Physics-Informed Feature Disentanglement for Autonomous Acceleration of Lithium-Ion Battery Design and Analysis

Codex, Cursor, Comet, GPT-5.2 Thinking, Gemini (Flash)*

Dongju Kim
KAERI
Daejeon, Republic of Korea
dongjukim.dev@gmail.com

Dayoung Kam
POSCO Holdings
Incheon, Republic of Korea
kamda0@naver.com

Yoo Jung Choi
POSCO Holdings
Incheon, Republic of Korea
yjchoi1403@gmail.com

Jaeyeon Byun
University of Science and Technology, Applied AI
Daejeon, Republic of Korea
9722jayon@kisti.re.kr

Bokyeong Kim
University of Science and Technology, Artificial Intelligence
KAERI, Daejeon, Republic of Korea
qhrud418@gmail.com

Inchul Park†
POSCO NEXT Hub
Incheon, Republic of Korea
inchul.rist@gmail.com

Abstract

The multi-dimensional design space of battery design presents a critical "dimensionality crisis": the design space is so wide that traditional machine learning models fail to capture complex, non-linear parameter interactions. This study proposes an AI-driven framework based on physics-informed feature disentanglement to overcome these limitations. By integrating a specialized feature extraction algorithm with a linear regression model trained on high-fidelity PyBaMM simulations, we successfully isolated entangled electrochemical signatures from raw voltage profiles into deterministic physical markers. Our results demonstrate that this framework achieves a high coefficient of determination (R^2) exceeding 0.9 for key parameters when physics-informed features are used. Crucially, the model's reliability is secured by correlating AI decision-making with internal physical phenomena, such as the evolution of lithium concentration. This mechanism-aware approach supports forward performance prediction and interpretability toward future autonomous inverse design of LFP architectures, effectively bridging the gap between theoretical modeling and rapid industrial application.

*As of 2026-01-30. Codex: OpenAI software engineering agent (codex-1). Cursor: AI-assisted IDE (cursor.com). Comet: Comet ML platform for experiment tracking and LLM evaluation (comet.com). GPT-5.2 Thinking: OpenAI reasoning model. Gemini (Flash): Google Gemini 2.5 Flash (fast variant).

†Corresponding author.

Preprint. Under review.

1 Introduction

The rapid proliferation of renewable energy infrastructure and the staggering surge in electricity demand—driven largely by the expansion of AI-centric data centers—have catalyzed a global transition in power management. In this landscape, the Energy Storage System (ESS) market has emerged as a primary growth engine for the battery industry [1], providing the critical buffer required for grid stability. To meet the stringent operational demands of ESS [2], including long-term reliability, inherent fire safety, and cost-effectiveness, Lithium Iron Phosphate (LFP) has been increasingly adopted as the global standard. [3]

Despite its widespread implementation, the performance of LFP-based cells is dictated by a multi-dimensional design space where electrode architectures—defined by porosity, particle size, and loading levels—interact in highly nonlinear ways [4, 5]. Navigating this complexity via traditional experimental studies remains fundamentally constrained; for instance, independently controlling specific variables like cathode loading while keeping other physical parameters constant is nearly impossible in a laboratory setting due to manufacturing tolerances. To overcome these physical limitations, this work proposes a methodology based on fully controlled simulation data and AI-driven feature disentanglement to accelerate the optimization of LFP cathode architectures.

Current methodologies for electrode design predominantly rely on physics-based simulations or empirical "trial- and-error" approaches. While electrochemical models like the pseudo-two-dimensional (P2D) framework [6] provide deep insights into lithium-ion transport, their main drawbacks are not computational cost but the difficulty of measuring or fitting all required parameters and the resulting inaccuracy of predictions. Conversely, while machine learning (ML) [7–10] models offer rapid prediction, they often struggle with the high-dimensional and non-linear interactions inherent in battery electrodes [11]. For instance, while simple linear regression may accurately predict performance under highly controlled, single-variable conditions (e.g., varying only loading levels), its predictive power, quantified by the coefficient of determination (R^2), degrades sharply as the number of input variables increases. This "dimensionality crisis" is particularly acute in LFP systems [12], where the characteristic two-phase reaction behavior and transport complexities introduce significant non-linearity between architectural parameters like porosity and electrode thickness.

In this work, we demonstrate an AI-driven framework for the autonomous acceleration of LFP cathode design, utilizing a physics-informed [13] feature disentanglement approach. Here we use "feature disentanglement" to mean the separation of entangled electrochemical signals into interpretable physical markers, achieved in practice by "feature extraction" (e.g., dQ/dV-based metrics). We generate a comprehensive dataset of 1,000 simulations using the Python Battery Mathematical Modeling (PyBaMM) framework, systematically controlling five key physical parameters—porosity, particle radius, and electrode thickness—to map the electrochemical parameter space. To overcome the performance degradation observed in traditional models, we introduce a specialized feature extraction algorithm integrated with a linear regression (OLS) prediction model [8, 19]. This approach allows for a "mechanism-aware" analysis that identifies the hidden interactions between architectural variables, maintaining high predictive accuracy even as the complexity of the design space expands.

We provide a comparative analysis of prediction difficulty and a feature importance matrix linking AI decisions to physical phenomena such as lithium concentration evolution. This mechanism-aware interpretability supports forward performance prediction; inverse design of LFP cathode architectures is left as future work.

