



A novel PIML architecture with innovative learning paradigm applied in battery prognostics

Weikun Deng, Khanh T P Nguyen, Christian Gogu, Kamal Mejaher, Jérôme Morio, Hung Le, Dazhong Wu

► To cite this version:

Weikun Deng, Khanh T P Nguyen, Christian Gogu, Kamal Mejaher, Jérôme Morio, et al.. A novel PIML architecture with innovative learning paradigm applied in battery prognostics. 2024 10th International Conference on Control, Decision and Information Technologies (CoDIT), Jul 2024, Vallette, Malta. pp.248-253, <10.1109/CoDIT62066.2024.10708278>. <hal-04816008>

HAL Id: hal-04816008

<https://hal.science/hal-04816008v1>

Submitted on 3 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

A Novel PIML Architecture with Innovative Learning Paradigm Applied in Battery Prognostics

1st Weikun DENG

Laboratoire Génie de Production
Université de technologie Tarbes Occitanie Pyrénées
Tarbes, France
weikun.deng@doctorant.utttop.fr

2nd Khanh T. P. NGUYEN

Laboratoire Génie de Production
Université de technologie Tarbes Occitanie Pyrénées
Tarbes, France
thi-phuong-khanh.nguyen@utttop.fr

3th Christian GOGU

ICA, ISAE-SUPAERO
Université de Toulouse
Toulouse, France
christian.gogu@gmail.com

4rd Kamal MEJAHHER

Laboratoire Génie de Production
Université de technologie Tarbes Occitanie Pyrénées
Tarbes, France
kamal.medjaher@utttop.fr

5th Jérôme Morio

ONERA/DTIS
Université de Toulouse
Toulouse, France
Jerome.Morio@onera.fr

6th Hung LE

Additive Manufacturing and Intelligent Systems Laboratory
University of Central Florida
Orlando, United States
Hung.Le@ucf.edu

7th Dazhong WU

Additive Manufacturing and Intelligent Systems Laboratory
University of Central Florida
Orlando, United States
Dazhong.Wu@ucf.edu

Abstract—Prognostics and health management (PHM) increasingly play a constructive role throughout the entire lifetime of industrial equipment, significantly benefiting from extensive research in physical modelling and machine learning techniques. This has led to the development of hybrid approaches that seamlessly integrate both domains through physics-informed machine learning (PIML). PIML ensures the generation of cohesive solutions encompassing various aspects of physics knowledge across different stages of the machine-learning pipeline, substantially contributing to detection, diagnostics, and prognostics. However, PIML's design relies heavily on expert experience and demands rigorous interdisciplinary expertise, requiring a profound understanding of machine learning and physical principles. Inadequate design of PIML often leads to suboptimal outcomes, where the combined effect is less than the sum of its parts. Currently, PIML lacks a scalable and engineered application architecture to effectively utilise its embedded results. To address this challenge, our paper introduces a novel parallel architectural approach that employs pre-training and fine-tuning strategies for optimising the different model parts. Its data-driven branch is first trained with zero output in the PI branch, then fine-tuning the physics-informed branch. It takes the frozen data-driven model as a fixed feature extractor to get physics-consistency prediction. This approach proposes a generic solution for embedding physics knowledge into ML that guarantees performance improvement. The effectiveness of our approach is validated in the context of Remaining Useful Life (RUL) prediction using MIT-Stanford battery data.

Index Terms—Physics-informed machine learning architectural approach, gainful learning strategy, prognostic and health management, pre-trained model, physics consistency alignment fine-tuning.

I. INTRODUCTION

Prognostics and Health Management (PHM) is an interdisciplinary engineering approach that uses data analysis and knowledge mining techniques to detect anomalies, forecast performance, and assess the health of equipment, ultimately aiming to predict and prevent failures before they occur. It employs two primary paradigms: data-driven (data-driven) approaches, which are facing the sparse & noise data as well as the physics-agnostic dilemma [1], and 2) physics model-based methods (PBM), where the physics understanding of degradation remains incomplete due to ambiguous analytical relationships and unclear parameter estimates for modeling system failure mechanism [2]. Consequently, PHM is viewed as working in the “small data, partial knowledge” region where the Physics-informed Machine Learning (PIML) emerges as a promising solution after 2017 [3]. PIML aims to integrate physics knowledge into machine learning (ML) data processing pipelines, harnessing the strengths of both data-driven and PBM approaches [4]. The data-driven methods, represented by ML, and the PBM methods, encompassing parsing, geometric topology, and logical relations, complement each other as key feature extractors, knowledge compensators, and learning behavior regulators for achieving a highly generalized and job-adaptable PHM solution. However, current PIML studies fall short in reliance on expert experience and ensure tangible performance improvements [5]. The effectiveness of PIML depends on the appropriateness of the physical knowledge and embedding approach [6], but there is no mature, industrial pipeline-like foundation

architecture and model learning & evaluation methodology to screen and evaluate the architectural logic and embedded knowledge of PIML.

