieves the lowest RMSE, representing a clear reduction relative to the baseline LNN. These findings quantitatively confirm that the synergy

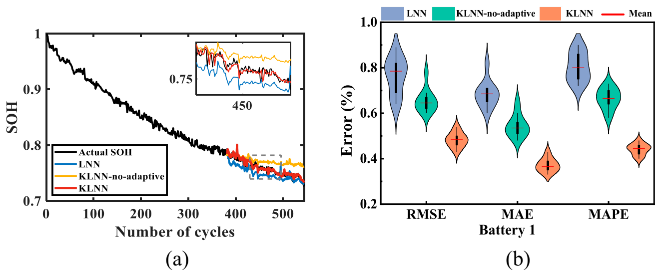

Fig. 10. Ablation study of KLNN components based on Battery 2. (a) SOH estimation results. (b) Distribution of relative error.

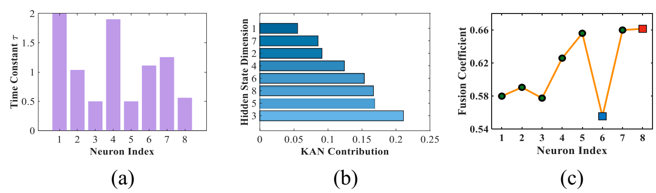

Fig. 11. Interpretability analysis. (a) Multitimescale dynamic capture. (b) Structured nonlinear modeling. (c) Adaptive gating mechanism.

between structured nonlinearity and adaptive gating is pivotal to KLNN's enhanced capability in modeling complex battery degradation dynamics.

2) Interpretability Analysis: According to the ablation results, we further examine KLNN's internal mechanisms from three complementary perspectives to elucidate its "white-box" characteristics and provide physical insights.

a) Multitimescale dynamic capture: The distribution of learned time constants, shown in Fig. 11(a), reveals neuronal temporal responses spanning multiple orders of magnitude. This enables KLNN to concurrently capture long-term gradual capacity fading and short-term disturbances, facilitating multiscale degradation modeling.

b) Structured nonlinear modeling: Analysis of the KAN module's hidden activations in Fig. 11(b) shows substantial functional variation across dimensions. Certain neurons specialize in local nonlinear fitting, while others maintain dynamic equilibrium, demonstrating KAN's capacity for structured representation of complex dynamics within a continuous-time framework.

c) Adaptive gating mechanism: As illustrated in Fig. 11(c), the gating coefficients dynamically adjust in response to input features. During stable health states, the LNN branch is prioritized to ensure prediction stability, whereas during capacity-transition or nonstationary phases, the KAN branch's contribution increases to enhance nonlinear responsiveness. This adaptive fusion effectively balances stability and flexibility across varying degradation rates.

Collectively, the interpretability analysis and ablation results provide consistent mechanistic evidence for KLNN's superior performance, validating the design rationale of integrating structured nonlinearity and adaptive gating within a continuous-time dynamical framework.

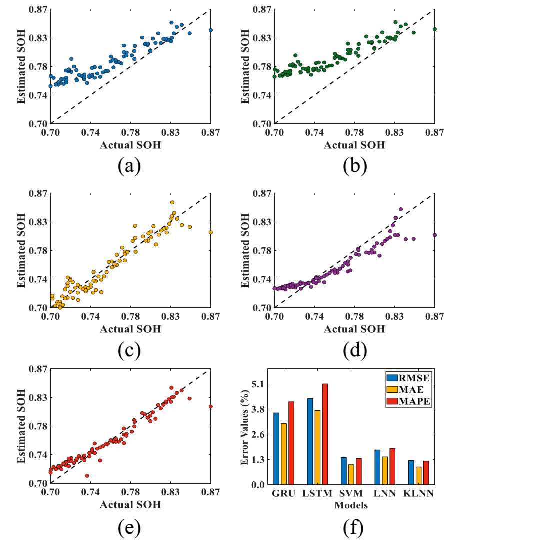

Fig. 12. Few-shot SOH estimation results for NASA cell B0005. (a)–(e) Estimated result. (f) Distribution of relative error.

