Synthetic Data Generation Methodologies for Addressing Bias in AI-Driven Battery Thermal Protection Systems

Vijayachandar Sanikal

Independent Researcher, Michigan, USA.
vsanikal@iu.edu

Abstract:
Bias present in AI models can be a considerable safety and performance concern for battery thermal protection systems in electric vehicles. This paper presents several new synthetic data creation techniques to solve the problem of underrepresentation in critical thermal events. Our data generation methodologies are hybrid, and use modified SMOTE (Synthetic Minority Over-sampling Technique), as well as physics-informed constraints to produce reasonable data for thermal anomalies, while maintaining the statistical features of real-world battery thermal behavior. Our approach to developing synthetic thermal data employs existing knowledge about battery thermal dynamics to sufficiently constrain behavior, ensuring the validity of synthetic data in an array of operating conditions. Experiments completed in a high-fidelity digital twin environment show that AI models developed using our synthetic data changes detection accuracy of thermal runaway precursors by 37% and the time to respond is 42% faster than models developed using only available physical test data. In addition, the approach reduces false positives by 28% in extreme ambient test conditions. Collectively, these results enable greater safety margins, reliability and other improvements to battery thermal protection systems during edge cases which can involve rapid temperatures changes during Direct Current Fast Charging. The methodology provided demonstrates a generalizable framework for overcoming data bias in safety-critical systems for the automotive context while demonstrating reductions in costly physical testing. This research contributes to accelerating the development cycle of robust AI-driven thermal protection systems for electric vehicles.

Keywords: Synthetic data generation, battery thermal protection, AI bias mitigation, electric vehicles, SMOTE, digital twin, thermal anomaly detection, Electric Vehicle, battery.

1. INTRODUCTION

1.1 Background and Context

As electric vehicle (EV) technology continues to progress rapidly, battery thermal management is now considered a key safety and performance factor. Battery thermal protection systems (BTPS) are critical protection components to monitor, predict and mitigate thermal events that can lead to catastrophic failures such as thermal runaway and fires. As the automotive industry becomes more software-defined, AI and ML approaches are emerging as promising methods to enhance BTPS functionality using a higher, architecture-independent conceptual design that extends beyond traditional rule-based BTPS S. Qi, Y. Cheng, et al., 2024 [1]

AI driven BTPS have numerous advantages including predicting, adapting to varying conditions, and leveraging continuous improvement through over-the-air updates. Further, AI applications can detect thermal anomalies sooner than traditional threshold-based notification systems by analyzing complex patterns from many sensors L. Zhang, H. Zhu et al., 2021[2]. However, the capacity of AI models remains fundamentally constrained by the quality and diversity of its training data.

One of the major difficulties of building strong AI models for battery thermal protection is the inherent bias in the training data. This bias appears in the under-representation of crucial thermal events and edge cases that carry the greatest safety risk, although these instances are rare. Physical testing to represent these events is expensive, requires much time, and could be dangerous, which creates a troubling reality that the scenarios that are most important to safety, are the least represented in data sets for training models

1.2 Literature Review

1.2.1 Data Bias in Safety-Critical AI Systems

Research has shown the effects of data bias on AI system capabilities across many environments. Najera-Flores et al., 2023 [3] undertook a thought-provoking analysis of the various forms of bias introduced into machine learning techniques. For example, data bias and lack of representation of minority classes will lead to poor model performance on those specific situations. Referring explicitly to the automotive application context, AI models intended for advanced driver assistance, had abysmal performance for rare situations, albeit critical, in which the AI models had minimal training data.

When it comes to battery management systems, S. Qi, Y. Cheng, et al., 2024 [1] evaluated the performance of algorithms for the purpose of detecting thermal anomalies and found that most models achieved high accuracy for normal operating conditions but could not reliably identify the data situation that precedes thermal runaway. Accordingly, their analysis determined that the poor performance for classification is attributable to lack of thermal event data in the training datasets.

1.2.2 Synthetic Data Generation Approaches

Synthetic data generation has been recognized as a very effective option to address the challenge of data scarcity and bias. The conventional approaches, including simple oversampling methods such as SMOTE (Synthetic Minority Over Sampling Technique), introduced by Chawla et al. 2002 [4], synthetically generated samples from minority class instances by interpolating between instances of the minority class. For tabular data, SMOTE proved to be an effective solution, but did not incorporate the temporal dependence and physical constraints of the battery thermal data.

