Maschinelles Lernen in der Batteriedatenanalyse

Ein datengetriebener Ansatz zur Bestimmung des Alterungsstands von Lithium-Ionen-Batterien

Masterarbeit
zur Erlangung des akademischen Grades

Master of Science (M.Sc)

Fachhochschule Vorarlberg
Nachhaltige Energiesysteme

Betreut von
Dr. Steffen Finck

Vorgelegt von
Rafael Pires Mimoso B.Sc.
Dornbirn, Januar 2023

Kurzreferat

Maschinelles Lernen in der Batteriedatenanalyse: ein datengetriebener Ansatz zur Bestimmung des Alterungsstands von Lithium-Ionen-Batterien

Ansätze des maschinellen Lernens werden sowohl in der Forschung als auch in der Praxis eingesetzt, um gewünschte Ausgabedaten anhand bekannter Eingabedaten vorherzusagen. In dieser Masterarbeit wird die Anwendung des maschinellen Lernens in der Batteriedatenanalyse zur Bestimmung des Alterungsstatus von Lithium-Ionen-Batterien untersucht.

Das Ziel dieser Arbeit besteht in der Vorhersage von Alterungskurven (englisch state of health - SoH) für Lithium-Ionen-Batterien über die Anzahl der Entladezyklen (Zeitachse). Dies erfolgt auf der Grundlage zuvor erfasster Daten für drei Typen von Lithium-Ionen-Batterien, die bei Temperaturen von 15 °C, 25 °C und 35 °C sowie C-Raten von 0,5C, 1C und 2C aufgenommen wurden. Im Zuge dessen wurden die angewandten Methoden des maschinellen Lernens analysiert und ihre Ergebnisse verglichen. Der Umfang dieser Arbeit hebt sich von anderen Ansätzen des maschinellen Lernens in der Batteriedatenanalyse ab, da dieselben Methoden in einem breiteren Spektrum von Daten mit unterschiedlichen Temperaturen und Kathodenmaterialien verwendet wurden. Dies ist für die Analyse von Unterschieden im Verhalten in der Praxis relevant.

Nach dem Erwerb und der Vorbereitung der Daten wurden Modelle mit vier ausgewählten Regressionsverfahren (lineare Regression, Ridge-Regression, Random-Forest-Regression und KNN-Regression) des überwachten Lernens trainiert und die Vorhersagen durchgeführt. Aus den Ergebnissen kann eine allgemeingültige Auslegungsgrundlage für weitere Untersuchungen und die praktische Anwendung abgeleitet werden, bei der die Vorhersagen von SoH-Kurven für Lithium-Ionen-Batterien mit linearer Regression und Ridge-Regression die höchste Genauigkeit aufweisen.

Schlüsselwörter: *maschinelles Lernen, Data Science, Alterungsstand, Ladezustand, Lithium-Ionen-Batterie, Regression*

Abstract

Machine learning in battery data analysis: A data-driven approach to determine the ageing status of lithium-ion batteries

Machine learning approaches are used in both research and practice to predict desired output data based on known input data. In this master's thesis, the application of machine learning in battery data analysis to determine the ageing curve of lithium-ion batteries is investigated.

The aim of this thesis is to predict the ageing curves (state of health - (SoH)) of lithium-ion batteries over the number of discharge cycles (time axis) based on previously collected data from three different types of lithium-ion batteries recorded at different temperatures (15°C, 25°C and 35°C) and C-rates (0.5C, 1C, 2C). The applied machine learning methods were analysed and their results compared. The scope of this work is distinct from other approaches to machine learning in battery data analysis as the same methods were used with a wider range of data obtained with different temperatures and cathode materials, which is important for analysing the different behaviour behaviour of different lithium-ion batteries in pracitcal applications.

After acquiring and preparing the data, four selected regression methods (linear regression, ridge regression, random forest regression and KNN regression) of supervised learning were trained and predictions were made. From the results, a general design basis for further research and practical application was obtained, with the SoH curves of lithium-ion batteries predicted by linear regression and ridge regression achieving the highest accuracy.

Keywords: *machine learning, data science, stat of health, state of charge, lithium-ion battery, regression*

Danksagung

An dieser Stelle möchte ich mich herzlich meinem Betreuer der Fachhochschule Vorarlberg Dr. Steffen Finck für die hilfreichen Vorschläge und zielorientierte Betreuung bedanken. Mein Dank gilt meinen Studienkollegen insbesondere für Karen Schwab und Nikolai Franzke für die Motivation und das Interesse an meiner Arbeit. Mein Dank gilt meiner Frau, Lisa Pires Mimoso und meinen Kinder Elias, Hanna und Gabriel Pires Mimoso. Sie brachten ebenfalls Geduld und motivierten mich dazu, das Masterstudium abzuschließen. Nicht zuletzt möchte ich mich bei meinen Eltern für ihre emotionale Unterstützung währen des Studiums bedanken.

Inhaltsverzeichnis

Abbildungsverzeichnis **VI**

Tabellenverzeichnis **VIII**

Abkürzungsverzeichnis **IX**

1 Einleitung **1**
1.1 Motivation 2
1.2 Zielsetzung 2

2 Stand von Wissenschaft und Technik **3**
2.1 Aufbau und Funktionsprinzip der Lithium Ionen Batterien und Akkumulatoren 3
2.2 Batteriesysteme 5
2.2.1 Kennwerten und Zustandsgrößen 6
2.2.2 Lebensdauer und Alterungsmechanismen 7
2.3 Methoden zur Bestimmung des Alterungsstands 9
2.3.1 White-Box-Methode 9
2.3.2 Black-Box-Methode 12
2.3.3 Grey-Box-Methode 12
2.4 Maschinelles Lernen 16
2.4.1 Unüberwachtes Lernen 16
2.4.2 Überwachtes Lernen 17

3 Methodik und Umsetzung **18**
3.1 Erwerb und Vorbereitung der Daten 19
3.1.1 Trainings- und Testdaten 21
3.1.2 Overfitting und Underfitting 21
3.2 Regressionsmodelle 25
3.3 Ablauf 31
3.4 Grundlagen zur Performancebewertung und Schätzungsqualität 33

4 Ergebnisse und Diskussion **35**
4.1 Ergebnisse der NMC-Datensätze 36
4.2 Ergebnisse der NCA-Datensätze 41

4.3 Ergebnisse der LCO-Datensätze 42

5 Fazit **44**
5.1 Zusammenfassung 44
5.2 Ausblick und kritische Bewertung 45

Literaturverzeichnis **45**

Eidesstattliche Erklärung **75**

Abbildungsverzeichnis

2.1 Aufbau einer Lithium-Ionen-Zelle 4
2.2 Lithium-Ionen-Batteriesystem 5
2.3 SoH-Bestimmung im Zusammenhang mit dem Box-Konzept . . 9
2.4 (a) IC und (b) DV-Kurven einer NCA-Batterie 10
2.5 Batteriemodelle hinsichtlich ihrer Komplexität 13
2.6 ML 16

3.1 Workflow maschinelles Lernen 18
3.2 Beispiele aus der Voranalyse der Rohdaten NMC 25° (a) und NMC 35°C (b) 20
3.3 Beispiele aus der Voranalyse der Rohdaten NMC 25° (a) und NMC 35°C (b) nach Umsetzung von Savitzy-Golay-Filter 20
3.4 Train-Test-Split 21
3.5 Erste Relevanzanalyse 22
3.6 Relevanzanalysen mit und ohne Lade- und Entladeenergie (a und b bzw.) 23
3.7 Modellfehler: Bias-Varianz-Dilemma 24
3.8 Funktionsprinzip des Grid Search 25
3.9 Abstände zu den nächsten Nachbarn 28
3.10 Struktur eines Random-Forest 30
3.11 Übersicht der verwendeten NMC-Datensätze aus SNL mit unterschiedlichen Temperaturen (15°C, 25°C und 35°C) und C-Rate (1C und 2C) bei DoD von 100% 31
3.12 Übersicht der verwendeten NCA-Datensätze aus SNL mit unterschiedlichen Temperaturen (15°C, 25°C und 35°C) und C-Rate (0.5C, 1C und 2C) bei DoD von 100% 32
3.13 Übersicht der verwendeten LCO-Datensätze aus CALCE mit 25°C Betreiebstemperatur und C-Rates von 0.5C und 1C bei DoD von 100% 32

4.1 Ergebnisse 15°C NMC 2, 3, 4 - Train-Test 80/20 mit Lade- und Entladeenergie 36
4.2 Ergebnisse 15°C NMC 2, 3, 4 - Train-Test 80/20 ohne Lade- und Entladeenergie 38

4.3 Ergebnisse 25°C NMC 6, 7 und 35°C NMC 26, 27 - Train-Test 80/20 mit Lade- und Entladeenergie. 40
4.4 Ergebnisse 15°C NCA 1, 2, 3, 4 - Train-Test 80/20 mit Lade- und Entladeenergie. 41
4.5 Ergebnisse 25°C NCA 5, 7 und 35°C NCA 21, 22 - Train-Test 80/20 mit Lade- und Entladeenergie. 42
4.6 Ergebnisse 25°C LCO 34, 36, 37, 38 - Train-Test 80/20 ohne Lade- und Entladeenergie. 43

Tabellenverzeichnis

2.1 Spezifische Ladung 4
2.2 Begriffe und Kennwerte von Batteriesysteme 6
2.3 Quantitativer Vergleich der SoH-Bestimmungsmethoden 15

3.1 Batteriedatensätze 19
3.2 Koeffizienten der linearen Regression mit Lade- und Entladeenergie 27
3.3 Koeffizienten der linearen Regression ohne Lade- und Entladeenergie 27
3.4 Hyperparameter der Random-Fores-Regression 30
3.5 Mathematische Grundlage von Bewertungsmetriken eines Regressionsmodells 33

4.1 Übersicht Kennzahlen zur Performancebewertung 35
4.2 Übersicht der Kennzahlen zur Performancebewertung 35
4.3 Übersicht der Ergebnisse 15°C NMC 2, 3, 4 - Train-Test 80/20 mit Lade- und Entladeenergie 37
4.4 Übersicht der Ergebnisse 15°C NMC 2, 3, 4- Train-Test 80/20 ohne Lade- und Entladeenergie 39

Abkürzungsverzeichnis

BMS	Batteriemanagementsystem
BoL	Beginn of Life
CALCE	Center for Advanced Life Cycle Engineering
CNN	Convolutional Neural Networks
DECDV	Duration of equal charging voltage difference
DEKF	Dual Extended Kalman Filter
DoD	Depth of Discharge
DV	Differentiellen Spannung
EIS	Electrochemical Impedance Spectroscopy
EKF	Extended Kalman Filter
EoL	End of Life
FNN	Feedfoward Neural Networks
IC	Inkrementellen Kapazität
KF	Kalman Filter
LCO	Lithium-Kobalt-Oxid
LIB	Lithium-Ionen-Batterien
LFP	Lithium-Ferrophosphat
ML	Maschinelles Lernen (engl. Machine Learning)
MAE	Mittlere absolute Fehler
MSE	Mittlere quadratische Fehler
RMSE	Mittlere quadratische Wurzelfehler
RSS	Residuenquadratsumme
NMC	Lithium-Nickel-Mangan-Cobalt-Oxide
NCA	Lithium-Nickel-Cobalt-Aluminium-Oxide
NN	Neuronales Netz (engl. Neural Networks)
OCV	Leerlaufspannung (engl. Open-Circuit Voltage)
SEI	Festelektrolyt-Zwischenphase (engl. Solid Electrolyte Interphase)
SNL	Sandia National Labs

SoC	State of Charge
SoF	State of Funktion
SoH	State of Health
SVM	Support-Vektor-Maschine engl. Support Vector Machine

1 Einleitung

Lithium-Ionen-Batterien sind fundamentale Komponenten moderner Technologien. Aufgrund ihrer hohen Energie- und Leistungsdichte sowie ihrer hohen Zellspannung ist sie annähernd die einzige Technologieklasse, die in Elektrofahrzeugen von Bedeutung ist [1]. Darüber hinaus stehen die Weiterentwicklung und Optimierung von Batteriesysteme sowohl in der Forschung als auch in der Industrie im Trend. Um die Ausnutzung der Batteriekapazität zu optimieren und die sicherheitsrelevanten Funktionen sicherzustellen, ist die Bestimmung der Zustandsgrößen (State of Charge SoC, State of Health SoH) von essenzieller Bedeutung. Der SoC beschreibt den momentanen Ladezustand im Vergleich zur maximalen Kapazität der Batterie und korreliert mit der Klemmspannung; der SoH beschreibt den aktuellen Zustand der Batterie im Vergleich zu einer neuen Batterie [2].

Das Entladungsverhalten und die zeitliche Veränderung der Zustandsgrößen werden in der Batterieentwicklung mittels zeit- und kostenaufwändiger elektrochemischer Lebensdauertests in einer kontrollierten Umgebung analysiert. Diese Art von experimentellen Verfahren erfordert ein hohes Maß an Systemwissen (Batteriebetrieb und Alterungsmechanismen). Die Ergebnisse von Echtzeitmessungen sind Input für die Parametrisierung von Batteriemodellen (z. B. RC-Modellen), die bei der Überwachung eines Batteriesystems zusammen mit Computeralgorithmen zur Schätzung des SoH-Werts eingesetzt werden. Die Kombination von parametrisierten Messergebnissen mit Batteriemodellen stellt einen Grey-Box-Lösungsansatz dar und wird häufig in der Praxis verwendet. Mit der zunehmenden Bedeutung von datengesteuerten Analysen wird die Anwendung von Algorithmen des maschinellen Lernens für die SoH Bestimmung sowohl in der Forschung als auch in der praktischen Anwendung untersucht. Ein rein datengetriebener Ansatz wird in der Literatur als Black-Box-Modell bezeichnet.

1.1 Motivation

Die SoH Black-Box Bestimmung könnte aufgrund ihrer Flexibilität und der Anwendung von vorhandenen Batteriedaten mögliche externe Faktoren in die Zustandsgrößenbestimmung aufnehmen und eine realitätsgetreue Schätzung ohne tatsächlichen vorherigen vertieften Wissen über einzelne Alterungsmechanismen liefern. Die Relevanz der Untersuchung und der ständigen Überwachung von Batterieparameter rechtfertigt sich, da die Batterie eine der teuersten Komponenten eines elektrisch betriebenen Gerätes betriebene Gerät ist und die bekannte White und Grey-Box verfahren zeit- und kostenintensiv sind. Abhängig von der genauen Schätzung des SoH ist es möglich, die Alterungsgeschwindigkeit der Batterie zu erkennen und anhand dieser zu entscheiden, ob die Batterie ersetzt werden sollte oder weiterverwendet werden kann [3]. Die bisher untersuchten Publikationen mit Black-Box Methoden (siehe Kapitel 2.3.2) liefern zufriedenstellende Ergebnisse in Bezug auf die Schätzgenauigkeit, wurden aber mit Daten desselben Batterietyps unter konstanten Betriebsbedingungen betrachtet. Ein umfassender Methodenvergleich unter Berücksichtigung unterschiedlicher Lithium-Ionen-Akkumulatoren ist bisher nicht veröffentlicht worden.

1.2 Zielsetzung

Im Rahmen der Masterarbeit sollen die SoH (Prozentwerte) der verwendeten Untersuchungsobjekte (Lithium-Ionen-Batterien) über die Anzahl der Entladezyklen (Zeitachse) auf Basis zuvor erfasster Daten (Spannung, Umgebungstemperatur, Ladeleistung, Entladeleistung, Anzahl von Lade- und Entladezyklen, Dauer eines Zyklus und Kapazität) mittels datengesteuerter Methoden und Algorithmen des maschinellen Lernens ermittelt und dargestellt werden. Dafür werden Datensätze aus drei verschiedenen Batterien aus unterschiedlichen Kathodenmaterialien für das Trainieren und Testen des Modells verwendet. Der Einfluss der jeweiligen Daten auf die Bestimmung der Zustandsgrößen wird untersucht und die relevanten Performancemerkmale (z. B. Genauigkeit, Rechenzeit) werden ausgewertet, miteinander verglichen und dargestellt. Aus den Zielsetzungen wurden folgende Forschungsfragen abgeleitet:

1. Wie gut können Algorithmen des maschinellen Lernens die SoH Verläufe von Lihtium-Ionen-Batterien mit unterschiedlichen Kathodenmaterialien und Einsatzbedingungen schätzen?
2. Welche Algorithmen des maschinellen Lernens rechtfertigen auf Basis ihrer Genauigkeit und ihrer Komplexität weitere Untersuchungen für eine praktische Anwendung?

2 Stand von Wissenschaft und Technik

In dem folgenden Kapiteln werden grundlegende Terminologien und Eigenschaften von Lithium-Ionen-Batterien (LIB) zum Verständnis dieser Arbeit erklärt. Anschließend werden Vorarbeiten hinsichtlich der Kernpunkte dieser Arbeit – eine datengesteuerte SoH-Bestimmung – eingeführt und evaluiert.

2.1 Aufbau und Funktionsprinzip der Lithium Ionen Batterien und Akkumulatoren

Lithium-Ionen-Batterien sind elektrochemische Energiespeicher auf Lithiumbasis, die verschiedene Aktoren mit Strom versorgen können. Batterien werden in Primär- und Sekundärbatterien unterschieden. Bei einer Primärbatterie ist die Umwandlung von chemischer Energie in elektrische Energie irreversibel. Der Begriff Akkumulator definiert, dass es sich um eine wiederaufladbare Batterie (Sekundärbatterie) handelt, bei der die interne chemische Reaktion umkehrbar ist. Lithium-Ionen-Zellen bestehen grundsätzlich aus zwei Elektroden, einer negativen (Anode) und einer positiven (Kathode)[4]. Zwischen der Anode und Kathode befinden sich ein Separator und ein ionenleitender Elektrolyt, der für den Ladungsaustausch zuständig ist. Der Separator ist eine poröse Barriere, die zur elektrischen Isolierung beider Elektroden dient und somit interne Kurzschlüsse vermeidet. Die einzelnen Lithium-Ionen bewegen sich beim Entladen und Laden zwischen den Elektroden hin und her. Gleichzeitig fließen die Elektronen als Elektrizitätsträger durch den äußeren Stromkreis.

Beim Laden wird das Aktivmaterial der positiven Elektrode oxidiert, wodurch Lithium-Ionen aus dem Wirtsgitter entfernt werden und über die Elektrolyte zur negativen Elektrode wandern. An der negativen Elektrode findet eine Reduktion statt. Dabei gehen Lithium-Ionen aus den Elektrolyten in das Aktivmaterial (Grafit) über und werden dort eingelagert [5]. Der Prozess läuft bei der Entladung in umgekehrter Richtung ab. Die Lithiumelektroden sind von einem Film aus Lithiumchlorid bedeckt, der als Festelektrolyt-Zwischenphase (engl. Solid electrolyte interphase, SEI) und als Schutzschicht dient. Diese verhindert

eine unerwünschte Zersetzung des Elektrolyten [6]. Die Anzahl der dabei freigesetzten oder aufgenommenen Elektronen pro Masse bzw. pro Volumeneinheit bestimmt die Speicherfähigkeit des aktiven Elektrodenmaterials und wird als spezifische Ladung (Ah/kg) angegeben. Je mehr Lithium-Ionen in dem Material gespeichert werden können, desto höher ist die Kapazität des Akkumulators. In Abbildung 2.1 ist der Aufbau einer zylindrischen Lithium-Ionen-Zelle dargestellt.

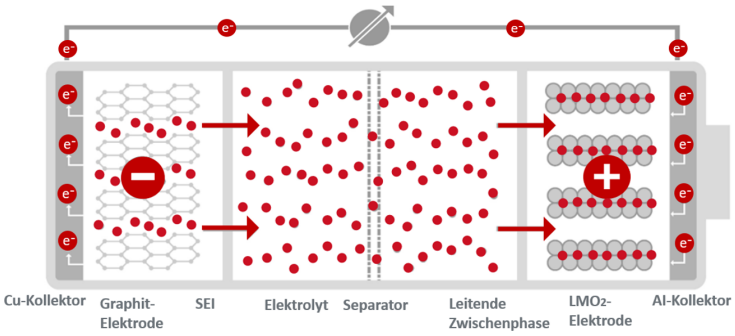


Abbildung 2.1: Aufbau einer Lithium-Ionen-Zelle
Quelle: eigene Ausarbeitung nach [4]

Anoden- und Kathodenmaterialien sind mit unterschiedlichen Elektrolyten kombiniert, um bestimmte Eigenschaften wie thermische Stabilität, gute elektrische/ionische Leitfähigkeit und Strukturstabilität zu betonen. Kobalt-Oxid, kurz LCO ($LiCoO_2$), war das erste kommerziell erhältliche Kathodenmaterial für Batterien in der Elektromobilität. Neue Kathodenmaterialien wie Nickel-Mangan-Kobaltoxid (NMC) ($LiNi_{0.33}Mn_{0.33}Co_{0.33}O_2$) haben die ursprünglichen Technologien aufgrund ihrer hohen Spannung und spezifischen Kapazität (maximale nutzbare spezifische Kapazität) ersetzt. Die Tabelle 2.1 stellt einen Vergleich der spezifischen Ladung von unterschiedlichen Kathodenmaterialien dar.