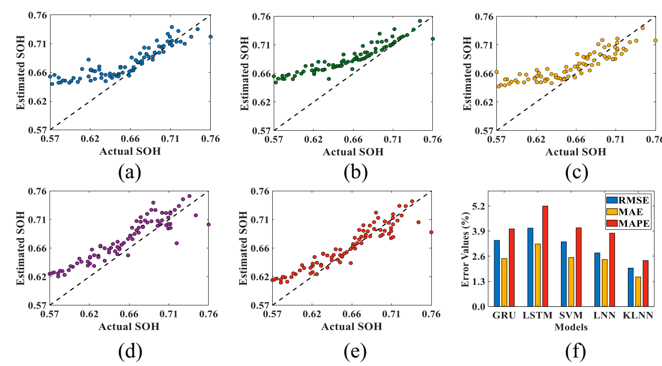

Fig. 13. Few-shot SOH estimation results for NASA cell B0006. (a)–(e) Estimated result. (f) Distribution of relative error.

D. Few-Shot Learning Capability Validation

To rigorously assess the proposed framework under extremely small-sample conditions, we conducted early-stage prediction experiments on Scenario 2. As illustrated in Figs. 12 and 13, KLNN exhibits consistently high estimation accuracy across both B0005 and B0006 cells. Quantitatively, KLNN achieves the lowest RMSE, MAE, and MAPE among all models, demonstrating consistent performance advantages over the other data-driven methods. The scatter plots in Figs. 12 and 13 further corroborate these quantitative results: KLNN's predictions cluster tightly around the identity line $y = x$, exhibiting high precision and minimal systematic bias. In contrast, models such as SVR, LSTM, and GRU display pronounced dispersion and larger deviations from the reference. These results indicate that KLNN can extract a stable and reliable degradation trajectory even when trained on only half of the early-cycle data, effectively addressing the challenge of accurate SOH estimation under severely limited data.

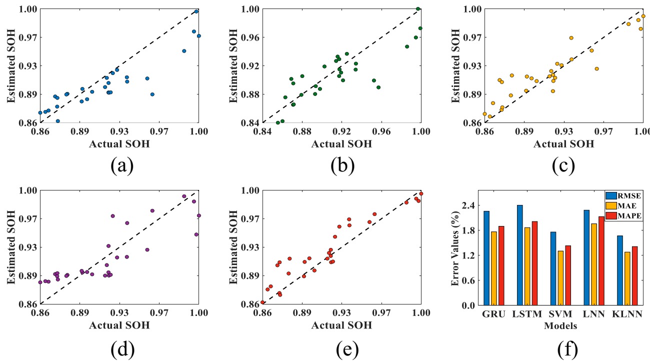

Fig. 14. Real-world EV operating data SOH estimation results. (a)–(e) Estimated result. (f) Distribution of relative error.

E. Validation on Real-World EV Operating Data

To assess the model's generalizability to high-dynamic operational profiles, the proposed framework is further evaluated under Scenario 3 using Dataset C. The estimation results obtained from the test vehicle are presented in Fig. 14. As shown in Fig. 14(a)–(e), the predictions of the proposed KLNN model exhibit the smallest deviation and tightest clustering along the reference line (y = x) compared with all baseline models, indicating higher accuracy and precision. The error distribution in Fig. 14(f) further confirms that KLNN achieves the lowest and most concentrated estimation errors. These findings underscore the robustness of the KLNN framework, which maintains high estimation accuracy even when applied to real-world EV data characterized by complex, high-dynamic currents, thus confirming its generalizability beyond controlled CC-CV cycling conditions protocols.

