State of Charge Imbalance Classification of Lithium-ion Battery Strings using Pulse-Injection-Aided Machine Learning

Alan Gen Li
Department of Electrical Engineering
Columbia University
New York, NY, USA
agl2142@columbia.edu

Matthias Preindl
Department of Electrical Engineering
Columbia University
New York, NY, USA
matthias.preindl@columbia.edu

***Abstract*—Lithium-ion battery strings are important modules in battery packs. Due to cell variation, strings may have imbalanced state of charge levels, reducing pack capacity and exacerbating degradation. While much research has been devoted to individual cells, string diagnostics using pulse-injection-aided machine learning can reduce sensing requirements and simplify computations. Experimental voltage response data from pulse perturbation of battery cells is used to generate virtual cell strings and 'design' the state of charge imbalance within the string. A feedforward neural network is trained on thousands of unique virtual string voltages and can distinguish between the balanced and imbalanced strings with up to 95% accuracy. Verification is performed using different string configurations and state of charge levels. The proposed technique has high promise and could be used to localize or regress the degree of imbalance.**

***Index Terms*—Lithium batteries, Neural networks, String estimation, State estimation**

I. Introduction

Lithium-ion battery (LIB) packs are typically composed of hundred of cells [1]. For proper functioning, the battery management system (BMS) must monitor each individual cell. Typically the voltage, current, and temperature are measured to yield information on the battery states, such as state of charge (SoC), state of health, and state of power. Battery models, such as the DNRC model in [2], often provide the basis of state estimation. Cell states inform the BMS how to operate the cells in the pack, and thus directly affect the performance of the battery. In an electric vehicle, this can affect the maximum driving range, the remaining driving range, and the maximum acceleration [3].

Managing cell imbalance is a key function of the BMS [4]–[6]. Even if each cell within a pack is cycled identically from the same initial states, the SoC of each cell will deviate due to internal electrochemical variation and external conditions such as unequal temperatures. If not managed properly, this imbalance can accelerate degradation and pack life. Decentralized cell-level BMS have been proposed to address this issue [7], [8], which has led to the concept of 'smart cells' [9]. Though smart cells have high promise, sensing and managing the voltage, current, and temperature of each cell requires costly amounts of power electronics, especially in large battery packs. Module sensing may offer a solution [10]. Modules refer to groups of cells connected in series or parallel. If the states of each cell in a module can be obtained from the modular voltage or current, then the number of sensors can be greatly reduced.

Of particular interest is the cell string [11], represented in Figure 1. In a string, each cell is charged and discharged at the same current and at the same time (neglecting high frequency behavior), but the cell voltages and SoC can have significant variation. This is an issue because the string's performance is limited by the weakest cell. If any one cell is empty, so is the string, even if the other cells have remaining charge. Since the string voltage response is the sum of all the cells' responses, the aim of modular sensing is to disaggregate the superposition – to obtain information about the individual cells from only the superimposed string voltage.

String sensing is non-trivial, and poses a difficult problem using conventional control theory techniques. Equivalent circuits are useful for single cells but multicell systems pose identifiability challenges [12]. In [13], switching phenomena in the cell balancing circuits are used to achieve observability. As switching occurs, the terminal voltage can yield information on individual states. In [14], nonlinearities in the OCV-SoC curve demonstrate observability in imbalanced cells. Kalman filters, often used for individual cell SoC estimation, can be applied to cell strings [15], but only for the extrema of the SoC and after balancing has been performed. In these studies, observability relies on certain conditions that may not be met at all times. Deviation matrices are shown to qualitatively identify high-impedance cells, but faces challenges for imbalanced strings [16]. A general-purpose string SoC estimation method using SoC differences demonstrated promising results but with high computational complexity [17]. Choosing a 'representative cell' is proposed to reduce computational burden but requires a carefully-designed selection process [18].

