Battery Degradation Modeling with Neural Ordinary Differential Equations: A State-of-the-Art Analysis

Executive Summary

The accurate prognosis of lithium-ion battery degradation is a critical enabler for the safety, reliability, and economic viability of electric vehicles and grid-scale energy storage. This report provides an exhaustive analysis of a transformative approach in this domain: the use of Neural Ordinary Differential Equations (NODEs) to model battery aging. This paradigm reframes the degradation process from a discrete sequence of health states into the continuous-time evolution of a dynamical system. This fundamental shift provides significant advantages over traditional discrete-time models, such as Recurrent Neural Networks (RNNs), particularly in handling irregularly sampled data and improving long-term forecasting accuracy. The analysis traces the evolution of these models from foundational concepts like the canonical NODE and Augmented NODE (ANODE) to the current state of the art, which is characterized by sophisticated hybrid architectures. Models such as ACLA, which integrates attention mechanisms, CNNs, and LSTMs into an ANODE framework, demonstrate superior feature abstraction and generalization to unseen battery types. Concurrently, novel architectures incorporating Graph Attention Networks and Kolmogorov-Arnold Networks (KANs) are pushing the frontier towards enhanced interpretability and the fusion of heterogeneous data sources. Quantitative benchmarking consistently shows that NODE-based models outperform traditional RNNs, with advanced variants achieving state-of-health estimation errors below 2.5% on entirely new datasets. The strategic implication of this research is clear: continuous-time deep learning models are poised to become a cornerstone of next-generation Battery Management Systems (BMS), offering a more physically principled, accurate, and ultimately more trustworthy method for battery prognostics.

I. A Paradigm Shift: Modeling Battery Degradation as a

Continuous Dynamical System

1.1 The Theoretical Underpinnings of an ODE-Based Approach to Battery Aging

The modeling of lithium-ion battery degradation has traditionally relied on methods that treat the battery's lifecycle as a discrete sequence of charge-discharge cycles. However, a more powerful and physically analogous paradigm frames the evolution of a battery's State of Health (SOH) as the trajectory of a continuous-time dynamical system.[1] This conceptual leap is at the heart of applying Neural Ordinary Differential Equations (NODEs) to battery prognostics.

The core principle is to represent the battery's state vector, y, which includes SOH and other relevant features, as evolving according to a governing ordinary differential equation (ODE):

$$\dot{y} = F(y, t)$$
Here, $\dot{y}$ represents the rate of change of the state vector with respect to time (or cycle number), and F is an unknown, generally non-linear function that encapsulates the complex electrochemical and physical dynamics of degradation.[1] The challenge then becomes learning this unknown function directly from observational data.

This is where neural networks provide a solution. By parameterizing the unknown dynamics function with a neural network, $F(y, t, \theta)$, the problem of prognostics is transformed into learning the network parameters θ that best describe the observed degradation data.[2] With a learned function F, predicting the future SOH from a known initial state $y(t_0)$ becomes an initial value problem (IVP). The future state $y(t_1)$ is found by integrating the learned dynamics over the desired time horizon:

$$y(t_1) = y(t_0) + \int_{t_0}^{t_1} F(y(\tau), \tau, \theta) d\tau$$
This integration is performed by a numerical ODE solver, which effectively becomes a layer within the deep learning architecture.[5]

1.2 Moving Beyond Discrete Time-Steps: Advantages over Traditional RNNs

This continuous-depth formulation offers several profound advantages over conventional discrete-layer architectures like Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU).[3]

First, the very concept of "model depth" is reframed. In traditional deep learning, depth is a discrete hyperparameter—the number of layers. A standard LSTM processes a sequence step-by-step, with its reasoning capacity tied to the sequence length and layer count.[1] The architecture of Residual Networks (ResNets) can be viewed as a series of discrete updates, $h_{t+1} = h_t + f(h_t, \theta_t)$.[6] NODEs emerge from taking the continuous limit of these discrete steps, resulting in the differential equation $dh(t)/dt = f(h(t), t, \theta)$.[3] Consequently, passing an input through the "network" is equivalent to solving this ODE over a time interval $[t_0, t_1]$.[4] Depth is no longer a layer count but a continuous integration time. This means a single trained model can perform both short-term and long-term predictions simply by adjusting the integration interval of the ODE solver, offering a more flexible and principled approach to variable-horizon forecasting.