VI. Conclusion

In this article, we addressed the challenge of accurately estimating battery SOH under mixed non-Gaussian measurement noise and limited training samples by proposing an integrated framework that combines robust feature extraction with nonlinear dynamic modeling. MCL-PCA suppresses mixed non-Gaussian disturbances and extracts stable low-dimensional health features through a Gaussian-Laplacian hybrid kernel, while KLNN enhances nonlinear representation and ensures stable training by embedding structured KAN mappings into continuous-time liquid dynamics with adaptive gating. The effectiveness of the proposed framework was validated through extensive experiments under controlled mixed-noise conditions and small-sample settings. Specifically, MCL-PCA significantly improved feature robustness, leading to substantially lower feature-level RMSE compared with conventional PCA, and KLNN consistently achieved the most accurate SOH predictions, yielding lower RMSE under both standard small-sample scenarios and challenging early-stage prediction tasks. These results indicate that the proposed design is capable of capturing multitimescale and strongly nonlinear degradation behaviors even when data availability is severely limited.

While the present investigation focuses on representative battery types and simulated mixed-noise scenarios, future research will seek to extend the generalizability and practical applicability of the proposed estimation framework. This includes systematic validation across large-scale and heterogeneous battery datasets, robustness evaluation under real-world operating conditions, and the development of uncertainty quantification mechanisms for partially labeled data. In addition, computationally efficient model structures and training strategies will be investigated to improve scalability and facilitate deployment in practical battery management systems.

References

[1] X. Hu, F. Feng, K. Liu, L. Zhang, J. Xie, and B. Liu, "State estimation for advanced battery management: Key challenges and future trends," *Renew. Sustain. Energy Rev.*, vol. 114, 2019, Art. no. 109334.

[2] X. Tang, F. Gao, K. Liu, Q. Liu, and A. M. Foley, "A balancing current ratio based state-of-health estimation solution for lithium-ion battery pack," *IEEE Trans. Ind. Electron.*, vol. 69, no. 8, pp. 8055–8065, Aug. 2022.

[3] K. Liu, Z. Wei, C. Zhang, Y. Shang, R. Teodorescu, and Q.-L. Han, "Towards long lifetime battery: AI-based manufacturing and management," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 7, pp. 1139–1165, Jul. 2022.

[4] X. Wu, X. Du, and Z. Song, "Evaluating the generality and effectiveness of equivalent circuit models for battery strings with heterogeneous cells connected in parallel," *IEEE Trans. Ind. Inform.*, vol. 21, no. 7, pp. 5644–5654, Jul. 2025.

[5] C. She, Z. Wang, F. Sun, P. Liu, and L. Zhang, "Battery aging assessment for real-world electric buses based on incremental capacity analysis and radial basis function neural network," *IEEE Trans. Ind. Inform*, vol. 16, no. 5, pp. 3345–3354, May 2020.

[6] Y. Li, M. Maleki, and S. Banitaan, "State of health estimation of lithium-ion batteries using EIS measurement and transfer learning," *J. Energy Storage*, vol. 73, no. C, Dec. 2023, Art. no. 109185.

[7] B. Gou, Y. Xu, and X. Feng, "An ensemble learning-based data-driven method for online state-of-health estimation of lithium-ion batteries," *IEEE Trans. Transp. Electrific.*, vol. 7, no. 2, pp. 422–436, Jun. 2021.

[8] Q. Li and W. Xue, "A review of feature extraction toward health state estimation of lithium-ion batteries," *J. Energy Storage*, vol. 112, Mar. 2025, Art. no. 115453.

[9] K. Liu, Y. Liu, S. Zhao, X. Li, and Q. Peng, "An ultrasonic wave-based method for efficient state-of-health estimation of li-ion batteries," *IEEE Trans. Ind. Electron.*, vol. 72, no. 4, pp. 3792–3802, Apr. 2025.

[10] P. Xu, T. Ouyang, B. Liu, and Z. Chen, "State-of-charge estimation and health prognosis for lithium-ion batteries based on temperature-compensated Bi-LSTM network and integrated attention mechanism," *IEEE Trans. Ind. Electron.*, vol. 71, no. 6, pp. 5586–5596, Jun. 2024.

