International Journal of Computer Techniques–IJCT Volume 13 Issue 2, March-April - 2026
Open Access and Peer Review Journal ISSN 2394-2231 https://ijctjournal.org/

Battery State-of-Health Prediction using Physics-Informed Feature Engineering and Group-Aware Machine Learning

Joseph Sujoy Pulivarthi, Suman Jana, Sai Sudheer Tadi, Lokeshwar Sai Gummidi, Veeranjaneyulu Rajamahendravarapu, P. Abdul Subhahanalla
*Department of Computer Science and Engineering, CSI Wesley Institute of Technology and Sciences, Hyderabad
Email: sujoypulivarthi@gmail.com
**Department of Computer Science and Engineering, CSI Wesley Institute of Technology and Sciences, Hyderabad
Email: p.subhan@gmail.com

Abstract:
Battery State-of-Health (SOH) prediction is critical for ensuring reliability and safety in energy storage systems. In this work, battery degradation is analyzed using Remaining Useful Life (RUL) trends and physics-informed features. A group-aware evaluation framework is proposed to analyze model performance under different data splits.

Three models, LSTM with attention, Linear Regression, and Random Forest are evaluated under both random and time-based splits. Experimental results show that while all models perform strongly under random split conditions ($R^2 > 0.99$), deep learning models exhibit severe performance degradation under distribution shift ($R^2 = -3.06$), whereas Linear Regression remains robust ($R^2 = 0.9975$).

A robustness score is introduced to quantify performance degradation across splits. The findings demonstrate that simpler models outperform deep learning when the underlying data exhibits strong monotonic degradation behavior. This highlights the importance of selecting models based on data characteristics rather than model complexity.

Keywords*—*Battery State-of-Health, Remaining Useful Life, Physics-Informed Features, LSTM, Random Forest, Distribution Shift, Robustness Analysis

I. INTRODUCTION

Battery health prediction is an essential component in modern energy systems such as electric vehicles and grid storage. State-of-Health (SOH) provides an estimate of battery degradation over time and is commonly derived from Remaining Useful Life (RUL).

Recent approaches have focused on deep learning models such as LSTM networks for capturing temporal dependencies. However, many studies rely on random data splits, which do not reflect real-world deployment scenarios where distribution shift occurs.

This paper evaluates multiple machine learning models under both random and time-based splits using a group-aware framework. The objective is to analyze robustness and understand how data characteristics influence model performance.

Unlike prior studies that primarily focus on improving predictive accuracy, this work emphasizes robustness under distribution shift and the impact of data characteristics on model performance.

II. METHODOLOGY

A. Data Processing

The dataset consists of lithium-ion battery cycle measurements from publicly available degradation datasets. Physics-informed features are derived to capture battery degradation dynamics, including voltage variation (ΔV), rolling averages, and degradation rate (Δcapacity). These features encode temporal trends and physical behavior of battery aging.

B. Sequence Modeling
Time-series sequences are generated using a fixed window size to capture temporal dependencies in battery degradation.

C. Models Used
The following models are evaluated:
1) LSTM with Attention
2) Linear Regression
3) Random Forest

D. Evaluation Strategy
Two evaluation strategies are used:
- Random Split (in-distribution)
- Time-Based Split (distribution shift)

E. Robustness Score
Robustness is defined as the difference in performance between random split and time-based split:

Robustness Score = R^2_random − R^2_time

A large robustness score indicates sensitivity to distribution shift, while smaller values indicate stable generalization.

Negative values indicate severe model failure. This metric provides a simple yet effective way to quantify model generalization.

Group-aware evaluation ensures that data from the same battery unit is not shared between training and testing sets. This prevents data leakage and provides a more realistic assessment of model generalization across different battery instances.

The dataset consists of battery cycle measurements collected from publicly available degradation datasets, including features such as voltage, current, and cycle count.

All experiments were conducted using Python with PyTorch and Scikit-learn libraries.

III. RESULTS AND DISCUSSION

A. Random Split Performance

Under random split conditions, all models achieve high predictive accuracy.

LSTM with attention achieved RMSE = 8.12, MAE = 6.39, and $R^2 = 0.9976$,

while Linear Regression outperformed all models with RMSE = 1.61, MAE = 1.23, and $R^2 = 0.9999$.

Random Forest also showed strong performance with RMSE = 2.25, MAE = 1.82, and $R^2 = 0.9998$.

These results indicate that when training and testing data follow similar distributions, classical machine learning models outperform deep learning models due to the near-linear degradation characteristics of the dataset.