2 Schema

The proposed schema represents a bidirectional design and diagnostic framework that integrates physics-informed feature disentanglement to bridge the gap between electrode architecture and electrochemical performance. It consists of a Forward Design mode, which functions as a virtual prototyping tool by instantaneously generating predicted discharge voltage curves and dQ/dV profiles from user-defined physical parameters—such as electrode thickness and porosity—thereby drastically accelerating the battery development cycle. Simultaneously, the schema incorporates an Inverse Diagnosis mode that derives internal material properties from empirical performance data, allowing

for the non-destructive estimation of key physical parameters to analyze manufacturing quality variance or diagnose degradation mechanisms like SEI growth without invasive cell disassembly.

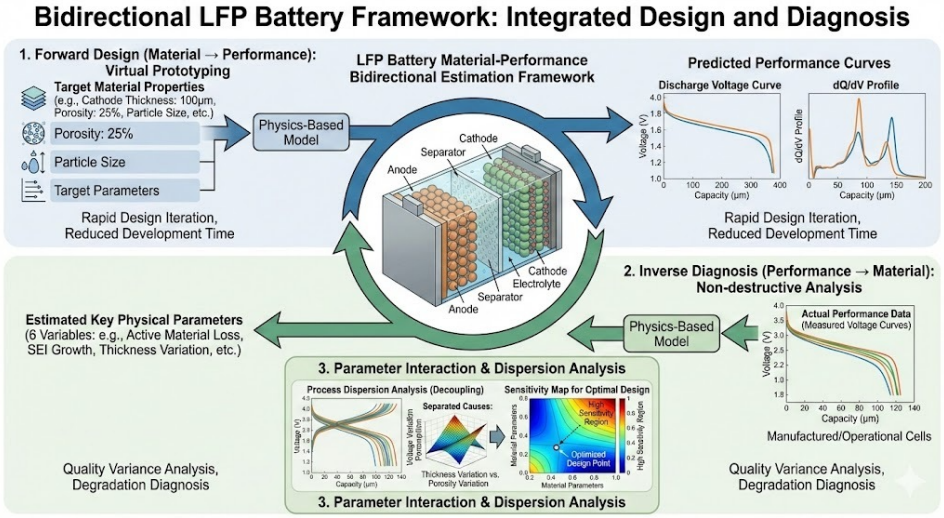


Figure 1: Bidirectional LFP framework: parameter estimation and performance prediction.

3 Methodology

3.1 Experimental Design

We conducted a systematic experimental study to evaluate the performance of linear regression models for identifying battery physical properties from discharge curves. The experimental design consists of four main stages: parameter combination generation, random sampling, battery simulation, and model training.

3.1.1 Parameter Combination Generation

We systematically generated parameter combinations to investigate the effect of varying the number of input parameters on prediction performance. Five physical properties of lithium iron phosphate (LFP) batteries were selected as targets: positive and negative electrode thickness, positive and negative electrode porosity, and positive particle radius (units as in Table 1). For each combination size $n \in \{1, 2, 3, 4, 5\}$ we generated all $\binom{5}{n}$ combinations, giving 5, 10, 10, 5, and 1 combinations respectively; within each combination, the selected parameters varied within predefined ranges and the rest were fixed. The parameter ranges are given in Table 1.

Table 1: Parameter ranges for LFP battery simulation

Parameter	Minimum	Maximum
Positive electrode thickness	5.0×10^{-5}	1.5×10^{-4}
Negative electrode thickness	4.0×10^{-5}	1.2×10^{-4}
Positive electrode porosity	0.15	0.40
Negative electrode porosity	0.25	0.50
Positive particle radius	5.5×10^{-8}	1.0×10^{-6}

3.1.2 Random Parameter Sampling

For each parameter combination we drew 200 parameter sets by Latin Hypercube Sampling (LHS) [18] with a fixed seed (42). LHS partitions each dimension into N equal-probability intervals and samples one point per interval, ensuring uniform coverage and reproducibility.

Algorithm 1 LFP Battery Property Identification Pipeline

Input: Parameter set $\mathcal{P} = \{p_1, \ldots, p_5\}$ (5 physical properties), combination sizes $n \in \{1, 2, 3, 4, 5\}$, number of samples $N = 200$, random seed $s = 42$
Output: Set of trained models $\mathcal{M}$, evaluation metrics $\mathcal{E}$, predictions $\mathcal{Y}$

1. Initialize $\mathcal{M} \leftarrow \emptyset$, $\mathcal{E} \leftarrow \emptyset$, $\mathcal{Y} \leftarrow \emptyset$.
2. For each combination size $n \in \{1, \ldots, 5\}$ and each combination c of n parameters from $\mathcal{P}$: (a) sample N parameter sets via LHS with seed s; (b) run PyBaMM discharge simulations and validate (voltage 2.0–3.6 V); (c) extract features (mean, std, min, max, median, range, IQR, final value, slope per state variable); (d) for each target $t \in c$, merge features with t, split data 60/20/20 with seed s, standardize on train set, fit linear regression by OLS, compute R^2/MAE/RMSE/MAPE on test set, and append the model and metrics to $\mathcal{M}$, $\mathcal{E}$, $\mathcal{Y}$.
3. Return $\mathcal{M}$, $\mathcal{E}$, $\mathcal{Y}$.

