

Abordagem Baseada em Dados para Predição de Vida Útil de Baterias via Aprendizado de Máquina

Matheus Henrique Pereira Soares¹,
Kevyn Robert do Vale Gandolfi¹,
Gustavo Hideki Scaranello¹,
Cauê Pais Justo¹,
Davi Paulino¹,
Arthur Emanuel de Oliveira Carosia ²

¹ Estudantes do curso de Bacharelado em Ciência da Computação, IFSP Campus São João da Boa Vista.

² Professor. Arthur Emanuel de Oliveira Carosia, IFSP Campus São João da Boa Vista.

Área de conhecimento (Tabela CNPq): 10300007

RESUMO:

Este estudo apresenta uma abordagem baseada em aprendizado de máquina supervisionado para estimar o Estado de Saúde (State of Health — SoH) de baterias de íons de lítio. O experimento utiliza uma base de dados estruturada derivada do repositório NASA Ames, composta por 636 registros de ciclos operacionais das células B0005, B0006, B0007 e B0018. Foram avaliados comparativamente os algoritmos Regressão Linear, Random Forest Regressor e Support Vector Regression (SVR) a partir de atributos elétricos, térmicos e temporais agregados por ciclo. Para respeitar a natureza sequencial dos dados, os primeiros 70% dos ciclos de cada bateria foram destinados ao treinamento e os 30% finais ao teste, utilizando GridSearchCV para a otimização de hiperparâmetros. A expectativa é identificar um modelo preditivo de baixa complexidade computacional que combine alta acurácia estatística com viabilidade de execução em tempo real em Sistemas de Gerenciamento de Bateria (BMS) embarcados.

PALAVRAS-CHAVE: Aprendizado de Máquina. Modelos de Regressão. Baterias de Íons de Lítio. Estado de Saúde (SoH). Predição de Capacidade.

1. INTRODUÇÃO

As baterias de íons de lítio consolidaram-se como a principal tecnologia de armazenamento energético para dispositivos eletroeletrônicos portáteis e veículos elétricos devido à sua elevada densidade de energia e vida útil prolongada. Contudo, durante os regimes dinâmicos de carga e descarga, essas células sofrem processos irreversíveis de degradação físico-química, destacando-se o crescimento da camada de interface de eletrólito sólido (SEI) no anodo e a perda de material ativo. Esses fenômenos reduzem progressivamente a capacidade máxima de retenção de carga e são severamente influenciados por variáveis operacionais externas, como temperatura, taxas de corrente e tempos de operação.

O monitoramento contínuo do Estado de Saúde (State of Health — SoH) é uma função crítica dos Sistemas de Gerenciamento de Bateria (BMS) para garantir a segurança e prever falhas prematuras. Todavia, métodos empíricos tradicionais, como a contagem de Coulomb, apresentam limitações severas por acumularem erros de medição ao longo do tempo e dependerem de calibrações frequentes. Como alternativa promissora, as abordagens baseadas em dados (data-driven) por Aprendizado de Máquina (Machine Learning) destacam-se por extrair padrões de envelhecimento diretamente de variáveis macroscópicas mensuráveis, eliminando a necessidade de modelagens eletroquímicas complexas e de difícil parametrização em tempo real. Embora arquiteturas baseadas em Aprendizado Profundo (Deep Learning) ofereçam precisão elevada em séries temporais complexas, o seu alto custo computacional e a demanda por hardware dedicado restringem sua aplicação prática em microcontroladores comerciais de baixo custo.

Diante desse cenário, este trabalho propõe uma avaliação comparativa de três algoritmos clássicos de regressão (Regressão Linear, Random Forest e SVR) utilizando um conjunto de dados consolidado com 636 ciclos do repositório NASA Ames. O objetivo principal é estruturar um framework preditivo otimizado via engenharia de atributos e ajuste de hiperparâmetros, selecionando o modelo que ofereça o melhor equilíbrio entre exatidão estatística (R^2 e RMSE) e viabilidade de implementação em sistemas embarcados restritos em memória e processamento.

2. FUNDAMENTAÇÃO TEÓRICA E TRABALHOS RELACIONADOS

O envelhecimento de baterias de íons de lítio está relacionado a mecanismos irreversíveis, como o crescimento da camada SEI e a perda de material ativo. O efeito acumulado desses processos reduz a capacidade disponível da célula. O SoH pode ser calculado pela razão entre a capacidade de descarga atual e a capacidade nominal, multiplicada por 100. No arquivo analisado, a capacidade nominal de referência é de 2 Ah, de modo que o SoH corresponde à capacidade de descarga dividida por 2 Ah e expressa em porcentagem.