Tabelle 2.1: Spezifische Ladung im Vergleich

Kathodenmaterial	**spez. Kapazität [Ah/kg]**	**Spannung[V]**
$LiMn_2O_4$ (LMO)	100-120 Ah/kg	4.1
$LiCoO_2$ (LCO)	130-150 Ah/kg	3.9
$LiFePO_4$ (LEP)	150-160 Ah/kg	3.4
$LiNi_{0.33}Mn_{0.33}Co_{0.33}O_2$ (NMC)	150-190 Ah/kg	3.7

Quelle: eigene Ausarbeitung nach [1].

2.2 Batteriesysteme

Lithium-Ionen-Batterien bestehen aus zusammengeschalteten Zellen, der zugehörigen Elektronik und dem umgebenden Batteriegehäuse. Die Zellen unterscheiden sich zwischen drei Bauformen: zylindrisch, prismatisch und Laminatzellen (Pouch-Zellen). Ein Modul besteht aus mehreren in Reihe geschalteten Zellen, die dann parallel zueinander geschaltet werden. Eine Parallelschaltung der Batteriemodule bewirkt eine Erhöhung der Ladungsmenge (Ah) der Batterie bei gleichbleibender Spannung sowie eine Erhöhung des maximalen Lade- und Entladestroms. Ein Batteriesystem besteht aus mehreren parallel geschalteten Modulen, dem Peripheriegerät (inklusive Kühlkreis) und einem Batteriemanagementsystem BMS. Die Kapazität eines Batteriesystems ergibt sich aus der Anzahl von verschalteter Module. Die Abbildung 2.2 zeigt den Aufbau eines Batteriepaketes.

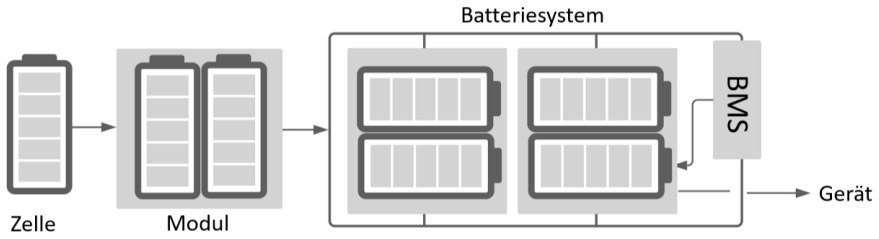


Abbildung 2.2: Lithium-Ionen-Batteriesystem
Quelle: eigene Ausarbeitung

Die primäre Aufgabe des BMS ist es, die Einzelzellen einer Batterie zu überwachen und die Lebenszeit sowie die Zyklenzahl zu erhöhen [1]. Besonders bei der Lithium-Ionen-Technologie in Speichersystemen ist der Schutz vor Überladung, Tiefentladung und Übertemperatur der Batteriezellen erforderlich. Dabei steuert das BMS den Lastausgleich und die Ladung der Batterie vom Stromnetz. Die Entladungsspannung gibt den Spannungswert in Volt an, bei dem die Batterieentladung beendet werden soll. Eine starke Entladung bzw. Tiefentladen bedeutet, dass das Entladeverfahren über die Entladungsspannung hinaus erfolgte. Das Tiefentladen führt zu irreversiblen Schäden sowie Kapazitätsverlust oder erhöhter Selbstentladung. Es kann die Lebensdauer einer Batterie reduzieren. Bei Batteriesystemen aus mehreren Zellen besteht das zusätzliche Risiko, dass einzelne Zellen während eines Ladevorgangs den Wert der Spannungsgrenze beim Laden (Ladeschlussspannung) früher als andere erreichen. Dabei können die Zellen zu einer Überspannung geführt werden oder der Ladevorgang kann vorzeitig abgebrochen werden [1]. Um diese Phänomene zu vermeiden und mögliche Betriebsstörungen zu überwachen, wird der Ladezustand sowie der Alterungsstand in der Steuerung des BMS berechnet.

2.2.1 Kennwerten und Zustandsgrößen

Die Tabelle 2.2 gibt einen Überblick über die wichtigsten Kennwerte und Begriffe von Batteriesystemen.

Tabelle 2.2: Begriffe und Kennwerte von Batteriesysteme

Kennwerten und Zustandsgrößen	**Erklärung**
Kapazität in Ah	Dient als Größe zur Angabe der gespeicherte Ladungsmenge beim Laden und Entladen einer Batterie verwendet berechnet sich durch das Produkt von Stromstärke I mit der Zeit t: $$K = I \cdot t \quad (2.1)$$
C-Rate (C) in $1/h$	Gibt die Stromstärke bezogen auf die Nennkapazität an. Es ist ein Maß für die Lade- und Entladegeschwindigkeiten. 1C entspricht dem Laden mit einem Ladestrom, der die Batterie in einer Stunde vollständig auf- bzw. entlädt. $$xC = x\frac{I}{K} = x\frac{1}{t} \quad (2.2)$$
Innenwiderstand in Ω	Setzt sich aus zwei Widerstände zusammen. Der elektrische Widerstand ist ein Maß für die Behinderung des Stromflusses in einem Schaltkreis. Der ionische Widerstand ist für die Begrenzung des Stromflusses aufgrund internen Faktoren zuständig. Z.B: Elektrolytleitfähigkeit.
Leerlaufspannung in V	Stationäre Spannung an den Batterieklemmen ohne Stromfluss (Leerlauf englisch *open-circuit*).
Klemmspannung in V	Wert der Spannung zwischen dem Pluspol und dem Minuspol einer Spannungsquelle bei geschlossenem Stromkreis. Die Klemmenspannung ist aufgrund des Innenwiderstands der elektrischen Quelle immer kleiner als die Leerlaufspannung.

Quelle: eigene Ausarbeitung nach [1][7][8].

Der Ladezustand (englisch *state of charge* abgekürzt SoC), gibt den momentanen Ladezustand K_t im Vergleich zur maximalen gespeicherten Kapazität K_0 der Batterie an (siehe Gleichung 2.3). Dieser Zustandswert beschreibt durch den prozentualen Wert die voraussichtliche Betriebsverfügbarkeit des Geräts bis es neu geladen werden muss. Es besagt wie viel Energie in einer Batterie verbleibt.

$$SoC = \frac{K_t}{K_0} \cdot 100\% \tag{2.3}$$

Der Alterungsstand (englisch *state of health* abgekürzt SoH) dient als Größe zur Beurteilung des Alterungsprozesses einer Batterie. Eine vollständig geladene Batterie hat eine Nennkapazität K, die von der maximale Kapazität K_{max} abhängig von der Nutzungsintensität und der Nutzungsdauer abnimmt. Das Verhältnis der derzeitigen Nennkapazität K zur maximalen Kapazität am Anfang wird für die Bewertung des SoH, der Batterie verwendet (siehe Gleichung 2.4).

$$SoH = \frac{K}{K_{max}} \cdot 100\% \tag{2.4}$$

2.2.2 Lebensdauer und Alterungsmechanismen

Obwohl eine lange Lebenszeit für große Batteriesysteme von großer Wichtigkeit ist, kann das Batteriesystem zum Ende seiner Lebensdauer aufgrund ungenügender Energie- oder Stromkapazität die Anforderungen des Fahrzeugs nicht mehr erfüllen. Die Lebensdauer einer LIB ist als Zeitspanne zwischen dem Auslieferungszeitpunkt (englisch *begin of life* abgekürzt BoL) und dem Zeitpunkt (englisch *end of life* abgekürzt EoL) definiert, zu dem die zuvor definierten Werte einer Batterie durch Alterung unterschritten werden [4]. In der Regel erreicht eine Batterie in Elektroautos und Großgeräten ihr EoL, wenn die nutzbare Kapazität auf 80 % des Nennwerts (SoH = 80%) gesunken ist.

Die Verkürzung der Lebensdauer von Batterien ist ein unvermeidlicher Prozess, der durch die Alterung der einzelnen Zellen verursacht wird. Er unterscheidet sich zwischen *zyklischer* und *kalendarischer Alterung*. Unter zyklischer Alterung versteht man die Alterung durch Lade- und Entladevorgänge. Diese führt z.B zur Erhöhung der SEI über wiederholte Lade- und Entladezyklen. Eine übermäßige SEI-Schicht ist nicht erwünscht, da sie die Leistung der Zelle verringert, die Diffusion kinetisch blockiert und die chemische Reaktion verhindert [6].

Der Begriff kalendarische Lebensdauer nimmt Bezug auf die Batterie, wenn sie sich im unbenutzten Zustand (z. B. Lagerung und Transport) befindet, d.

h. wenn sie nicht entladen und geladen wird. Auch wenn keine Zyklen gefahren werden, findet in der Zelle aufgrund der Wechselwirkung zwischen Elektrolyt und Lithium-Ionen ein elektrochemischer Degradationsprozess statt. [4].

Die Alterungsgeschwindigkeit ist von der Nutzung und Lagerung der Batterie abhängig. Verschiedene Faktoren, wie hohe Lade- und Entladeraten, Tiefentladungen oder extreme Ladezustände können den Kapazitätsverlust beschleunigen. Obwohl das BMS für einen sicheren Betrieb und die Vermeidung von außergewöhnlichen Ladungen sorgt, wird die Ladestrategie vom Verbraucher bzw. vom Kunden je nach Bedarf oder Empfehlung des Herstellers definiert. Das heißt, dass bspw. eine intensive Nutzung der Batterie ihre Lebensdauer verkürzen kann. Die Temperatur in der Batterie wird durch die Umgebung beeinflusst und steigt durch Zyklisieren der Zelle. Eine erhöhte Temperatur bedeutet eine erhöhte Zellspannung, die für die Zersetzung aller Zellkomponenten wie den Elektroden und dem Elektrolyten verantwortlich ist [4]. Je wärmer eine Batterie ist, desto schneller altert sie. Die Lebensdauer von Batteriezellen hängt vom Ladeumsatz pro Zyklus (englisch *depth-of-discharge* abgekürzt DoD) ab. Werden regelmäßig komplette Vollzyklen (100% DoD, d.h. SoC von 0 auf 100%) gefahren, reduziert sich der insgesamt mögliche Ladeumsatz um etwa 20 % gegenüber dem Regelbetrieb mit 80% DoD-Zyklen (d.h. Ladezustand immer zwischen 10 % und 90 %)[7][9]. Die Art, wie Lithiumionen-Batterien gelagert sind, beeinflusst die Alterungsgeschwindigkeit. Eine vollgeladene Lithiumionen-Batterie verliert etwa 6% der Kapazität bei 0 °C, 20% bei 25 °C, 35% bei 40 °C und 40% bei 60 °C pro Jahr. Bei Lagerung mit halber Ladung (50% SoC) ist der Kapazitätsverlust 2%, 4%, 15% bzw. 25% pro Jahr bei 0, 25, 40 und 60 °C [8].

Hauptursache für kalendarische Alterung ist die temperaturabhängige, stetige Zersetzung der Elektrolytkomponenten. Die mechanische Schädigung der Anode und Kathode sorgt für die Verkürzung der Lebensdauer bei Zyklenalterung. In der Praxis treten beide Phänomene parallel auf und sind nur schwierig voneinander zu trennen.

2.3 Methoden zur Bestimmung des Alterungsstands

Grundsätzlich sind SoH-Bestimmungsmethoden in drei Kategorien eingeteilt: experimentelle Methoden, modellbasierte und datengesteuerte Methoden. In der angewandten Mathematik und im modernen Softwareengineering beschreiben die Konzepte von White-, Black- und Grey-Box-Methoden unterschiedliche Methodologien zur Berechnung und Validierung von Parametern und Systemen. Die zuvor erwähnten SoH-Bestimmungsmethoden und ihre Eigenschaften lassen sich anhand des Box-Konzepts verstehen. Die Abbildung 2.3 stellt eine Übersicht der SoH-Bestimmungsmethoden im Zusammenhang mit dem Box-Konzept dar.

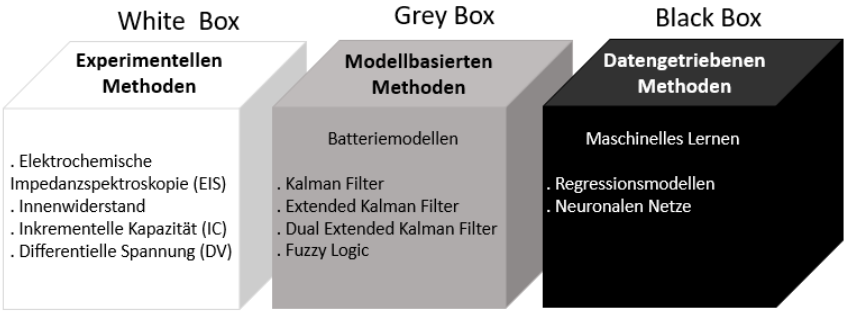


Abbildung 2.3: SoH-Bestimmung im Zusammenhang mit dem Box-Konzept
Quelle: eigene Ausarbeitung

2.3.1 White-Box-Methode

Ein White-Box-Modell ist ein Modell, bei dem die Struktur auf der Grundlage von Vorkenntnissen und physikalischen Erkenntnissen entwickelt wurde [10]. Bei White-Box-Modellen ist die Ursache-Wirkungs-Beziehung zwischen Variablen bekannt. Eine verbreitete experimentelle Methode der SoH-Bestimmung ist die elektrochemische Impedanzspektroskopie (englisch *electrochemical impedance spectroscopy* abgekürzt EIS). Dabei wird die Impedanz eines Systems in Abhängigkeit von der Frequenz des Wechselspannungspotentials gemessen [11]. Es ist erwiesen, dass die innere Impedanz einer Batterie mit der Nutzungsdauer zunimmt. Sie wird deshalb als Batterie-SoH-Indikator betrachtet. [12].

Der SoH kann zusätzlich durch den Innenwiderstand bestimmt werden. R_n ist der anfängliche Innenwiderstand der Batterie, R_c ist der aktuelle Innenwiderstand, und R_a ist der Innenwiderstand einer gealterten Batterie am Ende der Lebensdauer (siehe Gleichung 2.5).

Obwohl der Widerstand ein wichtiger Parameter für die SoH-Bestimmung ist, da die verfügbare Kapazität und Leistung der Batterie stark mit dem Innenwiderstand zusammenhängen, benötigt dieses Verfahren einen Referenzwert R_a von einem Widerstand am Ende der Batterielebensdauer [13].

$$SoH_R = \left| \frac{R_c - R_a}{R_n - R_a} \right| \cdot 100\% \tag{2.5}$$

Anhand der Leerlaufkurve über der Kapazität (OCV-Verlauf) können die Kurvenverläufe der inkrementellen Kapazität (IC) und der differentiellen Spannung (DV) ermittelt werden. Diese Methoden basieren auf der Aufnahme der Leerlaufspannung unter konstantem Gleichstrom, da somit eine konstante Ladungsänderung auftritt [14].

Die inkrementellen Kapazität Kurve (IC-Kurve) entspricht der Kapazitätsänderung über der Zellspannung und wird mit Hilfe der Gleichung 2.6 ermittelt [15].

$$IC = \frac{\partial K}{\partial V} = \frac{I.\partial t}{\partial V} \tag{2.6}$$

wobei t ist die Ladezeit, I ist der Ladestrom und V die OCV des Systems. Die Abbildung 2.4 a) stellt die IC-Kurve einer NCA-Batterie dar.

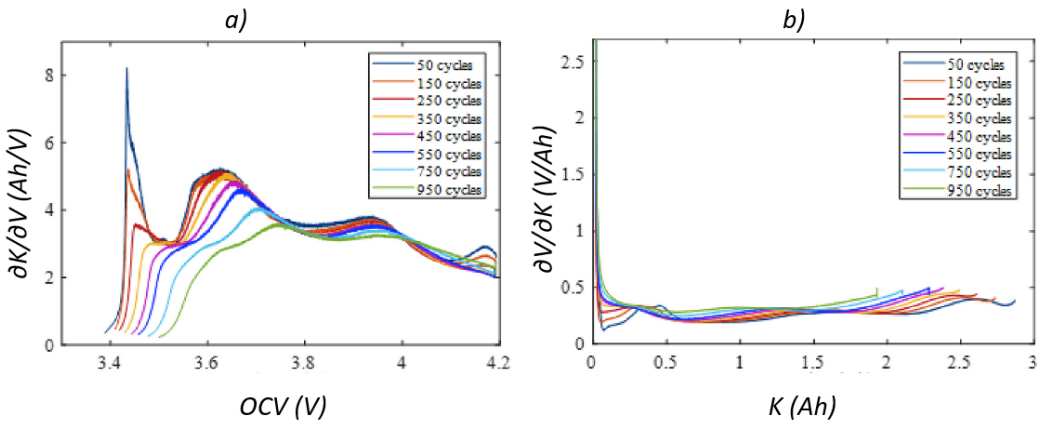


Abbildung 2.4: (a) IC und (b) DV-Kurven einer NCA-Batterie
Quelle:[16]

In ähnlicher Weise wird die differentielle Spannung durch Differenzierung der Batteriespannung in Bezug auf die Kapazität ermittelt (siehe 2.7). Die DV-Kurve wird zur Identifizierung der Alterungsreaktionen innerhalb einer Batterie verwendet [16].

$$DV = \frac{\partial V}{\partial K} = \frac{\partial V}{I.\partial t} \tag{2.7}$$

Die Alterungseffekte, die sich in der DV-Kurve widerspiegeln, stimmen mit denen der IC-Kurve überein. Daher werden die (IC)- und die (DV)-Methode zusammen verwendet, um sich gegenseitig zu überprüfen [16]. Die Abbildung 2.4 b) auf Seite 10 stellt die DV-Kurve einer NCABatterie dar.

Die "Coulomb-Counting"Methode wird verwendet, um die verbleibende Kapazität einer Batterie durch die Bestimmung des SoC sowie kumulierter Werte über mehrere Lade- und Entladevorgänge zu ermitteln. Die SoC Berechnung lässt sich durch die Integration des einer Batterie entnommenen und zugeführten Stroms über die Zeit realisieren (siehe Gleichung 2.8).

$$SoC(t) = SoC(t_0) - \int_t^{t_0} \frac{I}{C}.dt \tag{2.8}$$

Die Batteriezellen werden meistens mit niedrigen Strömen (z.B C-Rate $\leq$ 0.1C) in einer kontrollierten Umgebung mit konstanter Temperatur vollständig geladen und entladen. Nach jedem Zyklus wird der Kapazitätswert beim $SoC_(t)$ als Referenz für die maximale verfügbare Kapazität K benutzt, um mithilfe von 2.4 zusammen mit dem bekannten Wert von K_{max} den SoH Wert zu ermitteln. Es ist zu beachten, dass die "Coulomb-Counting"Methode ihre Grenzen hat. Erstens können sich ungenaue Strommessungen zu großen Fehlern summieren, und andere Faktoren, wie unterschiedliche Temperaturen und Zykluszeit, die ebenfalls die Genauigkeit der Zustandsbestimmung beeinflussen, werden nicht berücksichtigt [17].

2.3.2 Black-Box-Methode

Black-Box-Methoden betrachten die Entwicklung eines Modells, für das keine physikalischen Erkenntnisse verwendet werden oder verfügbar sind, aber die gewählte Modellstruktur wird aus einer bekannten Modellfamilie ausgewählt [10]. Das Ziel eines Black-Box-Modells ist die Vorhersage des Outputs basierend auf der Eingabekombination. Datengesteuerte Methoden sagen den SoH einer Batterie mithilfe statistischer oder maschineller Lernmodelle ohne elektrochemische Kenntnisse über den inneren Aufbau und die Alterungsmechanismen voraus. Die Support-Vektor-Maschine (englisch *support vector machine*, SVM), eine der datengestützten Methoden, wurde von Patil [18] untersucht, um den SOH-Wert mit Batteriezyklusdaten zu schätzen. Es handelt sich um eine Klassifikations- und Regressionsmethode des überwachten Lernalgorithmus, bei der diskrete Werte vorhergesagt werden können. Die Grundidee hinter SVM ist, die Lösungen als gewichtete Summe von Support-Vektoren, die nur eine Teilmenge der Trainingsinput sind, zu konstruieren. Shen [19] hat der SOH-Wert anhand der nutzbaren Kapazität einer Batterie mithilfe eines faltendes neuronales Netzwerk (englisch *convolutional neural network* abgekürzt CNN) eine Art der tiefen neuronalen Netzes, mit Strom- und Spannungsmessdaten geschätzt. Die vorgeschlagene neuronale Netzwerk Methode (englisch *neural network*) bietet eine höhere Genauigkeit und Robustheit bei der Kapazitätsschätzung von Lithium-Ionen-Batterien. Ähnliche Ansätze sowie Feed Forward Network (FNN) wurde von Xia und Abu Qahouq [20] für eine adaptive SOH-Schätzung untersucht. Ein standardmäßiges neuronales Netz (NN) besteht aus mehreren einfachen, miteinander verbundenen Neuronen und jeweils einer Folge von reell-wertigen Aktivierungen. Beim Lernen geht es darum, Gewichte zu finden, die das NN dazu bringen, das gewünschte Verhalten zu zeigen [21]. Je nach Problemkomplexität, werden mehreren Prozessoren benötigt, um die Algorithmen zu konfigurieren. Im Forschungsgebiet des maschinellen Lernens finden sich weitere Möglichkeiten, die für die Lösung der Herausforderungen im Rahmen der Masterarbeit zum Einsatz kommen können. In der Literaturrecherche werden Alternativen geprüft.

