

AI-Readiness von Batteriedaten: Datenvorverarbeitung für ML-basierte Batteriezellendiagnostik

Dominique Koster¹, Isabell Nagel¹, Daniela Fenske¹ and Julian Schwenzel¹

¹Abteilung Elektrische Energiespeicher, Fraunhofer-Institut für Fertigungstechnik und Angewandte Materialforschung

Email: dominique.koster@ifam.fraunhofer.de

Motivation & Zielsetzung

- In der Batteriezellendiagnostik findet aktuell eine Verschiebung von klassischen Verfahren hin zu vorausschauendem Monitoring mittels maschinellen Lernens (ML) statt
- ML-Modelle nutzen Muster in Zellbetriebsdaten, um Zustandsgrößen, wie den *State of Health (SoH)*, zu ermitteln → schnell, flexibel und oft präziser als experimentbasierte Ansätze
- Leistungsfähigkeit von ML-Modellen hängt unmittelbar von der Datenqualität ab → Batteriedaten weisen oft eine hohe Heterogenität auf, was eine große Herausforderung für die Entwicklung ML-basierter Modelle darstellt
- Ziel:** Entwicklung einer automatisierten Datenverarbeitungs-Pipeline, die Batteriedaten in robuste, ML-fähige Datensätze transformiert und die Grundlage für reproduzierbare, skalierbare und interpretierbare Diagnosemodelle schafft