This continuous nature yields several practical benefits:

- **Handling Irregularly Sampled Data:** Real-world battery usage data is often sampled at irregular intervals. NODEs are naturally equipped to handle such data, as the ODE solver can evaluate the dynamics at any required time point. In contrast, standard RNNs are designed for sequences with fixed, regular time steps.[5]
- **Adaptive Computation:** Modern ODE solvers are adaptive; they can dynamically adjust their step size based on the complexity of the learned dynamics. They can take larger, more efficient steps when the degradation trajectory is smooth and smaller, more precise steps during periods of rapid change, allowing for an explicit trade-off between computational cost and numerical precision.[3]
- **Constant Memory Cost:** A significant bottleneck in training very deep networks is the memory required to store activations for backpropagation. NODEs circumvent this by using the adjoint sensitivity method. This technique computes gradients by solving a second, augmented ODE backward in time, which has a constant memory cost with respect to the depth (i.e., the integration time), enabling the training of models with effectively infinite depth.[1]

II. Foundational Architectures: The Neural ODE Family

The application of NODE principles to battery degradation has led to the development of

several foundational architectures, each with distinct characteristics and trade-offs.

2.1 The Canonical Neural-ODE (NODE) Model

The standard Neural-ODE model, as first applied to battery prognostics, directly parameterizes the degradation dynamics function $F(y, \theta(t))$ using a relatively simple neural network, such as a multi-layer perceptron (MLP) with two linear layers and a tanh activation function.[1] The model's forward pass involves using a numerical ODE solver, such as the Dormand-Prince 5(4) method (Dopri5), to integrate the learned dynamics from an initial state to produce a predicted trajectory.[1] The backward pass for training leverages the computationally efficient adjoint sensitivity method to calculate parameter gradients, treating the ODE solver as a black box and avoiding the need to backpropagate through its internal operations.[1]

2.2 Augmented Neural-ODEs (ANODE): Overcoming Representational Limitations

A critical theoretical limitation of the canonical NODE architecture is that the continuous transformation it learns must be a homeomorphism. This means it cannot change the topology of the data space; for instance, it cannot map a high-dimensional space to a lower-dimensional one or cause trajectories to cross.[1] This can be a restrictive constraint when modeling complex degradation phenomena.

Augmented Neural-ODEs (ANODEs) were introduced to overcome this limitation. The solution is to augment the state vector y with additional dimensions a that are initialized to zero. The dynamics are then learned in this higher-dimensional space [1]:

$$\frac{d}{dt}\begin{bmatrix}y \\ a\end{bmatrix} = F\left(\begin{bmatrix}y \\ a\end{bmatrix}, \theta(t)\right), \quad \begin{bmatrix}y(0) \\ a(0)\end{bmatrix} = \begin{bmatrix}y_0 \\ 0\end{bmatrix}$$

By allowing the dynamics to evolve in this augmented space, the model can learn far more complex and expressive transformations, effectively resolving the representational bottleneck of standard NODEs. This has been shown to lower training losses and improve predictive performance in battery SOH forecasting.[1]

2.3 Discretization in Practice: The Predictor-Corrector Recurrent Neural Network (PC-RNN)

Bridging the conceptual gap between continuous-time ODEs and discrete-time RNNs, the Predictor-Corrector RNN (PC-RNN) is an architecture that explicitly discretizes the underlying ODE that governs battery degradation.[1] Instead of relying on a general-purpose black-box ODE solver, it implements a specific numerical integration scheme analogous to a second-order Runge-Kutta method. The process for advancing from cycle k to $k+1$ involves two steps:

1. **Predictor Step:** An initial estimate of the next state, $\hat{y}_{k+1}$, is calculated using a simple forward Euler step: $\hat{y}_{k+1} = y_k + \Delta t \cdot F(y_k, \theta)$.[1]
2. **Corrector Step:** This initial estimate is used to evaluate the gradient at the predicted future point. The final state, y_{k+1}, is then computed by averaging the gradient at the current state and the gradient at the predicted state: $y_{k+1} = y_k + \frac{1}{2}\Delta t (F(y_k, \theta) + F(\hat{y}_{k+1}, \theta))$.[1]

This structure provides a computationally simpler and more explicit alternative to a full ODE solver while retaining the core idea of modeling the rate of change of the system state.