We propose to use pulse-injection-aided machine learning (PIAML) for SoC imbalance classification in a cell string. We

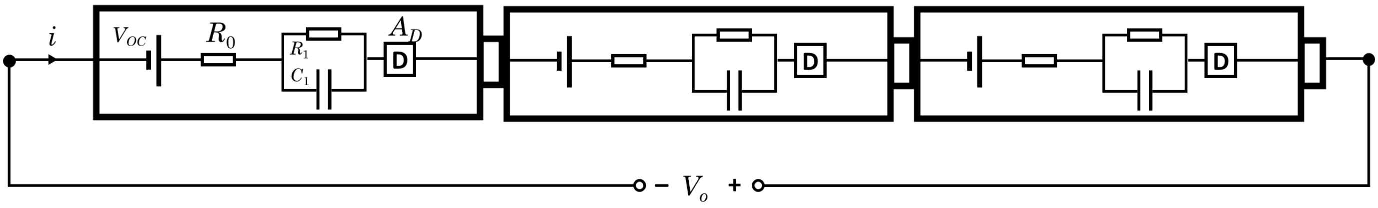
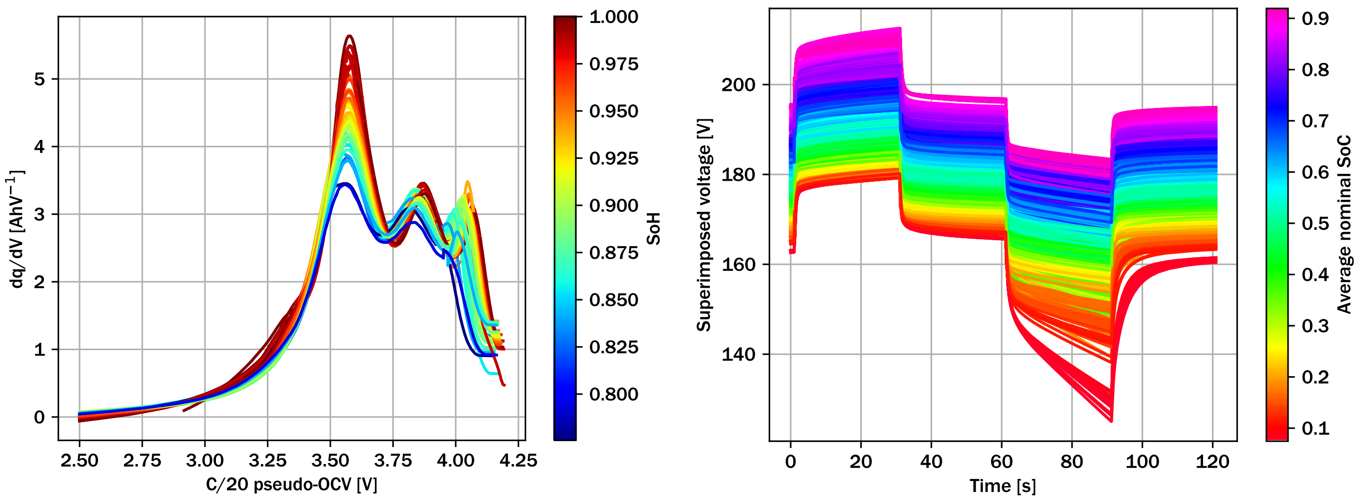


Fig. 1. Diagram of a 3-cell LIB string showing the DNRC equivalent circuit model [2]



(a)

(b)

Fig. 2. Experimental and simulated data, showing (a) Incremental capacity curves derived from pseudo-OCV discharge data collected from LIB cells at various levels of degradation, and (b) Samples of string simulation with $N = 48$ cells, colored by average nominal SoC