Methodik

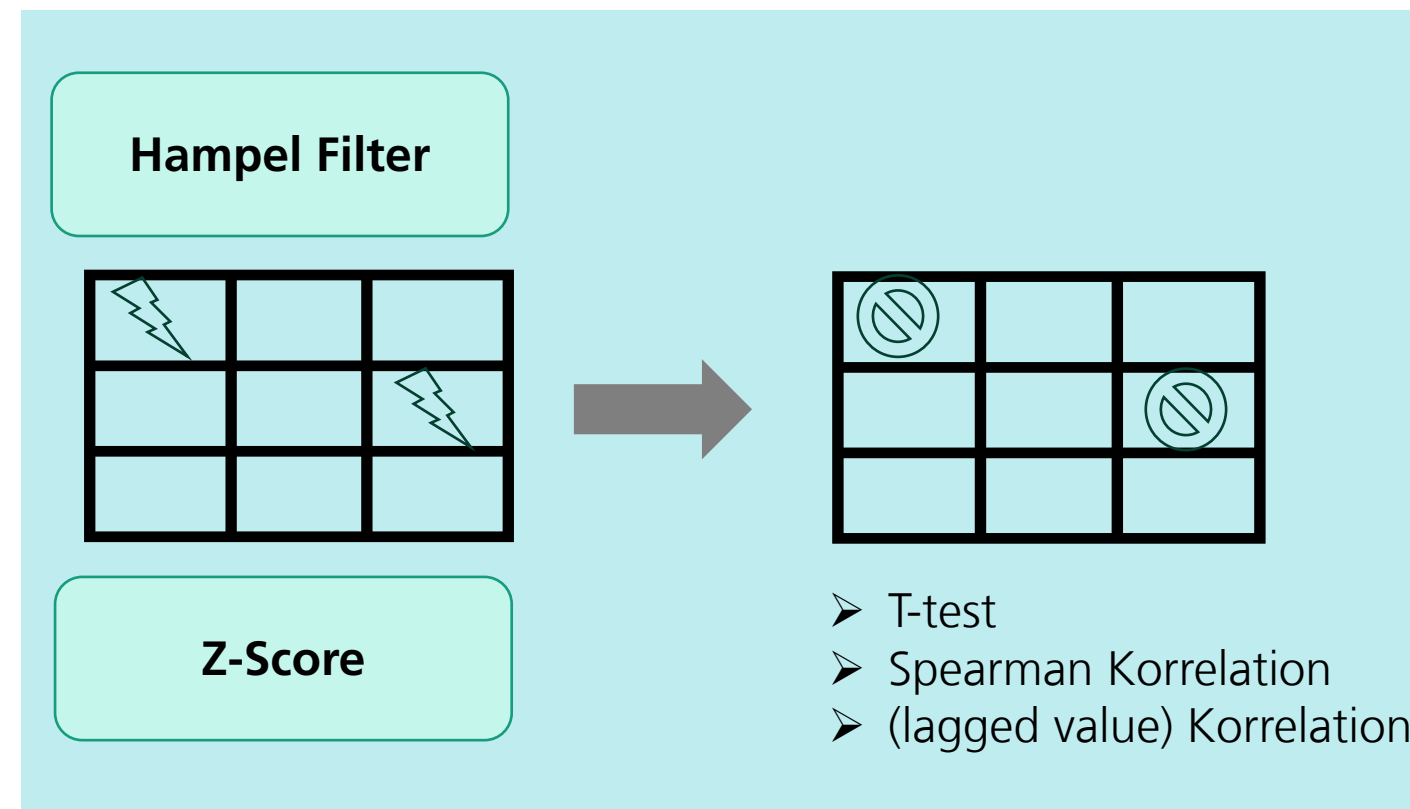


Ergebnisse

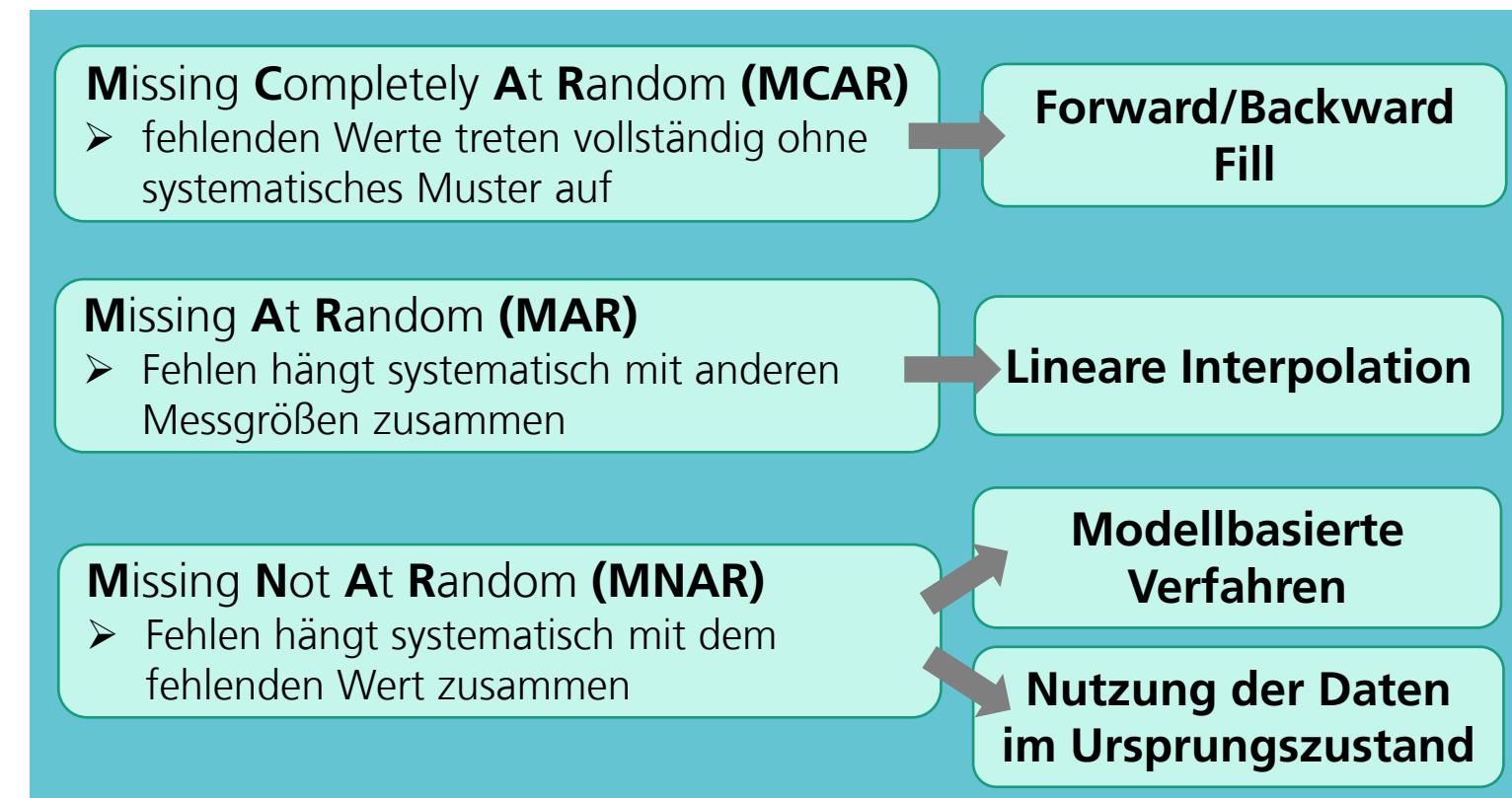
1. Untersuchung von Zeitreihencharakteristika