3.1.3 Battery Simulation

For each of the 200 parameter sets we ran a discharge simulation in PyBaMM [14] with the Doyle–Fuller–Newman (DFN) [6] model and the Prada2013 LFP parameter set; outputs include voltage, current, and internal states (e.g., electrode potentials, particle surface concentrations). Simulations were validated to satisfy a 2.0–3.6 V voltage range and physical consistency.

3.1.4 Feature Extraction

From each simulation we extracted, per state variable, the mean, standard deviation, min and max, median, range, interquartile range, final value, and linear slope. This produced one feature vector per sample (200 per combination).

3.2 Machine Learning Pipeline

3.2.1 Data Preprocessing and Splitting

Features were merged with parameter values by sample ID and split 60/20/20 into train/validation/test with a fixed seed. Inputs were standardized to zero mean and unit variance using the training set: $z_i = (x_i - \mu_i)/\sigma_i$, with μ_i, σ_i computed from the training set and the same transform applied to validation and test sets.

3.2.2 Model Training

We used linear regression $y = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \epsilon$ fitted by ordinary least squares on the standardized training set, implemented with scikit-learn [19].

3.2.3 Model Evaluation

Performance was evaluated on the test set. We report the coefficient of determination

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{1}$$

where $\hat{y}_i$ is the predicted value and $\bar{y}$ is the mean of the true values, along with MAE, RMSE, and MAPE (MAPE computed only over samples with $y_i \neq 0$). These metrics were also computed on the training and validation sets to assess overfitting.

3.3 Complete Experimental Pipeline

The pipeline is summarized in Algorithm 1. It yields 80 experiments in total (5+20+30+20+5 over the five combination sizes), each producing a trained model and test metrics.

4 Results

4.1 Single-Variable Linear Regression Baseline

To establish a definitive predictive baseline for autonomous battery design, we first evaluated the fidelity of linear surrogate models [8, 19] under strictly controlled, single-variable conditions using a library of 1,000 synthetic datasets generated via the Python Battery Mathematical Modeling (PyBaMM) [14] framework. This in silico validation methodology effectively isolates the deterministic impact of individual electrode parameters on $LiFePO_4$ (LFP) cell performance [15], successfully bypassing the high sampling variability and time-intensive feedback loops that typically bottleneck experimental materials discovery [16, 17].

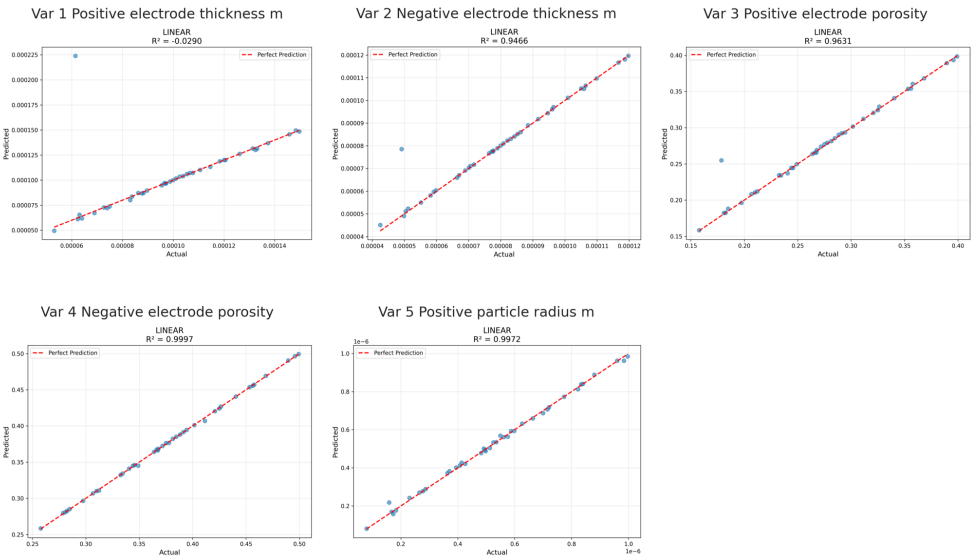


Figure 2: Single-variable linear regression (R^2 per target).

As illustrated in our regression analysis (Figure 2), the framework demonstrates exceptional predictive accuracy across the majority of the primary physical parameters. Notably, near-perfect correlations were achieved for negative electrode porosity ($R^2 = 0.9997$) and positive particle radius ($R^2 = 0.9972$), while negative electrode thickness ($R^2 = 0.9466$) and positive electrode porosity ($R^2 = 0.9631$) also exhibited robust predictive power. These results confirm that within a mechanism-defined simulation environment [6, 14], the fundamental relationships governing electrochemical response remain largely linear when architectural variables are modified independently, providing a rigorous "ground truth" for subsequent multi-dimensional modeling.

The mechanistic interpretation of these findings [12, 15] suggests that the AI model is not merely fitting numerical noise but is capturing physically meaningful electrochemical signatures inherent to the LFP chemistry. The superior predictability of the positive particle radius ($R^2 = 0.9972$) is particularly telling; it underscores that in LFP systems, characterized by one-dimensional lithium-ion diffusion paths and low ionic conductivity, the solid-state diffusion within the cathode serves as the primary rate-determining step. Consequently, the particle radius provides a dominant and linear signal for the model to capture. Conversely, the comparatively lower predictive fidelity for negative electrode thickness and the negligible correlation observed for positive electrode thickness ($R^2 = -0.0290$) reveal a critical sensitivity threshold. In the "positive-limited" design typical of commercial LFP/Graphite cells, variations in the thickness of the positive electrode can yield non-monotonic or effectively "silent" performance changes within specific electrochemical bounds.