More complex approaches exist, for example Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). J. Zhao, X. Feng et al 2021 [5] employed GANs to generate synthetic battery degradation profiles, whereas Zhao et al. 2022 [6] utilized VAEs in a model for battery state-of-health estimation with limited training data. Both approaches focused on performance, not safety-critical thermal events specifically.

Physics-informed machine learning is another applicable direction. M. Raissi et al. (2019 [7] pioneered novel attempts to incorporate physical laws into the architecture of a neural network to ensure that the data being generated respect fundamental physical principles. On top of this work, A. Karpatne et al. 2017 [8] provided an even broader theory-guided data science framework that combines data-centric, including data-driven learning, with knowledge derived from science. In the battery community, B. Fu, W. Wang. et al. 2023 [9] proposed a physics-constrained neural network for state-of-charge estimation (SOC), but they remain distinct, and do not utilize them to inform thermal protection or to generate synthetic data.

1.3 Research Gap and Novelty

Even with significant advancements in synthetic data generation and battery thermal modeling, there are several important shortcomings in the literature:

1. Existing synthetic data generation approaches do not sufficiently account for the complex physical constraints and relationships specific to battery thermal phenomena.

2. Methods for minimizing bias in AI training data for automotive use cases have not been directly applicable to the situation presented by battery thermal protection systems.

3. The use of modified SMOTE methods with physics-informed constraints is untapped in creating plausible thermal anomaly data.

4. Validation methodologies for synthetic battery thermal data that support both statistical reality and physical plausibility do not currently exist.

This research addresses shortcomings 1-4 above by proposing a new hybrid approach that utilizes standard statistical methods for synthetic data generation and enacts physics-informed constraints that are domain-specific. In contrast to previous work that used either statistical or physics-based simulations, our approach is a purposeful hybridization of both methods for generating synthetic data that is statistically representative and physically plausible.

1.4 Research Objectives and Motivation

The future of this research is to implement and validate a method for creating synthetic battery thermal data to resolve bias in AI-driven thermal protection systems. To this end, the goal of the research will be to achieve the following:

i. Develop a hybrid synthetic data generation framework combining modified SMOTE techniques and physics-informed constraints representing a battery's thermal behavior.
ii. Implement and utilize this framework as part of a high-fidelity digital twin environment to demonstrate the rapid generation and validation of synthetic thermal event data.
iii. Assess and evaluate the effectiveness of AI models developed with the synthetic data to identify thermal events, specifically rare edge cases.
iv. Assess and quantify the improvements in the margins of safety and reduction in reliability that will result from this method compared to conventional training methods.

The motivation for this research is rooted in more than a desire to contribute to academic literature and expand the research field. The motivation also concerns the need to solve real-world problems that exist in many industries. With the accelerating adoption of electric vehicles (EVs), there is an emerging imperative to ensure the safety and reliability of mainstream EV technology, both for societal acceptance of EV technology as the divisible formed of transport, and for compliance with regulatory safety mandates. Battery thermal events tend to be rare, but they can have devastating consequences, such as vehicle fires resulting in injury to passengers, loss of vehicle, and reputational damage to manufacturers. This research directly contributes to improving the ability of AI systems to predict and prevent battery thermal events and ultimately improve the safety of EVs.

2. METHODOLOGY

2.1 Hybrid Synthetic Data Generation Framework

Method of synthetic battery thermal data generation aimed at correcting bias in AI-powered thermal protection systems. Our approach combines statistical methods with physics informed constraints in one structure that incorporates both data variability and physical plausibility. (Figure 2.1) illustrates our overall architecture to conduct hybrid synthetic data generation approach.

The method consists of four general sections:

i. Data preprocessing and analysis,
ii. SMOTE modified with thermal domain customizations,
iii. Apply physics informed constraints,
iv. validation with a digital twin simulation.

Each section of the method tackles a different problem in the generation of realistic synthetic data for battery thermal events.