A NASA Ames Prognostics Center of Excellence disponibiliza dados de envelhecimento de baterias submetidas a ciclos de carga e descarga em condições laboratoriais (SAHA; GOEBEL, 2007). A base estruturada utilizada neste trabalho contém registros agregados por ciclo das baterias B0005, B0006, B0007 e B0018. Como o procedimento original de transformação do arquivo tabular não foi integralmente reconstruído, as conclusões deste estudo se restringem ao CSV analisado.

Modelos de aprendizado profundo apresentam resultados relevantes na previsão da degradação, mas geralmente exigem maior volume de dados e maior custo computacional (WANG et al., 2021).

Algoritmos clássicos, como Regressão Linear, Random Forest e SVR, constituem alternativas adequadas para bases menores e atributos agregados. A engenharia de atributos por ciclo também pode reduzir a dimensionalidade dos dados e facilitar a modelagem do envelhecimento (PAULSON et al., 2022). A adequação desses modelos a hardware embarcado, porém, depende de testes específicos de tempo, memória e consumo energético.

3. METODOLOGIA E MATERIAIS

A base utilizada contém 636 registros e 12 colunas, sem valores ausentes ou linhas duplicadas. Cada linha representa um ciclo de uma das quatro baterias. O alvo do experimento é a variável `soh_percent`. A capacidade de descarga foi mantida apenas para análise exploratória e não foi utilizada como entrada dos modelos, pois possui relação direta com o cálculo do SoH.

Foram utilizados oito atributos preditores: índice do ciclo, tensão média, tensão mínima, tensão máxima, corrente média, temperatura média, temperatura máxima e tempo de descarga. A temperatura ambiente foi removida da modelagem porque apresenta o valor constante de 24 em todos os registros e, portanto, não contém informação preditiva. A divisão dos dados preservou a ordem dos ciclos de cada bateria. Os primeiros 70% foram destinados ao treinamento e os 30% finais ao teste, resultando em 443 registros de treino e 193 de teste. O conjunto de teste permaneceu separado durante o ajuste dos modelos. No conjunto de treino, foi aplicada validação cruzada temporal com cinco divisões sucessivas.

A modelagem comparou Regressão Linear, Random Forest Regressor e Support Vector Regression com kernel de Função de Base Radial (RBF). A Regressão Linear foi empregada como referência. Os hiperparâmetros do Random Forest e do SVR foram selecionados por busca em grade, utilizando somente o conjunto de treino. Para o SVR, a padronização dos atributos foi incorporada a um Pipeline, evitando o uso de informações do teste durante o ajuste. O bloco a seguir resume a inicialização dos modelos:

4. RESULTADOS E DISCUSSÃO

Na validação cruzada temporal realizada no conjunto de treino, a Regressão Linear apresentou RMSE médio de 0,3472 ponto percentual e R^2 médio de 0,9854. O SVR obteve RMSE médio de 0,5010 e R^2 médio de 0,9662. O Random Forest apresentou RMSE médio de 2,1543 e R^2 médio de 0,4399. Os resultados indicam maior estabilidade da Regressão Linear nos segmentos temporais de validação.

Modelo	RMSE (média $\pm$ desvio)	MAE (média $\pm$ desvio)	R^2 (média $\pm$ desvio)
Regressão Linear	0,3472	0,3018	0,9854
Random Forest	2,1543	1,7552	0,4399
SVR	0,5010	0,3569	0,9662

Tabela 1 — Resultados da validação cruzada temporal no conjunto de treino

No teste composto pelos 30% finais dos ciclos, a Regressão Linear obteve MSE de 0,3403, RMSE de 0,5833 ponto percentual, MAE de 0,5012 ponto percentual e R^2 de 0,9800. Esse foi o menor RMSE entre os modelos avaliados, indicando boa capacidade de acompanhar a tendência temporal do SoH no conjunto analisado. O SVR apresentou MSE de 0,3565, RMSE de 0,5971 ponto percentual, MAE de 0,3569 ponto percentual e R^2 de 0,9790. Embora seu RMSE tenha sido ligeiramente superior ao da Regressão Linear, o modelo obteve o menor MAE, indicando menor erro absoluto médio. Assim, os dois modelos apresentaram desempenhos próximos, mas com vantagens em métricas diferentes. O Random Forest obteve MSE de 15,8343, RMSE de 3,9792 pontos percentuais, MAE de 3,1637 pontos percentuais e R^2 de 0,0689. O resultado foi inferior ao observado na divisão aleatória preliminar. Árvores de decisão apresentam dificuldade para extrapolar valores além das regiões observadas no treinamento, o que ajuda a explicar a perda de desempenho nos ciclos finais.