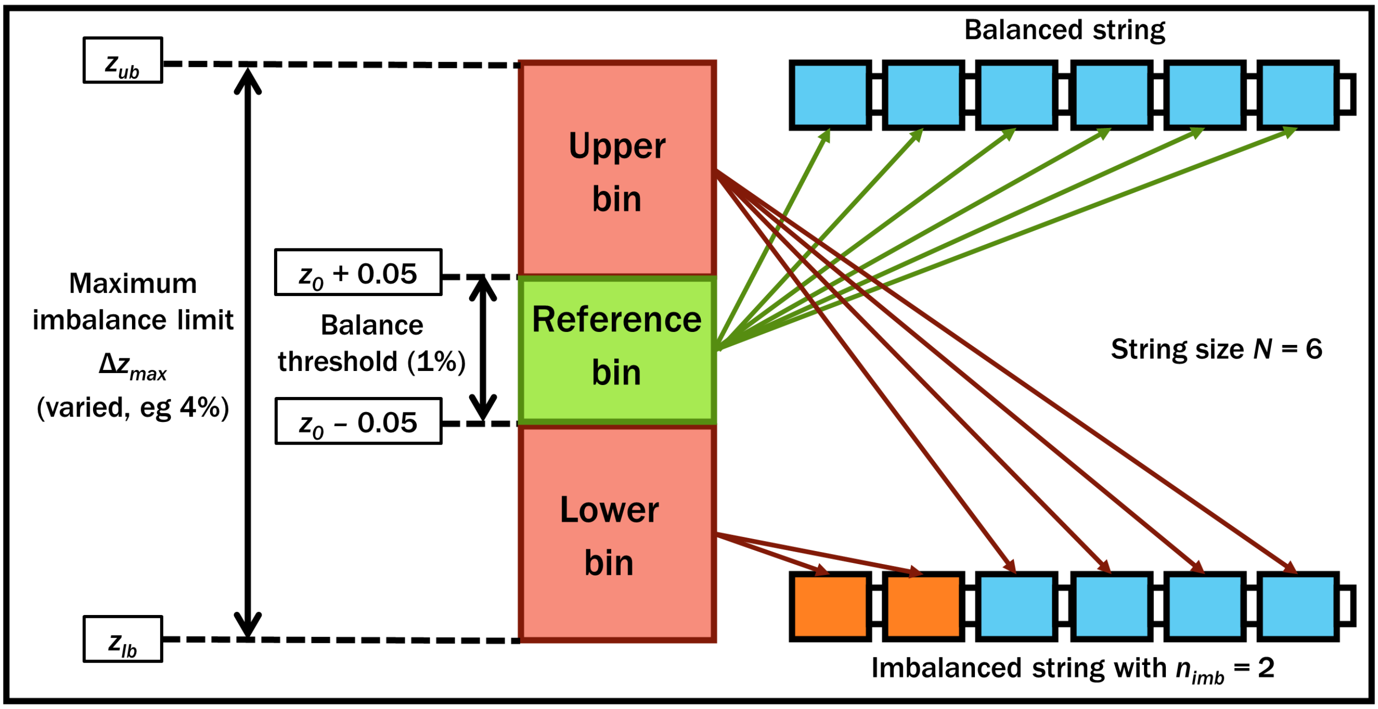


Fig. 3. Diagram of sampling process for simulating balanced and imbalanced cell strings

show that a neural networks (NN) can use pulse perturbation of a string of LIB cells to accurately classify a cell string as balanced or imbalanced, thus bypassing traditional control techniques. This could provide the basis for direct SoC estimation in a string using NN.

To date, only individual cells have been assessed using PIAML. Cell strings are important modules in battery packs that are well-suited for pulse perturbation. Understanding the SoC distribution of a string is important for the BMS. While locating and addressing the source of the imbalance is the ultimate task, identifying an imbalanced string is also important. A variety of scenarios are examined for validation, such as the degree of imbalance and the number of cells out of balance. There is potential for imbalance localization and state estimation.

The paper continues in Section II where key theoretical concepts are explained. Data collection and processing are described in Section III. Results are presented in Section IV. The paper is concluded and future work described in Section V.

II. Theoretical Concepts

PIAML has already shown high accuracy in estimating cell states [19] and it is a promising technique for characterizing degradation [20]. Offline, PIAML is developed by training a machine learning algorithm such as a feedforward NN. NN estimation consists of matrix multiplication between several layers of nodes and connections linked by network weights. The NN learns the optimal weights to match the target data given the observed voltage response of the cell. Targets can be cell states, as demonstrated in [19], but any cell characteristic could be used. Training is performed with repeated optimization cycles known as epochs. In each epoch the network weights and connections are readjusted based on the training data batch and optimizer function. Once the NN is trained, it estimates the targets given an unseen voltage response.

We hypothesize that imbalanced strings have a unique signature in the pulse voltage response that a NN can identify and learn from. We would expect this signature to be small for slight imbalance, and large for severe imbalance. We therefore train the PIAML algorithm on a string voltage response, with the classification boolean as the target: 0 for balanced, 1 for imbalanced.To label a string as balanced or imbalanced, the SoC of each cell must be known. Thus the design of the LIB strings must be controlled. While it is possible to collect experimental data of string pulse perturbation, it is costly and time-consuming to collect a sufficient amount of samples representative of real-world conditions. Thus we propose the simulation of cell strings using the experimental data from individual cells.