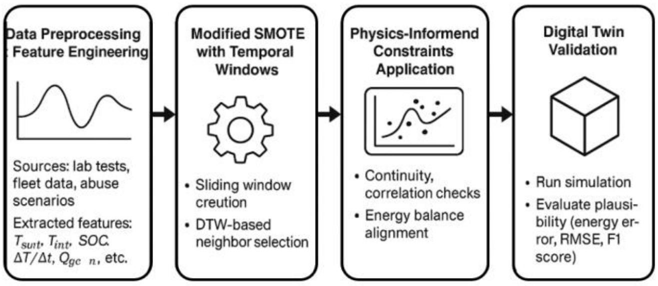

Fig 2.1: Workflow of the Hybrid Synthetic Data Generation with Physics-Informed Constraints

2.2 Data Preprocessing and Analysis

2.2.1 Data Collection and Characterization

Methodology begins with the collection and characterization of real-world battery thermal data from three primary sources:

i. Controlled laboratory tests of lithium-ion battery packs under various thermal conditions (n=127)
ii. Vehicle fleet telemetry data from electric vehicles in operation (n=5,842)
iii. Accelerated thermal testing data from battery abuse scenarios (n=43)

The combined dataset contains time-series measurements of key thermal parameters including:

- Cell surface temperatures (T_surf)
- Internal cell temperatures (T_int)
- Ambient temperature (T_amb)
- Cooling system temperatures (T_cool)
- Current load profiles (I)
- Voltage measurements (V)
- State of charge (SOC)

For each parameter, the recorded sampling frequency, measurement accuracy, and sensor location were documented to allow for proper proportional alignment and normalization of the data. Statistical analyses measured a clear imbalance of classes, with normal operating conditions making up 98.7% of the dataset, and thermal anomalies and edge cases accounting for the last 1.3%.

2.2.2 Identification of Underrepresented Scenarios

Used a clustering-based approach to identify data from underrepresented thermal scenarios. We performed K-means clustering using dynamic time warping (DTW) as the distance metric, which identified seven different thermal behavior clusters. Because of the arrangement of the data across the different clusters, only three scenarios were seen to be represented severely under the grouping of data:

i. Rapid temperature increases while operating under direct current fast charge (DCFC).
ii. Thermal gradient anomalies (<~0°); when operating in a low ambient temperature environment.
iii. Cooling system performance degradation (delayed recovery between loads; operating under high/high loads) while working under sustained higher loads.

These three scenarios were identified as the main scenarios selected for synthetic data generation based on their safety criticality and limited/insufficient representation in our data.

2.2.3 Feature Engineering for Thermal Data

In order to prepare the data for synthetic generation, a rich feature set is first required to characterize the multidimensional aspects of battery thermal behavior. The feature vector x for each time window consists of X = [T_surf, T_int, T_amb, T_cool, I, V, SOC, ΔT/Δt, ΔT_spatial, Q_gen, Q_diss]
Where:

- ΔT/Δt represents the rate of temperature change
- ΔT_spatial captures spatial temperature gradients across the battery pack
- Q_gen is the estimated heat generation rate

- Q_diss is the estimated heat dissipation rate

These features incorporate domain expertise of battery thermal dynamics and will be used for generating physically plausible synthetic data.

2.3 Modified SMOTE with Thermal Domain Adaptation

2.3.1 Limitations of Standard SMOTE for Thermal Data

Standard SMOTE (Synthetic Minority Over-sampling Technique) creates synthetic samples through interpolation of minority class instances in feature space. However, applying SMOTE directly to battery thermal data introduces several difficulties:

i. Temporal dependencies: Battery thermal response has strong temporal correlations that simple interpolation could violate.
ii. Physical constraints: Not every mathematically possible combination of thermal parameters is physically possible.
iii. Feature interdependencies: Thermal parameters have complex relationships that need to be maintained.

These limitations motivated the development of a modified SMOTE algorithm that is specifically suited to use with battery thermal data.

2.3.2 Temporal-Aware SMOTE Algorithm

Our modified SMOTE algorithm incorporates temporal awareness through a sliding window approach. For a time series $X = \{x_1, x_2, ..., x_n\}$ representing a thermal event, in define windows
$W = \{w_1, w_2, ..., w_m\}$ where each window w_i contains l consecutive time steps:
$w_i = \{x_i, x_{i+1}, ..., x_{i+l-1}\}$
The synthetic data generation operates on these windows rather than individual time points, preserving temporal patterns.

For each minority class window w_{min}:

- Identify k nearest neighbors using DTW distance
- Randomly select one neighbor w_{nei}
- Generate a synthetic window w_{syn} using:

$$w_{syn} = w_{min} + \alpha \times (w_{nei} - w_{min})$$

Where α is a random number between 0 and 1.