III. The State of the Art: Hybrid and Advanced Implementations

Building upon the foundational NODE family, recent research has focused on developing sophisticated hybrid architectures. These models combine the continuous-time dynamics of NODEs with powerful feature extraction modules to better capture the complex signatures of battery degradation.

3.1 The ACLA Framework: Synergistic Integration for Enhanced Feature Abstraction

The ACLA model represents a significant step forward by integrating an **A**ttention mechanism,

a **C**onvolutional Neural Network (CNN), and a **L**ong Short-Term Memory (LSTM) network into the **A**NODE framework.[1] This architecture features a sophisticated frontend designed to produce a highly informative latent representation of the battery's state, which is then evolved in continuous time by the ANODE backend.

The roles of the individual components are highly synergistic:

- **Attention Layer:** The model begins by applying an attention mechanism to the input features (normalized charging times). This allows the model to dynamically assign importance weights to different parts of the charging curve at each cycle, focusing on the most salient indicators of degradation.[1]
- **CNN Layers:** The attention-weighted features are then processed by two 1D convolutional layers. The CNNs excel at extracting local, time-invariant patterns from the charging curve data, such as the shape, slope, and curvature of different voltage segments.[1]
- **LSTM Unit:** The sequence of feature maps generated by the CNN is fed into an LSTM unit. The LSTM's role is to model the long-range temporal dependencies across many cycles, capturing how the degradation patterns identified by the CNN evolve over the battery's lifetime.[1]
- **ANODE Solver:** Finally, the rich, context-aware hidden state vector produced by the LSTM is used as the initial condition for the ANODE solver. The ANODE then models the continuous-time evolution of this latent health state to predict future degradation.[1]

The primary innovation of ACLA is its powerful, specialized feature engineering pipeline that precedes the ODE solver. This marks a strategic shift from using a simple MLP to define the ODE dynamics to employing a dedicated, multi-stage feature extractor to distill the most critical information from the raw data.

3.2 Topological and Dynamic Modeling: The GAT-CNN-KAN-ODE Architecture

A novel and distinct approach proposes a dual-stream architecture that processes degradation information from both Euclidean and non-Euclidean spaces, leveraging a Kolmogorov-Arnold Network (KAN) to parameterize the neuronal ODE.[1] This model aims not only for high accuracy but also for enhanced interpretability.

The architecture is composed of several key components:

- **Non-Euclidean Branch (GAT):** This branch models the battery's health indicators as a graph. Nodes in the graph represent measurements (e.g., voltage, Coulombic efficiency at different cycles), and the edges represent the relationships between them, quantified

by metrics like the Spearman correlation coefficient.[1] A Graph Attention Network (GAT) is then used to learn the implicit topological relationships, capturing complex interdependencies that a standard sequential model would overlook.

- **Euclidean Branch (CNN):** In parallel, a 2D CNN processes the raw time-series data (e.g., voltage vs. time for multiple cycles) as a grid or image, extracting local spatial and temporal features in the conventional manner.[1]
- **Neuronal ODE with KAN:** The feature vectors from both the GAT and CNN branches are concatenated to form a comprehensive initial state vector for a neuronal ODE. A crucial innovation is that the dynamics of this ODE are parameterized not by a standard MLP, but by a Kolmogorov-Arnold Network (KAN).[1] KANs are a new class of neural network that feature learnable activation functions on the network edges rather than fixed activations on the nodes. This architecture makes them more parameter-efficient and significantly more interpretable than traditional MLPs.[1]

The emergence of these two distinct state-of-the-art architectures, ACLA and GAT-KAN-ODE, reflects a divergence in prognostic philosophies. The ACLA model's sequential design (Attention → CNN → LSTM → ANODE) is predicated on perfecting the *temporal feature extraction* from a single primary data source—the charging curve.[1] It operates on the principle that all necessary information is contained within this signal, provided it is processed with sufficient sophistication before being handed to the ODE solver. In contrast, the GAT-KAN-ODE model's parallel, dual-stream architecture is built on the idea of *fusing heterogeneous information representations*—a time-series grid from the CNN and a relational graph from the GAT.[1] This approach assumes that different data views capture complementary aspects of degradation. Furthermore, by replacing the "black-box" MLP in the ODE with an interpretable KAN, this second philosophy prioritizes enhancing the transparency of the core dynamic model itself.[1] This divergence highlights a central question in the field: is superior performance best achieved through more powerful feature engineering or through a more fundamentally expressive and transparent core model? The former may yield higher accuracy on specific tasks, while the latter may lead to more robust and trustworthy models for safety-critical deployment.