To achieve this goal, we suggest examining the connection between embedded knowledge and the informed ML model as a plugin. We introduce a parallel branching architectural scheme, presented in Section II, which has the potential to function as a foundational PIML model. This, in conjunction with the training approach described in Section III, guarantees productive training of the embedded knowledge within the original ML model's framework. The proposed PIML architecture paradigm is validated in a battery remaining discharging life prediction case in Section. IV. Finally, the conclusion is discussed in Section V.

II. PARALLEL PIML ARCHITECTURE

The proposed architecture consists of two parallel branches: a data-driven (data-driven) branch and a physics-informed (PI) branch. The data-driven branch utilizes pre-trained models, known for their extensive parameters and intricate structures. In contrast, the PI branch employs a simpler shallow model. In this branch, physical knowledge guides input space construction, and hidden structure design, or serves as a regularization term for learning, as elaborated in our prior work [7]. These two branches process the input data in parallel and are subsequently fused to make PHM-related decisions. Their architecture is detailed in Section. II-A with an specific example shown in Section II-B.

A. Architecture details

The proposed architecture is shown in Fig. 1. To acquire implicit knowledge, the data-driven branch conducts nonlinear fitting of all available measured inputs denoted as x_{in} . It produces high-dimensional abstract degradation features denoted as \hat{h}_l . The performance of the selected pre-trained neural network model, serving as the purely data-driven branch, sets a performance baseline for the entire PIML model. Depending on the specific PHM task, various pre-trained models, either trained independently or obtained from an off-the-shelf library, are typically chosen. Examples include Long Short-Term Memory (LSTM) and Transformer models for regression-type tasks, Convolutional Neural Networks (CNN) and ResNet for recognition tasks, and Autoencoders for detection tasks.

In PI branch, based on the physics variables derived from the physics input-output relationships f , a limited set of signals \hat{x}_{in} was selected from the input space x_{in} and utilized as the inputs for the PI branch. Subsequently, independent hidden layers with the custom design layer structure, activation function, and layer connection are employed to process the selected physics-related inputs. This preparation aligns the chosen inputs with a shallow network architecture designed for modeling the system's physical characteristics (f). These connections are established using parametric relationships that mimic the connections between each sub-item in the embedded "physical analytical relation f ", as detailed in our

prior work [8]. The PI branch serves as a regulator based on physics, setting an upper bound on the model's performance.

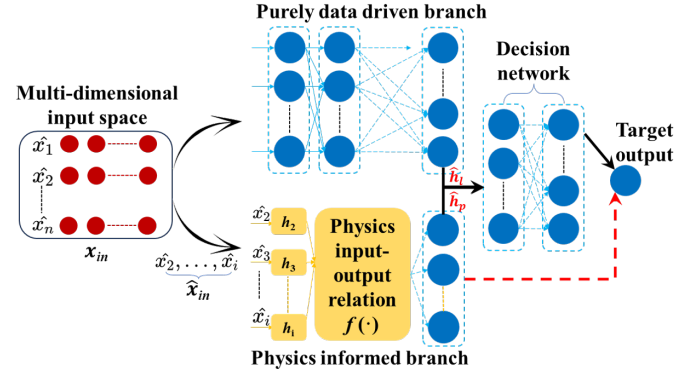


Fig. 1: Embedding physical knowledge into NN models in the form of parallel architectures.

The outputs from purely data-driven and PI branches, \hat{h}_l and \hat{h}_p , are coupled with each other and used as the inputs of the final module, named "Decision network", which is stacked by several full connection layers. The two branches operate autonomously within the intermediate hidden layers and collaborate only during the fusion of features \hat{h}_l and \hat{h}_p within the decision network. Furthermore, there is a potential to directly integrate the outputs of any existing physical resolution relations within the PI branch into the final module. The decision fusion is accomplished by combining the processing outcomes of the fused features with the output of the PI branch, as illustrated by the red dotted arrow in Figure 1.