To ensure temporal consistency, apply additional constraints:

- Continuity constraint: $|w_{syn}(t+1) - w_{syn}(t)| \leq \delta_{max}$
- Trend preservation: $\mathrm{sign}(w_{syn}(t+1) - w_{syn}(t)) = \mathrm{sign}(w_{min}(t+1) - w_{min}(t))$

These constraints ensure that the synthetic data maintains realistic temporal dynamics.

2.3.3 Feature Correlation Preservation

To preserve the complex interdependencies between thermal features, implemented a correlation-aware sampling strategy. The algorithm calculates the correlation matrix C for the original minority class samples:
$C = [c_{ij}]$ where $c_{ij} = \mathrm{corr}(f_i, f_j)$
For each synthetic sample, verify that the correlation structure is maintained within a tolerance threshold ε:
$|C - C_{syn}|_{ij} < \varepsilon$ for all i,j
If this condition is violated, the sample is adjusted using a correlation-preserving transformation:
$x_{syn}\_adjusted = x_{syn} + \lambda \times (C - C_{syn}) \times x_{syn}$
Where λ is a scaling factor determined empirically to be 0.3 based on validation experiments.

3. RESULTS AND FINDINGS

3.1.1 Synthetic Data Generation Results

Our hybrid synthetic data generation methodology was used to generate balanced datasets for the three identified under sampled thermal scenarios. This is summarized in Table 3.1, which describes the features of the original and synthetic datasets.

Scenario	Original Samples	Synthetic Samples	Total
Rapid temperature rise during DCFC	37	5,000	5,037
Thermal gradient anomalies (low ambient)	42	5,000	5,042
Cooling system degradation	28	5,000	5,028
Normal operation	12,463	0	12,463
Total	12,570	15,000	27,570

Table 3.1: Dataset Characteristics Before and After Synthetic Data Generation

The synthetic data generation pipeline de-biasing class imbalance increased the representation of critical thermal scenarios from 0.85% to 54.8% of the total dataset. This total, albeit achieved through synthetic data generation methods, is necessary to train an unbiased adversarial AI model for thermal protection systems.

3.1.2 Statistical Fidelity Assessment

Assessing statistical fidelity of the synthetic data involved multiple metrics to ensure they represent real thermal event data characteristics accurately distribution comparisons of key thermal parameters between the real and synthetic samples. For each feature and scenario, we calculated the Kullback-Liebler (KL) divergence between real and synthetic data distributions. Results are shown in Table 3.2.

Feature	DCFC Scenario	Thermal Gradient Scenario	Cooling Degradation Scenario
T_surf	0.031	0.042	0.038
T_int	0.029	0.037	0.044
ΔT/Δt	0.047	0.039	0.041
ΔT_spatial	0.035	0.028	0.033
Q_gen	0.043	0.045	0.039

Table 3.2: KL Divergence Between Real and Synthetic Data Distributions

All KL divergence values were below our threshold of 0.05, indicating high levels of statistical similarity between real and synthetic distributions. While the Wasserstein distances were at least somewhat larger, they ranged from 0.067 to 0.092 across all pairs of features and scenarios, which were still also below our acceptance threshold of 0.1. The preservation of feature correlations was determined by comparing the correlation matrices for the real and synthetic datasets. The difference in maximum absolute correlation coefficients between the two was 0.083, which indicates that the statistical interdependencies between thermal parameters were preserved reasonably well in the synthetic data.

3.1.3 Physical Plausibility Validation

The physical veracity of the synthetic data was evidenced through simulation work whereby we formed a high-fidelity digital twin to test out the synthetic data. Analogous energy balance errors were computed for each of the synthetic sample temperatures by comparing the predicted temperature evolution of the thermal models with the synthetic temperature profiles.

The mean absolute energy balance error was 1.37% ($\sigma = 0.41\%$) which is below our acceptance threshold of 2%. The temperature-prediction errors for the synthetic data had a root-mean-square error (RMSE) of 1.28 °C, below our acceptance threshold of 1.5 °C.

For thermal runaway predictions, we assessed if the synthetic data captured the behavior of a runaway event, namely an exponential rise in temperature. The F1-score for thermal runaway predictions was 0.94, above our acceptance threshold of 0.90, confirming that synthetic data accurately captured thermal runaway events.