2.3.3 Grey-Box-Methode

Eine Grey-Box-Methode ist eine Kombination aus Black-Box- und White-Box-Methoden [22]. Der Ansatz eines Grey-Box-Modells besteht darin, Eigenschaften von Black-Box-Modellen (Daten) und White-Box-Modellen (physikalischen Erkenntnissen) zu nutzen, um mit den Vorteilen von beiden ein effizienteres zusammengesetztes Modell zu erstellen. Der SoH gibt die Degradation der Batterie durch Alterungsindikatoren (z. B. Kapazitätsverlust und Innenwi-

derstandserhöhung) an. Wenn diese Informationen nicht bekannt sind, werden Batteriemodelle verwendet, um das Batterieverhalten zu beschreiben und somit die Alterungsindikatoren zu identifizieren. Bei Batteriemodellen wird in der Regel eine Spannungsquelle,die die Ruhespannung der Batterie wiedergibt, mit Widerständen, Induktivitäten und Kapazitäten kombiniert, die zusätzlich das dynamische Batterieverhalten nachbilden [23]. Im Allgemeinen gilt, dass je komplexer das Modell ist, desto höher ist seine Genauigkeit, andererseits nimmt die Geschwindigkeit der Berechnung mit der Komplexität des Modells ab. Das Innenwiderstandsmodell besteht aus einem Widerstand aus einer Leerlaufspannungsquelle (OCV) und einem in Reihe geschalteten Widerstand. Die Umsetzung des Modells ist einfach, berücksichtigt aber nicht die schlechten Auswirkungen des elektrochemischen Degradationsprozess (z. B. die Erhöhung der Stromdichte), die im Inneren der Batterie auftreten [24].RC-Modelle der ersten und zweiten Ordnung werden eingesetzt, um die subtilen Variationen des Lade- und Entladeverhaltens von Batterien zu charakterisieren [25].

Die Abbildung 2.5 stellt unterschiedlichen Ersatzschaltbildmodell einer Batterie hinsichtlich ihrer Komplexität dar.

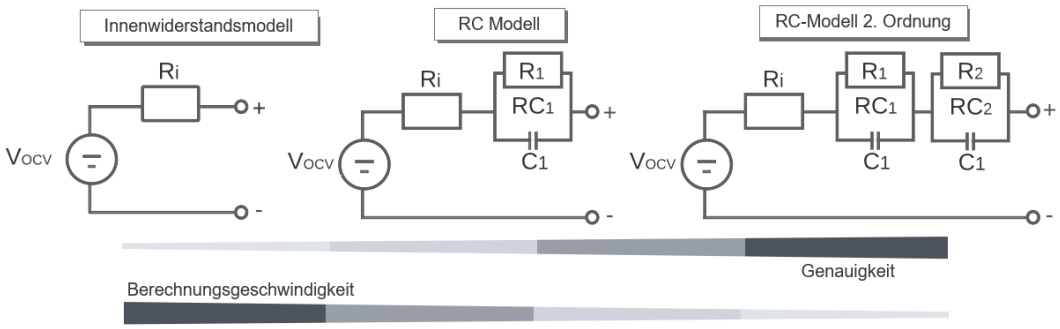


Abbildung 2.5: Batteriemodelle hinsichtlich ihrer Komplexität
Quelle: eigene Ausarbeitung nach [23]

Die Zustandsgrößen von Lithium-Ionen Batterien werden mithilfe von fortgeschrittenen Filtertechniken geschätzt. Der Kalman-Filter (KF) ist eine weit verbreitete Schätzungsmethode und wird gern für die Schätzung von Modellparametern und allgemeinen Zuständen verwendet [25].

Der Kalman-Filter ist im Wesentlichen eine Reihe von mathematischen Gleichungen, die als ein effizientes rechnerisches Mittel zur Schätzung des Zustands eines Prozesses dient und dabei die geschätzte Fehlerkovarianz (Varianz des geschätzten Wertes) minimiert [26] [27]. Zur Bestimmung des SoH werden mehrere RC-Batteriemodelle der zweiten Ordnung eingesetzt, um das dynamische Batterieverhalten beim Laden und Entladen als Prozess abzubilden und daraus die Zustandsgrößen zu schätzen. Da die Batterie ein nichtlineares System ist, wur-

den Variationen der Kalman-Filter erprobt, um die Kapazität der Batterie zu schätzen. Der Erweiterte Kalman-Filter (EKF) liefert genauere Ergebnisse, da er nichtlineare Systeme in jedem Zustand linearisiert und zusätzlich das Rauschen in der Messung und im Prozess kompensiert [28]. Der zweifach erweiterte Kalman-Filter (englisch *dual extended Kalman-filter* abgekürzt DEKF) iteriert zwischen zwei EKF. Der erste EKF wird für die Schätzung der Systemzustände verwendet, der zweite bestimmt die Parameter [29]. Weitere Varianten von Kalman-Filtern wurden angewandt und liefern bedeutende Ergebnisse. Im Allgemeinen werden Kalman-Filter-Varianten für die Bestimmung der Zustände in Echtzeit verwendet.

Weiterhin wurden andere Kombinationen von experimentellen Methoden und Algorithmen zur Schätzung von Zustandsvariablen verwendet. Die Fuzzylogik ist eine Problemlösungsmethode, die auf der Grundlage mehrdeutiger, ungenauer, verrauschter oder fehlender Eingangsinformationen zu einer eindeutigen Schlussfolgerung gelangt [30]. Die Fuzzylogik kann im Gegensatz zu einer booleschen Logik (falsch = 0, wahr = 1) mehrere Möglichkeiten einer Variablen zwischen den Wahrheitswerten zu schätzen. Diese Werte repräsentieren den Grad der Zugehörigkeit der Variablen zu einer bestimmten Kategorie [30]. In der SoH-Bestimmung verwendet Fuzzylogik den Strom, die Leerlaufspannung, die Temperatur, den SoC und die Anzahl der Zyklen der Batterie, um ihre Kapazität zu schätzen. In [31] wurde aus der EIS Modellierung mithilfe den Messergebnissen (Impedanzwerte und Phasenverschiebung) von neuen und gealterten Zellen als Input eine Fuzzy-Logik verwendet, um die Zykluszahl der Batterie entsprechend SoH zu schätzen.

Algorithmen des maschinellen Lernens können mit vorherigem Wissen über Batteriesysteme kombiniert werden, um ein Grey-Box Vorhersagemodell zu entwickeln. Die Dauer des Intervalls gleicher Klemmspannung während des Entladevorgangs (DECDV) einer Lithium-Ionen-Batterie wird als Gesundheitsindikator zur Messung des Kapazitätsverlusts verwendet [32] und ist wie in Gleichung 2.9 definiert.

$$DECVD(V_{min}, V_{max}) = t_{V_{max}} - t_{V_{min}} \tag{2.9}$$

Dabei sind $t_{V_{min}}$ und $t_{V_{max}}$ die vergangenen Zeiten zwischen den Werten von V_{min} und V_{max} (minimale bzw. maximale Klemmspannung) während des Ladevorgangs. In [33] wurde ein Random Vector Functional Link Neural Network, eine Weiterentwicklung von neuronalen Netzen, verwendet, um die Korrelation zwischen DECDV und den Kapazitätswerten zu ermitteln und den Trend der Kapazitätsverlustkurve vorherzusagen.

Die Tabelle 2.3 zeigt einen quantitativen Vergleich zwischen White-, Grey- und Black-Box-Konzepten in der SoH-Bestimmung mit ihren Vor- und Nachteilen.

Tabelle 2.3: Quantitativer Vergleich der SoH-Bestimmungsmethoden

Kategorie	**Vorteile**	**Nachteile**
White-Box	Hohe Genauigkeit, geringer Rechenaufwand	Notwendigkeit vertiefter Systemkenntnisse, erfordern eine komplexe Ausrüstung für die Durchführung, zeitintensiv
Black-Box	Hohe Genauigkeit, einfache Umsetzung im Gegensatz zu Grey-Box-Methoden, erfordern keine Systemkenntnisse, es besteht noch viel Forschungspotenzial	Die Modellqualität ist stark von der Qualität der Trainingsdaten (Datenmenge, Betriebsbedingungen, Anzahl von Zyklen) abhängig, die Genauigkeit und Rechenzeit hängen vom Umfang des Modells ab
Grey-Box	Liefern relativ präzise Ergebnisse, relative einfache Umsetzung im Gegensatz zu White-Box-Methoden	Erfordern eine experimentelle Vorabvalidierung, empfindlich auf die Modellstruktur und Genauigkeit

Quelle: eigene Ausarbeitung nach [12].

Obwohl eine datenbasierte SoH-Bestimmung ein aktuelles Forschungsthema ist, wurde ein umfassender Methodenvergleich mit verschiedenen maschinellen Lernalgorithmen zur SoH-Bestimmung von Lithium-Ionen-Batterien unter diversen Randbedingungen wie extremen Temperaturunterschieden, unterschiedlichen C-Raten, Ladestrategien, Batteriezusammensetzungen und deren Auswirkung auf die Modellgenauigkeit bis zum Beginn dieser Arbeit nicht als abgeschlossen betrachtet. Aus diesem Forschungspotential ergibt sich ein Bedarf für die vorliegende Masterarbeit.

2.4 Maschinelles Lernen

Maschinelles Lernen (ML) ist eine Unterkategorie der künstlichen Intelligenz, die sich auf die Entwicklung von Modellen zur Vorhersage und Mustererkennung auf der Grundlage von Daten bezieht [34]. Die korrekte Anwendung eines Algorithmus des maschinellen Lernens hängt von der Art des zu lösenden Problems, der Anzahl der Variablen und der Art der Datensätze ab. Dabei werden die Algorithmen hauptsächlich zwischen *überwachtem Lernen* (englisch *supervised learning*) und *unüberwachtem Lernen* (englisch *unsupervised learning*) differenziert. Die Abbildung 2.6 gibt einen Überblick über die Kategorien des maschinellen Lernens und ihre Aufgabengebiete. In Rot ist der für diese Masterarbeit gewählte Algorithmentyp hervorgehoben. In den nächsten Absätzen werden die oben genannten Kategorien erklärt und die Entscheidung für Regressionsalgorithmen des überwachten Lernens begründet.

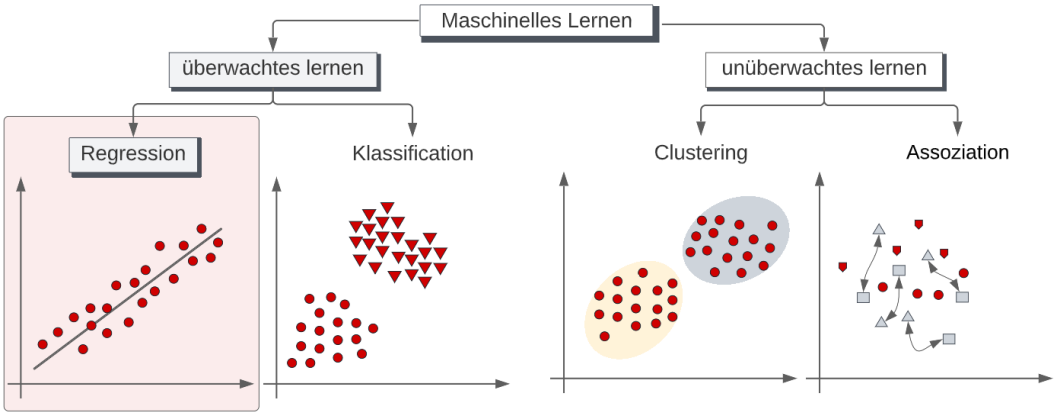


Abbildung 2.6: Übersicht der Aufgabengebiete des maschinellen Lernens
Quelle: eigene Ausarbeitung nach [35]

2.4.1 Unüberwachtes Lernen

Beim unüberwachten Lernen verwenden Algorithmen Merkmale, um aus einem Datensatz zu lernen und es wird erwartet, dass sie anhand dessen neue Datensätze mit der höchstmöglichen Genauigkeit zuordnen können [36].Dabei sind die Zielwerte nicht bekannt [37]. Diese Art von Algorithmus wird hauptsächlich für die Gruppierung (Clustering) und Assoziation von Daten genutzt. Bei Clusteringaufgaben besteht das Ziel darin, die Strukturen der Eingangsdaten selbst herauszufinden und sie eigenständig in Gruppen (Clusters) zusammenzufassen [37]. Die Assoziation beim unüberwachten Lernen ist ein regelbasierter Ansatz zur Erkennung von Kausalitätsbeziehungen (Wenn-dann-Anweisungen)

zwischen Variablen in großen Datensätzen [38]. Assoziationsregeln werden in vielen Anwendungen eingesetzt, z. B. bei der Analyse von Kundennutzungsverhalten und in der Cybersicherheit [39].

2.4.2 Überwachtes Lernen

Im Gegensatz zum unüberwachten Lernen werden beim überwachten Lernen dem Lernenden (in der Regel ein Computerprogramm oder Skript) zwei Arten von Datensätzen zur Verfügung gestellt: ein *Trainings*- und ein *Testdatensatz*. Die Idee ist, aus einer Reihe von gekennzeichneten Beispielen im Trainingssatz die Algorithmen zu trainieren, sodass die nicht gekennzeichneten Beispiele im Testdatensatz mit der höchstmöglichen Genauigkeit von den Algorithmen identifiziert werden [40]. Somit können die eingelernten Algorithmen die Zielwerte neuer Datensätze vorhersagen. Klassifizierungsaufgaben sind ein Teilbereich des überwachten maschinellen Lernens, bei dem das Hauptziel darin besteht, auf der Grundlage gelernter Daten kategorische Ausgangsbezeichnungen für die Eingabedaten vorherzusagen [41].Häufige Anwendung von Klassifikationsalgorithmen sind medizinische Diagnosen anhand von Krankheitssymptomen [41], das Filtern von Spammails und die Bildklassifizierung [34].

Bei der Regressionsanalyse geht es darum, ein Modell für die Abhängigkeit der quantitativen Zielgröße von einem oder mehreren anderen quantitativen Merkmalen (Einflussgrößen) zu erstellen [42]. Der Grundgedanke dabei ist, einen bestehenden Zusammenhang oder Beziehung zwischen den Daten und der Ergebnisvariablen zu finden [41]. Beispielhafte Anwendungen von Regressionsverfahren beim maschinellen Lernen sind die Vorhersage des Preises beim Verkauf einer Immobilie anhand marktrelevanter Faktoren [42] und die kurzfristige Leistungsprognose der generierten elektrischen Energie eines Laufkraftwerks [43].

In dieser Arbeit halten nur Algorithmen des überwachten Lernens für eine Regressionsaufgabe Einzug, da das Modell für die SoH-Bestimmung anhand vorhandener Daten trainiert und getestet wird und das Ergebnis einer reellwertigen Prognose entspricht. In den folgenden Kapiteln werden angewendete Algorithmen vorgestellt und erläutert.

3 Methodik und Umsetzung

Im Kapitel *Methodik und Umsetzung* werden die notwendigen Schritte zur Eingrenzung des Systems bis hin zum Aufbau des maschinellen Lernmodells zur Schätzung des SoH vorgestellt. Als Erstes wird die Vorgehensweise zum Erwerb und zur Aufbereitung der Daten erklärt. Dazu gehören auch die Aufteilung der Eingabedaten in Lern- und Testdaten und weitere Anpassungen, um die Genauigkeit und Stabilität des Modells zu gewährleisten, z. B. Vermeidung von Über- und Unteranpassungen (englisch *overfitting* und *underfitting*). Zunächst wird die Konstruktion und Optimierung des maschinellen Lernens-Modell (Modellbildung und Modellevaluation) schrittweise beschrieben. Anschließend werden die Parameter zur Performanceauswertung erklärt und die Grundlagen zur Bewertung des Ergebnisses und für den Vergleich zwischen den angewendeten Methoden definiert. Die Abbildung 3.1 zeigt den herkömmlichen Workflow eines überwachten maschinellen Lernmodells.

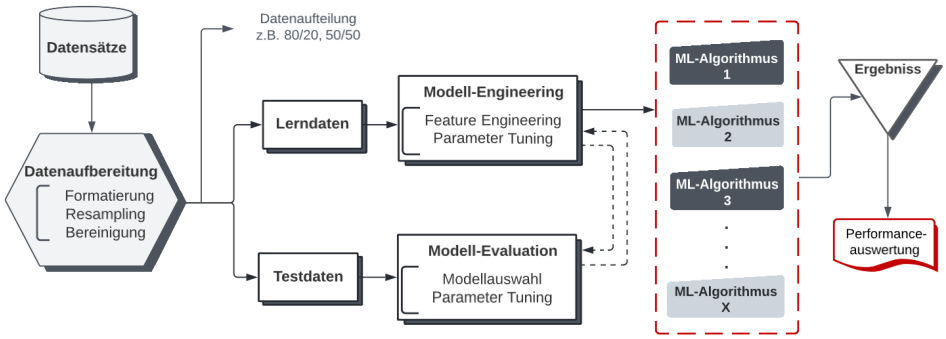


Abbildung 3.1: ML Workflow.
Quelle: eigene Ausarbeitung nach [35]

Die Untersuchungen im Rahmen dieser Masterarbeit wurden auf einem Intel(R) Core(TM) i7-10510U CPU Rechner mit 1,80 GHz Grundtaktfrequenz des Prozessors, aufgeteilt in 4 Kerne und 8 Threads, durchgeführt und mit Python Version 3.9.7 kompiliert. Die Skriptsprache Python wurde aufgrund ihrer Praxisrelevanz im Bereich Forschung von Data Science und maschinelles Lernen eingesetzt. Python wird durch umfassende Module (Bibliotheken) unterstützt, sodass viele Programmcodeteile verwendet werden können [44].

3.1 Erwerb und Vorbereitung der Daten

Im Folgenden werden der zum Vortraining des ML-Modells verwendete Quell- und Zieldatensatz vorgestellt und ihr Aufbereitungsprozess erklärt. Angewendet wurden die Informationen von Entladezyklen der unten genannten LIB aus öffentlichen Quellen. Die Batteriedatensätze wurden entweder bis zum Erreichen eines EoL-Wertes (z. B. bis 80%) oder einer bestimmten Anzahl von Zyklen aufgezeichnet. Die SNL (Sandia National Labs) Daten bestehen aus handelsüblichen 18650er NCA (Lithium-Nickel-Cobalt-Aluminium-Oxide), und NMC Zellen, die auf EoL zwischen 80% bis 60% ihrer Kapazität gebracht wurden [45]. Die Datensätze aus CALCE (englisch *Center for Advanced Life Cycle Engineering*) von der Universität Maryland betrachten die Lade- und Endladezyklen von mehreren prismatischen LCO-Zellen mit C-Raten von 0.5 und 1C, die bis zu einem EoL $\approx$ 0% geladen und entladen wurden [46]. Die Tabelle 3.1 gibt einen Überblick über die Batteriedaten.

Tabelle 3.1: Batteriedatensätze

Quelle	**Type**[a]	**Temperatur [°C]**[b]	**C-Rate [C]**	**DoD [%]**	**K [Ah]**
SNL [45]	NMC	15, 25, 35	0.5, 1, 2, 3	100	3
SNL [45]	NCA	15, 25, 35	0.5, 1, 2	100	3,2
CALCE [46]	LCO	25	0.5, 1	100	1.35

Quelle: eigene Ausarbeitung nach [45],[46].

[a] Die Abkürzungen beziehen sich auf die Kathodenmaterialien der Batterien.
[b] Umgebungstemperatur mit ± 0,5°C Toleranz.

Da die Datensätze aus unterschiedlichen Quellen erworben wurden, ist der erste Schritt, diese nach gemeinsamen Eigenschaften zu filtern. Für die Erstellung von Trainings- und Lernmodellen wurden die Umgebungstemperatur, die minimalen und maximalen Spannungs- und Stromwerte während der Entladezyklen sowie die Dauer und Anzahl der Entladezyklen erfasst. Angaben wie das Datum der Datenerfassung und Textinformationen zu rein dokumentarischen Zwecken haben in dieser Betrachtung keine Bedeutung und wurden nicht genutzt. Da die Datensätze von verschiedenen Batterien stammen und die Informationen über die ursprünglichen SoH-Werte nach der Batterieherstellung sowie die tatsächliche Lebensdauer dieser Batterien nicht bekannt sind, wird der gemessene Kapazitätswert des ersten Zyklus als maximale Kapazität verwendet. Mit diesem Wert wird die Gesamt-SoH-Kurve für den Trainingsdatensatz mit der Gleichung 2.4 berechnet. Die Voranalyse der Rohdaten weisen Messfehler in den Datensätze auf (siehe Abbildung 3.2). Die Kapazität einer

LIB nimmt mit der Zeit ab, daher sollte die Kapazität-Kurven kontinuierlich abnehmen. Messunsicherheit und Messfehler sind Faktoren, die in der Praxis auftreten können, daher wurden fehlerhafte Messpunkte nicht entfernt. Um den Einfluss dieser Faktoren zu minimieren und Ausreißer aus der SoH-Kurven zu entfernen, wurde ein Filter umgesetzt.