III. Data Collection and Simulation

Since cell strings are typically subject to identical cycling conditions, pulse data was collected similarly. Three nickel-magnesium-cobalt Panasonic NCR18650PF cells were cycled at 8°C between 0 to 0.5 state of charge (SoC) at 1 C-rate, 2.7 A. Pulses last 2 min with 1 C-rate amplitude and are composed of a charge and discharge portion. Capacity checks are performed every 100 cycles using a 1/20 C-rate discharge from full, the inverse derivative of which is shown in Fig. 2a as the incremental capacity (IC) curve. After the capacity check, pulses are applied from rest at various SoC, with 1 hour rest between each SoC level. In total, 363 unique pulses are obtained.

TABLE I
Numbers of imbalanced cells for different string sizes

String size N	Values of n_{imb}
3	1
6	1, 2, 3
12	1, 2, 4, 6
24	1, 4, 8, 12
48	1, 4, 8, 16, 24
96	1, 8, 16, 32, 48

Strings may be simulated with high fidelity using the pulse data collected from individual cells. In a string, it is known that each cell receives the same pulse current. By the superposition principle, the string voltage is equal to the sum of the voltages from all the cells. Since the same pulse current was applied to all the individual cells during cycling, we can sum the voltage responses from individual cells to obtain a ‘virtual string’.

There multiple variables to consider when designing a virtual string: string size N, the number of imbalanced cells n_{imb}, and the maximum imbalance limit Δz_{max}. These 3 factors control the string simulation process, represented in Fig. 3. A string is considered imbalanced if the SoC level of any cell is more than 1% away from the SoC of another cell in the string. We do not explicitly control clusters, a scenario in which there are multiple groups of balanced cells but the total string is imbalanced.

The simulation method mirrors random variation in cells by sampling the entire experimental dataset to generate cell strings. First, a SoC value corresponding to a pulse is randomly selected (with replacement) from the 363 real values. This allows us to define 3 sampling regions named the reference, upper, and lower bins. The reference bin is centered around the reference SoC z_0 and has a width of 1%, with the bounds given by $z_0 \pm 0.05$. The upper and lower bins extend beyond the limits of the reference bin, and stop at the upper and lower bounds z_{ub} and z_{lb}. The maximum possible imbalance is then defined as

$$\Delta z_{max} = z_{ub} - z_{lb} \tag{1}$$

To generate a balanced string, N samples are drawn from the reference bin. To generate an imbalanced string, n_{imb} samples are drawn from the lower and bin and $(N - n_{imb})$ from the upper bin. The voltage responses corresponding to each sample are then summed to obtain the string voltage.

We generate imbalanced strings for each unique combination of string size and maximum imbalance, where

$$\begin{aligned} N &= \{3, 6, 12, 24, 48, 96\} \\ \Delta z_{max} &= \{0.02, 0.04, 0.06, 0.08, 0.10\} \end{aligned} \tag{2}$$