B. The architecture of SEI-DCN

To enhance method clarity, we utilize a Solid Electrolyte Interphase growth model-informed Dilated Convolution Neural Network (SEI-DCN) as an illustrative example for predicting the remaining discharging cycles in batteries. The detailed presentation of this architecture is shown in Fig. 2. **The end-to-end DCN model structure used as the data-driven branch is taken from the research by Hong et al. [9]. Further upgrading by mounting PI branch on DCN, to ensure the comparison of control variables.** data-driven branch takes the raw monitoring series: voltage, temperature, and current as inputs. The inputs are combined by concatenation with the reshape operation to get the feature tensor with the shape [Training batch size, time-steps, feature channels]. The feature tensors undergo processing through hidden convolution layers with varying dilation rates. By progressively increasing the dilation rate in subsequent layers, DCN can efficiently expand its input data scope, gaining a more detailed and comprehensive understanding of data across various time scales without added computational cost. The outputs of the data-driven branch serve as the part of the inputs for the final decision network to determine the

remaining discharging cycles. The work in the data-driven branch is similar to the pre-trained Dilated CNN. But in PIML, it is combined with the PI branch's output for feature fusion-based decision-making.

The embedded SEI growth model (See Eq. (1)) is used to design a "physics-informed layer". This SEI growth model is sourced from the dataset in Attia et al. [10], with empirical constraint (Eq. (2)) proposed by Lui et al. [11] employed as the loss function layer for regulating PI branch's output behavior. In Eq. (1), where D represents the SEI degradation indicator. D_0 signifies the baseline degradation value. E_a corresponds to the effective activation energy for SEI growth, which is determined by the Arrhenius relation. k_B denotes Boltzmann's fundamental physical constant, and T signifies the operational temperature of the battery, influencing the SEI growth rate. In SEI-DCN, D is the output of the custom design layer which uses Eq. (1) as an activation function. D_0 and k_B are set as the default parameters of the custom design layer while the E_a are the trainable weights.

$$D = D_0 e^{-\frac{E_a}{k_B T}} \quad (1)$$

The empirical model used as the activation function in "Physics informed loss layer" to replace the traditional loss function. Minimizing the Mean Squared Error between the loss layer output and the label "0" ensures fulfillment of the optimization objective.

$$M(t) = 1 - \alpha_{\Pi} t^{\beta_{\Pi}} \quad (2)$$

As shown in Eq. (2), where $M(t)$ represents the degradation at time t , α_{Π} is a coefficient that scales the capacity loss, and β_{Π} is an exponent that dictates the rate of the capacity fade. These parameters are determined empirically from battery measurement data. α_{Π} is set as trainable weights in the "Physics informed loss layer". β_{Π} is set to 0.5, in which the model closely mimics the typical behavior of battery capacity fade observed caused by SEI growth in many studies.

III. GAINFUL PIML LEARNING STRATEGY

The challenge of effectively integrating small-parameter physical models with large-parameter data-driven models within parallel architectures becomes apparent because the PI branch and the data-driven branch exhibit notable disparities in the size and complexity of their parameter spaces. Consequently, these differences can lead to inconsistent optimization rates during training, thereby presenting a formidable obstacle to achieving simultaneous convergence of the two models. Furthermore, while large-parameter models are susceptible to overfitting, their small-parameter counterparts may suffer from underfitting or in reverse conditions. Hence, numerous approaches in the realm of PIML may fall short of achieving the anticipated synergy between these models, even impeding preserving the original data-driven model's fitting capability.

This paper claims that an effective PIML paradigm should

entail leveraging the capabilities of machine learning algorithms while incorporating insights from the domain of physics instead of commencing the learning process from a rudimentary state. The shift in this learning paradigm holds the potential to expand the applicability of PIML. It enables PIML, similar to existing natural language processing models, to utilize numerous well-established, data-driven, pre-trained models. These models can then be readily employed to create diverse PIML models by incorporating PI branches infused with varying knowledge.