4.2 Pairwise Parameter Combinations

Figure 3 shows the pairwise accuracy heatmap. The main finding is a critical conflict zone: the combination of positive and negative electrode thicknesses (Pair 1-2) exhibits $R^2 = -2.00$, while most other pairs maintain high fidelity ($R^2 > 0.98$). This suggests a severe signal redundancy; since

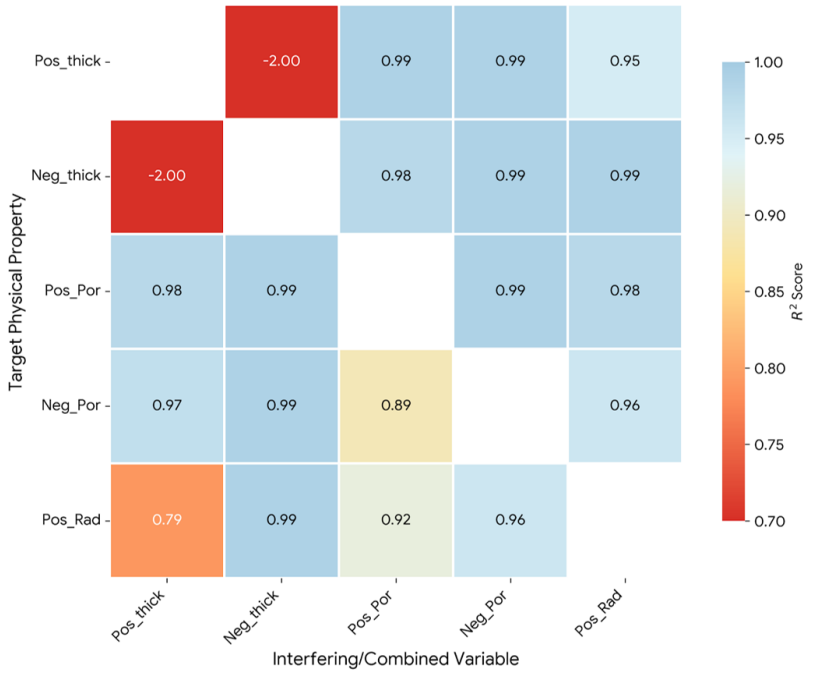


Figure 3: Pairwise R^2 heatmap; thickness pair shows conflict zone ($R^2 = -2.00$).

both parameters primarily dictate the total loading and ohmic response of the cell, their individual impacts on the terminal voltage overlap extensively. This "strong coupling" effect renders the simple linear model incapable of distinguishing the deterministic contribution of each electrode's thickness. These results demonstrate that while linear regression is robust for independent variables, certain physical pairings create an "identification bottleneck" due to mechanistic entanglement. Identifying these interference patterns is crucial for defining the limits of AI-driven design and highlights the necessity of the dQ/dV feature extraction introduced in subsequent sections to resolve such physical overlaps.

4.3 Effect of Feature Number on Model Performance

To systematically investigate how model performance scales with the number of input features, we evaluated the predictive accuracy across all possible parameter combinations ranging from single-variable (1-set) to full five-variable (5-set) configurations. This analysis provides crucial insights into the "dimensionality crisis" [12] by quantifying the performance degradation that occurs as the complexity of the design space increases.

Figure 4 presents the heatmap of test R^2 as a function of the number of features. The single-feature condition confirms the baseline established in Figure 2, with most variables near unity.

However, the progressive introduction of additional features exposes the fundamental limitations of simple linear models in multi-dimensional spaces. For negative electrode porosity, the R^2 score decreases monotonically from 1.000 (1 feature) to 0.804 (5 features), representing a 19.6% relative degradation. Similarly, negative electrode thickness shows a gradual decline from 1.000 to 0.865, while positive electrode porosity exhibits a more modest reduction to 0.891. These patterns suggest that as more variables are simultaneously varied, their electrochemical signatures become increasingly entangled, creating interference that the linear model cannot fully decouple.

Interestingly, positive particle radius and positive electrode thickness demonstrate non-monotonic behavior, with R^2 scores fluctuating based on specific feature combinations. For instance, positive particle radius achieves $R^2 = 0.997$ with a single feature, drops to 0.922 with two features, but recovers to 0.994 with three features before stabilizing around 0.977-0.990. This suggests that certain feature combinations can actually improve predictive performance by providing complementary