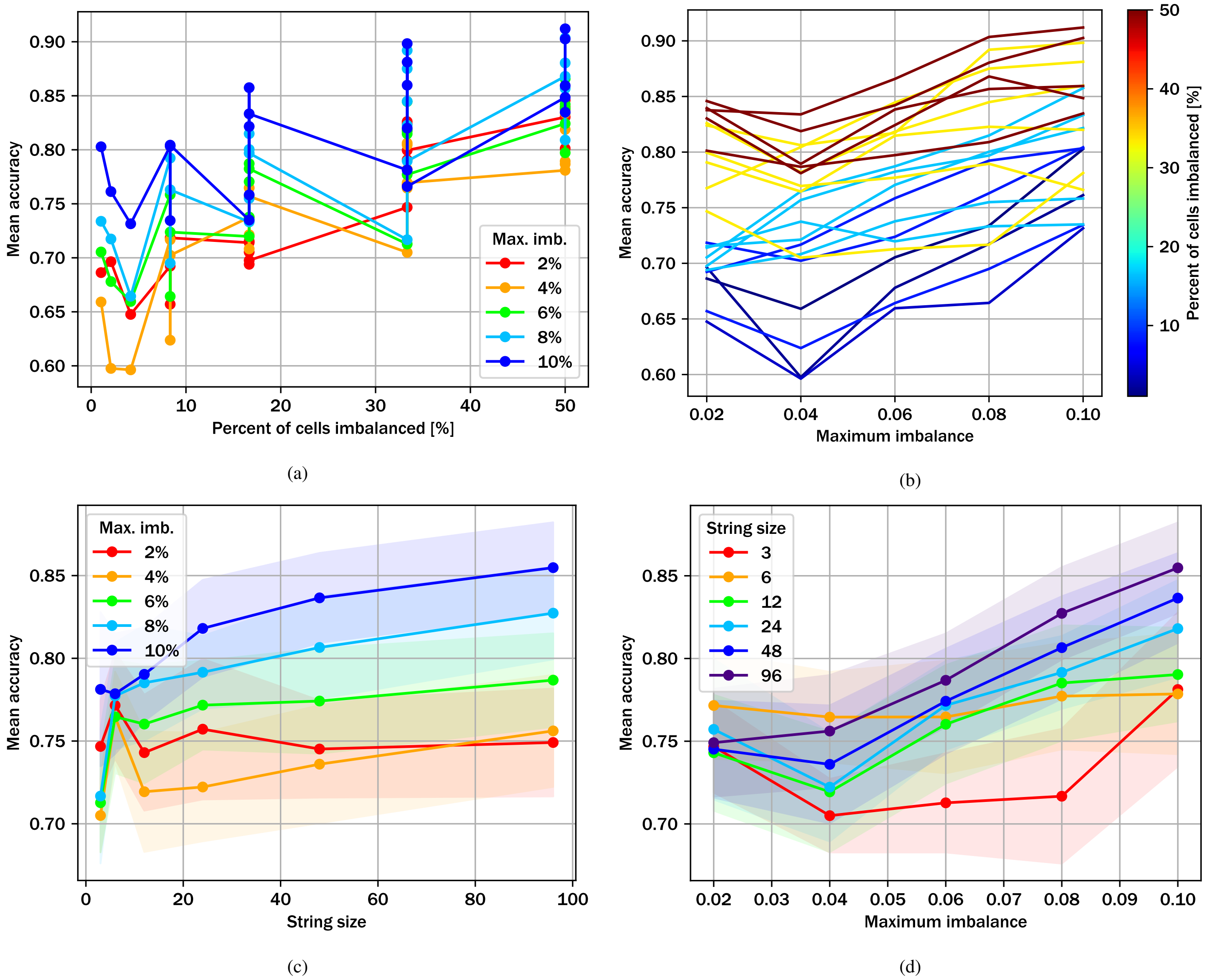


Fig. 4. Accuracy across 10 trials shown in various dimensions showing (a) Accuracy against the percent of imbalanced cells $\frac{n_{imb}}{N}$, (b) Accuracy against maximum imbalance for all string sizes, color-coded by the percent of imbalanced cells, and (c)-(d) Accuracy against string size and maximum imbalance with standard deviation across different values of n_{imb}

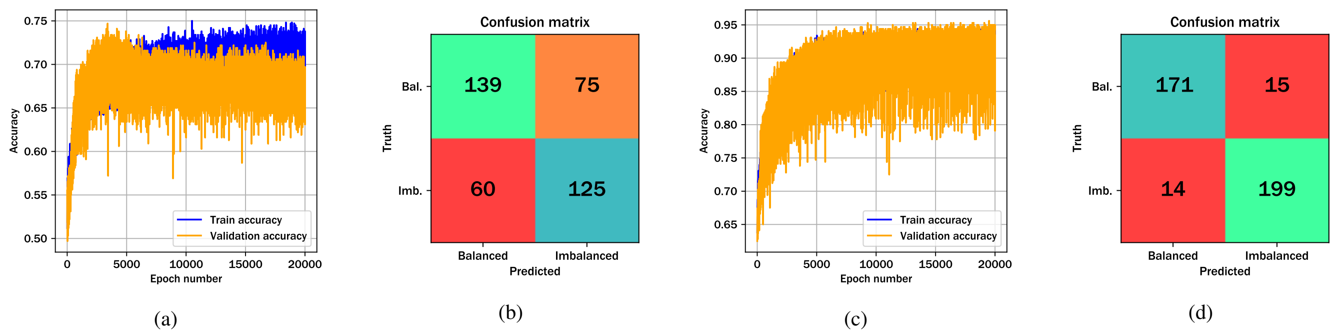
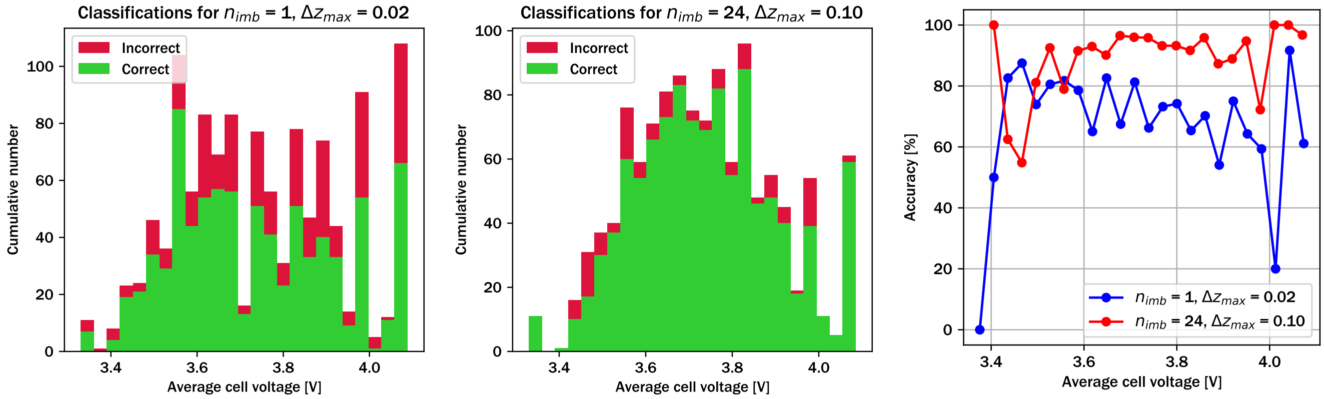