Table 1 provides a structured comparison of the key NODE-based architectures discussed.

Table 1: Architectural Comparison of NODE-based Models for Battery Prognostics

Model Name	Base Framework	Feature Extractor	Dynamics Parameterization	Key Innovation	Targeted Problem
NODE	ODE	None (Direct	MLP	Continuous-depth model	Basic time-series

		Input)		using ODE solver	forecasting
ANODE	ODE	None (Direct Input)	MLP	State augmentation to overcome topological constraints	Learning complex, non-homeomorphic transformations
PC-RNN	RNN	None (Direct Input)	MLP	Discretized ODE using a predictor-corrector scheme	Computationally efficient ODE approximation
ACLA	ODE (ANODE)	Attention + 1D CNN + LSTM	MLP	Synergistic feature extraction frontend for ANODE	Generalization across diverse battery datasets
GAT-CNN-KAN-ODE	ODE	GAT (Topological) + 2D CNN (Euclidean)	KAN	Fusion of Euclidean/non-Euclidean features; Interpretable dynamics	Multi-modal data fusion and model interpretability

IV. Performance Benchmarking and Comparative Analysis

Quantitative evaluation across standardized datasets reveals the consistent advantages of NODE-based models and highlights the performance gains achieved by advanced hybrid architectures.

4.1 Quantitative Comparison: $RMSE_{SOH}$ and EOL Prediction Accuracy

Multiple studies have benchmarked NODE-based models against traditional RNNs (LSTM, GRU) on the widely used Oxford and NASA battery datasets. The key performance metrics are the Root Mean Squared Error on SOH prediction ($RMSE_{SOH}$) and the error in End-of-Life (EOL) prediction.

A foundational study by Pepe et al. found that on the Oxford dataset, ANODE and PC-RNN generally outperformed LSTM and GRU, with ANODE achieving an average $RMSE_{SOH}$ below 3% across all prediction windows.[1] On the more challenging NASA dataset, which is characterized by irregular degradation patterns, the performance of all models degraded. However, the simpler NODE and PC-RNN models proved most robust, whereas ANODE exhibited high variance.[1] A notable characteristic of the ODE-based models was their tendency to produce more conservative (pessimistic) EOL predictions, a potentially beneficial trait for safety-critical applications.[1]

Subsequent work by Li et al. demonstrated the superiority of the ACLA architecture. On the same datasets, the ACLA model and its variant without attention (ACL) consistently outperformed the baseline ANODE and NODE models.[1] For instance, when trained on 90% of the NASA dataset, ACLA achieved an $RMSE_{SOH}$ of 1.19%, a remarkable 74.1% error reduction compared to the baseline ANODE model.[1] Similarly, the GAT-KAN-ODE model was shown to achieve optimal minimum RMSE while reducing the number of model parameters by 39-49% compared to benchmarks like stacked CNNs and LSTMs.[1]

Table 2 consolidates the performance metrics from these studies on the NASA and Oxford datasets, providing a direct comparison across all major models.

Table 2: Consolidated Performance Metrics on NASA & Oxford Datasets ($RMSE_{SOH}$ %)

Data Split	GRU	LSTM	NODE	PC-RNN	ANODE	ACL	ACLA
Oxford -50%	3.99	4.73	2.37	1.74	2.02	2.12	1.74

Oxford -70%	1.66	2.01	1.13	1.05	1.77	1.89	1.74
Oxford -90%	1.79	1.40	1.46	1.08	1.17	0.97	0.93
NASA-50%	7.57	4.06	4.49	9.54	10.88	7.41	8.87
NASA-70%	4.56	5.88	3.97	3.81	5.36	3.96	3.54
NASA-90%	2.39	4.97	1.85	1.87	4.59	1.41	1.19
Note: Data synthesized from [1] (Table 2, using $\mathcal{L}_{F-norm}$ results) and [1] (Table 3). Lower values are better.							