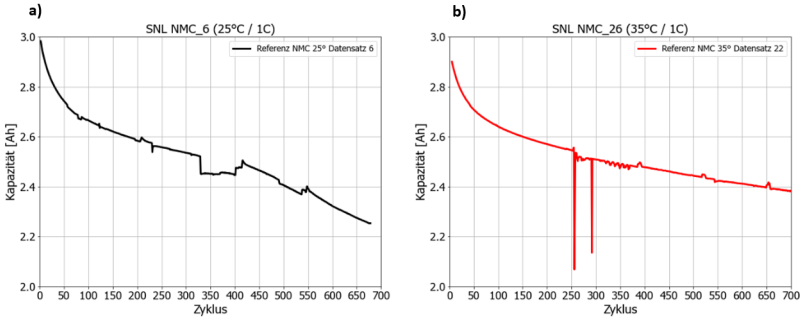


Abbildung 3.2: Beispiele aus der Voranalyse der Rohdaten NMC 25° (a) und NMC 35°C (b)
Quelle: eigene Ausarbeitung

Der Savitzy-Golay-Filter ist eine klassische Signalglättungsmethode, die auf der lokalen Annäherung des analysierten Signals beruht [47]. Dieser Filter bewahrt die Form und die Länge des Signals und entfernt gleichzeitig kleine Störgeräusche, indem lokaler Signalabtastpunkte mithilfe einer Polynomfunktion anpasst werden [48].

Die Abbildung 3.3 zeigt die Kapazitätskurven von NMC 25° (a) und NMC 35°C (b) nach Umsetzung von Savitzy-Golay-Filter. Dieser Filter wurde im Rahmen der Datenaufbereitung auf alle Rohdaten angewendet.

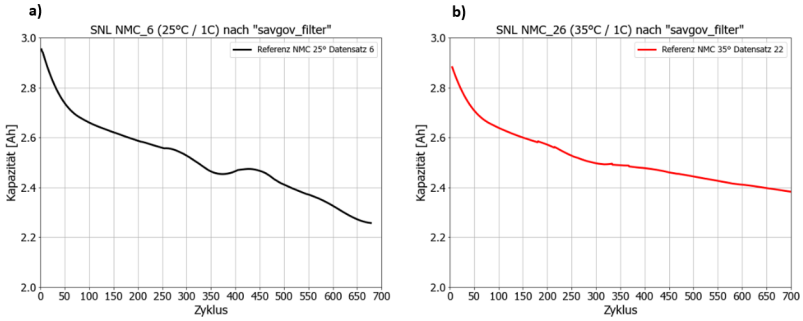


Abbildung 3.3: Beispiele aus der Voranalyse der Rohdaten NMC 25° (a) und NMC 35°C (b) nach Umsetzung von Savitzy-Golay-Filter
Quelle: eigene Ausarbeitung

Die Datensätze mit mehr als 30% Messfehlern wurden aussortiert und nicht verwendet.

3.1.1 Trainings- und Testdaten

Als Referenz für das Training wurden die Datensätze von NMC LIB aus SNL verwendet, da die folgende Bedingungen wie Temperatur von 15°C mit DoD von 100% bei C-Rate von 0.5C einer idealen Nutzung eines Akkumulators entspricht und dieser Datensatz kein Fehler oder Ausreißer in den Daten hat. Das Modelltraining ist der wichtigste Schritt beim maschinellen Lernen und führt zu einem funktionierenden Modell Vorhersagen zu treffen [35]. Um die Performance des Modells zu bewerten, werden die Daten in der Regel in zwei Teile aufgeteilt. Ein Teil der Daten, die so genannten*Trainingsdaten*, wird zur Erstellung des maschinellen Lernmodells verwendet. Die *Testdaten* dienen dazu, die Funktionsfähigkeit des Modells zu beurteilen [35]. Die Daten wurden mittels der Funktion "`train_test_split`"des Pakets `scikitlearn` aufgeteilt [49]. In Rahmen dieser Arbeit wurden zwei Quoten, 50/50 und 80/20, zwischen Training- und Testdaten definiert. Die Abbildung 3.4 stellt die verwendete Aufteilung (englisch *train test split*) dar.

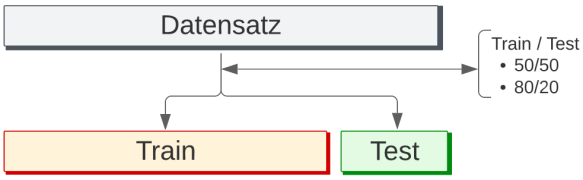


Abbildung 3.4: Train-Test-Split
Quelle: eigene Ausarbeitung nach [35]

In den ersten Untersuchungen mit Datensatz aus 15° NMC wurde festgestellt, dass eine Quote 80/20 eine bessere Train-und Testwertung ergab. Deshalb wurden alle folgenden Modelle in einem Verhältnis von 80/20 trainiert.

3.1.2 Overfitting und Underfitting

Beim Trainieren der Daten muss darauf geachtet werden, dass das Modell auf der einen Seite lauffähig ist, aber auf der anderen Seite nicht zu einer Verallgemeinerung führt, wenn eine nicht relevante Eingabe (z. B. Ausreißer, Rausch, Messfehler) präsentiert wird. Jedoch ist diese ideale Bedingung nicht immer leicht zu finden, und es ist wichtig, zwei verschiedene Gefahren zu berücksichtigen: die Überanpassung (englisch *overfitting*) und die Unteranpassung (englisch

underfitting) [34]. *Overfitting* tritt auf, wenn das Modell angesichts der Menge an Trainingsdaten und der Menge an Rauschen (vernachlässigbare Daten) zu komplex ist [50]. Das heißt, das Modell weist eine niedrige Fehlerquote bei den Trainingsdaten, aber eine hohe Fehlerquote bei den Testdaten auf. Dieses Problem kann mit folgenden Maßnahmen behoben werden [50] [41] [34].

- Anzahl von Merkmalen im Trainingsdatensatz verringern,
- Anpassung von Hyperparameter und Crossvalidation der Train/Test Daten,
- Resampling der Daten.

In dieser Arbeit wurden zuerst redundante Datensätze herausgefiltert und die Relevanz der Eigenschaften in dem Trainingssatz bewertet. Dabei wird Bedeutung der einzelnen Eigenschaften (z. B. Zeitdifferenz, Zyklenzahl) mithilfe von `Random-Forest-Regressor` quantifiziert, um das Modell nur mit den relevantesten Merkmalen zu trainieren. Random-Forest-Regressor ist ein Algorithmus, der binäre Unterteilungen mittels Entscheidungsbäumen (englisch *decision trees*) von Variablen zur Bestimmung von Ergebnisvorhersagen verwendet [51]. Eine Reihe von Entscheidungsbäumen werden auf verschiedene Unterstichproben des Datensatzes anpasst, um die Vorhersagegenauigkeit zu verbessern und das Overfitting zu kontrollieren [52]. Je höher der Koeffizient eines Random-Forest-Regressor, desto relevanter ist die Eigenschaft. Die Abbildung 3.5 zeigt die erste Relevanzanalyse mit allen Features (Eigenschaften).

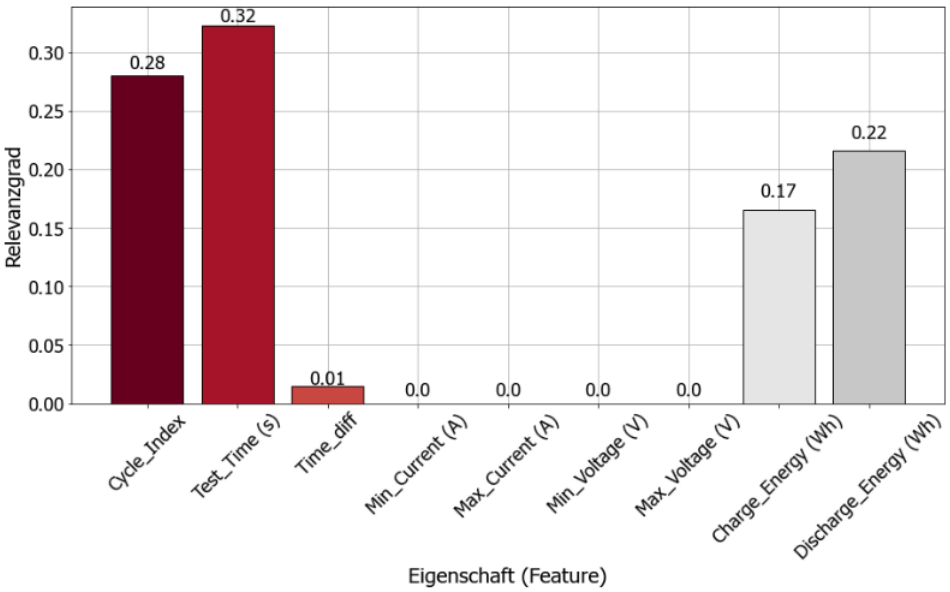


Abbildung 3.5: Erste Relevanzanalyse
Quelle: eigene Ausarbeitung

Als nächstes wurde der Anzahl von Features reduziert, um ihre Relevanz zu analysieren (siehe Abbildung 3.6. **a**). Anschließend wurden nur die zeitrelevanten Features betrachtet (siehe Abbildung 3.6. **b**).

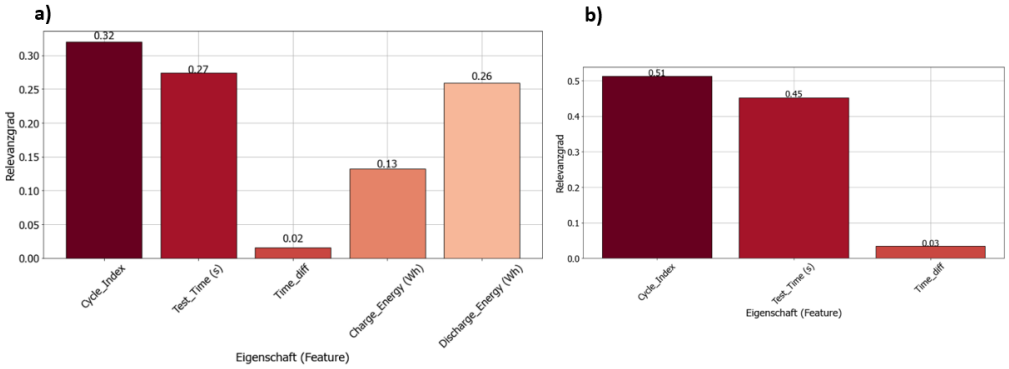


Abbildung 3.6: Relevanzanalysen mit und ohne Lade- und Entladeenergie (a und b bzw.)
Quelle: eigene Ausarbeitung

Aus dieser Analyse lassen sich die folgenden Schlussfolgerungen ziehen: 1) Die minimale und maximale Werte von Spannung und Strom sind zwar wichtige physische Eigenschaften einer LIB aber weniger relevant für das Training des Modells. Deshalb wurden sie aus der folgenden Betrachtungen entnommen. 2) Die zeitgesteuerten Features haben höchsten Relevanz. Da eine Zeitmessung weniger aufwand als die Messung von Lade- und Entladeenergien fordert, wird die Möglichkeit SoH-Kurven nur auf der Grundlage von zeitgesteuerten Merkmalen vorherzusagen in Betrachtung gezogen.

Im Gegensatz zum Overfitting, tritt das *Underfitting* auf, wenn das Modell zu einfach ist, um die in den Daten enthaltene Struktur zu erlernen [50]. Das heißt, wesentliche Einflussfaktoren für die Modellerstellung wurden nicht berücksichtigt und das Modell weist eine schlechte Leistung bei den Trainingsdaten auf und führt zu unzuverlässigen Vorhersagen. Dabei helfen folgende Ansätze [50] [41] [34]:

- Anwendung von komplexeren Modellen;
- Restriktionen im Modell zu verringern;

Die Anwendung von komplexeren Modellen wird nicht in Betrachtung gezogen, da solche Modellen (z. B. Neuronales Netz) in vorherigen Arbeiten verwendet wurden (siehe Kapitel 2.3.2 *Black-Box*).

Bias und Varianz

Die Konzepte von *Bias* und *Varianz* in ML sind von entscheidender Bedeutung, um das Modellverhalten zu verstehen. Das Bias misst die Abweichung des Erwartungswerts der Vorhersage zum wahren Wert [53]. Bei einem Modell mit hohem Bias ist die Wahrscheinlichkeit groß, dass es die Trainingsdaten nicht richtig abbilden würden (Underfitting) [54]. Die Varianz in ML gibt an, wie stark die Vorhersage von schwankt [53]. Ein Modell mit hohen Sensitivität wird wahrscheinlich eine hohe Varianz aufweisen und somit die Trainingsdaten übererfüllen (Overfitting) [54]. Die Abbildung 3.7 zeigt die Beziehung zwischen dem Modellverhalten und den Ergebnisse.

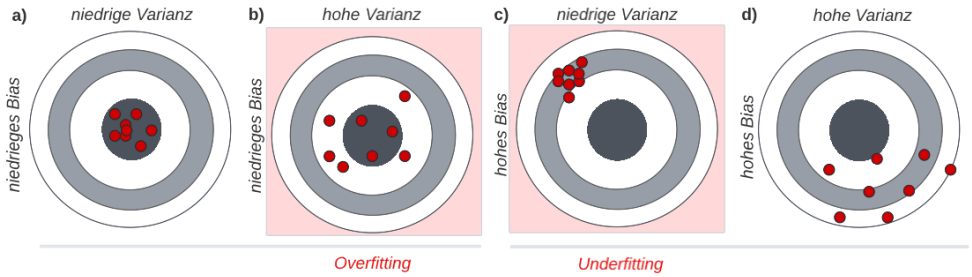


Abbildung 3.7: Modellfehler: Bias-Varianz-Dilemma
Quelle: eigene Ausarbeitung nach [35]

Das Modell der linken Zielscheibe (**a**) wiederholt die richtigen Ergebnisse mit hoher Genauigkeit. Die rechte Zielscheibe (**d**) bildet den schlimmsten Fall ab, der ein Modell Vorhersagen mit geringer Genauigkeit (hohem Bias) und sehr weit voneinander entfernt (hohe Varianz) liefert. Die mittleren Zielscheiben (**b** und **c**) zeigen Ergebnisse von Modellen mit Überanpassung bzw. Unteranpassung.

3.2 Regressionsmodelle

Nach dem Abschluss der Vorverarbeitung des Modells wurden die Algorithmen trainiert und Vorhersagen mit den unten beschriebenen Regressionsverfahren durchgeführt. Diese Verfahren unterscheiden sich grundsätzlich zwischen parametrisch und nichtparametrisch Modellen.Erstere beinhalten einen zweistufigen modellbasierten Ansatz, in dem zunächst eine Annahme über die funktionale Form oder Gestalt der unbekannten Funktion getroffen wird (z. B. wird eine unbekannte Funktion f in linearen Regressionsmodellen als lineare Funktion betrachtet). Im Anschluss werden die Werte der Koeffizienten mit Hilfe des Trainings bestimmt [55]. Nichtparametrische Methoden machen keine expliziten Annahmen über die funktionale Form von eine unbekannten Funktion *(f)*. Stattdessen suchen sie nach einer Schätzung von *f*, die den Datenpunkten möglichst nahe kommt [55].

Um ein Modell für maschinelles Lernen an verschiedene Probleme anzupassen, müssen sogenannte Hyperparameter eingestellt werden. Die Auswahl der besten Hyperparameterkonfiguration für Modelle des maschinellen Lernens hat direkte Auswirkungen auf die Leistung des Modells [56]. Die Hyperparameter müssen angepasst werden, da die optimalen Werte nicht im Voraus bekannt sind. Um die Hyperparameterauswahl zu optimieren, wurde die Suche mittels eines Gitters (Grid-Search) automatisiert.

Grundsätzlich wird hierbei der Wertebereich der Hyperparameter in ein diskretes Raster unterteilt (siehe Abbildung 3.8) und alle Wertekombinationen dieses Gitters werden ausprobiert. Der Punkt des Gitters, an dem bei der Kreuzvalidierung der maximale Durchschnittswert erreicht wird, repräsentiert die optimale Wertekombination für die Hyperparameter.

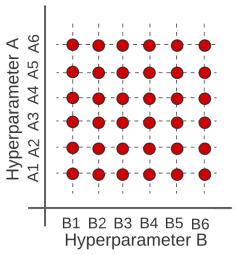


Abbildung 3.8: Funktionsprinzip des Grid Search
Quelle: eigene Ausarbeitung nach [56]

In den folgenden Abschnitten werden die Funktion und die zentralen Merkmale der ausgewählten Modelle erläutert. Hierbei handelt es sich um zwei parametrische Modelle, lineare Regression und Ridge-Regression, sowie zwei nichtparametrische Modelle, KNN-Regression und Random-Forest-Regression.

Lineare Regression

Die *Lineare Regression* bezeichnet die lineare Beziehung zwischen der unabhängigen x_i und der abhängigen y_i Variable besteht [35]. Die Gleichung 3.1 zeigt die Beziehung zwischen den unabhängigen Variablen x zu den abhängigen Variablen y aus einer lineare Regression.

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{3.1}$$

wobei $\hat{y}_i$ ist die zu vorherzusagende Variable Y, x_i ist die unabhängige Variabel $\hat{\beta}_0$ ist der Achsenabschnitt $\hat{\beta}_i$ ist der Regressionskoeffizient. Der Zirkumflex, $\hat{}$, wird hier verwendet, um den geschätzten Wert für einen unbekannten Parameter oder Koeffizienten oder den vorhergesagten Wert der Antwort zu bezeichnen [55].

$\hat{\beta}_i$ kann als ein Gewichtungsvektor angesehen werden, der den Einfluss der einzelnen Features von x_i auf $\hat{y}_i$ beschreibt. Das Ziel des linearen Regressionsalgorithmus ist es, die besten Werte für $\hat{\beta}_0$ und $\hat{\beta}_i$ zu finden. Dabei werden die Parameter $\hat{\beta}_0$ und $\hat{\beta}_i$ ermittelt, die die Residuenquadratsumme RSS zwischen der Vorhersage und dem wahren Regressionsziel y auf den Trainingsmengen minimieren (siehe Gleichung 3.2) [57]. Die RSS ist ein Fehlermaß, das die Abweichung quadratisch aufsummiert [57].

$$RSS = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \sum_{j=1}^{p} \hat{\beta}_j x_{ij})^2 \tag{3.2}$$

Die optimalen Werte für $\hat{\beta}_0$ und $\hat{\beta}_i$ lassen sich mithilfe von Gleichungen 3.3 und 3.4 berechnen:

$$\hat{\beta}_i = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \tag{3.3}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_i \bar{x} \tag{3.4}$$

wobei bezeichnen $\bar{x}$ und $\bar{y}$ jeweils die arithmetischen Mittel von x_i und y_i.

Die Tabellen 3.2 und 3.3 zeigen die Regressionskoeffizienten für die trainierte Modelle mit und ohne Lade- und Entladeenergie. Schließlich ist der Einsatz komplexer Verfahren nicht gerechtfertigt, wenn bereits durch ein lineares Modell gleich gute Ergebnisse erzielt werden können [57].

Tabelle 3.2: Koeffizienten der linearen Regression mit Lade- und Entladeenergie

Zyklenanzahl	**Kumulierte Zeit**	**Zeitdifferenz**	**Ladeenergie**	**Entladeenergie**
β_1	β_2	β_3	β_4	β_5
3.5665e-04	-3.5843e-08	3.3946e-06	2.2980e-01	4.8328e-02

Quelle: eigene Ausarbeitung

Tabelle 3.3: Koeffizienten der linearen Regression ohne Lade- und Entladeenergie

Zyklenanzahl	**Kumulierte Zeit**	**Zeitdifferenz**
β_1	β_2	β_3
7.2422e-03	5.0369e-07	-1.4351e-04

Quelle: eigene Ausarbeitung

Ridge-Regression

Die Ridge-Regression ist der linearen Regression sehr ähnlich, mit dem Unterschied, dass die Koeffizienten der Ridge-Regression Regression durch Minimierung (siehe Gleichung 3.5) anderen Größe geschätzt werden [55].

$$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \sum_{j=1}^{p}\hat{\beta}_j x_i)^2 - \lambda \sum_{i=1}^{n}\hat{\beta}_j^{\,2} = RSS + \lambda \sum_{i=1}^{n}\hat{\beta}_j^{\,2} \tag{3.5}$$

λ wird durch einen Alpha-Parameter in der Ridge-Funktion bezeichnet und ist ein der Strafterm der RSS einer Ridge-Regression. Dieser Term dient dazu, den relativen Einfluss auf die Regressionskoeffizientenschätzungen zu kontrollieren. Das heißt, der Werte von Alpha steuert den Strafterm. Je höher die Werte von Alpha sind, desto größer ist die Strafe und desto geringer ist die Größe der Koeffizienten.

Schließlich ist die Ridge-Regression empfindlich gegenüber Ausreißern und kann zu instabilen Ergebnissen führen, wenn die Daten Ausreißer enthalten [57]. Alpha ist ein Hyperparameter von der Ridge-Regression. Ein optimaler Wert für Alpha von 0,01 wurde gemäß der Grid Search festgelegt.

KNN-Regression

Das Prinzip der Nearest-Neighbour-Methoden besteht darin, eine vordefinierte Anzahl von Trainingsstichproben zu finden, die einem neuen einzuordnenden Punkt am nächsten liegen, und anhand dieser das Label vorherzusagen [58]. Die K-Nearest-Neighbour-Regression basiert auf dem K-Nearest-Neighbour-Algorithmus, der für die Klassifikation von Daten eingesetzt wird [57]. Um zu bestimmen, welche Datenpunkte einem festgelegten Abfragepunkt am nächsten liegen, muss der Abstand zwischen dem Abfragepunkt und anderen Datenpunkten berechnet werden (siehe 3.9) [59].