Modelo	MSE	RMSE	MAE	R^2
Regressão Linear	0,3403	0,5833	0,5012	0,9800
Random Forest	15,8343	3,9792	3,1637	0,0689
SVR	0,3565	0,5971	0,3569	0,9790

Tabela 2 — Comparativo das métricas no conjunto de teste temporal

A comparação entre valores reais e previstos mostrou que a Regressão Linear e o SVR acompanharam a redução do SoH nos ciclos finais das quatro baterias. A importância por permutação permitiu identificar quais atributos mais contribuíram para o Random Forest, sem recorrer apenas à importância interna baseada nas árvores. Esses resultados devem ser interpretados no contexto da base utilizada, composta por quatro baterias de uma única fonte laboratorial. O experimento não avaliou tempo de inferência, consumo de memória, consumo energético ou compatibilidade com microcontroladores. Portanto, os resultados não comprovam viabilidade para execução em BMS embarcados. Também não é possível generalizar o desempenho para outras químicas, temperaturas ambientes ou condições de operação sem novos testes.

5. CONCLUSÃO

Este estudo comparou três algoritmos de regressão para estimar o SoH de quatro baterias do conjunto NASA Ames. A separação temporal permitiu avaliar os modelos nos ciclos posteriores aos utilizados no treinamento, oferecendo uma condição mais próxima da previsão de degradação futura do que uma divisão aleatória dos registros. A Regressão Linear apresentou o menor RMSE e o maior R^2 no teste temporal, enquanto o SVR obteve o menor MAE. O Random Forest apresentou desempenho inferior, indicando dificuldade para extrapolar a tendência de degradação nos ciclos finais. Os resultados mostram que modelos clássicos podem estimar o SoH com baixo erro no conjunto analisado, mas não comprovam generalização para outras baterias ou aplicação direta em sistemas embarcados. Como trabalhos futuros, recomenda-se ampliar a quantidade de baterias, incluir condições ambientais variadas, testar conjuntos externos e avaliar os modelos no hardware-alvo. Também é necessário documentar integralmente o processo de transformação dos dados brutos da NASA para o formato tabular utilizado.

CRÉDITOS DE AUTRALIDADE (CONTRIBUIÇÃO DOS AUTORES)

[Nome do Aluno 1]: Concepção do projeto, Engenharia de atributos, Desenvolvimento e execução do código de Machine Learning, Escrita do manuscrito original. [Nome do Aluno 2]: Coleta e padronização dos datasets públicos, Condução dos ensaios caseiros nos dispositivos portáteis, Formatação de dados. [Nome do Orientador]: Orientação acadêmica, Revisão crítica da metodologia científica, Validação dos resultados estatísticos e Coordenação da pesquisa. Todos os autores contribuíram ativamente na revisão textual e aprovaram formalmente a versão submetida ao congresso.

REFERÊNCIAS

BATTERYML. BatteryML: An Open-source Platform for Machine Learning on Battery Degradation. Microsoft Research, GitHub Repository, 2024. Disponível em: <<https://github.com/microsoft/BatteryML>>. Acesso em: 15/06/2026.

SAHA, B.; GOEBEL, K. Battery Data Set. NASA Prognostics Center of Excellence Data Repository, 2007. Disponível em: <<https://c3.ndc.nasa.gov/dashlink/resources/133/>>. Acesso em: 15/06/2026.

GOTZ, J. D. Aplicação de Machine Learning para a identificação, tratamento e antecipação de falhas em baterias de íons de lítio. 2023. 175 f. Tese (Doutorado em Engenharia de Manufatura) - Universidade Tecnológica Federal do Paraná, Curitiba, 2023.

LI, A. G.; WEST, A. C.; PREINDL, M. Towards unified machine learning characterization of lithium-ion battery degradation across multiple levels: a critical review. Applied Energy, Elsevier, v. 316, p. 119030, 2022.

NASA AMES PROGNOSTICS CENTER OF EXCELLENCE. Li-ion Battery Aging Datasets. NASA Dashlink Repository, 2010. Disponível em: <<https://c3.ndc.nasa.gov/dashlink/resources/133/>>. Acesso em: 15/06/2026.

PAULSON, N. H. et al. Feature engineering for machine learning enabled early prediction of battery lifetime. *Journal of Power Sources*, Elsevier, v. 527, p. 231127, 2022.

SULZER, V. et al. The challenge and opportunity of battery lifetime prediction from field data. *Joule*, Cell Press, v. 5, n. 8, p. 1934-1955, 2021.

VAJJA, J. V. R. et al. Predicting rapid degradation onset in lithium-ion batteries during real-time operation using machine learning. In: *In Operando Health Monitoring for Lithium-Ion Batteries*. [S.l.]: Academic Press, 2024. p. 1-40.

WANG, X. et al. A critical review of improved deep learning methods for the remaining useful life prediction of lithium-ion batteries. *Energy Reports*, Elsevier, v. 7, p. 5562-5574, 2021.