The proposed training paradigm divides the original training process into two steps: In Step 1, the pre-trained model's structure and weights are loaded and frozen for the data-driven branch, servicing as the static feature extractor for PI branch. The parameters of the PI branch are randomly initialized and made trainable. The feature fusion layer in the decision network, which is also part of the pre-trained model, is relearned. In this configuration, the data-driven branch functions as a static feature extractor, permitting the PI branch to adapt and optimize itself using the extracted data-driven features. The training process is efficiently streamlined by exclusively updating the weights and biases of the PI branches. This focused training approach harnesses the capabilities of the pre-trained data-driven branch, which has already reached convergence. The physics-consistency alignment strategy in "Step 1" is designed to ensure that the PI branch aligns with the physical knowledge while leveraging the feature extraction capabilities of the pre-trained data-driven branch. By freezing the data-driven branch and only updating the weights and biases of the PI branch, the model can efficiently learn to incorporate physical insights without compromising the performance of the pre-trained data-driven component. This alignment process allows the PI branch to adapt and optimize itself based on the extracted data-driven features, resulting in a more physically consistent model. Given that the PI branch consists of a relatively shallow network with fewer parameters, it simplifies the overall training complexity. This not only conserves training resources but also takes advantage of publicly available pre-trained models for the data-driven branch, ensuring a resource-efficient and effective training process.

Upon completing "Step 1," the evaluation of the efficacy of the knowledge and approach utilized in the informed component becomes feasible. Specifically, by comparing its performance with that of the pre-trained data-driven branch operating independently, we can ascertain the suitability of the learned information. **This principle serves as a valuable criterion for informed knowledge, informed approaches, and data-driven model selection and evaluation, allowing for a rigorous recognition of the informed component's contribution.**

Building upon the preceding steps, this paper proposes "Step 2," denoted as the joint training process. The fundamental objective of this approach, known as PIML extends beyond rectifying the inconsistency between the physical insights of the data-driven approach and the embedded physical knowl-

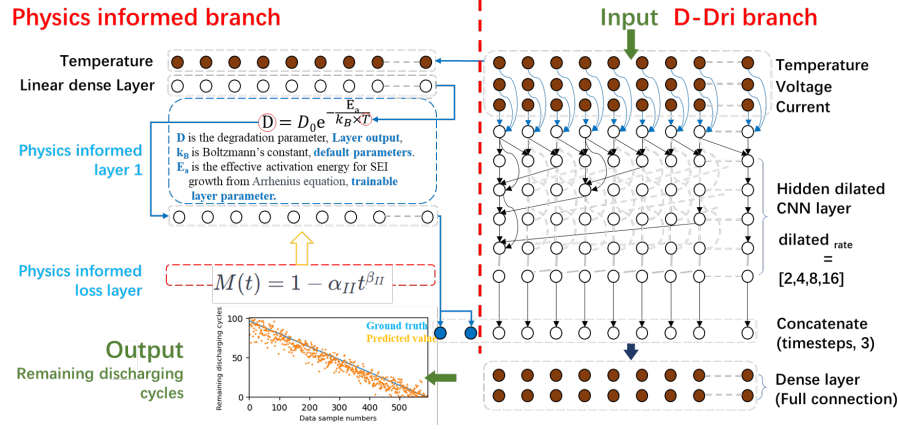


Fig. 2: Architecture of the SEI-DCN.

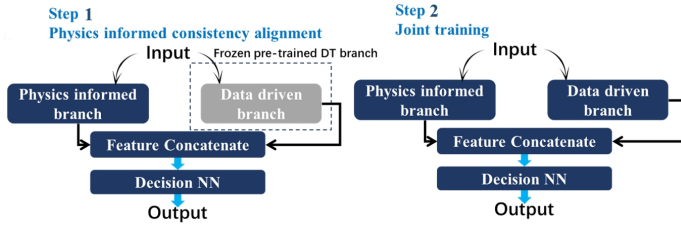


Fig. 3: Novel PIML learning mechanism: a multi-step learning process based on pre-training and then physical alignment in a parallel architecture.

edge. It also addresses enhancing the incomplete physical perspectives through data mining within the data-driven framework. Consequently, once the data-driven branch has been fine-tuned to a local optimum, demonstrating strong performance, and the PI branch has been optimized to a local optimum, considering its alignment with the data-driven branch's performance, the pre-training weights of the PI branch are loaded. Subsequently, the weights of the data-driven branch are configured as trainable, enabling the entire model to undergo more lifting retraining. This iterative process aims to bring the model closer to achieving a global optimal solution for the problem at hand.