Fig. 5. NN results for a single trial of $N = 48$ strings, showing training loss (left) and confusion matrix (right) for (a)-(b) 2% maximum imbalance and 1 imbalanced cell, and (c)-(d) 10% maximum imbalance and 24 imbalanced cells.



(a) (b) (c)

Fig. 6. Voltage-varying plots showing how the number of correct classifications varies with the average cell voltage across multiple trials for 48 cell string, with (a) Histogram for n_{imb} = 1, Δz_{max} = 0.02 and (b) Histogram for $n_{imb} = 24$, $\Delta z_{max} = 0.10$, and (c) Plot of voltage-varying accuracy for the cases in (a) and (b)

TABLE II
HYPERPARAMETERS FOR NN TRAINING

Hyperparameter	**Value**
Hidden Layers	4
Nodes	32
Activation Function	Swish
Network Weight Constraint	5
Dropout Rate	0.1
Optimizer	Adam
Learning Rate	0.001
Batch Size	64
Training Epochs	20000
Patience	10000
Trials	10

We also consider varying values of n_{imb}, as shown in Table I. For comparison between different-size strings we can define the percent of imbalanced cells as

$$\text{percent imbalanced} = 100 \times \frac{n_{imb}}{N} \tag{3}$$

given the string size. Note that typically $\frac{n_{imb}}{N}$ does not exceed 0.5 for an imbalanced string. The chosen ranges of parameters represent typical real-world values [1].

For each combination of parameters, approximately 1000 balanced strings and 1000 imbalanced strings are generated. This yields 2000 simulated strings for 110 unique scenarios of imbalance. The pulse voltages for $N = 48$, $n_{imb} = 16$, and $\Delta z_{max} = 0.1$ are shown in Fig. 2b color-coded by the average nominal SoC of all the cells during the pulse. It can be seen that there is no obvious visual signature of an imbalanced string. The feedforward NN uses 64% of the data for training, 16% for training validation, and 20% for testing. The testing set is withheld from the NN during training so it makes predictions on unseen data. NN hyperparameters for network size, node activation, and optimization are shown in Table II. Due to the stochastic nature of NN optimization, the overall training and testing process is repeated for 10 trials.

IV. Results and Discussion

Results for all 110 scenarios are shown in Fig. 4. The maximum recorded accuracy for a single trial was 95% for $N = 96$, $n_{imb} = 48$, and $\Delta z_{max} = 0.10$. The lowest was 49% for $N = 48$, $n_{imb} = 1$, and $\Delta z_{max} = 0.04$. This suggests that PIAML for SoC imbalance classification is highly effective for large strings with many highly imbalanced cells, but ineffective for large strings with a very slight imbalance. This agrees with our hypothesis: the more severe the imbalance, the greater the signature in the pulse voltage and thus the more accurate the NN.