[11] G. Niu, X. Wang, E. Liu, and B. Zhang, "Lebesgue sampling based deep belief network for lithium-ion battery diagnosis and prognosis," *IEEE Trans. Ind. Electron.*, vol. 69, no. 8, pp. 8481–8490, Aug. 2022.

[12] Y. Jiang, Y. Ke, F. Yang, and J. Zhang, "State of health estimation for second-life lithium-ion batteries in energy storage system with partial charging-discharging workloads," *IEEE Trans. Ind. Electron.*, vol. 71, no. 10, pp. 13178–13188, Oct. 2024.

[13] F. Wang, Z. Zhai, Z. Zhao, and L. Sun, "Physics-informed neural network for satellite battery degradation estimation based on an extreme segment data within 2 min," *IEEE Trans. Ind. Electron.*, vol. 72, no. 9, pp. 9748–9757, Sep. 2025.

[14] P. Lee, S. Kwon, D. Kang, I. Cho, and J. Kim, "Principle component analysis-based optimized feature extraction merged with nonlinear regression model for improved state-of-health prediction," *J. Energy Storage*, vol. 48, Apr. 2022. Art. no. 104026.

[15] M. Wu, Y. Zhong, J. Wu, and L. Wang, "State of health estimation of the lithium-ion power battery based on the principal component analysis-particle swarm optimization-back propagation neural network," *Energy*, vol. 283, 2023. Art. no. 129061.

[16] R. He, W. Zheng, and B. Hu, "Robust principal component analysis based on maximum correntropy criterion," *IEEE Trans. Image Process*, vol. 20, no. 5, pp. 1485–1494, May 2011.

[17] N. Li, M. Wang, Y. Lei, B. Yang, X. Li, and X. Si, "Remaining useful life prediction of lithium-ion battery with nonparametric degradation modeling and incomplete data," *Rel. Eng. Sys. Saf.*, vol. 256, 2025. Art. no. 110721.

[18] X. Zhang, Y. Meng, S. Yan, and R. Xu, "State-of-Health estimation of lithium titanate batteries under locomotive conditions with meta-learning based on long short-term memory network," *IEEE Trans. Transport. Electrific.*, vol. 11, no. 6, pp. 12964–12973, Aug. 2025.

[19] J. Tian, J. Xie, J. Zhang, C. Huang, Y. Jiang, and H. Luo, "A transfer gated recurrent unit model for battery health monitoring under cross-domain conditions," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–12, 2025.

[20] L. Ca, N. Cui, H. Jin, J. Meng, S. Yang, and J. Peng, "A unified deep learning optimization paradigm for lithium-ion battery state-of-health estimation," *IEEE Trans. Energy Convers.*, vol. 39, no. 1, pp. 589–600, Mar. 2024.

[21] Y. Wu, L. Zhang, H. Gao, and Z. Li, "Data-driven transfer-stacking-based state of health estimation for lithium-ion batteries," *IEEE Trans. Ind. Electron.*, vol. 69, no. 12, pp. 13154–13164, Dec. 2022.

[22] G. Zhao, C. Zhang, B. Duan, Y. Shang, Y. Kang, and R. Zhu, "State-of-health estimation with anomalous aging indicator detection of lithium-ion batteries using regression generative adversarial network," *IEEE Trans. Ind. Electron.*, vol. 70, no. 3, pp. 2685–2695, Mar. 2023.

[23] R. Hasani, M. Lechner, A. Amini, D. Rus, and R. Grosu, "Liquid time-constant networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 9, pp. 7657–7666, May 2021.

[24] R. Hasani, M. Lechner, A. Amini, D. Rus, and R. Grosu, "Closed-form continuous-time neural networks," *Nat. Mach. Intell.*, vol. 4, pp. 992–1003, Nov. 2022.