While the validation case in this paper focuses on battery RUL prediction, the proposed learning strategy integrated with the parallel architecture framework is designed to be generalizable to various PHM challenges. The approach leverages pre-existing pre-trained models and partially completed physical knowledge resources, making it adaptable to different PHM domains. By incorporating domain-specific physical knowledge into the PI branch and utilizing relevant pre-trained data-driven models, the proposed methodology can be applied to a wide range of PHM problems, such as remaining useful life prediction for other critical components, fault diagnosis, and prognostics for various industrial systems. The flexibility of the parallel architecture and the multi-step learning process allows for the integration of different

physical models and data-driven techniques, enabling the framework to address the unique challenges of each PHM application.

IV. VALIDATION CASE

The method's effectiveness is evaluated using the fast-charging lithium batteries dataset developed by the MIT-Stanford research team [10]. In Section IV-B, we conduct a comparative analysis of SEI-DCN against various state-of-the-art machine learning models employed in battery datasets and discussed in the work by Hong et al. [9]. It is worth noting that our SEI-DCN model's data-driven branch uses the same end-to-end model as in study [9], thus performing a practical demonstration of how we inform the physics based on the existing model. Section IV-C delves into the decision mechanisms of the proposed methods.

A. Dataset introduction and data preprocessing

The dataset comprises three experiments, focusing on the primary battery sets from "2017-05" and "2017-06" for model training and testing. A randomized data splitting process is initially employed, dividing the dataset equally into training and 1st testing subsets (20 batteries for training, 20 for testing). This process involves shuffling the dataset indices and then dividing them into two groups, ensuring an appropriate partition. The training and testing datasets include corresponding indices to extract data samples and time information. This meticulous approach guarantees that both datasets accurately capture the diverse conditions present in the comprehensive dataset.

Before training, the sliding window technique is applied to sequentially segment continuous time-series data into overlapping windows from the beginning to the end of the cycle life. For example, in the sliding truncation process of temperature measurements, the parameter $\Delta_i, i = 1, 2, \dots, n$ represents the window length, determining the fixed analysis period within each window. Simultaneously, the step size δ determines the extent of overlap and the displacement between successive windows. Within each window segment,

the sequence information $(t_i, t_i + t_{\Delta_i})$, the termination time point of each window $t_i + t_{\Delta_i}$, and the corresponding number of loops c_i are extracted as potential input samples to the model.

B. Comparison with benchmark model in SOTA

The performance of ML models is assessed through a multifaceted approach which encompasses the determination of the necessary prediction cycles, the calculation of the mean absolute error (MAE) for cycle life prediction, and an assessment of the model's complexity based on its parameter count. The number of required cycles data is a crucial metric for remaining discharging cycles prediction, signifying the model's ability to make reliable predictions with limited data, essential for timely decision-making and efficient resource allocation. Simultaneously, the mean absolute error (MAE) quantifies the alignment between the model's predictions and actual cycle life, reflecting its reliability. Additionally, the model's complexity, gauged by its parameter count, influences the balance between overfitting risk, computational efficiency, and interpretability. This directly impacts the model's practical deployment and maintenance in real-world scenarios. These criteria seek to reflect both accuracy and practical applicability in various settings. Table. I shows com-

TABLE I: Performance of different models

Model	Required cycles	MAE
Shallow MLP [9]	4	150
Variance model [10]	100	112
CNN [9]	4	82
CNN + LSTM [9]	4	72
D-CNN in research [9]	4	65
D-CNN in our research	4	55
SEI-DCN	4	27

parison results. We find that the Shallow MLP model, despite having a relatively short required cycle length of 4, exhibits high MAEs, indicating less precise predictions. Conversely, the Variance model requires a significantly higher number of cycles (100) but demonstrates improved MAE, suggesting a trade-off between prediction speed and accuracy. Notably, the CNN and CNN + LSTM models balance this trade-off well, offering lower MAEs with shorter required cycle lengths. The D-CNN model stands out with the lowest MAE among models requiring only 4 cycles, though its parameter count is substantial. Finally, the proposed SEI-DCN model distinguishes itself by achieving the lowest MAE (27) while requiring a minimal cycle length. On a 16 GB NVIDIA T4 Tensor Core GPU platform, one sample's predicting time takes roughly 3.5×10^{-5} seconds.