There are four similar but distinct perspectives with which to comprehend the variation in accuracy. First, Fig. 4a shows the accuracy against the percent of imbalanced cells $\frac{n_{imb}}{N}$ for various levels of severity Δz_{max}. As the number of imbalanced cells in the string increases, the NN becomes more accurate. This may be due to an exacerbated signature in the pulse data. Next, Fig. 4b shows the accuracy against the severity of imbalance. It is clear that accuracy increases linearly with the maximum imbalance. In Fig. 4c, the average of the mean trial accuracies for all values of n_{imb} is plotted against the string size. Accuracy increases approximately logarithmically with the size of the string for large values of Δz_{max} but is mostly constant for smaller values. This indicates that a large string with a higher maximum imbalance yields a greater pulse signature than the same string with a smaller maximum imbalance, a conclusion also supported by Fig. 4d.

An intriguing characteristic of the final results is the drop in accuracy between $\Delta z_{max} = 0.02$ and $\Delta z_{max} = 0.04$, as observed in Figs. 4a-b. This may be attributed to the simulation process. Imbalanced strings were created by sampling from the lower and upper SoC bins. Since the bins are 1% apart, this guarantees that the string will be imbalanced. When the maximum imbalance is 2%, the bin widths are 0.5%. This means that all the samples within each bin will be within 0.5% SoC of each other and the simulated string has two clusters of SoC levels. Once the Δz_{max} reaches 4% and above, the maximum SoC difference between samples within each bin exceeds 1%. Therefore it is possible for multiple SoC clusters to exist. The drop in accuracy corresponds to the onset of increased SoC clustering in the string. For $\Delta z_{max} > 0.04$, all strings are subject to clustering and thus the trends are affected by other variables.

Training and validation curves and post-testing confusion matrix for 2 unique scenarios are shown in Fig. 5. In Figs. 5a-b, $N = 48$, $n_{imb} = 1$, and $\Delta z_{max} = 0.02$. The validation accuracy ceases to decrease from approximately 6000 epochs, indicating a lack of improvement and the potential for overfitting the training data. Indeed, this scenario resulted in low accuracy, as shown in the confusion matrix. In Figs. 5c-d, $N = 48$, $n_{imb} = 24$, and $\Delta z_{max} = 0.10$. This scenario resulted in much higher accuracy and higher-quality validation curves that could improve with further training.

Finally, we examine the variation of classification accuracy with the average cell voltage in Fig. 6. This adds yet another dimension to the results so multiple trials from 2 specific scenarios were considered. Similar trends are expected for other scenarios. As expected, the voltages are clustered towards the middle, reflecting the simulation process. Since accuracy is measured across the entire dataset, the number of correct classifications as a portion of the total number of samples is plotted against the voltage range. Accuracy is lowest at the voltage extremes but this may be due to the small numbers of available samples at low and high voltages. There may be a correlations between classification accuracy and the OCV characteristic shown in Fig. 2a, where peaks in the IC curve represent phase changes in the LIB cell. Further research is needed to verify these links.

V. Conclusion

PIAML was shown to be a promising technique for classifying imbalance in LIB strings. It uses a NN to classify a string using the string's voltage response to a current pulse perturbation. We assess the performance of PIAML by simulating cell strings using experimental data collected from individual cells. String design is governed by 3 parameters: string size, the maximum imbalance, and the number of imbalanced cells. In total, we examine 110 unique combinations of parameters. PIAML performs the best when examining a 96-cell string with a high proportion of cells that are up to 10% imbalanced. This suggests that imbalanced strings leave a 'signature' in the voltage response that becomes easier to observe the greater the imbalance. There numerous opportunities for further improvement and extension of our work.

New string sizes and imbalance designs can be examined to validate the results for multiple configurations. Additionally, the effect of voltage levels on the classification accuracy can be quantitatively described for multiple scenarios. Wider real-world conditions could be examined, such as variations in the external and internal temperature. The NN could also regress the SoC, directly estimating the SoC in the string rather than performing classification. Direct string SoC estimation using PIAML could rival the existing state-of-the-art methods.

References

[1] M. F. R. Zwicker, M. Moghadam, W. Zhang, and C. V. Nielsen, "Automotive battery pack manufacturing –a review of battery to tab joining," *J. Advanced Joining Processes*, vol. 1, no. 100017, 2020.

[2] A. G. Li, K. Mayilvahanan, A. C. West, and M. Preindl, "Discrete-time modeling of li-ion batteries with electrochemical overpotentials including diffusion," *J. Power Sources*, vol. 500, no. 229991, 2021.