[25] M. Chahine et al., "Robust flight navigation out of distribution with liquid neural networks," *Sci. Robot.*, vol. 8, no. 77, Apr. 2023, Art. no. eadc8892.

[26] Z. Liu et al., "KAN: Kolmogorov-Arnold networks," 2024, *arXiv:2404.19756*.

[27] B. Koenig, S. Kim, and S. Deng, "KAN-ODEs: Kolmogorov-Arnold network ordinary differential equations for learning dynamical systems and hidden physics," *Comput. Methods Appl. Mech. Eng.*, vol. 432, 2024, Art. no. 117397.

[28] I. Jarraya et al., "SOH-KLSTM: A hybrid Kolmogorov-Arnold network and LSTM model for enhanced lithium-ion battery health monitoring," *J. Energy Storage*, vol. 122, Jun. 2025. Art. No. 116541.

[29] B. Chen, X. Wang, N. Lu, S. Wang, J. Cao, and J. Qin, "Mixture correntropy for robust learning," *Pattern Recognit.*, vol. 79, pp. 318–327, Jul. 2018.

[30] Z. Deng, L. Xu, H. Liu, X. Hu, Z. Duan, and Y. Xu, "Prognostics of battery capacity based on charging data and data-driven methods for on-road vehicles," *Appl. Energy*, vol. 339, Jun. 2023. Art. no. 120954.

Wentao Ma (Senior Member, IEEE) received the B.S. degree in mathematics and applied mathematics from Shaanxi University of Technology, Hanzhong, China, in 2003, the M.S. degree in computing mathematics from Huazhong University of Science and Technology, Wuhan, China, in 2008, and the Ph.D. degree in information and communication engineering from Xi'an Jiaotong University, Xi'an, China, in 2015.

He is currently an Associate Professor with the School of Electrical Engineering, Xi'an University of Technology, China. His research interests include machine learning, information theory, and state estimation of power systems.

Zhuo Li received the B.S. degree in 2023 from Xi'an University of Technology, China, where he is currently working toward the M.S. degree, both in electrical engineering.

His research interests include battery system modeling, state estimation, and nonlinear filter design.

Peng Guo received the bachelor's degree in electrical engineering from Xi'an Polytechnic University, Xi'an, China, in 2020. He is currently working toward the Ph.D. degree in electrical engineering with Xi'an University of Technology, Xi'an.

His research interests include battery system modeling, state estimation, and nonlinear filter design.

Yang Li (Senior Member, IEEE) received the B.E. degree in electrical engineering from Wuhan University, Wuhan, China, in 2007, and the M.Sc. and Ph.D. degrees in power engineering from Nanyang Technological University (NTU), Singapore, in 2008 and 2015, respectively.

He was affiliated with the Energy Research Institute, NTU; the School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, QLD, Australia; the School of Automation, Wuhan University of Technology, Wuhan; and the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden. He is currently a Professor with the School of Electrical Engineering and Automation, Wuhan University. His research interests include modeling and control of energy storage systems in the power and transport sectors.

Dr. Li serves as an Associate Editor for several journals, such as IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS and IEEE TRANSACTIONS ON TRANSPORTATION ELECTRIFICATION.

Badong Chen (Senior Member, IEEE) received the B.S. and M.S. degrees in control theory and engineering from Chongqing University, Chongqing, China, in 1997 and 2003, respectively, and the Ph.D. degree in computer science and technology from Tsinghua University, Beijing, China, in 2008.

He was a Postdoctoral Researcher with Tsinghua University from 2008 to 2010 and a Postdoctoral Associate with the Computational Neuro Engineering Laboratory, University of Florida, Gainesville, FL, USA, from 2010 to 2012. He visited Nanyang Technological University, Singapore, in 2015, as a Visiting Research Scientist for one month. He is currently a Professor with the Institute of Artificial Intelligence and Robotics (IAIR), Xi'an Jiaotong University, Xi'an, China. His research interests include signal processing, information theory, machine learning, and their applications to cognitive science and neural engineering.