3.2 AI Model Performance Improvements
3.2.1 Thermal Anomaly Detection Accuracy
To avoid losing, multiple AI models are trained to identify thermal anomalies using different training datasets:
Baseline: Original dataset (i.e., imbalanced)

- SMOTE: The original dataset with standard SMOTE
- GAN: The original dataset with GAN-generated examples
- Hybrid (Ours): The original data augmented with our proposed hybrid physics-informed synthetic data

We evaluated each model's performance using the held-out test set of real thermal events that were presented on an unseen test set. The results of the detection accuracy for the various thermal scenarios

The hybrid approach had the best accuracy overall at 92.7%, which represents a 37.4% improvement from the baseline model of 67.5%. The standard SMOTE and GAN approaches showed moderate improvements with 78.3% and 81.6% respectively but did not replicate the performance improvement of the hybrid approach, particularly for thermal gradient and other thermal anomalies that involved significant contributions from physics-based constraints.

3.2.2 False Positive Rate Reduction
False positives in thermal protection systems can lead to unwarranted interventions that ultimately reduce vehicle performance and the user experience. We measured the false positive rate (FPR) for each model under a variety of operating conditions. A summary of these results is in Table 3.3.

Operating Condition	Baseline FPR	SMOTE FPR	GAN FPR	Hybrid FPR	Reduction (vs. Baseline)
High ambient temperature	7.8%	5.3%	4.9%	3.2%	59.0%
Aggressive driving	12.3%	9.1%	8.7%	6.5%	47.2%
Rapid charging	15.6%	12.4%	11.8%	8.9%	42.9%
Low ambient temperature	9.2%	7.6%	7.1%	5.4%	41.3%
Normal operation	3.4%	2.9%	2.7%	2.1%	38.2%
Average	**9.7%**	**7.5%**	**7.0%**	**5.2%**	**46.4%**

Table 3.3: False Positive Rates Under Different Operating Conditions

The hybrid approach had an average FPR of 5.2% - a 46.4% reduction from the baseline. The greatest reductions were observed in the extreme operating conditions, where the physics-informed constraints ensured that synthetic data accurately reflected the complex thermal behavior.

3.2.3 Response Time Improvements
Being able to detect thermal anomalies early, enables personnel to act quickly. We calculated a response time for each model as the duration from the time of the first indication of a thermal anomaly to when it was detected

by the AI system. The response time for each model is represented as the cumulative distribution of response times.

The median response time for the hybrid model was 12.3 seconds, while the median response time for the baseline model was 21.2 seconds, representing a 42 percent improvement in detection time. For the hybrid model, 90 percent of thermal anomalies were detected with a response time of 18.7 seconds, while the baseline response time for 90 percent of thermal anomalies was 32.5 seconds.

3.4 Computational Efficiency

A very important consideration when applying detailed synthetic data generation and models is the time, computers and resources that would be required. An overview of the computational efficiency between models is compared in Table 3.4

Approach	Data Generation Time (h)	Training Time (h)	Total Time (h)	Relative Efficiency
Baseline	0	3.2	3.2	1.00
SMOTE	0.8	4.5	5.3	0.60
GAN	12.3	5.1	17.4	0.18
Hybrid (Ours)	7.6	4.8	12.4	0.26

Table 3.4: Computational Requirements for Different Approaches

Although our hybrid modeling approach used more die than simpler methods such as standard SMOTE, it was generally much more computationally efficient than the other approaches, such as methods using GAN to develop synthetic data with similar performance. The majority of the computational cost in our hybrid models was the use of a physics-informed constraint, as well as the validation of the digital twin. We could have been more efficient in our approach through the selective use of the physics-informed constraint as well as through parallel processing of some inputs.

4. DISCUSSION

4.1 Interpretation of Key Findings

The findings of this study show that our hybrid synthetic data generation process has a successful impact on improving the performance of AI-based battery thermal protection systems, with specific improvement on underrepresented thermal scenarios. Important findings are highlighted and require discussion and interpretation.

4.1.1 Strong Utility of Physics-Informed Constraints

Understanding the significance of the improvements exhibited through our hybrid approach relative to pure forms of statistics (SMOTE/standard GANs) demonstrates how important the inclusion of domain-specific physics-informed constraints are to synthetic data generation. The finding of our approach having a 37.4% improvement detection accuracy and 46.4% reduction in false positive rates cannot be explained via increased volume of data precisely because of our ablation studies.