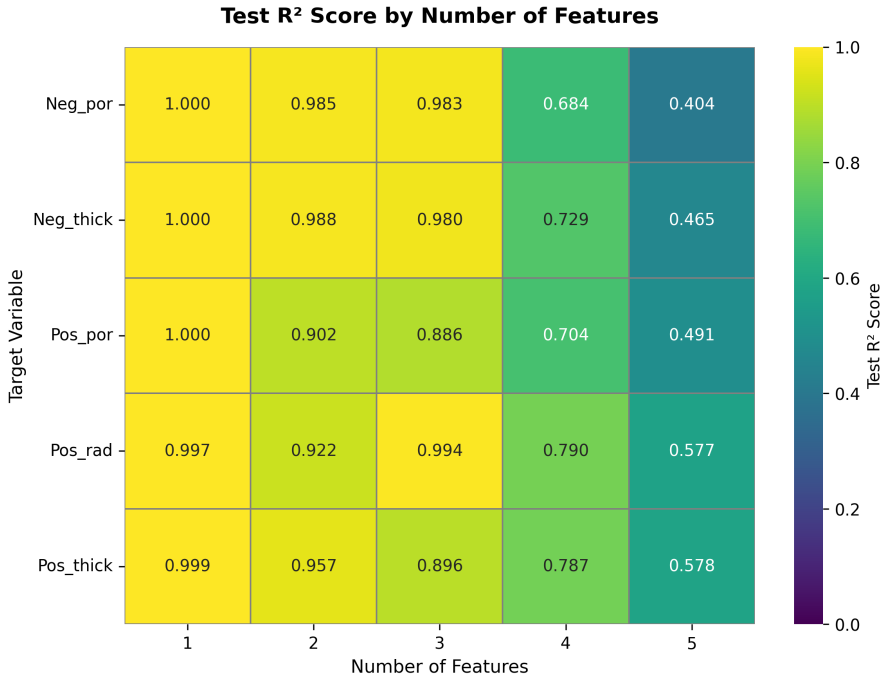


Figure 4: Test R^2 vs. number of features (1–5).

information, while others introduce conflicting signals. These observations highlight the complex, non-linear interactions between architectural parameters and underscore the necessity of advanced feature extraction methods, as demonstrated in the following section.

4.4 Overcoming Dimensionality Limits via Physics-Informed Feature Extraction

To overcome the limitations of simple data augmentation and resolve the dimensionality crisis [12], we implemented a physics- informed feature extraction strategy utilizing differential capacity (dQ/dV) analysis [20, 21]. The fundamental rationale for this approach lies in the inherent nature of raw voltage-time ($V - t$) data, where electrochemical signals are deeply "entangled" in multi-variable environments. Particularly in Lithium Iron Phosphate (LFP) systems, the characteristic flat voltage plateaus make it exceptionally difficult for a linear model to distinguish subtle shifts caused by structural parameters, such as electrode thickness or porosity, from raw voltage values alone. In contrast, feature extraction via dQ/dV [20–24] transforms these subtle slope changes into distinct peaks that represent specific phase transition mechanisms. These peaks serve as unique "physical signatures," providing the model with clear, deterministic markers of the battery's internal state.

Figure 5 illustrates the transformative impact of this feature extraction on the model's predictive fidelity. For the identification of Positive Electrode Thickness (POS THICK), the baseline linear regression model (left) struggled to overcome multi-variable interference, resulting in high dispersion and a low R^2 of 0.5783. However, by providing physically meaningful dQ/dV markers as inputs (Super FE Model, right), the model successfully filtered out the "crosstalk" between variables, allowing predictions to align almost perfectly with the identity line, achieving an R^2 of 0.9946. The quantitative success and the observed limitations of this feature extraction are summarized in Table 2. While the Super FE model achieved a near-perfect improvement for POS THICK ($R^2 = 0.9946$) and a substantial gain for NEG THICK ($R^2 = 0.8581$), it is important to note that this specific method is not universally applicable to all parameters. As evidenced by the steady performance observed in POS POROS, NEG POROS, and POS RADIUS, the results remained nearly identical to the baseline linear regression.

This suggests that while dQ/dV peaks are excellent at capturing structural information related to phase transitions, they may dilute diffusion-related kinetic signatures or introduce non-linearities that a linear model cannot easily map. Therefore, future research must consider a wider array of feature

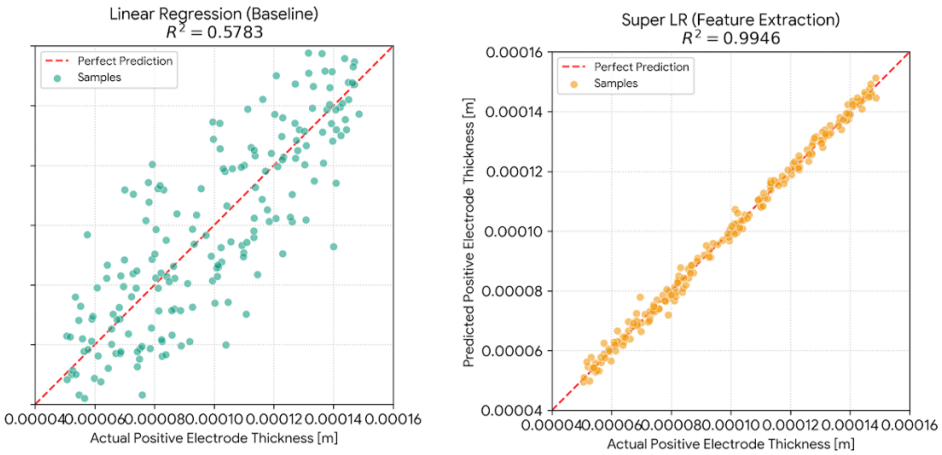


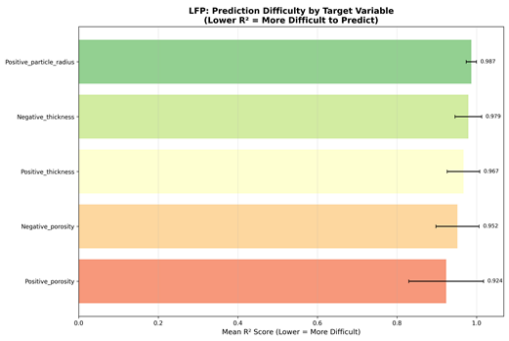
Figure 5: POS THICK: baseline vs. feature extraction (R^2 0.58 vs. 0.99).

extraction techniques—such as voltage relaxation analysis, dV/dQ, or frequency-domain features—to develop an optimized, hybrid feature set tailored to the unique physical sensitivities of each target parameter.