B. Time-Based Split Performance

Under time-based split conditions, significant performance differences are observed.

LSTM performance degraded drastically with RMSE = 68.74, MAE = 64.47, and $R^2 = -3.06$, indicating severe performance degradation under distribution shift.

In contrast, Linear Regression remained stable with RMSE = 1.69, MAE = 1.45, and $R^2 = 0.9975$, demonstrating strong generalization.

This highlights that deep learning models are highly sensitive to distribution changes, whereas simpler models retain robustness.

C. Robustness Analysis

Robustness is evaluated using the difference in R^2 between random and time-based splits.

LSTM exhibits a robustness score of -3.07, with a performance drop of 4.06 in R^2, indicating extreme sensitivity to distribution shift.

Linear Regression shows a robustness score of 0.9976 with a minimal performance drop of 0.0023, demonstrating high stability.

These findings confirm that model performance is strongly influenced by data distribution, and simpler models can outperform complex models in structured degradation scenarios.

D. Performance Drop Analysis

To further quantify model stability, performance drop is computed as the difference between R^2 under random and time-based splits.

LSTM shows a performance drop of 4.06, indicating extreme sensitivity to distribution shift.

In contrast, Linear Regression shows a negligible drop of 0.0023, demonstrating strong robustness.

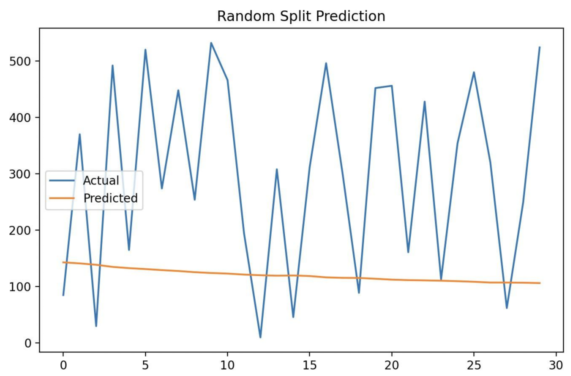

Fig. 1 Actual vs Predicted RUL under random split conditions

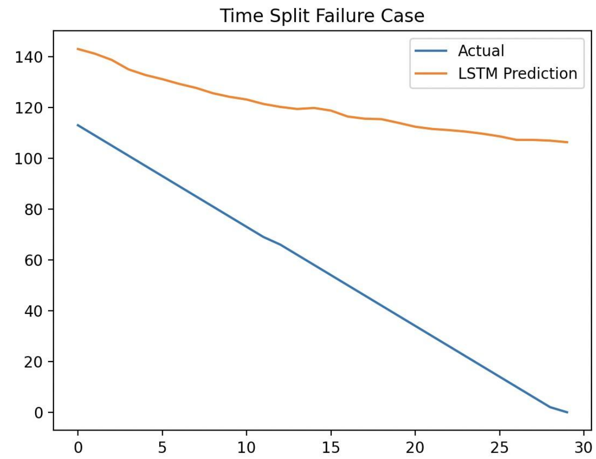

Fig. 2 Model failure under time-based split demonstrating LSTM inability to generalize under distribution shift

Table I. Model Performance Comparison under Random and Time-Based Splits

Model	RMSE	MAE	R^2
LSTM (Random)	8.12	6.39	0.9976
Linear Regression	1.61	1.23	0.9999
Random Forest	2.25	1.82	0.9998
LSTM (Time)	68.74	64.47	-3.06
Linear (Time)	1.69	1.45	0.9975

IV. CONCLUSION

This study demonstrates that model performance is strongly influenced by data characteristics rather than model complexity. While deep learning models perform well under random splits, they fail under temporal distribution shift. Linear models exhibit strong robustness, making them more suitable for real-world deployment.

These results highlight that model robustness is a critical factor for real-world deployment in battery health monitoring systems.

Future work includes extending the approach to larger and more diverse datasets, and exploring hybrid modeling techniques for improved robustness under distribution shift.

This work demonstrates that model selection should prioritize robustness over complexity in real-world battery health prediction systems.

ACKNOWLEDGMENT

The authors would like to thank the faculty and institution for their support in completing this research work.

REFERENCES

[1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[2] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, 1997.

[3] X. Zhang et al., “Battery RUL Prediction Using Machine Learning,” IEEE, 2020.

[4] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” JMLR, 2011.

[5] Y. Zhang et al., “Data-driven battery degradation modeling,” IEEE Transactions on Industrial Electronics, 2021.

[6] J. Liu et al., “Battery remaining useful life prediction using machine learning,” Applied Energy, 2020.