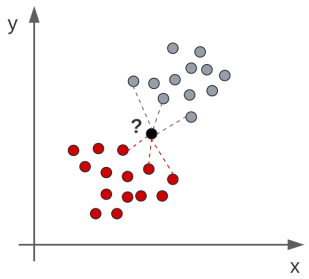


Abbildung 3.9: Abstände zu dem nächsten Nachbarn
Quelle: eigene Ausarbeitung nach [59]

Mit dem Euklidischen Abstand (siehe Gleichung 3.6) wird eine gerade Linie zwischen einem Abfragepunkt und einem Datenpunkt gemessen:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{3.6}$$

Der Manhattan-Abstand (siehe Gleichung 3.7) repräsentiert hingegen den absoluten Wert zwischen zwei Datenpunkten:

$$d(x, y) = (\sum_{i=1}^{m} |x_i - y_i|) \tag{3.7}$$

Der Minkowski-Abstand stellt eine verallgemeinerte Form der Euklidischen und der Manhattan-Abstandsmetrik dar (siehe Gleichung 3.8).Wenn p = 2 ist, wird der Euklidische Abstand durch diese Formel dargestellt. Den Manhattan-Abstand repräsentiert sie bei p = 1. In Rahmen dieser Arbeit wurde die Abstandmessung gemäß Minkowski verwendet:

$$d(x, y) = (\sum_{i=1}^{n} |x_i - y_i|)^{1/p} \tag{3.8}$$

wobei y_i ist der Zielwert und x_i ist der Eigenschaftswert.

Bei der KNN-Regression ist die Zielfunktion jedoch eine reelle Funktion, keine mit diskreten Werten. Dies wird der Mittelwert oder Durchschnittszielwert über die k nächsten Nachbarn berechnet:

$$\hat{y}_i = \frac{1}{k} \sum_{i=1}^{k} x_i \tag{3.9}$$

wobei K die Anzahl von Nachbarn ist. Ein kleiner Wert für K bietet die flexibelste Anpassung, die eine geringe Verzerrung, jedoch eine hohe Varianz aufweist. Letztere ist darauf zurückzuführen, dass die Vorhersage in einer bestimmten Region vollständig von einer einzelnen Beobachtung abhängig ist. Im Gegensatz dazu bieten größere Werte für K eine glattere und weniger variable Anpassung. Die Vorhersage in einer Region ist ein Durchschnitt aus mehreren Punkten, sodass die Änderung einer Beobachtung einen geringeren Effekt hat [55]. In dieser Arbeit wurden K Werten von 2, 3, 5, 7 und 10 getestet und $K = 2$ entspricht gemäß Grid-Search die beste Einstellung diesem Hyperparameter.

Im Vergleich zu der Mehrheit der Algorithmen des maschinellen Lernens hat der grundlegende KNN-Algorithmus mit lediglich K und der Abstandsmetrik wenige Hyperparameter [59]. Dadurch ist die Implementierung dieser Regression einfach. Andererseits ist die KNN-Regression empfindlich gegenüber Rauschen im Datensatz, sodass fehlende Werte manuell hinzugefügt und Ausreißer entfernt werden, um eine besserer Performance zu erzielen [55].

Random-Forest-Regression

Das Konzept des Entscheidungsbaums wurde im Kapitel 3.1.2 (*Overfitting und Underfitting*) im Kontext der Relevanzanalyse aller Features vorgestellt. In den nächsten Abschnitten werden die Nutzung von Entscheidungsbäumen in einem Regressionsverfahren und die Bedeutung der jeweiligen Hyperparameter erläutert. In Abbildung 3.10 ist die Struktur eines Random Forest dargestellt.

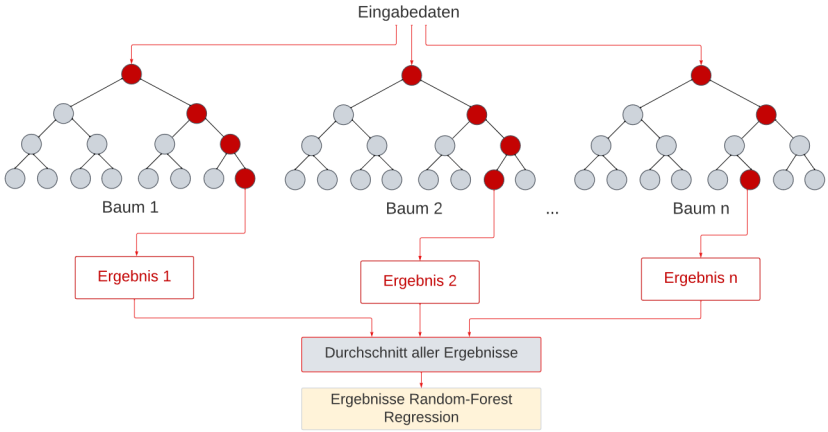


Abbildung 3.10: Struktur eines Random-Forest
Quelle: eigene Ausarbeitung nach [34] [60]

Für einen Random Forest werden während der Trainingszeit mehrere Entscheidungsbäume berechnet. Der Durchschnitt aller Ergebnisse stellt die Vorhersage aus diesem Random Forest dar [34]. In Tabelle 3.4 sind die zentralen Hyperparameter dargestellt. Die letzte Spalte enthält die Ergebnisse aus der Grid-Search für die Random-Forest-Regression.

Tabelle 3.4: Hyperparameter der Random-Fores-Regression

Hyperparameter	**Bedeutung**	**Ergebnis**
n-estimators	Anzahl der Entscheidungsbäume im Modell	400
criterion	Funktion zur Messung der Qualität eines Splits	MSE
max depth	maximal mögliche Tiefe jedes Baums	7
max features	maximale Anzahl von Merkmalen, die bei der Bestimmung einer Aufteilung berücksichtigt werden	20

Quelle: eigene Ausarbeitung nach [34] [60]

Um die Rechenzeit gering zuhalten, wurden zusätzliche Hyperparameter wie die Anzahl an enthaltenen Datenpunkten, ab der ein Knoten geteilt wird (*min samples split*), und die Mindestanzahl der in einem Blattknoten zulässigen Datenpunkte (*min samples leaf*) nicht in der Grid-Search betrachtet. Stattdessen wurden die Standardwerte übernommen.

3.3 Ablauf

m folgenden Unterkapitel wird der Ablauf der Implementierung erläutert. Wie in Kapitel 3.1 (*Erwerb und Vorbereitung der Daten*) erwähnt , wurde der gemessene Kapazitätswert für den ersten Zyklus als maximaler Kapazitätswert eines Datensatzes angenommen und als Basis zur Berechnung der SoH-Werte für den Trainingsdatensatz verwendet. Die gleiche Vorgehensweise wurde für die SoH-Bestimmung von neuen Datensätzen genutzt. Das heißt, die Kapazitätskurve wird mit Hilfe der erwähnten Regressionsmodelle bestimmt und die SoH-Kurve mit Gleichung 2.4 gerechnet.

In erster Linie wurden die Kapazitätskurven von weiteren Daten aus SNL-NMC mit und ohne Lade- und Entladeenergie in den Trainingsdaten bestimmt und die SoH-Kurven berechnet. Die Abbildung 3.11 enthält eine Übersicht der SNL-Daten.

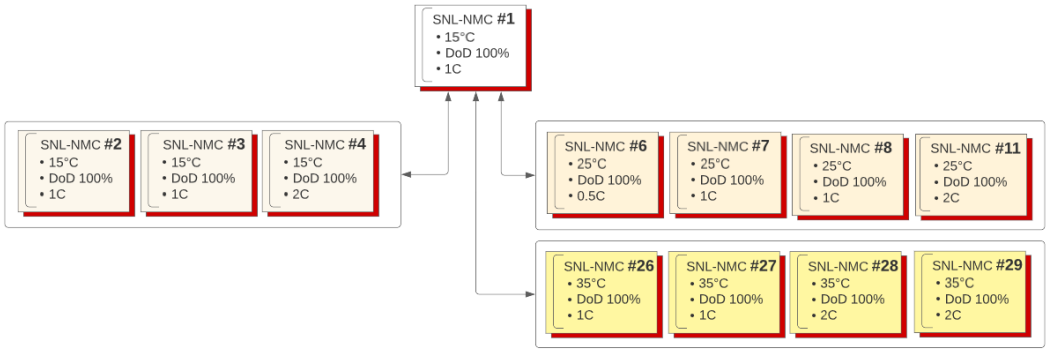


Abbildung 3.11: Übersicht der verwendeten NMC-Datensätze aus SNL mit unterschiedlichen Temperaturen (15°C, 25°C und 35°C) und C-Rate (1C und 2C) bei DoD von 100%
Quelle: eigene Ausarbeitung nach [45]

Als nächstes wurden die Kapazitätskurve aus den SNL-NCA bestimmt und auf SoH-Kurven umgerechnet (siehe Abbildung 3.12).

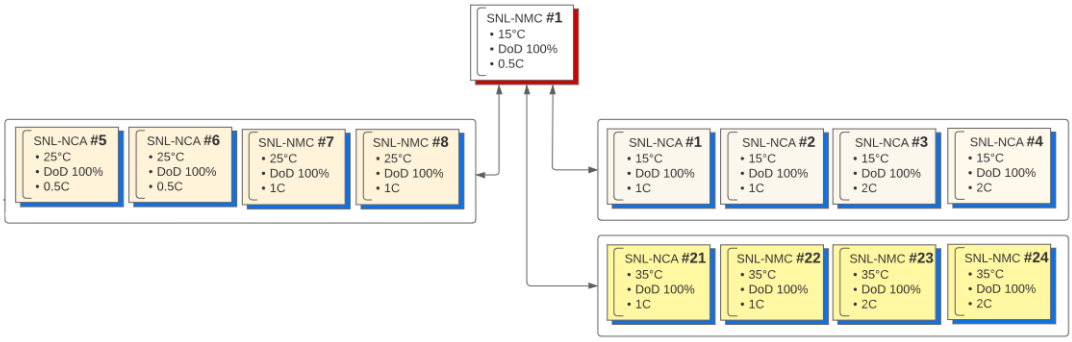


Abbildung 3.12: Übersicht der verwendeten NCA-Datensätze aus SNL mit unterschidlichen Temperaturen (15°C, 25°C und 35°C) und C-Rate (0.5C, 1C und 2C) bei DoD von 100%
Quelle: eigene Ausarbeitung nach [45]

Anschließend wurden die Datensätze aus CALCE-LCO betrachtet. Diese sind auf 3.13 ersichtlich. In Anhang A ‚*Datensätze*' befindet sich eine detaillierte Liste aller Datensätze.

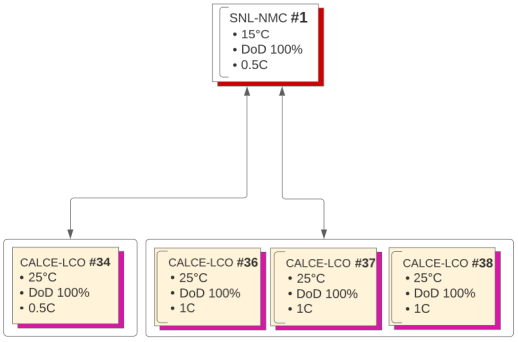


Abbildung 3.13: Übersicht der verwendeten LCO-Datensätze aus CALCE mit 25°C Betreiebstemperatur und C-Rates von 0.5C und 1C bei DoD von 100%
Quelle: eigene Ausarbeitung

3.4 Grundlagen zur Performancebewertung und Schätzungsqualität

Einer der wichtigsten Schritte bei der Anwendung jedes Modells des maschinellen Lernens ist die Bewertung der Modellgenauigkeit. Das R-Quadrat (R^2), der mittlere quadratische Fehler MSE, der mittlere absolute Fehler MAE und der mittlere quadratische Wurzelfehler RMSE werden zur Bewertung der Leistung des Modells in der Regressionsanalyse verwendet. Die Tabelle 3.5 enthält grundlegende mathematische Konzepte zum Verständnis der verwendeten Bewertungsmetriken.

Tabelle 3.5: Mathematische Grundlage von Bewertungsmetriken eines Regressionsmodells

Term	**Erklärung und Formel**
$\hat{y}_i$	Vorhergesagter Wert von y_i
$\bar{y}_i$	Mittelwert von y_i $$\bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i \quad (3.10)$$
Residuum	Abweichung eines tatsächlichen Wertes y_i vom vorhergesagten Werte $\hat{y}_i$. $$\hat{e}_i = y_i - \hat{y}_i \quad (3.11)$$

Quelle: eigene Ausarbeitung nach [61] [62]

Das R-Quadrat (siehe 3.12) liegt zwischen 0 und 1 und misst den Anteil der Variation in den Daten, der im Modell berücksichtigt wird. Diese Metrik ist nützlich, um zu beurteilen, wie gut das Modell zu den Daten passt [62]. Das heißt, es zeigt, wie nahe die Daten an der Regressionslinie liegen. Zusätzlich wird R-Quadrat auch als Bestimmtheitsmaß bezeichnet. Es ist Kennzahl zur Beurteilung der Anpassungsgüte einer Regression.

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \bar{y}_i)^2} \quad (3.12)$$

Der MAE ist der Durchschnitt der absoluten Differenz zwischen den tatsächlichen und den vorhergesagten Werten des Datensatzes. Er misst den Durchschnitt der Residuen im Datensatz [62].

$$MAE = \frac{1}{m} \sum_{i=1}^{m} |y_i - \hat{y}_i| \tag{3.13}$$

Der MSE misst die durchschnittliche quadratische Differenz zwischen den vorhergesagten Werten und der abhängigen Variable.

$$MSE = \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 \tag{3.14}$$

Der RMSE ist die Fehlerrate durch die Quadratwurzel des MSE. Es misst die Standardabweichung der Residuen und wird gemeinsam mit R-Quadrat für den Vergleich der Genauigkeit zwischen verschiedenen linearen Regressionsmodellen verwendet [62].

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2} \tag{3.15}$$

Die RMSE-, MAE- und MSE-Werte haben den Wert 0, wenn das lineare Regressionsmodell die Daten perfekt abbildet, und einen positiven Wert, wenn es die Daten mit Abweichungen abbildet [61]. Dieses Kriterium wird für die Performancebewertung des Regressionsmodelles in Rahmen dieser Arbeit verwendet. Je kleiner der Wert ist, desto besser ist die Genauigkeit des Modells.

Kriterien für die Zielerreichung

Um die Forschungsfragen zu beantworten, müssen weitere Bewertungskriterien für die allgemeine Performance der festgelegt werden. Die gesamte Rechnenzeit für die Vorhersage der SoH Kurve wird betrachtet, da dieses Kriterium für eine praktische Umsetzung des Modells relevant ist. Je weniger die Komplexität und Rechnenzeit des Modells ist, desto besser ist die Handhabung in der Praxis. Wie im Kapitel 2.2.2 erwähnt wurde, der Referenzwert EoL definiert wie lange ein Akkumulator in seinem ersten Lebenszyklus verwendet wird. Die Differenzen zwischen realen und vorhergesagten SoH bei 95%, 90%, 85%, 80% der unterschiedlichen Algorithmen wird gegenübergestellt. Eine geringe Differenz weist auf, dass die Regressionsverfahren mehrere SoH-Kurven aus unterschiedlichen Kathodenmaterialien bis zum Ende der ersten Leben (EoL = 80%) einer LIB vorhersagen kann.

4 Ergebnisse und Diskussion

In diesem Kapitel werden die Ergebnisse vorgestellt. Zuerst werden die Kennzahlen zur Performancebewertung und Schätzungsqualität gezeigt. Anschließend werden die vorhergesagten SoH-Kurven mit Hilfe der genannten Regressionsalgorithmen dargestellt. In Tabelle 4.1 sind die Kennzahlen zur Performancebewertung des Trainingsmodells mit Lade- und Entladeenergie aufgeführt.

Tabelle 4.1: Übersicht Kennzahlen zur Performancebewertung

Regression	**R2**	**MAE**	**RMSE**
Lineare Regression	0.99	0.0012	0.0014
Ridge-Regression	0.99	0.0012	0.0014
knn-Regression	0.99	0.0007	0.0012
RFR-Regression	0.99	0.0011	0.0014

Quelle: eigene Ausarbeitung

Die Kennzahlen des Modells ohne Energieangaben sind in Tabelle 4.2 dargestellt.

Tabelle 4.2: Übersicht Kennzahlen zur Performancebewertung

Regression	**R2**	**MAE**	**RMSE**
Lineare Regression	0.98	0.0356	0.0385
Ridge-Regression	0.98	0.0356	0.0385
knn-Regression	0.98	0.0056	0.0136
RFR-Regression	0.98	0.0038	0.0063

Quelle: eigene Ausarbeitung

In beiden Tabellen deuten die Werte von MAE, RMSE auf Modelle hin, die die Daten sehr gut abgebildet haben. Hierbei wird MAE als der durchschnittliche Fehler bei einer Vorhersage mit dem Modell interpretiert. Der RMSE

hingegen kann als der durchschnittliche Fehler interpretiert werden, den die Vorhersagen des Modells im Vergleich zu den tatsächlichen Werten aufweisen. Im Gegensatz zu MAE und RMSE, ist R^2 kein Maß für die Genauigkeit der Vorhersagen, , sondern für das Fitting. Die Werten von R^2 aus den obigen Tabellen nähern sich 1,0 an. Höhere Werte für R^2 bedeuten,dass ein größerer Teil der Variation der abhängigen Variable durch die unabhängigen Variablen im Regressionsmodell erklärt wird. Obwohl ein hoher R^2-Wert grundsätzlich als positiver Hinweis interpretiert wird, kann er das Overfitting des Modells zeigen. Aus der ersten Analyse der Kennzahlen zur Performancebewertung ist ersichtlich, dass die lineare und die Ridge-Regression bessere Modelle für die Vorhersage der SoH-Kurven sind.

4.1 Ergebnisse der NMC-Datensätze

SoH 15°C NMC 2, 3, 4 mit Train-Test 80/20 mit Lade- und Entladeenergie.

Auf Abbildung 4.1 ist zu erkennen, dass die lineare Regression unter den angewendeten Regressionsverfahren mit Lade- und Entladeenergien die besten Ergebnisse in den Trainingsdaten erreicht hat. Beim Datensatz **1** hat die Random-Forest-Regression die beste Performance gezeigt

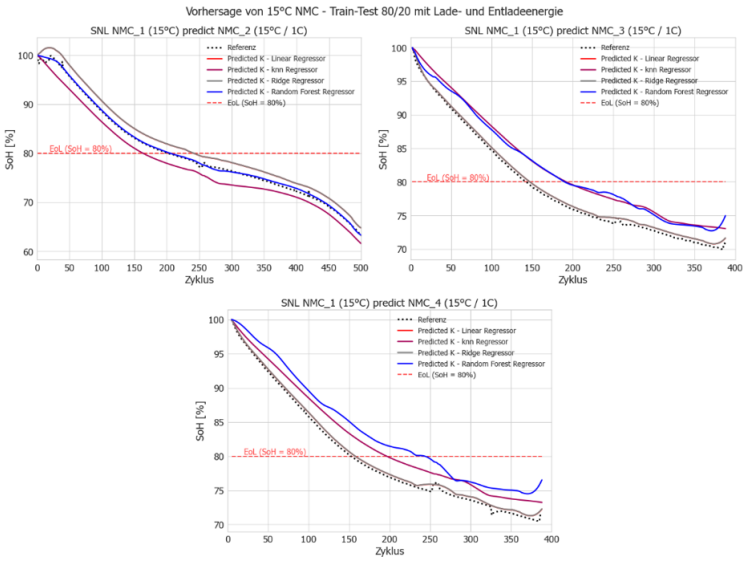


Abbildung 4.1: Ergebnisse 15°C NMC 2, 3, 4 - Train-Test 80/20 mit Lade- und Entladeenergie
Quelle: eigene Ausarbeitung

Jedoch wurde die Random-Forest-Regression ausschließlich für diesen Fall als bestes Regressionsmodell ausgewiesen. Es ist zu beachten, dass sich die Kurven der linearen und der Ridge-Regression überlappen, da die Ergebnisse aus diesen Modellen überaus ähnlich sind. Bei den Datensätzen **2** und **3** beträgt die Differenz der Zyklenzahl bei EoL = 80% zwischen der Referenzkurve und vorhergesagte Kurve mithilfe der Linear Regression 4 und 2 Zyklen bzw. Die Tabelle 4.3enthält eine Zusammenfassung der Ergebnisse. Die KNN- und die Random-Forest-Regression weisen zwar eine große Differenz zwischen Vorhersage und Referenz auf, aber die Ergebnisse bewegen sich noch im Rahmen von 50 Zyklen.

Tabelle 4.3: Übersicht der Ergebnisse 15°C NMC 2, 3, 4 - Train-Test 80/20 mit Lade- und Entladeenergie

SoH-Kurve 15°C-NMC(2)	95%	90%	85%	80%
Referenzkurve	54	88	129	201
Lineare Regression	69	104	150	240
Ridge-Regression	69	104	150	241
knn-Regression	35	72	110	162
RFR-Regression	55	90	131	204
SoH-Kurve 15°C-NMC(3)	95%	90%	85%	80%
Referenzkurve	19	56	94	141
Lineare Regression	19	58	99	144
Ridge-Regression	19	58	99	144
knn-Regression	39	83	127	190
RFR-Regression	33	78	123	188
SoH-Kurve 15°C-NMC(4)	95%	90%	85%	80%
Referenzkurve	27	64	101	140
Lineare Regression	28	66	105	152
Ridge-Regression	32	71	111	153
knn-Regression	44	87	134	191
RFR-Regression	46	93	145	237

Quelle: eigene Ausarbeitung

SoH 15°C NMC 2, 3, 4 mit Train-Test 80/20 ohne Lade- und Entladeenergie.