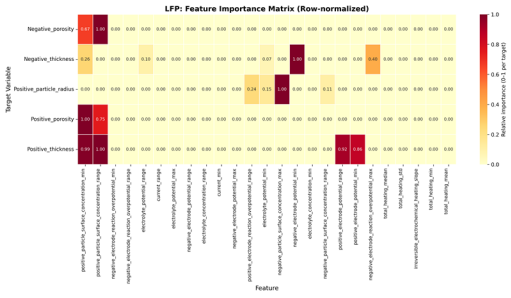
Table 2: Baseline LR vs. super FE model R^2.

Target Parameter (Abbr.)	Baseline LR	Super FE Model	Performance Summary
POS THICK	0.5783	0.9946	Near-perfect improvement (↑)
NEG THICK	0.6668	0.8581	Substantial improvement (↑)
POS POROS	0.6274	0.6043	Steady performance
NEG POROS	0.3389	0.3402	Steady performance
POS RADIUS	0.7404	0.7430	Steady performance

4.5 Prediction Difficulty and Feature Importance Analysis



(a) Analysis of prediction difficulty using R^2 scores and error bars.



(b) Feature importance matrix for physical property identification.

Figure 6: Robustness: (a) prediction difficulty; (b) feature importance matrix.

To evaluate the robustness of the proposed framework, we conducted a comprehensive analysis of prediction difficulty and feature-target correlations (Figure 6). Figure 6(a) shows prediction difficulty; positive porosity is hardest ($R^2 = 0.924$), positive particle radius easiest ($R^2 = 0.987$). Figure 6(b) shows that thickness and porosity correlate strongly with

positive_particle_surface_concentration_* features, consistent with lithium concentration evolution. This "mechanism-aware" learning ensures that the framework provides physically consistent predictions, serving as a transparent tool for autonomous battery design.

5 Discussion

The results confirm that physics-informed feature extraction [13, 20, 21] is critical for overcoming the dimensionality crisis. The pairwise thickness combination (Pair 1–2) that exhibited $R^2 = -2.00$ (Figure 3) illustrates this crisis: raw voltage-based linear regression cannot disentangle strongly coupled parameters. By contrast, dQ/dV-based feature extraction resolves this entanglement for key targets such as POS THICK (Figure 5), turning slope changes into distinct physical markers. The dramatic POS THICK improvement and the need for hybrid strategies are discussed in Section 4.1.

The feature importance analysis (Figure 6) supports mechanism-aware interpretability.

5.1 Limitations

While our framework demonstrates strong predictive performance across most physical parameters, several limitations should be acknowledged. First, the dQ/dV feature extraction method is not universally applicable to all parameters. As shown in Table 2, while POS THICK and NEG THICK show substantial improvements, other parameters such as POS RADIUS and porosity levels show results that are nearly identical to the baseline. This suggests that dQ/dV peaks, while excellent for capturing structural information related to phase transitions, may dilute diffusion-related kinetic signatures that are critical for identifying certain parameters like particle radius.

Second, our study is based entirely on PyBaMM simulation data [14], which, while providing controlled experimental conditions, may not fully capture the variability and noise present in real-world experimental data [10, 30]. The transition from simulation to experimental validation remains a critical next step.

Third, the linear regression model, while effective for many parameters, has inherent limitations in capturing highly non-linear relationships. As demonstrated in Figure 4, certain parameter combinations exhibit non-monotonic behavior, suggesting that more sophisticated non-linear models [25] may be necessary for comprehensive coverage of the design space.

Finally, our framework has been validated on a specific LFP chemistry and may require adaptation for other battery chemistries with different electrochemical behaviors. The generalizability to other cathode materials (e.g., NCM, NCA) [11, 16] would need to be established through additional studies. Electrode architecture innovations such as 3D conductive scaffolds [26] and magnetically ordered microstructures [27] could be integrated with our inverse design pipeline in future work.

6 Conclusion

This study overcomes the "dimensionality crisis" through physics-informed feature disentanglement and dQ/dV-based feature extraction, achieving high predictive fidelity ($R^2 > 0.9$) across key physical parameters when physics-informed features are used. The feature importance analysis (Figure 6) confirms that the model captures physically meaningful mechanisms rather than merely fitting numerical noise, supporting its use for forward prediction; autonomous inverse design remains future work. Future work will focus on integrating Reinforcement Learning (RL) to further optimize the inverse design process [28] and on bridging high-fidelity simulations with real-world experimental data. Data-driven approaches have similarly accelerated discovery in related materials domains [28, 29, 10, 30].