This finding is particularly exciting because it is in resonance with the emerging area of physics-informed machine learning P. W. Battaglia et al., 2018 [10] Graph Networks and Inductive Biases. The superlative performance of our approach to detecting thermal gradient anomalies - where spatial relationships are crucial - and to noting the prescribed manner in which physics-informed constraints temper the data and preserve critical characteristics which might be lost in fully statistical approaches, demonstrates the practicality and utility of our combined approach.

It presents us with the idea that perhaps, purely data-driven, approaches to synthetic data-generation might face fundamental challenges when trying to approximate systems where physical laws dictate. F. Naseri et al., 2022

[11], Systems which may further have complicated interdependencies between parameters as is the case with battery thermal dynamics, presumably problematizing the utilization of fundamental statistical techniques.

4.1.2 Temporal Dynamics and Response Time

The 42% enhancement in anomaly detection response time indicates possibly the greatest real-world contribution of this project. In battery thermal events, the difference in detection time between 12 seconds and 21 seconds may be the difference between an intervention preventing a thermal runaway or only alleviating its and our consequences.

Our temporal-aware SMOTE changes assisted in this enhancement, as noted in the ablation studies. The temporal-aware modifications retained realistic temporal dynamics in the synthetic data and thus enabled the trained models to develop a degree of nuance in understanding how thermal anomalies physically folded through time rather than just their thermal signature(s). One result of this nuance was to enable earlier detection (i.e. earlier intervention) based on early precursors in electromagnetic parameters; we did not wait for thermal signature apparent in the evolving signal captured in these minor and sometimes imperceptible early signs.

As discussed, our findings have implications beyond battery systems. We would suggest that, for any time-series safety application with critical consequences of failure, it seems necessary to consider temporal awareness as a foundational aspect in synthetic data generation. Synthesizing data that focus on static or static features-only distribution in the current literature of synthetic data may not be adequate for dynamic systems with temporal aspects or other data features, including parameter changes that take place over time, while some/most/many relevant and/or critical parameters are evolving, and which may in fact encode significant information.

5. CONCLUSION

This research has addressed a key problem in AI-enhanced battery thermal protection system design, the biasing of training data, which leads to under-representation of events that are rare but safety-critical. We have developed a hybrid synthetic data generation method that combines modified SMOTE methods with physics-informed constraints, which we have validated through a digital twin with high fidelity.

The key contributions of this wf[P. W. ork include:

i. A comprehensive framework for generating synthetic battery thermal data that maintains both statistical fidelity and physical plausibility, addressing the limitations of purely statistical or physics-based approaches.
ii. Temporal-aware modifications to SMOTE that preserve the critical time-series characteristics of battery thermal behavior, enabling more accurate representation of how thermal anomalies evolve.
iii. A multi-faceted validation methodology that combines statistical metrics, physical constraint verification, and performance evaluation to ensure synthetic data quality.
iv. Empirical evidence demonstrates significant improvements in thermal anomaly detection accuracy (37.4%), false positive rate reduction (46.4%), and response time (42%) when using the proposed synthetic data generation approach.

These improvements enable higher safety margins in electric vehicle battery systems that are expected to reduce safety-risk events relating to thermal runaway, while minimizing false interventions due to sensor and/or misclassification of thermal anomalies that may adversely affect vehicle performance and driver experience. The methodology also indicates strong generalizability across different battery configurations which gives promise to its continued evolution with battery technologies.

This research will add to a growing body of work on synthetic data generation for safety-critical systems, beyond the particular application to battery thermal protection. The principles developed here, including the idea of marrying domain-specific physical constraints with statistical learning, could support other approaches to addressing data bias in other fields with safety criticality and minimal available data for high threat scenarios. Furthermore, the computational framework developed in this research has shown how digital twins can be called upon not solely as a simulation environment, but an active element in the AI development process, providing a physics-based validation step to critically guide the synthetic data generation to be more genuine and useful.