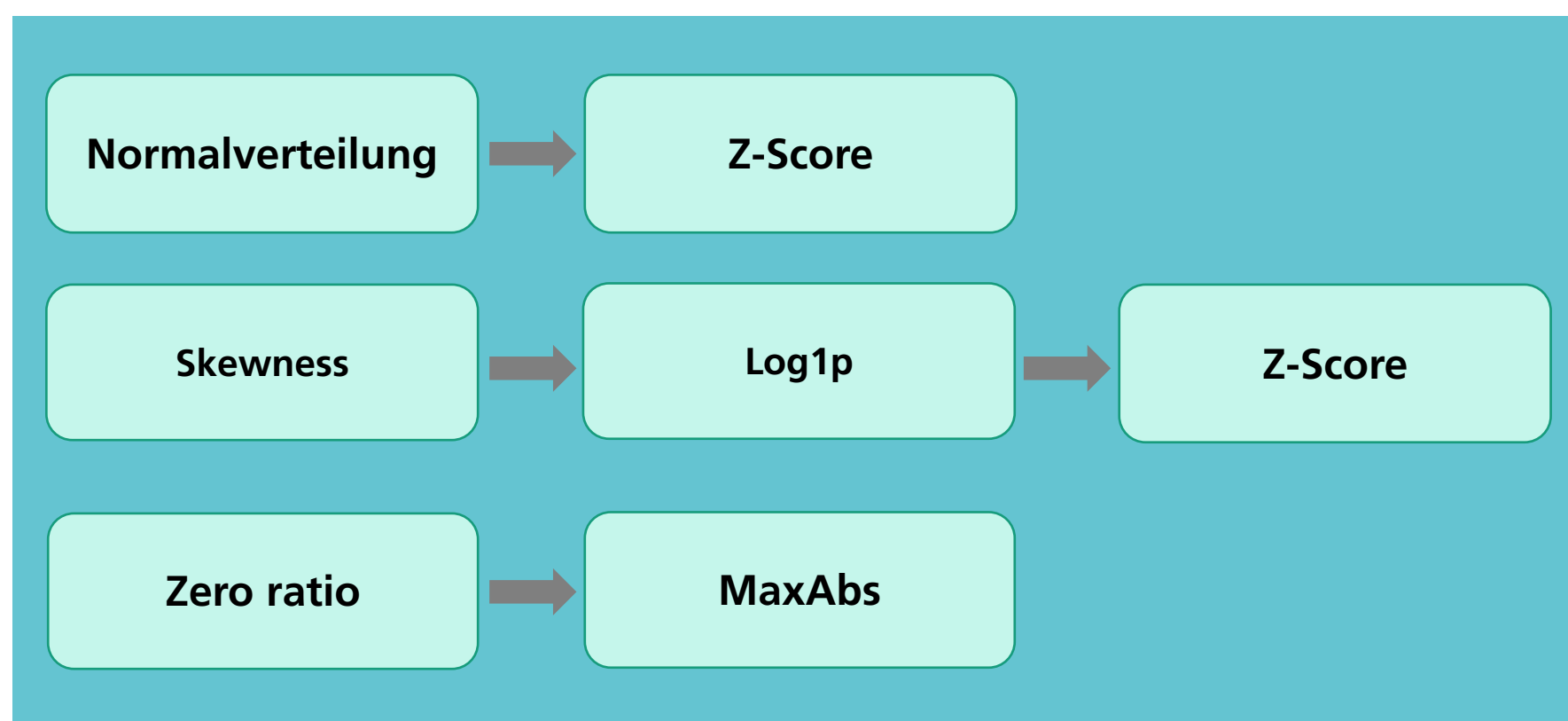
2. Ausreißererkenennung & Ersetzen durch Fehlstellen