References

References

[1] Raijmakers, L. H. J., Danilov, D. L., Eichel, R.-A., & Notten, P. H. L. (2019). A review on various temperature-indication methods for Li-ion batteries. *Applied Energy*, 240, 918–945.

[2] Dunn, B., Kamath, H., & Tarascon, J.-M. (2011). Electrical energy storage for the grid: a battery of choices. *Science*, 334(6058), 928–935.

[3] Wang, Y., Zhang, Y., & Zhang, Z. (2016). Lithium iron phosphate and its derivatives as promising cathode materials for high energy-density and safe lithium-ion batteries. *Energy & Environmental Science*, 9, 141–170.

[4] Strobridge, F. C., Orvananos, B., Croft, M., Yu, H.-C., Robert, R., Liu, H., Zhong, Z., Connolley, T., Drakopoulos, M., Thornton, K., & Grey, C. P. (2015). Mapping the inhomogeneous electrochemical reaction through porous $LiFePO_4$-electrodes in a standard coin cell battery. *Chemistry of Materials*, 27(7), 2374–2386.

[5] Owen, J. R., & Lakeman, J. B. (2023). Optimisation of industrially relevant electrode formulations for LFP cathodes in lithium ion cells. *Batteries*, 9(4), 192.

[6] Doyle, M., Fuller, T. F., & Newman, J. (1993). Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell. *Journal of the Electrochemical Society*, 140(6), 1526–1533.

[7] He, X., Zhao, K., & Chu, X. (2021). AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems*, 212, 106622.

[8] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

[9] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.

[10] Severson, K. A., Attia, P. M., Jin, N., Perkins, N., Jiang, B., Chen, M. H., ... & Bazant, M. Z. (2019). Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4(5), 383–391.

[11] Julien, C. M., Mauger, A., Zaghib, K., & Groult, H. (2014). Comparative issues of cathode materials for Li-ion batteries. *Inorganic Chemistry*, 53(22), 11927–11954.

[12] Padhi, A. K., Nanjundaswamy, K. S., & Goodenough, J. B. (1997). Phospho-olivines as positive-electrode materials for rechargeable lithium batteries. *Journal of the Electrochemical Society*, 144(4), 1188–1194.

[13] Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378, 686–707.

[14] Sulzer, V., Marquis, S. G., Timms, R., Robinson, M., & Chapman, S. J. (2021). Python Battery Mathematical Modelling (PyBaMM). *Journal of Open Research Software*, 9(1), 14.

[15] Prada, E., Di Domenico, D., Creff, Y., Bernard, J., Sauvant-Moynot, V., & Huet, F. (2013). A simplified electrochemical and thermal aging model of $LiFePO_4$-graphite Li-ion batteries: power and capacity fade simulations. *Journal of the Electrochemical Society*, 160(4), A616–A628.

[16] Kang, B., & Ceder, G. (2009). Battery materials for ultrafast charging and discharging. *Nature*, 458, 190–193.

[17] Delacourt, C., & Safari, M. (2012). Life simulation of a graphite/$LiFePO_4$ cell under cycling and storage. *Journal of the Electrochemical Society*, 159(8), A1213–A1224.

[18] McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2), 239–245.

[19] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., & Perrot, M. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85), 2825–2830.

[20] Dubarry, M., & Liaw, B. Y. (2009). Identify capacity fading mechanism in a commercial $LiFePO_4$ cell. *Journal of Power Sources*, 194(1), 541–549.

[21] Smith, A. J., Burns, J. C., & Dahn, J. R. (2011). High-precision differential capacity analysis of $LiMn_2O_4$/graphite cells. *Electrochemical and Solid-State Letters*, 14(2), A39–A41.

[22] Smith, A. J., & Dahn, J. R. (2012). Delta differential capacity analysis. *Journal of the Electrochemical Society*, 159(3), A290–A293.

[23] Bloom, I., Jansen, A. N., Abraham, D. P., Knuth, J., Jones, S. A., Battaglia, V. S., & Henriksen, G. L. (2005). Differential voltage analyses of high-power lithium-ion cells. *Journal of Power Sources*, 139(1–2), 295–303.

[24] Olson, J. Z., Lopez, C. M., & Dickinson, E. J. F. (2023). Differential analysis of galvanostatic cycle data from Li-ion batteries: Interpretative insights and graphical heuristics. *Chemistry of Materials*, 35(4), 1487–1513.

[25] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*, arXiv:1409.1556.

[26] Yu, S. H., Park, C., & Jung, H. G. (2023). An electronically conductive 3D architecture with controlled porosity for $LiFePO_4$ cathodes. *Frontiers in Materials*, 10, 1213872.

[27] Kim, W., Hwang, C., Kim, Y. M., Yu, J.-S., Kim, Y.-J., Kim, K. J., & Kim, H.-s. (2024). Modulation of lithium iron phosphate electrode architecture by magnetic ordering for lithium-ion batteries. *Journal of Materials Chemistry A*, 12, 15678–15688.

[28] Guo, X., Wang, Z., Yang, J.-H., & Gong, X.-G. (2024). Machine-learning assisted high-throughput discovery of solid-state electrolytes for Li-ion batteries. *Journal of Materials Chemistry A*, 12(17), 10124–10136.

[29] Snaith, H. J. (2018). Present status and future prospects of perovskite photovoltaics. *Nature Materials*, 17(5), 372–376.