4.2 The Generalization Challenge: Performance on Unseen Datasets

A critical weakness of many data-driven models is their inability to generalize to new data that

differs from the training set in chemistry, manufacturing, or operating conditions.[1] The ACLA framework was explicitly designed and tested to address this challenge. After being trained exclusively on the NASA and Oxford datasets, the ACLA model was evaluated directly on two entirely independent datasets: Tianjin University (TJU) and Huazhong University of Science and Technology (HUST).[1]

The results were compelling. On the challenging HUST dataset, which contains 77 cells with a different chemistry (LFP), ACLA achieved a low average $RMSE_{SOH}$ of 2.24% and an average EOL absolute error (AE_{EOL}) of 5.33%. This performance represents a 57% and 54.7% reduction in error, respectively, compared to the baseline NODE model.[1] This provides strong evidence that the sophisticated feature extraction within the ACLA architecture allows it to learn more fundamental degradation patterns that are transferable across different battery types.

The performance of these models reveals a subtle trade-off between theoretical expressiveness and practical robustness. While ANODE is theoretically more powerful than NODE due to its ability to learn more complex functions, its performance on the stochastic NASA dataset was highly unstable, with an $RMSE_{SOH}$ exceeding 10% in one scenario.[1] The simpler NODE and discretized PC-RNN models, however, remained more stable. This suggests that the increased flexibility afforded by ANODE's augmented dimensions can lead to overfitting or learning unstable dynamics when the underlying degradation process is noisy or irregular—a critical consideration for real-world deployment. The success of the ACLA model, which uses an ANODE backend, indicates that this instability can be mitigated. Its powerful Attention-CNN-LSTM frontend acts as a sophisticated filter, transforming the raw, potentially noisy input data into a smoother, more informative latent representation. This frontend effectively regularizes the problem, allowing the powerful ANODE solver to model the core degradation trend without being derailed by noise. This highlights a key design principle: the more complex and expressive the ODE solver, the more critical the quality and stability of its initial state vector become.

Table 3 summarizes the generalization performance of the NODE family of models on the unseen TJU and HUST datasets.

Table 3: Generalization Performance on Unseen TJU & HUST Datasets

Dataset	Metric	NODE	ANODE	ACL	ACLA
HUST	$RMSE_{SOH}$ % (Mean ± Std)	5.22 ± 0.64	3.45 ± 0.68	2.34 ± 0.91	**2.24 ± 0.99**

	AE_{EOL} % (Mean ± Std)	11.76 ± 1.62	8.05 ± 1.55	5.60 ± 2.20	**5.33 ± 2.45**
TJU	$RMSE_{SOH}$ % (Mean ± Std)	**0.94 ± 0.24**	1.17 ± 0.25	1.04 ± 0.27	1.04 ± 0.27
	AE_{EOL} % (Mean ± Std)	1.23 ± 0.59	1.28 ± 0.76	1.24 ± 0.76	**1.01 ± 0.72**
Note: Data from [1] (Table 5). The best-performing model for each metric is in bold.					

V. Critical Implementation Details: Data, Features, and Training Strategies

The performance of any machine learning model is intrinsically linked to the data it is trained on and the methods used for training. This section details these critical aspects for NODE-based battery prognostics.

5.1 An Overview of Benchmark Datasets

The research in this area relies on several public datasets, whose heterogeneity provides a robust testbed for model evaluation:

- **NASA Dataset:** Contains data from 18650-type LCO/NCA cells and is known for its irregular degradation patterns, including capacity regeneration phenomena, making it a challenging benchmark.[1]
- **Oxford Dataset:** Features LCO/NMC pouch cells and generally exhibits smoother, more predictable degradation curves compared to the NASA set.[1]
- **TJU Dataset:** Comprises cells with a blended NCM+NCA cathode chemistry.[1]
- **HUST Dataset:** A large-scale dataset with 77 LFP cells, providing significant diversity in cycle life and degradation behavior, making it an excellent test for model generalization.[1]

The distinct chemistries, form factors, and resulting degradation behaviors across these datasets underscore the primary challenge in the field: creating universal, generalizable battery models.[1]