[3] A. Masias, J. Marcicki, and W. A. Paxton, "Opportunities and challenges of lithium ion batteries in automotive applications," *ACS Energy Lett.*, vol. 6, pp. 621–630, 2021.

[4] H. Dai, B. Jiang, X. Hu, X. Lin, X. Wei, and M. Pecht, "Advanced battery management strategies for a sustainable energy future: Multilayer design concepts and research trends," *Renew. Sustain. Energy Rev.*, vol. 138, no. 110480, 2021.

[5] Z. Wei, J. Zhao, H. He, G. Ding, H. Cui, and L. Liu, "A comprehensive overview of the dc-dc converter-based battery charge balancing methods in electric vehicles," *Renew. and Sus. Energy Rev.*, vol. 489, no. 229462, 2021.

[6] Z. B. Omariba, L. Zhang, and D. Sun, "Review of battery cell balancing methodologies for optimizing battery pack performance in electric vehicles," *IEEE Access*, vol. 9, pp. 129 335–129 352, 2019.

[7] D. F. Frost and D. A. Howey, "Completely decentralized active balancing battery management system," *IEEE Trans. Power Electronics*, vol. 33, no. 1, pp. 376–386, 2018.

[8] B. G. Carkhuff, P. A. Demirev, and R. Srinivasan, "Impedance-based battery management system for safety monitoring of lithium-ion batteries," *IEEE Trans. Industrial Electronics*, vol. 65, no. 8, 2018.

[9] Z. Wei, J. Zhao, H. He, G. Ding, H. Cui, and L. Liu, "Future smart battery and management: Advanced sensing from external to embedded multi-dimensional measurement," *J. Power Sources*, vol. 489, no. 229462, 2021.

[10] M. Naguib, P. Kollmeyer, and A. Emadi, "Lithium-ion battery pack robust state of charge estimation, cell inconsistency, and balancing: Review," *IEEE Access*, vol. 9, pp. 50 570–50 582, 2021.

[11] J. Xu, H. Wang, H. Shi, and X. Mei, "Multi-scale short circuit resistance estimation method for series connected battery strings," *Energy*, vol. 202, no. 117647, 2020.

[12] L. D. Couto, D. Zhang, A. Aitio, S. Moura, and D. Howey, "Estimation of parameter probability distributions for lithium-ion battery string models using bayesian methods," *Proceedings of the ASME 2020 Dynamic Systems and Control Conference*, 2022.

[13] L. Y. Wang, M. P. Polis, G. G. Yin, W. Chen, Y. Fu, and C. C. Mi, "Battery cell identification and soc estimation using string terminal voltage measurements," *IEEE Trans. Vehicular Technology*, vol. 61, no. 7, pp. 2925–2935, 2012.

[14] X. Lin, A. G. Stefanopoulou, Y. Li, and R. D. Anderson, "State of charge imbalance estimation for battery strings under reduced voltage sensing," *IEEE Trans. Control Systems Technology*, vol. 23, no. 3, pp. 1052–1062, 2015.

[15] J. Wei, G. Dong, Z. Chen, and Y. Kang, "System state estimation and optimal energy control framework for multicell lithium-ion battery system," *Appl. Energy*, vol. 187, pp. 37–49, 2017.

[16] M. A. Roscher, O. S. Bohlen, and D. U. Sauer, "Reliable state estimation of multicell lithium-ion battery systems," *IEEE Trans. Energy Conversion*, vol. 26, no. 3, 2011.

[17] B. Jiang, H. Dai, and X. Wei, "A cell-to-pack state estimation extension method based on a multilayer difference model for series-connected battery packs," *IEEE Trans. Transportation Electrification*, 2022, in press.

[18] S. Zhang, N. Peng, H. Liu, R. Li, and X. Zhang, "A systematic and low-complexity multi-state estimation framework for series-connected lithium-ion battery pack under passive balance control," *J. Energy Storage*, vol. 48, 2022.

[19] A. G. Li, W. Wang, A. C. West, and M. Preindl, "Health and performance diagnostics in li-ion batteries with pulse-injection-aided machine learning," *Appl. Energy*, vol. 315, no. 119005, 2022.

[20] A. G. Li, , A. C. West, and M. Preindl, "Towards unified machine learning characterization of lithium-ion battery degradation across multiple levels: A critical review," *Appl. Energy*, vol. 316, no. 119030, 2022.