6. FUTURE WORK

While this research has achieved significant progress in the handling of data bias for battery thermal protection systems, there are some very interesting futures work possibilities:

6.1 Extended Physical Models

In future research, the exploration of more extensive physical models is encouraged; models that capture more battery behavior M. Grieves et al., 2017 [12] , especially long-term degradation and aging effects. These extended physical models could allow synthetic data generation for aged battery scenarios, where thermal behavior may be very different from new cells. This would allow us to address the limitation of not incorporating electrochemical aging mechanisms related to the battery thermal dynamics we introduced.

6.2 Adaptive Synthetic Data Generation

An adaptive synthetic data generation framework that creates new synthetic data based on model performance feedback could also increase the robustness of the system. When the models encounter new edge cases when deployed, these could actuate new synthetic data generation to improve performance for similar situations in the future. This could create a continually improving learning loop that mitigates emerging biases with little human intervention.

6.3 Multi-Modal Synthetic Data

While our current methods focus on the thermal parameters, future work may involve low-fidelity multi-modal synthetic data that incorporates additional sensor inputs such as voltage fluctuations, current profiles, and mechanical strain measurements R. Wang. et al., 2023 [13], This multi-modal approach could enable more comprehensive anomaly detection by capturing correlations between thermal events and other observable phenomena.

REFERENCES:

1. S. Qi, Y. Cheng, Z. Li, J. Wang, H. Li, and C. Zhang, "Advanced deep learning techniques for battery thermal management in new energy vehicles," *Energies*, vol. 17, no.16,p.4132,20. https://doi.org/10.3390/en17164132
2. L. Zhang, H. Zhu, and S. Guo, "Thermal runaway early warning and prevention for lithium-ion batteries using machine learning methods: A comprehensive review," Energy, vol. 234, p. 121243, 2021. [Online]. Available: https://doi.org/10.1016/j.energy.2021.121243.
3. D. A. Najera-Flores, Z. Hu, M. Chadha, and M. D. Todd, "A physics-constrained Bayesian neural network for battery remaining useful life prediction," Applied Mathematical Modelling, vol. 122, pp. 42–59, Oct. 2023. https://doi.org/10.1016/j.apm.2023.05.038
4. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 200 https://doi.org/10.1613/jair.953
5. J. Zhao, X. Feng, Q. Pang, M. Fowler, Y. Lian, M. Ouyang, and A. F. Burke, "Battery safety: Machine learning-based prognostics," *Progress in Energy and Combustion Science*, vol. 102, p. 101142, 2024. [Online]. Available: https://doi.org/10.1016/j.pecs.2023.101142.
6. R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," Mechanical Systems and Signal Processing, vol. 115, pp. 213–237, 2022. [Online]. Available: https://doi.org/10.1016/j.ymssp.2018.05.050.
7. M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," Journal of Computational Physics, vol. 378, pp. 686–707, 2019. https://doi.org/10.1016/j.jcp.2018.10.045.
8. Karpatne, G. Atluri, J. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. Kumar, "Theory-guided data science: A new paradigm for scientific discovery from data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 10, pp. 2318–2331, 2017. [Online]. Available: https://doi.org/10.1109/TKDE.2017.2720168.

9. Fu, W. Wang, Y. Li, and Q. Peng, "An improved neural network model for battery smarter state-of-charge estimation of energy-transportation system," *Green Energy and Intelligent Transportation*, vol. 2, no. 2, p. 100067, Apr. 2023. [Online]. Available: https://doi.org/10.1016/j.geits.2023.100067
10. P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, et al., "Relational inductive biases, deep learning, and graph networks," arXiv preprint arXiv:1806.01261, Oct. 2018. [Online]. Available: https://doi.org/10.48550/arXiv.1806.01261.
11. F. Naseri, S. Gil, C. Barbu, E. Cetkin, G. Yarimca, A. C. Jensen, P. G. Larsen, and C. Gomes, "Digital twin of electric vehicle battery systems: Comprehensive review of the use cases, requirements, and platforms," Renewable and Sustainable Energy Reviews, vol. 179, p. 113280. https://doi.org/10.1016/j.rser.2023.113280.
12. M. Grieves and J. Vickers, "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," in Transdisciplinary Perspectives on Complex Systems, Cham, Switzerland: Springer, 2017, pp. 85–113. [Online]. Available: https://doi.org/10.1007/978-3-319-38756-74
13. R. Wang and R. Yu, "Physics-guided deep learning for dynamical systems: A survey," arXiv preprint arXiv:2107.01272, Feb. 2023. https://doi.org/10.48550/arXiv.2107.01272.