5.2 Feature Extraction from Electrochemical Signals

A consistent feature engineering pipeline is employed across the primary studies, demonstrating a consensus on effective feature selection.[1]

- **Data Source:** The primary source of information is the voltage curve during the constant current (CC) charging phase.[1] This phase is chosen because charging protocols are typically well-controlled and standardized in applications, leading to more consistent and comparable features than variable discharge profiles.
- **Extraction Process:** The process converts the continuous voltage-time curve into a fixed-size feature vector for each cycle. First, the voltage range during CC charging is divided into a set of N_V equispaced voltage points. The features are then the normalized times required to reach each of these voltage points.[1] For a given cycle k, this creates a time/voltage feature vector V_k. This vector is combined with the cycle's SOH value to form the complete state vector $y_k =^T$, which serves as the input to the prognostic model.[1]

5.3 The Impact of Training Methodology

The strategy used to train the models can significantly influence their performance, particularly for early-life predictions.

- **Single-Battery vs. Multi-Battery Learning:** Research has shown that leveraging data from multiple cells during training can substantially improve prognostic accuracy. A

multi-battery approach, which includes data from fully aged cells in the training set, allows the model to learn a more general representation of the entire degradation trajectory.[1] This is especially effective for early-stage predictions. For example, when using a GRU model on the NASA dataset to predict SOH from only the first 20% of a battery's life, the multi-battery training approach reduced the $RMSE_{SOH}$ from 18.08% to 6.55%.[1]

- **Influence of Loss Function:** The choice of loss function can also impact learning by changing the relative importance of different features. A study compared two loss functions: one that gave equal weight to all features ($\mathcal{L}_F$) and another that normalized the contribution from the time/voltage features ($\mathcal{L}_{F-norm}$).[1] The results indicated that model performance was not highly sensitive to this specific weighting choice. While some marginal improvements were observed in specific cases, no strong, consistent trend emerged, suggesting that the model architecture is the dominant factor in overall performance.[1]

VI. Emerging Frontiers and Strategic Outlook

The field of battery prognostics using NODEs is rapidly evolving, with current research pointing towards a future defined by the integration of physical knowledge, enhanced interpretability, and the pursuit of universal models.

6.1 The Path to "Grey-Box" Models: Integrating Physical Constraints

A significant limitation of purely data-driven models is that their predictions are not necessarily constrained by the laws of physics, which can lead to unrealistic forecasts and limit their trustworthiness. The emerging frontier is the development of "grey-box" models that fuse the flexibility of neural networks with the rigor of physics-based principles.

The continuous-time framework of NODEs is exceptionally well-suited for this fusion. Research explicitly highlights the high potential of ODE-based codes in **Physics-Informed Neural Network (PINN)** applications.[1] In a PINN framework, the model's loss function is augmented with a term that penalizes violations of known physical laws, which are often expressed as differential equations.[13] For batteries, these could include simplified electrochemical models or conservation laws. This approach guides the learning process, ensuring that the learned dynamics are not only consistent with the data but also physically plausible. This can lead to more robust, data-efficient models that generalize better to unseen

conditions.[15] A related concept is the development of constrained NODEs, where stability guarantees or physical bounds (e.g., SOH must be non-increasing) are enforced directly within the model's architecture.[17]

6.2 Enhancing Interpretability with Novel Architectures

The "black-box" nature of deep learning remains a significant barrier to its adoption in safety-critical systems like BMS, where understanding the reasoning behind a prediction is crucial. Consequently, there is a strong push towards more interpretable architectures. The introduction of **Kolmogorov-Arnold Networks (KANs)** as a parameterization for the ODE dynamics is a prime example of this trend.[1] Unlike MLPs with fixed non-linearities at the nodes, KANs have learnable activation functions on the edges. This structure offers a potential pathway to visualizing and symbolically analyzing the learned degradation dynamics, transforming the core of the model from an opaque black box into a more transparent grey box.[1]