To further present SEI-DCN model performance, we used box plots to visualize predictions for each group in Fig. 4. Fig. 4(a) displays the predicted Mean Absolute Percentage

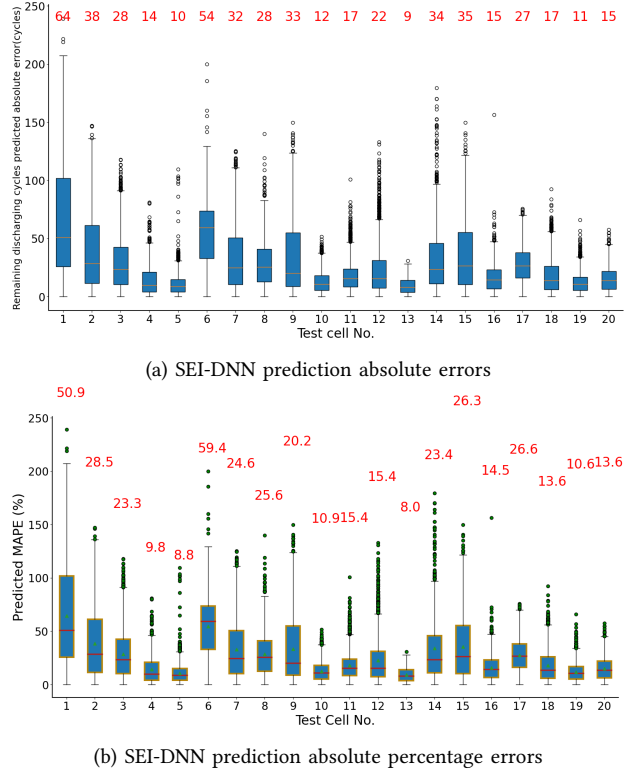


Fig. 4: Box plots of prediction performance (The red markers on the graph represent the MAE).

Error (MAPE), with average MAPE values ranging significantly from below 15% to over 50%. In this case, after reviewing the experimental records, the serial numbers of the cells in the group with a prediction error of 50 per cent were consistent with the samples reported in the dataset with recording errors. Fig. 4(b) shows the absolute error in predicting remaining discharge cycles, with averages per cell varying between 9 and 64 cycles, further demonstrating the model's variable performance across different cells.

C. Decision mechanistic exploration in learning process

In SEI-DCN, the importance of branch-specific features is vital for effectively merging empirical insights and fundamental principles, bolstering predictive accuracy and resilience, especially in intricate systems. This section lies in the examination of the decision fusion process within the proposed PIML model, particularly handling combined features in the concatenate layer (using a batch size of 80 dimensions), as shown in Fig. 5. To elucidate the reasoning behind feature prioritization and its influence on prediction outcomes, we employ a channel weights map in the initial dense layer following the concatenate layer, which reveals the relative significance attributed to features from each branch. The map after the first training step suggests the model's consideration of various features, and many of the weights in the PI channel are reversed from the data-driven channel, the PI branch plays a role similar to that of a regulariser. Following the

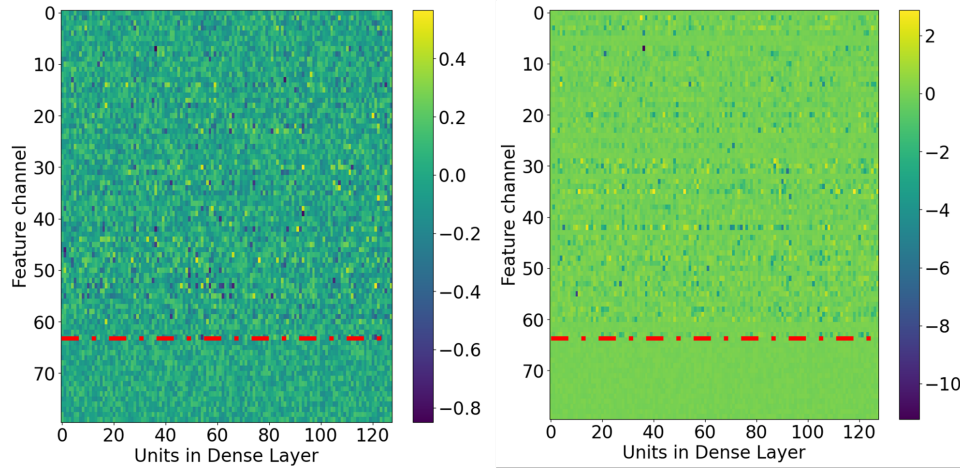


Fig. 5: Channel weights shift from data-driven (channels 1-64, above the red dotted line) to PI (channels 65-80, below the red dotted line) after “Physics informed alignment”(left) and “joint training”(right).

second joint training, the model exhibits increased selectivity. Notably, it is intriguing to observe that, at this stage, the knowledge from the PI branch, characterized by smaller parameters, appears to integrate with the more complex data-driven branch. Consequently, the entire model becomes sparser, with an emphasis on channels within the data-driven branch.