Die Kurven, für die die Regressionsmodelle ohne Lade- und Entladeenergie trainiert wurden, liefern quantitativ und qualitativ schlechtere Ergebnisse. Dies lässt sich durch die Berechnung der Kapazität einer Batterie durch die Lade- und Entladeenergie erklären (siehe Gleichung 4.1). Die Lade- und Entladeenergien nehmen wie die Kapazität einer Batterie mit der Zeit ab. Dieses Verhalten ist bei den Ergebnissen aller Vorhersagen sowohl bei NMC, NCA als auch bei LCO zu finden.

$$E = K.V \tag{4.1}$$

wobei K die Kapazizät und V die Spannung ist.

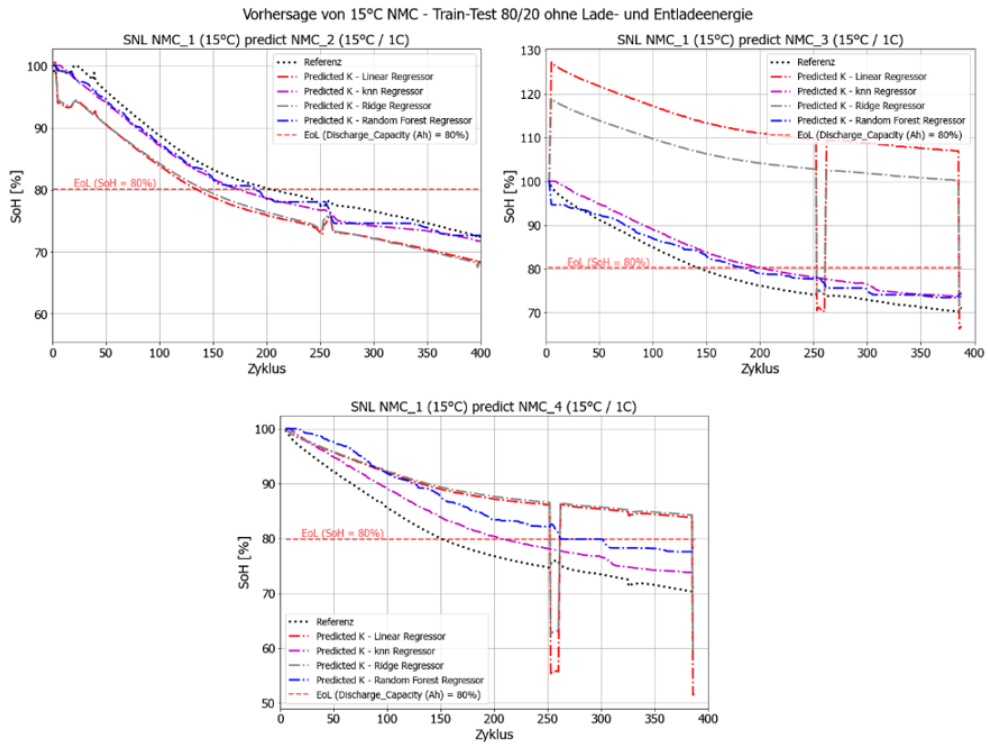


Abbildung 4.2: Ergebnisse 15°C NMC 2, 3, 4 - Train-Test 80/20 ohne Lade- und Entladeenergie
Quelle: eigene Ausarbeitung

In Tabelle 4.4 werden die Ergebnisse aus der oben genannten Vorhersage vorgestellt.

Tabelle 4.4: Übersicht der Ergebnisse 15°C NMC(2) - Train-Test 80/20 ohne Lade- und Entladeenergie

SoH-Kurve 15°C-NMC(2)	95%	90%	85%	80%
Referenzkurve	54	88	129	200
Lineare Regression	19	52	89	134
Ridge-Regression	19	52	92	141
knn-Regression	41	77	117	174
RFR-Regression	45	79	120	189
SoH-Kurve 15°C-NMC(3)	95%	90%	85%	80%
Referenzkurve	19	56	95	141
Lineare Regression	> 400	> 400	> 400	> 400
Ridge-Regression	> 400	> 400	> 400	> 400
knn-Regression	15	72	130	186
RFR-Regression	51	83	178	201
SoH-Kurve 15°C-NMC(4)	95%	90%	85%	80%
Referenzkurve	27	64	101	148
Lineare Regression	54	126	309	> 400
Ridge-Regression	56	132	317	> 400
knn-Regression	44	87	134	200
RFR-Regression	70	121	177	260

Quelle: eigene Ausarbeitung

Im Anhang B ‚*Ergebnisse*‘ befindet sich alle Abbildungen und Ergebnissehtabellen aus dieser Arbeit. Da die Vorhersagen ohne Lade- und Entladeenergie eindeutig schlechte bzw. ungenauere Ergebnisse lieferten, ist eine Tabelle mit einzelnen SoH-Werten nicht notwendig.

SoH 25°C NMC und 35°C NMC mit Train-Test 80/20 ohne Lade- und Entladeenergie.

Durch die Bestimmung der SoH-Kurve von Datensätzen mit höheren Temperaturen wird die gute Performance der linearen und der Ridge-Regression bestätigt. Abbildung 4.3 enthält eine Übersicht von vier Datensätzen (zwei aus 25°C und zwei aus 35°C) dar. Im Anhang B befinden sich die komplette Ergebnisse.

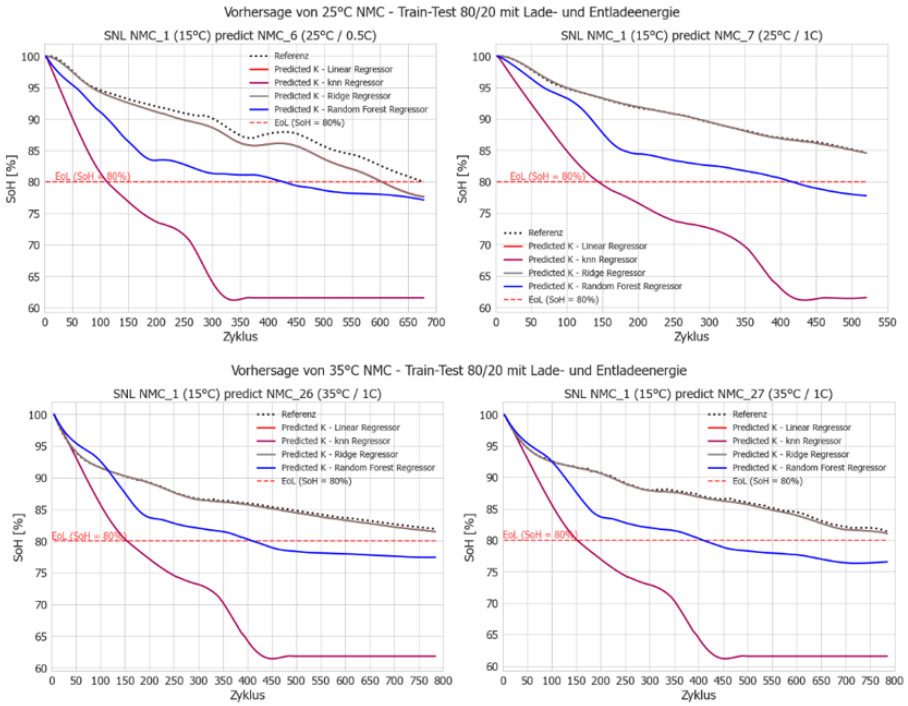


Abbildung 4.3: Ergebnisse 25°C NMC 6, 7 und 35°C NMC 26, 27 - Train-Test 80/20 mit Lade- und Entladeenergie
Quelle: eigene Ausarbeitung

4.2 Ergebnisse der NCA-Datensätze

SoH 15°C NCA mit Train-Test 80/20 mit Lade- und Entladeenergie.

Obwohl die NCA- und NMC-Batterien unterschiedliche chemische Zusammensetzung und Kapazitäten (siehe Tabelle 3.1) haben, konnten die SoH-Kurven der NCA-Batterien mit einem Trainingsmodell aus NMC mit der linearen und der Ridge-Regression bestimmt werden (siehe Abbildung 4.4). Im Vergleich zu den Daten aus NCA-1 und NCA-2 enthielten die Daten der Kurven von NCA-3 und NCA-4 mehr Ausreißer. Der Savitzy-Golay-Filter konnte die Ausreißer zwar unterdrücken, ihre Effekte sind aber anhand der Kurven zu sehen.

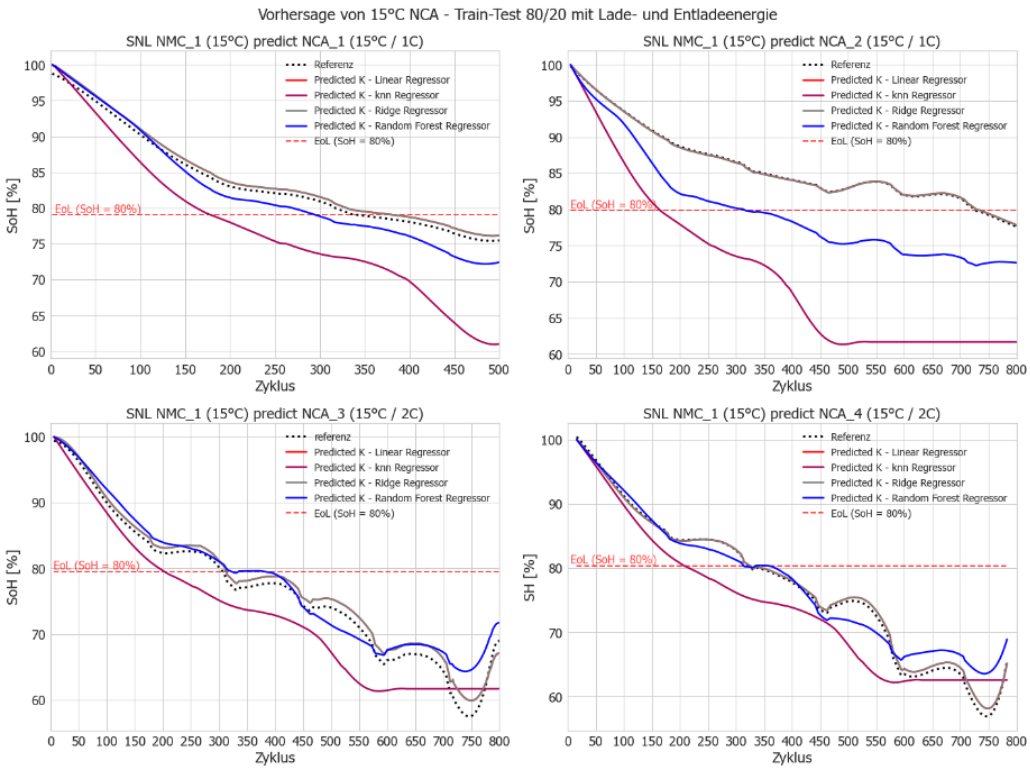


Abbildung 4.4: Ergebnisse 15°C NCA 1, 2, 3, 4 - Train-Test 80/20 mit Lade- und Entladeenergie
Quelle: eigene Ausarbeitung

4.3 Ergebnisse der LCO-Datensätze

SoH 25°C NCA und 35°C NCA mit Train-Test 80/20 mit Lade- und Entladeenergie.

Ähnlich positive Ergebnisse sind bei den Kurven von NCA bei Betriebstemperaturen 25°C und 35°C zu finden.

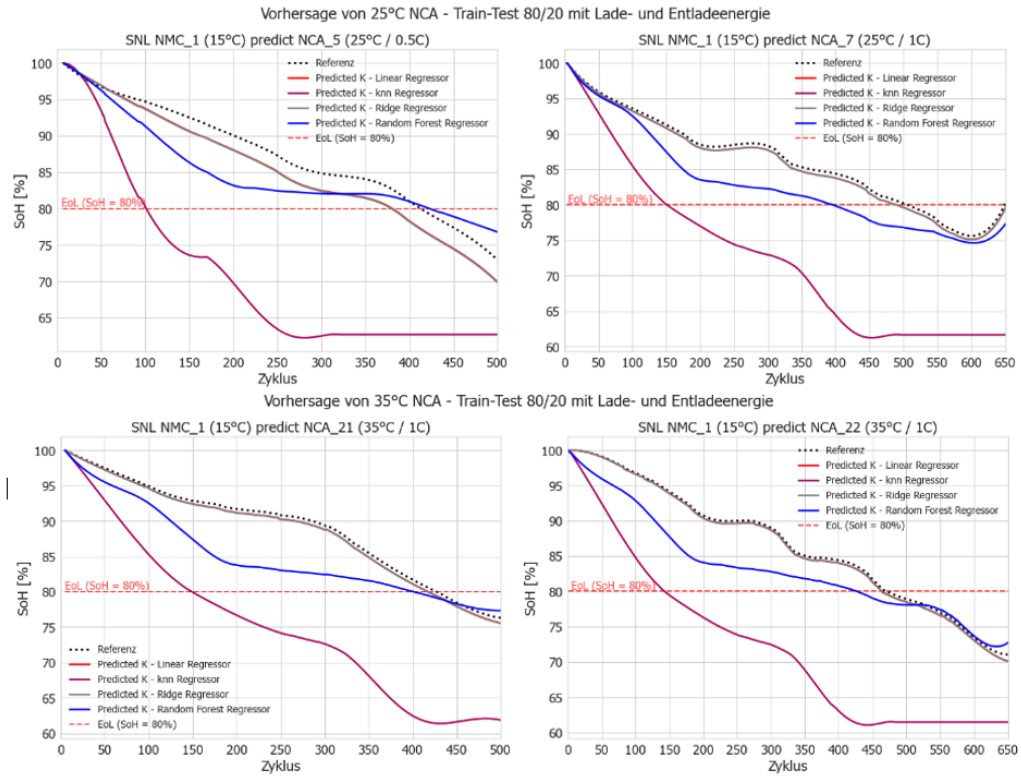


Abbildung 4.5: Ergebnisse 25°C NCA 5, 7 und 35°C NCA 21, 22 - Train-Test 80/20 mit Lade- und Entladeenergie
Quelle: eigene Ausarbeitung

SoH 25°C LCO mit Train-Test 80/20 mit Lade- und Entladeenergie.

Die SoH-Kurven aus den LCO-Batterien folgen der gleichen Tendenz wie die bei der Bestimmung von NMC- und NCA-Batterien. Das heißt, die lineare und die Ridge-Regression weisen bessere Ergebnisse auf. Der Unterschied liegt daran, dass die Differenz zwischen Referenzkurve und Vorhersagen gemäß linearer und Ridge-Regression größer ist.

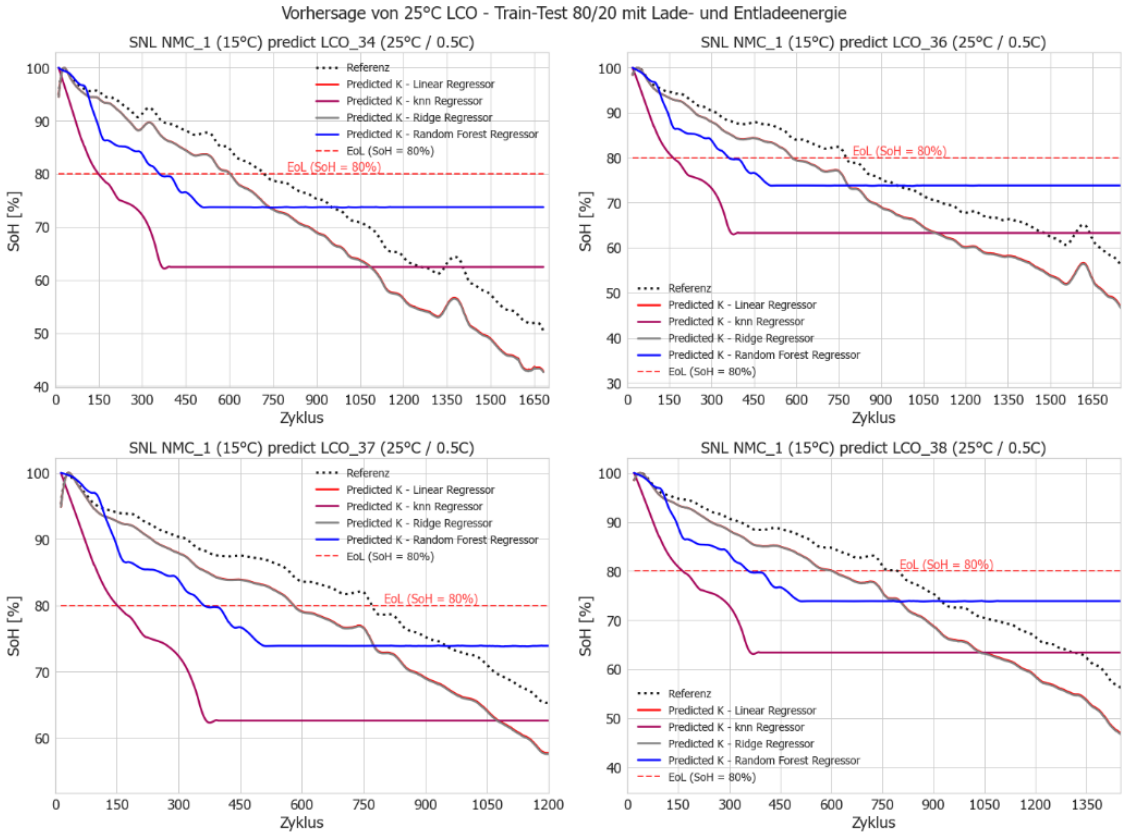


Abbildung 4.6: Ergebnisse 25°C LCO 34, 36, 37, 38 - Train-Test 80/20 mit Lade- und Entladeenergie
Quelle: eigene Ausarbeitung

Die parametrische Regression sowie die lineare Regression und die Ridge-Regression benötigen weniger Datenpunkte als nichtparametrische Lernmethoden und liefern mit weniger Daten bessere Ergebnisse. Die Modellstruktur wurde in dieser Arbeit besser durch eine Funktion mit festen Parametern definiert, z. B. eine multiple lineare Regression. Dies erklärt zum Beispiel, warum die nichtparametrischen Regressionsmodelle nicht gut abschnitten.

5 Fazit

In den nachfolgenden Unterkapiteln wird die untersuchte Black-Box-Ansatz zur SoH-Schätzung von LIB abschließend zusammengefasst. Dabei wird kritisch erläutert, welche Ziele erreicht wurden und welche Perspektive sich für die weitere Forschung auf diesem Themengebiet bietet.

5.1 Zusammenfassung

Die Bestimmung der Zustandsgröße SoH ist essenziell, um die Nutzung der Batteriekapazität zu optimieren und die Erfüllung der sicherheitsrelevanten Funktionen sicherzustellen. Aufgrund der zunehmenden Bedeutung datengetriebener Analysen wird die Anwendung von Algorithmen des maschinellen Lernens für die SoH-Bestimmung sowohl in der Forschung als auch in der praktischen Anwendung untersucht.

Das Ziel der Masterarbeit war es, die SoH (Prozentwerte) der verwendeten Untersuchungsobjekte (Lithium-Ionen-Batterien mit unterschidlichen Kathodenmateriallen: NMC, NCA und LCO) unter verschiedenen Betriebskonditionen (Temperaturen und C-Rate) zu bestimmen. Dies wurde auf Basis zuvor erfasster Daten (z. B. Umgebungstemperatur, Ladeleistung, Entladeleistung, Anzahl von Lade- und Entladezyklen sowie Dauer eines Zyklus) auf die Anzahl der Entladezyklen bezogen und mittels datengesteuerter Methoden und Algorithmen des maschinellen Lernens ausgeführt.

Die zugrundeliegende Fragestellung dieser Masterarbeit, 'Wie gut können Algorithmen des maschinellen Lernens die SoH Verläufe von Lihtium-Ionen-Batterien mit unterschiedlichen Kathodenmaterialien und Einsatzbedingungen schätzen?', wurde anhand einer praktischen Untersuchung in Rahmen dieser Arbeit beantwortet. Im besten Fall können die SoH-Verläufe von Lihtium-Ionen-Batterien auf exakte Zykluszahlen bei EoL = 80% für dasselbe Kathodenmaterial sowie dieselbe Temperatur und mit einer Differenz von 50 Zyklen für unterschiedliche Kathodenmaterialien bestimmt werden. Je größer der Unterschied zwischen den Betriebstemperaturen und der chemischen Zusammensetzung der Batterien ist, desto ungenauer ist die Vorhersage. Die zweite Forschungsfrage, 'Welche Algorithmen des maschinellen Lernens rechtfertigen weitere Untersuchungen für eine praktische Anwendung?', wurde ebenfalls beantwortet. Die pa-

rametrische Regressionsverfahren lineare Regression und die Ridge-Regression für die praktische Anwendung am besten geeignet.

