Title: Generative data augmentation for improving state estimation and prognostics in lithium-ion batteries: Advances, Challenges, and Future directions

Author names: Md. Sulyman Islam Sifat , Ummy Sawda , Md Alamgir Kabir , Mobyen Uddin Ahmed , M.S. Hossain Lipu , M.M. Manjurul Islam

Abstract: Lithium-ion battery state estimation and prognostics are crucial for the safe and reliable operation of electric vehicles (EVs), renewable energy systems, and portable devices. However, nonlinear battery behavior complicates State of Charge (SOC) estimation, limited and imbalanced data hinder State of Health (SOH), and scarce degradation trajectories with domain shift challenge Remaining Useful Life (RUL) prediction. Generative Adversarial Networks (GANs) offer an effective approach by generating realistic synthetic data that mitigates scarcity, improves diversity, and enhances model robustness. This study aims to systematically review and consolidate the current literature on GAN-based data augmentation for battery state estimation and prognostics. Specifically, it examines the effectiveness of different GAN architectures and techniques in improving SOC, SOH, and RUL estimation, assesses synthetic data quality and reliability, identifies technical challenges and limitations, and outlines evidence-based guidelines and future research directions. A systematic literature review was conducted following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines, incorporating relevant studies obtained from four prominent digital libraries. Analysis of 31 primary studies reveals that widely used public datasets (National Aeronautics and Space Administration (NASA), Center for Advanced Life Cycle Engineering (CALCE) and Oxford) dominate the field, enabling reproducibility and benchmarking. GAN-based augmentation achieves 17% to 90% error reductions across root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and mean squared error (MSE) metrics. Time-series GANs and Wasserstein GANs (WGANs) emerge as most effective, with Adam optimizer learning rate (LR) = 0.001 and gradient penalty ($\lambda GP=10$) providing improved stability. Major challenges include data scarcity, synthetic data quality concerns, GAN instability, and domain shift. Six priority areas are identified for advancing the field: physics-informed constraints, domain adaptation, uncertainty quantification, real-time deployment, multimodal learning, and data efficiency. This review establishes GAN-based augmentation as significant for battery state estimation and prognostics. It provides evidence-based insights and a roadmap towards developing reliable, interpretable, and deployable battery management systems.

Keywords: Generative Adversarial Networks; Lithium-ion batteries; Battery state estimation; Prognostics; Data augmentation; Systematic literature review; Deep learning; Battery health management

https://www.sciencedirect.com/science/article/abs/pii/S2352152X26006705