V. CONCLUSION

Our research streamlines the development of PIML in a dual branch parallel model framework. In this framework, we suggest a concise learning approach: leveraging pre-trained ML models to embed physical knowledge as plugins through multi-step alignment. We take an end-to-end model from SOTA as an informed object and use it as a branch of our model to practice our approach to develop a SEI growth model informed dilated convolution neural network, which is validated on the MIT-Stanford battery dataset. The results show that our method outperforms SOTA models by offering more accurate and consistent predictions of remaining discharging cycles. Specifically, the proposed model achieves a MAE of 27 for predicting remaining discharging cycles with just 4 cycle data. Additionally, we unveil a distinctive learning phenomenon within our framework, where the PI branch functions as a regularizer during the physical alignment process. This phenomenon leads to a sparser model and improved consolidation of knowledge during joint training. Future work should focus on deeper integration, where PI outputs are seamlessly incorporated within the data-driven branch at multiple levels to ensure the better application of PIML. Moreover, While the results are promising, the scalability of the proposed architecture to larger and more complex datasets requires investigation.

REFERENCES

- [1] S. Kohtz, Y. Xu, Z. Zheng, and P. Wang, “Physics-informed machine learning model for battery state of health prognostics using partial charging segments,” *Mechanical Systems and Signal Processing*, vol. 172, p. 109002, 2022.
- [2] J. M. Reniers, G. Mulder, and D. A. Howey, “Review and performance comparison of mechanical-chemical degradation models for lithium-ion batteries,” *Journal of The Electrochemical Society*, vol. 166, no. 14, p. A3189, 2019.
- [3] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, “Physics-informed machine learning,” *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.
- [4] R. Giorgiani do Nascimento, F. Viana, M. Corbetta, and C. S. Kulkarni, “Usage-based lifing of lithium-ion battery with hybrid physics-informed neural networks,” in *AIAA AVIATION 2021 FORUM*, 2021, p. 3046.
- [5] Y. Xu, S. Kohtz, J. Boakye, P. Gardoni, and P. Wang, “Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges,” *Reliability Engineering & System Safety*, p. 108900, 2022.
- [6] D. Weikun, K. T. Nguyen, K. Medjaher, G. Christian, and J. Morio, “Physics-informed machine learning in prognostics and health management: State of the art and challenges,” *Applied Mathematical Modelling*, vol. 124, pp. 325–352, 2023.
- [7] W. Deng, K. T. Nguyen, C. Gogu, J. Morio, and K. Medjaher, “Physics-informed lightweight temporal convolution networks for fault prognostics associated to bearing stiffness degradation,” in *PHM Society European Conference*, vol. 7, no. 1, 2022, pp. 118–125.
- [8] W. Deng, K. T. Nguyen, K. Medjaher, C. Gogu, and J. Morio, “Rotor dynamics informed deep learning for detection, identification, and localization of shaft crack and unbalance defects,” *Advanced Engineering Informatics*, vol. 58, p. 102128, 2023.
- [9] J. Hong, D. Lee, E.-R. Jeong, and Y. Yi, “Towards the swift prediction of the remaining useful life of lithium-ion batteries with end-to-end deep learning,” *Applied energy*, vol. 278, p. 115646, 2020.
- [10] P. M. Attia, A. Grover, N. Jin, K. A. Severson, T. M. Markov, Y.-H. Liao, M. H. Chen, B. Cheong, N. Perkins, Z. Yang et al., “Closed-loop optimization of fast-charging protocols for batteries with machine learning,” *Nature*, vol. 578, no. 7795, pp. 397–402, 2020.
- [11] Y. H. Lui, M. Li, A. Downey, S. Shen, V. P. Nemani, H. Ye, C. VanElzen, G. Jain, S. Hu, S. Laflamme et al., “Physics-based prognostics of implantable-grade lithium-ion battery for remaining useful life prediction,” *Journal of Power Sources*, vol. 485, p. 229327, 2021.