5.2 Ausblick und kritische Bewertung

Die Ergebnisse dieser Arbeit dienen als Auslegungsgrundlage für weitere Forschungsarbeiten und für die Nutzung von Lithium-Ionen-Batterien in elektrisch angetriebenen Systemen. Dennoch lassen sich die gewonnenen Erkenntnisse durch weitere Untersuchungen ergänzen. Als Beispiel ist die Frage zu nennen, ob das Training mit Datensätzen mit drei Betriebstemperaturen (15 °C, 25 °C und 35 °C) zu besseren allgemeinen Ergebnisse führen würde. Zudem könnten weitere Kathodenmaterialien in die Betrachtung eingeschlossen werden und die Verwendung anderer Regressionsalgorithmen wie Lasso-Regression, Gaußprozess-Regression und einfacher neuronaler Netze könnte Einfluss auf die Performance haben. Ergänzend lässt aus den vorgestellten Ergebnissen die weiterführende Frage ableiten, warum die Performance der KNN-Regression und der Random-Forest-Regression schlechter ausfällt als jene der linearen und der Ridge-Regression, obwohl die Kennzahlen zur Performancebewertung kaum voneinander abweichen.

In diesem Zusammenhang kann die Entwicklung eines Prüfstands sinnvoll sein, um die Ergebnisse der Arbeit in der Praxis zu validieren und die Nutzung neuer Kathodenmaterialien und verschiedener Betriebsbedingungen zu erforschen. Dabei können die Untersuchungen auf unterschiedliche Lade- und Entladetiefen (z. B. DoD von 20 oder 40) und den Betrieb bei extremen Temperaturen (z. B. 0°C, 5°C, 45°C) erweitert werden. Ein Prüfstand bietet zusätzlich die Möglichkeit, einen direkten Vergleich zwischen den Ergebnissen aus White-, Grey- und Black-Box-Methoden für die gleichen Untersuchungsobjekte zu ziehen. Dabei kann untersucht werden, ob die experimentellen und die modellbasierten Methoden durch einen auf maschinellem Lernen basierenden datengetriebenen Ansatz ersetzt werden können. Dabei wäre es möglich, auf die für Grey- und White-Box-Methoden nötige experimentelle Vorabvalidierung und komplexe Ausrüstung zu verzichten, um Investitionskosten zu reduzieren und Zeit zu gewinnen.

Darüber hinaus können mit einem validierten datengetriebenen Ansatz Funktionen eines robusten BMS für Systeme mit definierten Anwendungen ersetzt und ein EoL kann vorhergesagt werden. Im medizinischen Bereich könnte hiermit z. B. bestimmt werden, wann ein Defibrillator ersetzt werden sollte. Außerdem könnte das Wartungsintervall für Brandmeldeanlagen definiert und mit stationären Batterien für eine unterbrechungsfreie Stromversorgung könnte die Zuverlässigkeit erhöht werden.

Literaturverzeichnis

[1] R. Korthauer, Hrsg., *Handbuch Lithium-Ionen-Batterien*, de. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, ISBN: 978-3-642-30652-5 978-3-642-30653-2. DOI: 10.1007/978-3-642-30653-2. Adresse: http://link.springer.com/10.1007/978-3-642-30653-2 (besucht am 07. 10. 2022).

[2] C. Vidal, P. Malysz, P. Kollmeyer und A. Emadi, „Machine Learning Applied to Electrified Vehicle Battery State of Charge and State of Health Estimation: State-of-the-Art," *IEEE Access*, Jg. 8, S. 52 796–52 814, 2020, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.2980961. Adresse: https://ieeexplore.ieee.org/document/9036949/ (besucht am 22. 11. 2022).

[3] H. Tschöke, P. Gutzmer und T. Pfund, Hrsg., *Elektrifizierung des Antriebsstrangs: Grundlagen - vom Mikro-Hybrid zum vollelektrischen Antrieb*, de. Berlin, Heidelberg: Springer Berlin Heidelberg, 2019, ISBN: 978-3-662-60355-0 978-3-662-60356-7. DOI: 10.1007/978-3-662-60356-7. Adresse: http://link.springer.com/10.1007/978-3-662-60356-7 (besucht am 07. 10. 2022).

[4] L. Dorrmann, „Kompendium: Li-Ionen-Batterien," de, S. 30,

[5] M. Sterner und I. Stadler, Hrsg., *Energiespeicher: Bedarf, Technologien, Integration*, ger, 2., korrigierte und ergänzte Auflage. Berlin: Springer Vieweg, 2017, ISBN: 978-3-662-48892-8.

[6] U. S. Meda, L. Lal, S. M und P. Garg, „Solid Electrolyte Interphase (SEI), a boon or a bane for lithium batteries: A review on the recent advances," en, *Journal of Energy Storage*, Jg. 47, S. 103 564, März 2022, ISSN: 2352152X. DOI: 10.1016/j.est.2021.103564. Adresse: https://linkinghub.elsevier.com/retrieve/pii/S2352152X21012433 (besucht am 07. 10. 2022).

[7] M. Doppelbauer, *Grundlagen der Elektromobilität: Technik, Praxis, Energie und Umwelt* (Lehrbuch), ger. Wiesbaden [Heidelberg]: Springer Vieweg, 2020, ISBN: 978-3-658-29729-9.

[8] P. Kurzweil und O. K. Dietlmeier, *Elektrochemische Speicher*, de. Wiesbaden: Springer Fachmedien Wiesbaden, 2015, ISBN: 978-3-658-10899-1 978-3-658-10900-4. DOI: 10.1007/978-3-658-10900-4. Adresse: http:

//link.springer.com/10.1007/978-3-658-10900-4 (besucht am 07. 10. 2022).

[9] H. Proff, Hrsg., *Entscheidungen beim Übergang in die Elektromobilität*, de. Wiesbaden: Springer Fachmedien Wiesbaden, 2015, ISBN: 978-3-658-09576-5 978-3-658-09577-2. DOI: 10.1007/978-3-658-09577-2. Adresse: http://link.springer.com/10.1007/978-3-658-09577-2 (besucht am 07. 10. 2022).

[10] L. Ljung, „Black-box models from input-output measurements," in *IMTC 2001. Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference. Rediscovering Measurement in the Age of Informatics (Cat. No.01CH 37188)*, Bd. 1, Budapest, Hungary: IEEE, 2001, S. 138–146, ISBN: 978-0-7803-6646-6. DOI: 10.1109/IMTC.2001.928802. Adresse: http://ieeexplore.ieee.org/document/928802/ (besucht am 07. 10. 2022).

[11] M. Galeotti, L. Cinà, C. Giammanco, S. Cordiner und A. Di Carlo, „Performance analysis and SOH (state of health) evaluation of lithium polymer batteries through electrochemical impedance spectroscopy," en, *Energy*, Jg. 89, S. 678–686, Sep. 2015, ISSN: 03605442. DOI: 10.1016/j.energy.2015.05.148. Adresse: https://linkinghub.elsevier.com/retrieve/pii/S0360544215007756 (besucht am 07. 10. 2022).

[12] N. Noura, L. Boulon und S. Jemeï, „A Review of Battery State of Health Estimation Methods: Hybrid Electric Vehicle Challenges," en, *World Electric Vehicle Journal*, Jg. 11, Nr. 4, S. 66, Okt. 2020, ISSN: 2032-6653. DOI: 10.3390/wevj11040066. Adresse: https://www.mdpi.com/2032-6653/11/4/66 (besucht am 07. 10. 2022).

[13] S. Park, J. Ahn, T. Kang u. a., „Review of state-of-the-art battery state estimation technologies for battery management systems of stationary energy storage systems," en, *Journal of Power Electronics*, Jg. 20, Nr. 6, S. 1526–1540, Nov. 2020, ISSN: 1598-2092, 2093-4718. DOI: 10.1007/s43236-020-00122-7. Adresse: https://link.springer.com/10.1007/s43236-020-00122-7 (besucht am 07. 10. 2022).

[14] C. Deatcu, D. Lückerath, O. Ullrich und U. Durak, *Proceedings ASIM SST 2020 25. ASIM Symposium Simulationstechnik: 14.10.-15.10.2020 : Online-Tagung*, ger. Wien: ARGESIM, 2020, OCLC: 1226115554, ISBN: 978-3-901608-93-3.

[15] X. Li, C. Yuan, X. Li und Z. Wang, „State of health estimation for Li-Ion battery using incremental capacity analysis and Gaussian process regression," en, *Energy*, Jg. 190, S. 116 467, Jan. 2020, ISSN: 03605442. DOI: 10.1016/j.energy.2019.116467. Adresse: https://linkinghub.elsevier.com/retrieve/pii/S0360544219321620 (besucht am 07. 10. 2022).

[16] R. Xiong, Y. Pan, W. Shen, H. Li und F. Sun, „Lithium-ion battery aging mechanisms and diagnosis method for automotive applications: Recent advances and perspectives," en, *Renewable and Sustainable Energy Reviews*, Jg. 131, S. 110 048, Okt. 2020, ISSN: 13640321. DOI: 10.1016/j.rser.2020.110048. Adresse: https://linkinghub.elsevier.com/retrieve/pii/S1364032120303397 (besucht am 07. 10. 2022).

[17] H. Mu und R. Xiong, „Modeling, Evaluation, and State Estimation for Batteries," en, in *Modeling, Dynamics and Control of Electrified Vehicles*, Elsevier, 2018, S. 1–38, ISBN: 978-0-12-812786-5. DOI: 10.1016/B978-0-12-812786-5.00001-X. Adresse: https://linkinghub.elsevier.com/retrieve/pii/B978012812786500001X (besucht am 07. 10. 2022).

[18] M. A. Patil, P. Tagade, K. S. Hariharan u. a., „A novel multistage Support Vector Machine based approach for Li ion battery remaining useful life estimation," en, *Applied Energy*, Jg. 159, S. 285–297, Dez. 2015, ISSN: 03062619. DOI: 10.1016/j.apenergy.2015.08.119. Adresse: https://linkinghub.elsevier.com/retrieve/pii/S0306261915010557 (besucht am 07. 10. 2022).

[19] S. Shen, M. Sadoughi, X. Chen, M. Hong und C. Hu, „A deep learning method for online capacity estimation of lithium-ion batteries," en, *Journal of Energy Storage*, Jg. 25, S. 100 817, Okt. 2019, ISSN: 2352152X. DOI: 10.1016/j.est.2019.100817. Adresse: https://linkinghub.elsevier.com/retrieve/pii/S2352152X19302233 (besucht am 07. 10. 2022).

[20] Z. Xia und J. A. Abu Qahouq, „Adaptive and Fast State of Health Estimation Method for Lithium-ion Batteries Using Online Complex Impedance and Artificial Neural Network," in *2019 IEEE Applied Power Electronics Conference and Exposition (APEC)*, Anaheim, CA, USA: IEEE, März 2019, S. 3361–3365, ISBN: 978-1-5386-8330-9. DOI: 10.1109/APEC.2019.8721906. Adresse: https://ieeexplore.ieee.org/document/8721906/ (besucht am 07. 10. 2022).

[21] J. Schmidhuber, „Deep Learning in Neural Networks: An Overview," 2014, Publisher: arXiv Version Number: 4. DOI: 10.48550/ARXIV.1404.7828. Adresse: https://arxiv.org/abs/1404.7828 (besucht am 07. 10. 2022).

[22] *Practical Grey-box Process Identification* (Advances in Industrial Control), en. Springer London, 2006, ISBN: 978-1-84628-402-1. DOI: 10.1007/1-84628-403-1. Adresse: http://link.springer.com/10.1007/1-84628-403-1 (besucht am 07. 10. 2022).

[23] P. Keil und A. Jossen, „Aufbau und Parametrierung von Batteriemodellen," de, S. 12,

[24] V. Johnson, „Battery performance models in ADVISOR," en, *Journal of Power Sources*, Jg. 110, Nr. 2, S. 321–329, Aug. 2002, ISSN: 03787753. DOI:

10.1016/S0378-7753(02)00194-5. Adresse: https://linkinghub.elsevier.com/retrieve/pii/S0378775302001945 (besucht am 07. 10. 2022).

[25] J. Meng, D.-I. Stroe, M. Ricco, G. Luo und R. Teodorescu, „A Simplified Model-Based State-of-Charge Estimation Approach for Lithium-Ion Battery With Dynamic Linear Model," *IEEE Transactions on Industrial Electronics*, Jg. 66, Nr. 10, S. 7717–7727, Okt. 2019, ISSN: 0278-0046, 1557-9948. DOI: 10.1109/TIE.2018.2880668. Adresse: https://ieeexplore.ieee.org/document/8536907/ (besucht am 07. 10. 2022).

[26] G. Welch und G. Bishop, „An Introduction to the Kalman Filter," en, S. 16, 2006.

[27] Q. Li, R. Li, K. Ji und W. Dai, „Kalman Filter and Its Application," in *2015 8th International Conference on Intelligent Networks and Intelligent Systems (ICINIS)*, Tianjin, China: IEEE, Nov. 2015, S. 74–77, ISBN: 978-1-4673-8221-2. DOI: 10.1109/ICINIS.2015.35. Adresse: http://ieeexplore.ieee.org/document/7528889/ (besucht am 07. 10. 2022).

[28] G. L. Plett, „Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs," en, *Journal of Power Sources*, Jg. 134, Nr. 2, S. 252–261, Aug. 2004, ISSN: 03787753. DOI: 10.1016/j.jpowsour.2004.02.031. Adresse: https://linkinghub.elsevier.com/retrieve/pii/S0378775304003593 (besucht am 07. 10. 2022).

[29] R. Bustos, S. A. Gadsden, P. Malysz, M. Al-Shabi und S. Mahmud, „Health Monitoring of Lithium-Ion Batteries Using Dual Filters," en, *Energies*, Jg. 15, Nr. 6, S. 2230, März 2022, ISSN: 1996-1073. DOI: 10.3390/en15062230. Adresse: https://www.mdpi.com/1996-1073/15/6/2230 (besucht am 07. 10. 2022).

[30] B. Johann, *Informatik*, eng. Place of publication not identified: Springer Vienna, 2013, OCLC: 1196195379, ISBN: 978-3-7091-4051-2.

[31] A. Zenati, P. Desprez und H. Razik, „Estimation of the SOC and the SOH of li-ion batteries, by combining impedance measurements with the fuzzy logic inference," in *IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society*, Glendale, AZ, USA: IEEE, Nov. 2010, S. 1773–1778, ISBN: 978-1-4244-5225-5. DOI: 10.1109/IECON.2010.5675408. Adresse: http://ieeexplore.ieee.org/document/5675408/ (besucht am 10. 10. 2022).

[32] Datong Liu, Jianbao Zhou, Haitao Liao, Yu Peng und Xiyuan Peng, „A Health Indicator Extraction and Optimization Framework for Lithium-Ion Battery Degradation Modeling and Prognostics," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Jg. 45, Nr. 6, S. 915–928, Juni 2015, ISSN: 2168-2216, 2168-2232. DOI: 10.1109/TSMC.2015.2389757.

Adresse: http://ieeexplore.ieee.org/document/7018028/ (besucht am 07. 10. 2022).

[33] B. Gou, Y. Xu, S. Fang, R. A. Pratama und S. Liu, „Remaining Useful Life Prediction for Lithium-ion Battery Using Ensemble Learning Method," in *2019 IEEE Power & Energy Society General Meeting (PESGM)*, Atlanta, GA, USA: IEEE, Aug. 2019, S. 1–5, ISBN: 978-1-72811-981-6. DOI: 10.1109/PESGM40551.2019.8973811. Adresse: https://ieeexplore.ieee.org/document/8973811/ (besucht am 07. 10. 2022).

[34] G. Bonaccorso, *Machine learning algorithms: a reference guide to popular algorithms for data science and machine learning*, eng. Birmingham Mumbai: Packt, 2017, ISBN: 978-1-78588-962-2.

[35] A. C. Müller und S. Guido, *Introduction to machine learning with Python: a guide for data scientists*, First edition. Sebastopol, CA: O'Reilly Media, Inc, 2016, OCLC: ocn895728667, ISBN: 978-1-4493-6941-5.

[36] P. Xu, „Review on Studies of Machine Learning Algorithms," *Journal of Physics: Conference Series*, Jg. 1187, Nr. 5, S. 052 103, Apr. 2019, ISSN: 1742-6588, 1742-6596. DOI: 10.1088/1742-6596/1187/5/052103. Adresse: https://iopscience.iop.org/article/10.1088/1742-6596/1187/5/052103 (besucht am 07. 10. 2022).

[37] A. Y. Al-Omary und M. S. Jamil, „A new approach of clustering based machine-learning algorithm," en, *Knowledge-Based Systems*, Jg. 19, Nr. 4, S. 248–258, Aug. 2006, ISSN: 09507051. DOI: 10.1016/j.knosys.2005.10.011. Adresse: https://linkinghub.elsevier.com/retrieve/pii/S0950705106000189 (besucht am 18. 10. 2022).

[38] R. Agrawal, T. Imieliński und A. Swami, „Mining association rules between sets of items in large databases," en, in *Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93*, Washington, D.C., United States: ACM Press, 1993, S. 207–216, ISBN: 978-0-89791-592-2. DOI: 10.1145/170035.170072. Adresse: http://portal.acm.org/citation.cfm?doid=170035.170072 (besucht am 18. 10. 2022).

[39] I. H. Sarker, „Machine Learning: Algorithms, Real-World Applications and Research Directions," en, *SN Computer Science*, Jg. 2, Nr. 3, S. 160, Mai 2021, ISSN: 2662-995X, 2661-8907. DOI: 10.1007/s42979-021-00592-x. Adresse: https://link.springer.com/10.1007/s42979-021-00592-x (besucht am 18. 10. 2022).

[40] E. G. Learned-Miller, „Introduction to Supervised Learning," en, S. 5,

[41] D. Sarkar, R. Bali und T. Sharma, *Practical Machine Learning with Python: A Problem-Solver's Guide to Building Real-World Intelligent Sys-*

tems, 1st ed. 2018. Berkeley, CA: Apress : Imprint: Apress, 2018, ISBN: 978-1-4842-3207-1. DOI: 10.1007/978-1-4842-3207-1.

[42] M. von der Hude, „Prädiktive Verfahren: Klassifikation und Regression,“ de, in *Predictive Analytics und Data Mining*, Wiesbaden: Springer Fachmedien Wiesbaden, 2020, S. 95–97, ISBN: 978-3-658-30152-1 978-3-658-30153-8. DOI: 10.1007/978-3-658-30153-8_7. Adresse: http://link.springer.com/10.1007/978-3-658-30153-8_7 (besucht am 18. 10. 2022).

[43] C. Klingler, M. Feigl, T. Linsbichler, S. Frey und K. Schulz, „Potenzial von Machine Learning bei der kurzfristigen Leistungsprognose innerhalb einer Laufkraftwerkskette,“ de, *Österreichische Wasser- und Abfallwirtschaft*, Jg. 74, Nr. 5-6, S. 224–240, Juni 2022, ISSN: 0945-358X, 1613-7566. DOI: 10.1007/s00506-022-00849-6. Adresse: https://link.springer.com/10.1007/s00506-022-00849-6 (besucht am 18. 10. 2022).

[44] B. M. Abdel-Karim, Hrsg., *Data Science: Best Practices mit Python*, ger. Wiesbaden [Heidelberg]: Springer Vieweg, 2022, ISBN: 978-3-658-33459-8.

[45] *BatteryArchive.org*. Adresse: https://www.batteryarchive.org/snl_study.html (besucht am 09. 10. 2022).

[46] *Battery Data | Center for Advanced Life Cycle Engineering*. Adresse: https://calce.umd.edu/battery-data (besucht am 09. 10. 2022).

[47] M. J. Niedźwiecki, M. Ciołek, A. Gańcza und P. Kaczmarek, „Application of regularized Savitzky–Golay filters to identification of time-varying systems,“ en, *Automatica*, Jg. 133, S. 109 865, Nov. 2021, ISSN: 00051098. DOI: 10.1016/j.automatica.2021.109865. Adresse: https://linkinghub.elsevier.com/retrieve/pii/S000510982100385X (besucht am 11. 12. 2022).

[48] J. M.-T. Wu, M.-H. Tsai, Y. Z. Huang u. a., „Applying an ensemble convolutional neural network with Savitzky–Golay filter to construct a phonocardiogram prediction model,“ en, *Applied Soft Computing*, Jg. 78, S. 29–40, Mai 2019, ISSN: 15684946. DOI: 10.1016/j.asoc.2019.01.019. Adresse: https://linkinghub.elsevier.com/retrieve/pii/S1568494619300237 (besucht am 11. 12. 2022).

[49] *sklearn.model_selection.train_test_split*, en. Adresse: https://scikit-learn/stable/modules/generated/sklearn.model_selection.train_test_split.html (besucht am 24. 10. 2022).

[50] A. Géron, *Praxiseinstieg Machine Learning mit Scikit-Learn, Keras und TensorFlow: Konzepte, Tools und Techniken für intelligente Systeme*, ger, 2. Auflage, übers. von K. Rother und T. Demmig. Heidelberg: O'Reilly, 2020, ISBN: 978-3-96009-124-0.

[51] L. Breiman, J. H. Friedman, R. A. Olshen und C. J. Stone, *Classification And Regression Trees*, en, 1. Aufl. Routledge, Okt. 2017, ISBN: 978-1-

315-13947-0. DOI: 10.1201/9781315139470. Adresse: https://www.taylorfrancis.com/books/9781351460491 (besucht am 11. 12. 2022).

[52] *sklearn.ensemble.RandomForestRegressor*, en. Adresse: https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html (besucht am 11. 12. 2022).

[53] P. Mehta, M. Bukov, C.-H. Wang u. a., „A high-bias, low-variance introduction to Machine Learning for physicists," en, *Physics Reports*, Jg. 810, S. 1–124, Mai 2019, ISSN: 03701573. DOI: 10.1016/j.physrep.2019.03.001. Adresse: https://linkinghub.elsevier.com/retrieve/pii/S0370157319300766 (besucht am 26. 10. 2022).