3. Analyse der Fehlstellenverteilung & Imputation^[6]



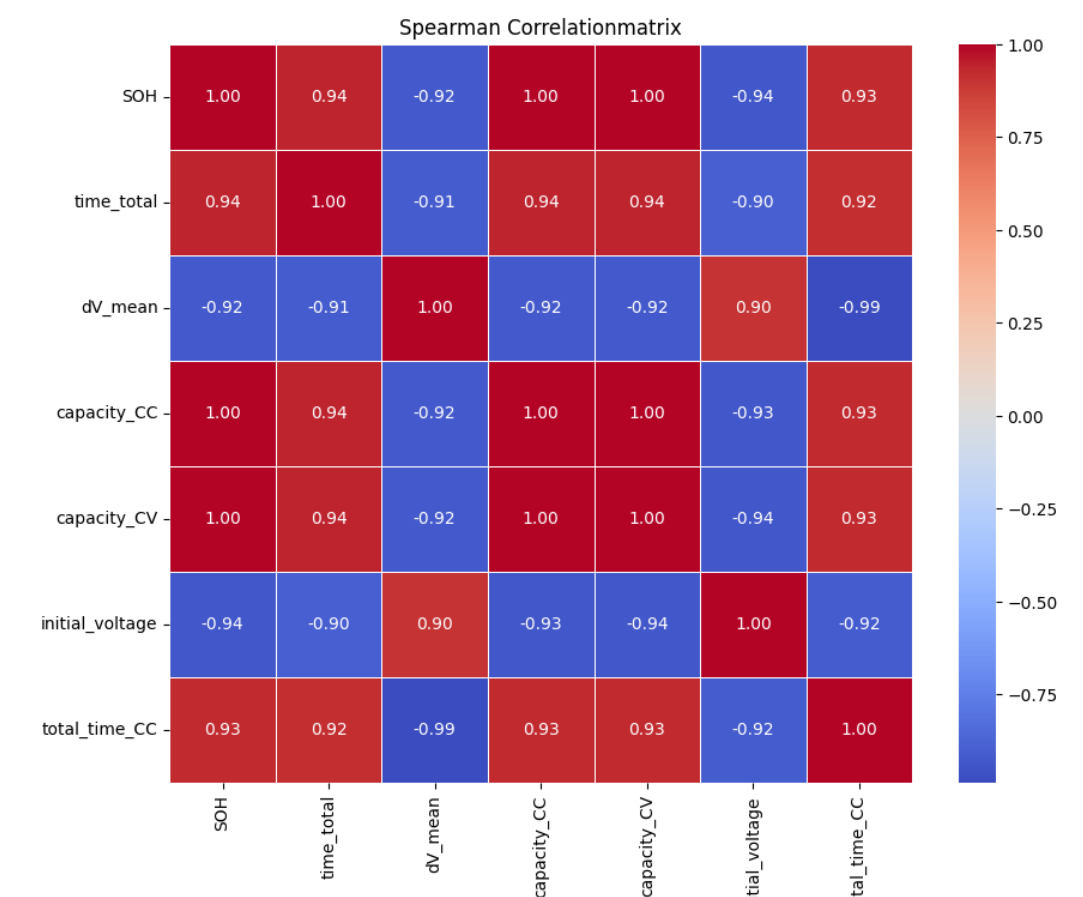
4. Datennormalisierung



5. Feature Extraktion

Nr.	Feature Beschreibung	Quelle
1	Mean of voltage slope	(Hu et al., 2014)
2	Max of voltage slope	
3	First voltage value	
4	Time for CC	(Guo et al., 2019)
5	Time for CV	
6	t _{CC} /t _{CV}	
7	t _{CC} /t _{total}	
8	total time	
9	total capacity during CC	
10	total capacity during CV	(Deng et al., 2019)

6. Feature Selektion



7. Modelltraining und Evaluierung

Diskussion und Schlussfolgerung

- Eine umfassende Datenexploration und Vorverarbeitung sind essenziell, um
 - Transparenz über die Datenqualität zu schaffen und die
 - Belastbarkeit von ML-Modellen besser einschätzen zu können
 - Signalspezifisches Profiling ermöglicht eine automatisierte Erkennung und Imputation von Ausreißern, ohne dass zyklische Muster und erwartbare Alterungstrends fälschlicherweise als Anomalien markiert werden
 - wenn systematisch auftretende Ausreißer/Fehlstellen (MNARs) durch nicht-funktionierende Messungen, d.h. echte Ausreißer verursacht werden, wird eine bessere Modellperformance erzielt, wenn betroffene Datensätze vollständig aus Trainingsdatensatz entfernt werden → **Vorverarbeitung wenig effektiv**
 - wenn MNARs durch reale physikalische Effekte verursacht werden, mindert das Entfernen der betroffenen Datensätze die Modellleistung → **Vorverarbeitung sinnvoll**
- Korrekte Handlungsstrategie setzt richtige Einordnung der Ursache der Systematik durch **domänenspezifisches Verständnis** voraus → **Automatisierung nur eingeschränkt möglich**

[1] NASA – Ames Prognostics Center of Excellence (2010), <https://data.mendeley.com/datasets/nsc7hnsq4s/2>

[2] HUST – Huazhong University of Science and Technology, <https://c3.ndc.nasa.gov/dashlink/resources/133/>

[3] Hu, C., Jain, G., Zhang, P., Schmidt, C., Gomadam, P., & Gorka, T. (2014). *Applied Energy*, 129, 49–55. <https://doi.org/10.1016/j.apenergy.2014.04.077>

[4] Deng, Y., Ying, H., E, J., Zhu, H., Wei, K., Chen, J., Zhang, F., & Liao, G. (2019). *Energy*, 176, 91–102. <https://doi.org/10.1016/j.energy.2019.03.177>

[5] Guo, P., Cheng, Z., & Yang, L. (2019). *Journal of Power Sources*, 412, 442–450. <https://doi.org/10.1016/j.jpowsour.2018.11.072>

[6] Little, R. J. A., & Rubin, D. B. (2020). *Statistical analysis with missing data* (3rd edition). Wiley series in probability and statistics. Wiley. <https://doi.org/10.1002/9781119482260>