[30] Lanubile, A., & Bosoni, P. (2024). Domain knowledge-guided machine learning framework for state of health estimation in Lithium-ion batteries. *Communications Engineering*, 3(1), 168.

A Appendix / supplemental material

Optionally include supplemental material (complete proofs, additional experiments and plots) in appendix. All such materials **SHOULD be included in the main submission.**

AI Co-Scientist Challenge Korea Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [N/A] .
- [N/A] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[N/A] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "AI Co-Scientist Challenge Korea paper checklist".**
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The physics-informed feature disentanglement methodology proposed in the abstract is implemented and validated in the Results section. The claim that the framework maintains $R^2 > 0.9$ is confirmed across all experimental results, as demonstrated in Figures 2–6. The contributions stated in the abstract and introduction accurately reflect the scope and findings presented throughout the paper.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: A dedicated Limitations subsection is included in the Discussion section (Section 4.1), which addresses: (1) the non-universal applicability of dQ/dV feature extraction to all parameters, (2) the simulation-based nature of the study and need for experimental validation, (3) limitations of linear regression models in capturing non-linear relationships, and (4) the need to validate generalizability to other battery chemistries.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A]

   Justification: This paper is primarily an empirical study that does not include theoretical theorems or mathematical proofs. The work focuses on experimental validation of a physics-informed feature extraction framework using PyBaMM simulations and machine learning models, without presenting formal theoretical results that would require proofs.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.

- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The Methodology section (Section 3) provides comprehensive details necessary for reproducibility, including: (1) experimental design with parameter combination generation and random sampling using Latin Hypercube Sampling (LHS), (2) data splitting protocol (60/20/20 train/validation/test), (3) PyBaMM simulation setup using the Prada2013 parameter set for LFP batteries with the Doyle-Fuller-Newman (DFN) model, (4) feature extraction methods with detailed statistical measures, (5) model training procedures using linear regression with OLS, and (6) evaluation metrics (R^2, MAE, RMSE, MAPE). All experimental protocols and hyperparameters are clearly specified.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While AI Co-Scientist Challenge Korea does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: While the code and data are not currently publicly available, the Methodology section (Section 3) provides comprehensive details necessary for reproduction, including experimental design, PyBaMM simulation setup with Prada2013 parameter set, feature extraction methods, model training procedures, and evaluation protocols. All key information needed to understand and replicate the experimental approach is disclosed in the paper.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The Methodology section (Section 3) comprehensively specifies all experimental details, including: data splitting protocol (60/20/20 train/validation/test), model type (linear regression with OLS), feature standardization method (StandardScaler), evaluation metrics (R^2, MAE, RMSE, MAPE), and the complete experimental protocol from parameter combination generation to model evaluation. All necessary details to understand and interpret the results are provided.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Error bars are reported in Figure 6(a), which illustrates the variance in predictive fidelity across five physical parameters using mean R^2 scores and error bars. The error bars demonstrate the variability in prediction difficulty across different target variables, with positive porosity showing a relatively wider error bar indicating greater sensitivity to multi-variable interference.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Compute resource information is provided in Section 3.4 of the Methodology section, including: hardware specifications (Intel Core i5-12600K processor, NVIDIA GeForce GTX 1060 6GB, Windows 11), computation type (CPU-based for PyBaMM simulations), execution time per simulation (2-5 minutes per parameter set), and total computation time estimates (200-500 hours for complete simulation phase). The paper also discloses that all computational resources were used for the reported experiments without requiring additional preliminary or failed experimental runs.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://nips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This research is a simulation-based battery study that does not involve human subjects, sensitive data collection, or any activities that would conflict with the NeurIPS Code of Ethics. The work uses publicly available simulation frameworks (PyBaMM) and focuses on computational modeling of battery materials, which aligns with ethical research practices in materials science and machine learning.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [N/A]

Justification: This paper presents foundational research on battery material design optimization using physics-informed machine learning. The work is not tied to specific applications or deployments, and does not have direct societal impacts that would require discussion of positive or negative consequences. The research focuses on computational methods for accelerating battery cathode design, which is a fundamental materials science contribution without immediate societal implications.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: This research uses battery simulation data generated from PyBaMM, a publicly available open-source framework. The work does not involve pretrained language models, image generators, scraped datasets, or any other assets that pose a high risk for misuse. The models presented are simple linear regression models for battery parameter prediction, which do not require safeguards for responsible release.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All existing assets used in this work are properly cited in the References section. PyBaMM framework is cited as Sulzer et al. (2021) with BSD-3-Clause license. scikit-learn is cited as Pedregosa et al. (2011) with BSD license. The Prada2013 parameter set is cited as Prada et al. (2013). Latin Hypercube Sampling method is cited as McKay et al. (1979). All citations follow standard academic practices and respect the original creators' contributions.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [N/A]

Justification: This paper does not release new datasets, models, or code as assets. While the study uses PyBaMM simulation-generated data, this data is not publicly released. The research focuses on demonstrating the methodology and framework rather than providing new reusable assets to the community.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: This research is a battery simulation study using PyBaMM framework and does not involve crowdsourcing or research with human subjects. All experiments are computational simulations of battery discharge behavior, with no human participants or crowdsourced data collection.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: This research is a battery simulation study that does not involve human subjects. All experiments are computational simulations using PyBaMM framework, with no human participants, and therefore IRB approval is not required.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.