6.3 The Generalization Challenge and Universal Models

The ultimate goal of battery prognostics is to develop a single, universal model that can accurately predict the lifetime of any battery, regardless of its chemistry, design, or operating conditions. This remains a formidable challenge.[1] The demonstrated success of the ACLA model in generalizing to unseen datasets is a significant step in this direction.[1] Future work will likely build on this by explicitly incorporating operational and battery-specific parameters into the model. As proposed in the literature, this can be achieved by reformulating the governing ODE to include an input vector, $\alpha(t)$, that represents variables like temperature, C-rate, and battery chemistry. The learned dynamics would then take the form $\dot{y} = F(y, \theta(t), \alpha(t), t)$, allowing the model to adapt its predictions based on these contextual inputs.[1]

The trajectory of this research field points towards a convergence of data-driven and physics-based modeling paradigms. The initial success of data-driven methods demonstrated their power but also exposed their limitations in terms of generalization and trustworthiness.[12] The progression from simple NNs to LSTMs and then to NODEs provided more sophisticated mathematical frameworks but did not fundamentally break out of the "black-box" paradigm.[1] The current emphasis on generalization, as seen in the ACLA study [1], and the explicit calls for

incorporating physical constraints via PINNs [1] signal a maturation of the field. It is a direct acknowledgment that data alone is insufficient for building truly robust prognostic tools. By grounding data-driven models in physical reality, researchers aim to regularize the learning process, prevent non-physical predictions, and improve performance in low-data regimes. The state-of-the-art model of the future will likely combine a powerful feature extractor, like that of ACLA, with a core ODE solver that is constrained by physical laws, making it both highly accurate and scientifically plausible.

VII. Concluding Remarks and Recommendations

7.1 Synthesis of Key Findings

This analysis of battery degradation modeling using Neural Ordinary Differential Equations reveals several key findings that mark a significant advancement in the field of battery prognostics.

- **Superiority of the Continuous-Time Paradigm:** By modeling degradation as a continuous dynamical process, NODE-based architectures consistently outperform traditional discrete-time RNN models like LSTM and GRU in both SOH and EOL prediction accuracy, particularly for long-term forecasting horizons.
- **Architectural Evolution is Key:** The current state of the art is defined by advanced hybrid architectures. Models like ACLA achieve superior accuracy and, critically, enhanced generalization to unseen battery types by coupling a sophisticated feature extraction frontend (Attention-CNN-LSTM) with a continuous-time ANODE solver.
- **Emergence of Specialized Models:** The field is diversifying, with different architectures targeting different challenges. While ACLA focuses on maximizing predictive power through feature abstraction, models like the GAT-KAN-ODE framework prioritize the fusion of heterogeneous data sources and the pursuit of model interpretability.
- **Identifiable Performance Trade-offs:** There exists a clear trade-off between a model's theoretical expressiveness and its practical robustness. More flexible models like ANODE can be susceptible to instability on noisy data, a challenge that can be mitigated by robust feature extraction or the incorporation of physical constraints.

7.2 Strategic Recommendations for R&D and Industrial Application

Based on these findings, the following strategic recommendations are proposed:

For Research and Development:

1. **Prioritize Physics-Informed Models:** The primary research focus should shift from purely data-driven models to the development of universal, physics-informed neural ODEs. Integrating physical constraints via PINN methodologies or direct architectural constraints is the most promising path to creating models that are robust, data-efficient, and can generalize across a wide range of battery chemistries and operating conditions.
2. **Invest in Interpretable Architectures:** To build trust and facilitate adoption in safety-critical applications, research into interpretable dynamics learners, such as KANs, should be accelerated. The ability to inspect and understand the learned degradation function is a crucial next step for the field.
3. **Develop Universal Training Strategies:** The effectiveness of multi-battery training for improving early-life prediction should be further explored. Research should focus on creating large, diverse, and standardized training datasets and developing methodologies to train universal models that can ingest battery-specific parameters as inputs.

For Industrial Application and BMS Development:

1. **Adopt a Phased Implementation Approach:** While complex models like ACLA demonstrate state-of-the-art performance, simpler and more computationally efficient variants like PC-RNN or a well-regularized NODE may be more suitable for initial deployment on resource-constrained embedded systems in current-generation BMS.
2. **Leverage Conservative EOL Predictions:** A key operational advantage of ODE-based models is their tendency to provide conservative (pessimistic) EOL estimates. This should be treated as a valuable safety feature, enabling more prudent maintenance scheduling and reducing the risk of unexpected battery failure.
3. **Emphasize Data Quality and Feature Engineering:** The performance of all models, particularly the more advanced ones, is highly dependent on the quality of the input data. Industrial efforts should focus on standardized data collection protocols, especially for the CC charging phase, to enable the effective application of these advanced prognostic techniques.

Works cited

1. neural-ordinary-differential-equations-and-recurrent-neural-networks-for-predicting-the-state-of-health-of-batteries.pdf
2. Dynamical System Modeling Using Neural ODE - MATLAB & Simulink - MathWorks, accessed October 17, 2025, https://uk.mathworks.com/help/deeplearning/ug/dynamical-system-modeling-using-neural-ode.html

3. Neural Ordinary Differential Equations, accessed October 17, 2025, http://papers.neurips.cc/paper/7892-neural-ordinary-differential-equations.pdf
4. Understanding Neural ODE's - Jonty Sinai, accessed October 17, 2025, https://jontysinai.github.io/jekyll/update/2019/01/18/understanding-neural-odes.html
5. Chapter 3: Neural Ordinary Differential Equations - Deep Implicit Layers, accessed October 17, 2025, http://implicit-layers-tutorial.org/neural_odes/
6. Neural Ordinary Differential Equation (Neural ODE) - Ayan Das, accessed October 17, 2025, https://ayandas.me/blogs/2020-03-20-neural-ode.html
7. [1806.07366] Neural Ordinary Differential Equations - arXiv, accessed October 17, 2025, https://arxiv.org/abs/1806.07366
8. Neural Ordinary Differential Equations (Neural ODEs): Rethinking Architecture - Medium, accessed October 17, 2025, https://medium.com/@justygwen/neural-ordinary-differential-equations-neural-odes-rethinking-architecture-272a72100ebc
9. What are the practical uses of Neural ODEs? - Cross Validated - Stack Exchange, accessed October 17, 2025, https://stats.stackexchange.com/questions/445562/what-are-the-practical-uses-of-neural-odes
10. A Novel Neural-Ode Model for the State of Health Estimation of Lithium-Ion Battery Using Charging Curve - ResearchGate, accessed October 17, 2025, https://www.researchgate.net/publication/391581592_A_Novel_Neural-Ode_Model_for_the_State_of_Health_Estimation_of_Lithium-Ion_Battery_Using_Charging_Curve
11. A novel Neural-ODE model for the state of health estimation of lithium-ion battery using charging curve - arXiv, accessed October 17, 2025, https://arxiv.org/pdf/2505.05803
12. Neural ordinary differential equations and recurrent neural networks for predicting the state of health of batteries - ChemRxiv, accessed October 17, 2025, https://chemrxiv.org/engage/api-gateway/chemrxiv/assets/orp/resource/item/620b6faabd05a030d9077c9a/original/neural-ordinary-differential-equations-and-recurrent-neural-networks-for-predicting-the-state-of-health-of-batteries.pdf
13. Differential Equation-Informed Neural Networks For State-of-Charge Estimation - Scribd, accessed October 17, 2025, https://www.scribd.com/document/698138291/Differential-Equation-Informed-Neural-Networks-for-State-of-Charge-Estimation
14. Hybrid Modeling of Lithium-Ion Battery: Physics-Informed Neural Network for Battery State Estimation - MDPI, accessed October 17, 2025, https://www.mdpi.com/2313-0105/9/6/301
15. Differential Equation-Informed Neural Networks for State of Charge Estimation | Request PDF - ResearchGate, accessed October 17, 2025, https://www.researchgate.net/publication/376010373_Differential_Equation-Informed_Neural_Networks_for_State_of_Charge_Estimation
16. PINN surrogate of Li-ion battery models for parameter inference, Part I: Implementation and multi-fidelity hierarchies for the single-particle model

(Journal Article) - OSTI, accessed October 17, 2025, https://www.osti.gov/pages/biblio/2429276
17. CONSTRAINED NEURAL ORDINARY DIFFERENTIAL EQUATIONS WITH STABILITY GUARANTEES - OpenReview, accessed October 17, 2025, https://openreview.net/pdf?id=mTmgaxwynS
18. MINN: Learning the dynamics of differential-algebraic equations and application to battery modeling - arXiv, accessed October 17, 2025, https://arxiv.org/html/2304.14422v2