[54] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems*, eng, Second edition. Sebastopol, CA: O'Reilly, 2019, OCLC: 1099548239, ISBN: 978-1-4920-3261-8.

[55] G. James, D. Witten, T. Hastie und R. Tibshirani, Hrsg., *An introduction to statistical learning: with applications in R* (Springer texts in statistics 103). New York: Springer, 2013, OCLC: ocn828488009, ISBN: 978-1-4614-7137-0.

[56] L. Yang und A. Shami, „On hyperparameter optimization of machine learning algorithms: Theory and practice," en, *Neurocomputing*, Jg. 415, S. 295–316, Nov. 2020, ISSN: 09252312. DOI: 10.1016/j.neucom.2020.07.061. Adresse: https://linkinghub.elsevier.com/retrieve/pii/S0925231220311693 (besucht am 02. 01. 2023).

[57] T. de Graaff, „Realisierung einer Kurzfrist-Prognose des Intraday-Markts mittels verschiedener Machine-Learning-Techniken," de, S. 122,

[58] *1.6. Nearest Neighbors*, en. Adresse: https://scikit-learn/stable/modules/neighbors.html (besucht am 01. 01. 2023).

[59] *Was ist der „k-nearest neighbors algorithm"? | IBM*, de-de. Adresse: https://www.ibm.com/de-de/topics/knn (besucht am 01. 01. 2023).

[60] *sklearn.ensemble.RandomForestRegressor*, en. Adresse: https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html (besucht am 03. 01. 2023).

[61] D. Chicco, M. J. Warrens und G. Jurman, „The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," en, *PeerJ Computer Science*, Jg. 7, e623, Juli 2021, ISSN: 2376-5992. DOI: 10.7717/peerj-cs.623. Adresse: https://peerj.com/articles/cs-623 (besucht am 29. 10. 2022).

[62] A. Bruce und P. Bruce, *ESTATÍSTICA PRÁTICA PARA CIENTISTAS DE DADOS*, und. S.l.: Alta Books, 2019, OCLC: 1162803941, ISBN: 978-85-508-0603-7.

Anhang

Anhang A - Datensätze

Übersicht Datensätze - SNL

Quelle	Type[a]	Sys. Code	Temperatur [°C][b]	C-Rate [C]	DoD [%]
SNL	NMC	1	15	1	100
SNL	NMC	2	15	1	100
SNL	NMC	3	15	2	100
SNL	NMC	4	15	2	100
SNL	NMC	5	25	0.5	100
SNL	NMC	6	25	0.5	100
SNL	NMC	7	25	1	100
SNL	NMC	8	25	1	100
SNL	NMC	9	25	1	100
SNL	NMC	10	25	1	100
SNL	NMC	11	25	2	100
SNL	NMC	12	25	2	100
SNL	NMC	13	25	3	100
SNL	NMC	14	25	3	100
SNL	NMC	24	35	1	100
SNL	NMC	25	35	1	100
SNL	NMC	26	35	1	100
SNL	NMC	27	35	1	100
SNL	NMC	28	35	2	100
SNL	NMC	29	35	2	100

SNL	NCA	1	15	1	100
SNL	NCA	2	15	1	100
SNL	NCA	3	15	2	100
SNL	NCA	4	15	2	100
SNL	NCA	5	25	0.5	100
SNL	NCA	6	25	0.5	100
SNL	NCA	7	25	1	100
SNL	NCA	8	25	1	100
SNL	NCA	9	25	1	100
SNL	NCA	10	25	1	100
SNL	NCA	11	25	2	100
SNL	NCA	12	25	2	100
SNL	NCA	19	35	1	100
SNL	NCA	20	35	1	100
SNL	NCA	21	35	1	100
SNL	NCA	22	35	1	100
SNL	NCA	23	35	2	100
SNL	NCA	24	35	2	100

Quelle: eigene Ausarbeitung nach [45]

[a] Die Abkürzung bezieht sich auf das Kathodenmaterial der Batterien.

[b] Umgebungstemperatur mit ± 0,5°C Toleranz.

Übersicht Datensätze - CALCE

Quelle	**Type**[a]	**Sys. Code**	**Temperatur [°C]**[b]	**C-Rate [C]**	**DoD [%]**
CALCE	LCO	34	25	0.5	100
CALCE	LCO	35	25	1	100
CALCE	LCO	36	25	1	100
CALCE	LCO	37	25	1	100
CALCE	LCO	38	25	1	100

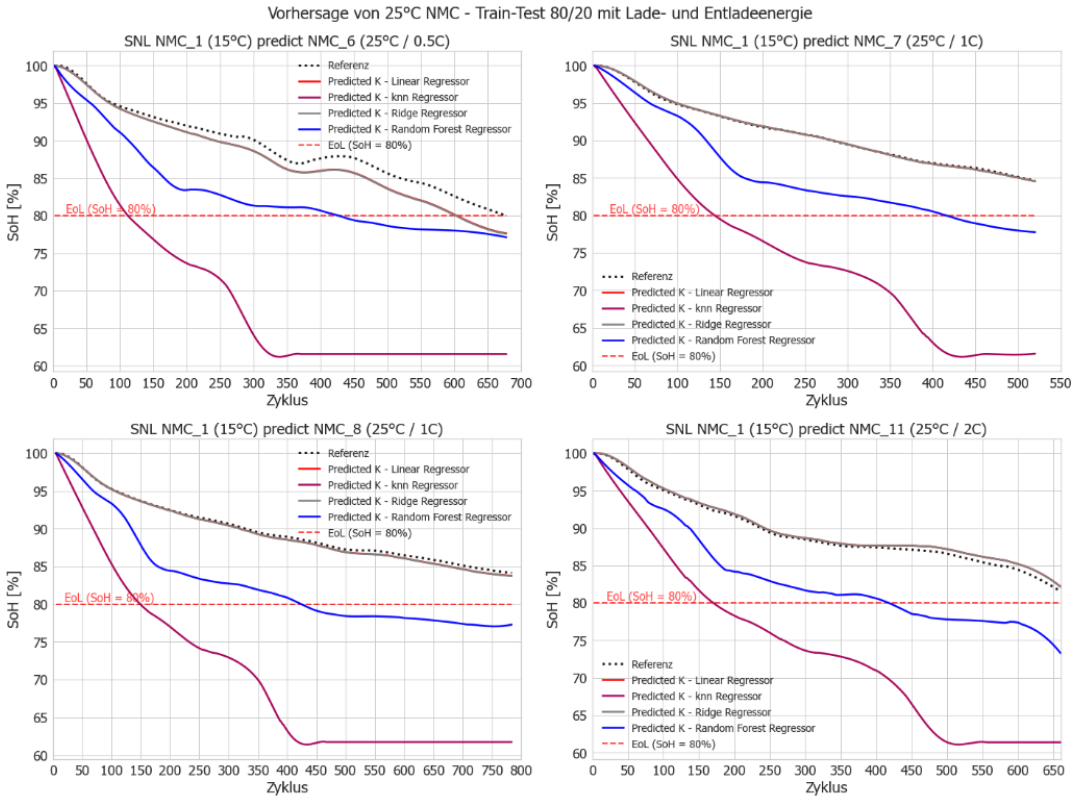
Quelle: eigene Ausarbeitung nach [46]

[a] Die Abkürzung bezieht sich auf das Kathodenmaterial der Batterien.
[b] Umgebungstemperatur mit ± 0,5°C Toleranz.

Anhang B - Ergebnisse

Ergebnisse von SNL NMC-Datensätze

SoH 25°C NMC 6, 7, 8, 11 mit Train-Test 80/20 mit Lade- und Entladeenergie.



Ergebnisse 25°C NMC 6, 7, 8, 11 - Train-Test 80/20 mit Lade- und Entladeenergie Quelle: eigene Ausarbeitung

Übersicht der Ergebnisse 25°C NMC(2) - Train-Test 80/20 mit Lade- und Entladeenergie

SoH-Kurve 25°C-NMC(6)	95%	90%	85%	80%
Referenzkurve	86	295	511	668
Lineare Regression	82	235	461	594
Ridge-Regression	82	235	460	592
knn-Regression	25	49	75	110
RFR-Regression	54	11	164	419

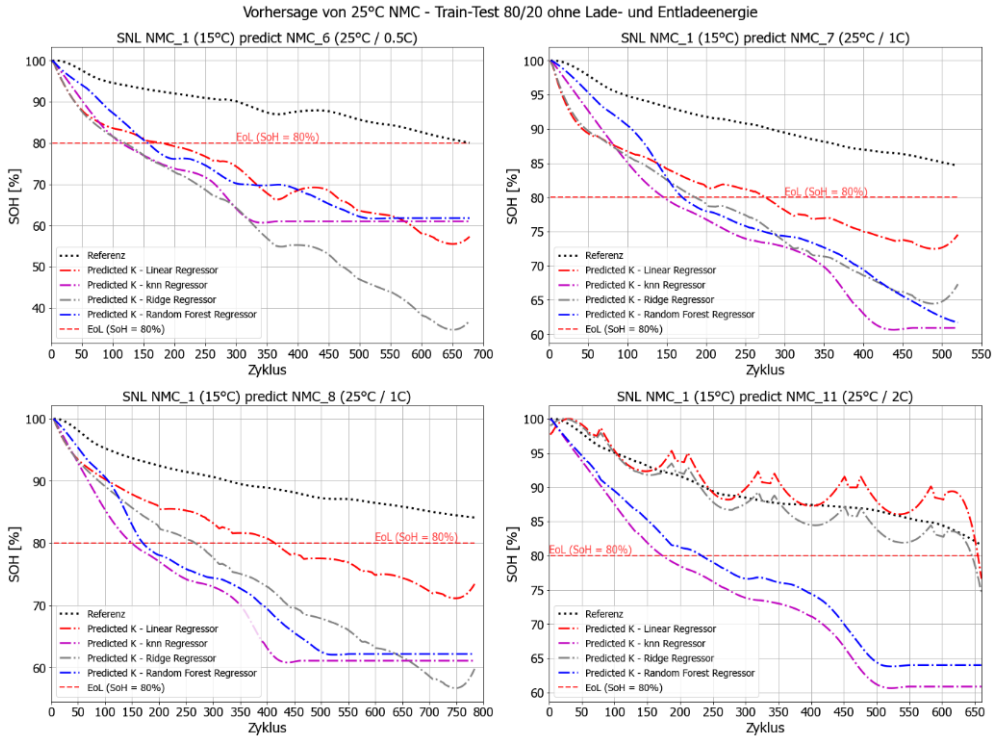
SoH-Kurve 25°C-NMC(7)	95%	90%	85%	80%
Referenzkurve	93	275	495	> 500
Lineare Regression	96	276	483	> 500
Ridge-Regression	96	276	492	> 500
knn-Regression	31	63	97	139
RFR-Regression	65	131	176	401

SoH-Kurve 25°C-NMC(8)	95%	90%	85%	80%
Referenzkurve	102	320	691	> 750
Lineare Regression	101	310	659	> 750
Ridge-Regression	101	310	657	> 750
knn-Regression	32	65	98	145
RFR-Regression	65	130	174	415

SoH-Kurve 25°C-NMC(11)	95%	90%	85%	80%
Referenzkurve	99	232	558	> 650
Lineare Regression	103	237	594	> 650
Ridge-Regression	103	237	594	> 650
knn-Regression	37	76	116	165
RFR-Regression	59	132	174	408

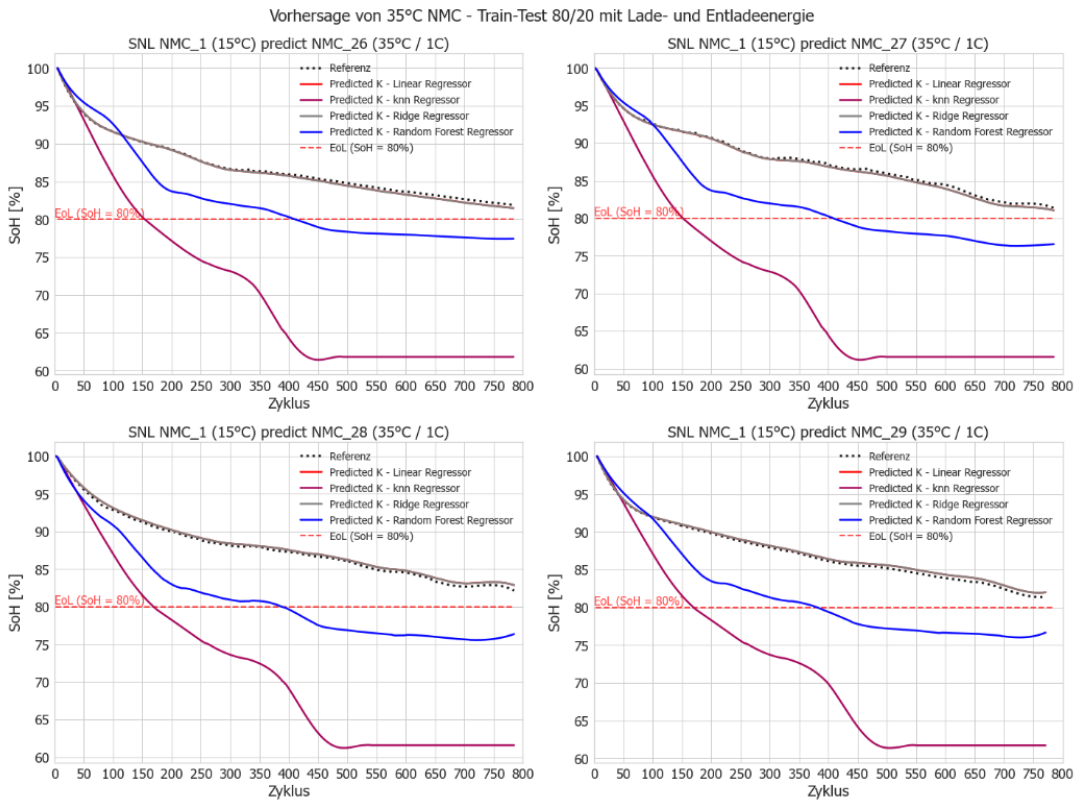
Quelle: eigene Ausarbeitung

SoH 25°C NMC 6, 7, 8, 11 mit Train-Test 80/20 ohne Lade- und Entladeenergie.



Ergebnisse 25°C NMC 6, 7, 8, 11 - Train-Test 80/20 mit Lade- und Entladeenergie Quelle: eigene Ausarbeitung

SoH 35°C NMC 26, 27, 28, 29 mit Train-Test 80/20 mit Lade- und Entladeenergie.



Ergebnisse 35°C NMC 26, 27, 28, 29 - Train-Test 80/20 mit Lade- und Entladeenergie Quelle: eigene Ausarbeitung

Übersicht der Ergebnisse 35°C NMC 26, 27, 28, 29 - Train-Test 80/20 mit Lade- und Entladeenergie

SoH-Kurve 35°C-NMC(26)	95%	90%	85%	80%
Referenzkurve	35	152	472	> 800
Lineare Regression	36	155	449	> 800
Ridge-Regression	36	155	448	> 800
knn-Regression	34	65	101	150
RFR-Regression	54	122	172	401

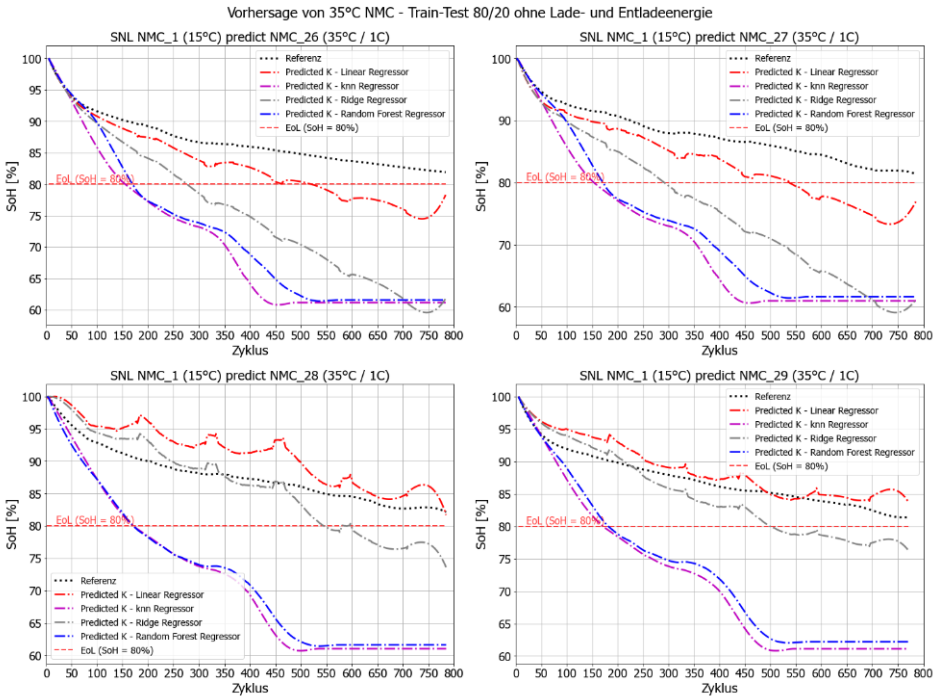
SoH-Kurve 35°C-NMC(27)	95%	90%	85%	80%
Referenzkurve	43	221	540	> 800
Lineare Regression	43	217	525	> 800
Ridge-Regression	43	217	525	> 800
knn-Regression	35	67	103	150
RFR-Regression	55	124	174	399

SoH-Kurve 35°C-NMC(28)	95%	90%	85%	80%
Referenzkurve	55	194	533	> 800
Lineare Regression	59	203	558	> 800
Ridge-Regression	59	203	558	> 800
knn-Regression	37	74	113	166
RFR-Regression	38	108	164	381

SoH-Kurve 35°C-NMC(29)	95%	90%	85%	80%
Referenzkurve	37	184	502	> 780
Lineare Regression	38	192	534	> 780
Ridge-Regression	38	193	535	> 780
knn-Regression	37	74	113	165
RFR-Regression	48	115	169	372

Quelle: eigene Ausarbeitung

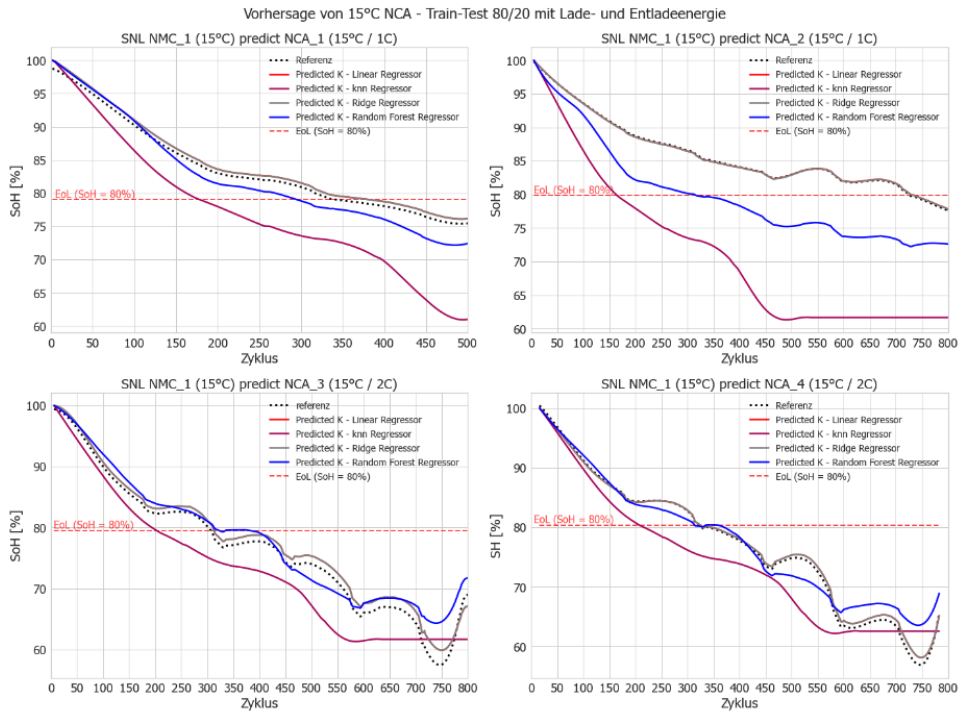
SoH 35°C NMC 26, 27, 28, 29 mit Train-Test 80/20 ohne Lade- und Entladeenergie.



Ergebnisse 35°C NMC 26, 27, 28, 29 - Train-Test 80/20 ohne Lade- und Entladeenergie Quelle: eigene Ausarbeitung

Ergebnisse von SNL NCA-Datensätze

SoH 15°C NCA 1, 2, 3, 4 mit Train-Test 80/20 mit Lade- und Entladeenergie.



Ergebnisse 15°C NCA 1, 2, 3, 4 - Train-Test 80/20 mit Lade- und Entladeenergie Quelle: eigene Ausarbeitung

Übersicht der Ergebnisse 15°C NCA 1, 2, 3, 4 - Train-Test 80/20 mit Lade- und Entladeenergie

SoH-Kurve 15°C-NCA(1)	95%	90%	85%	80%
Referenzkurve	47	100	164	309
Lineare Regression	56	109	174	321
Ridge-Regression	56	108	174	321
knn-Regression	36	71	109	160
RFR-Regression	55	105	149	260

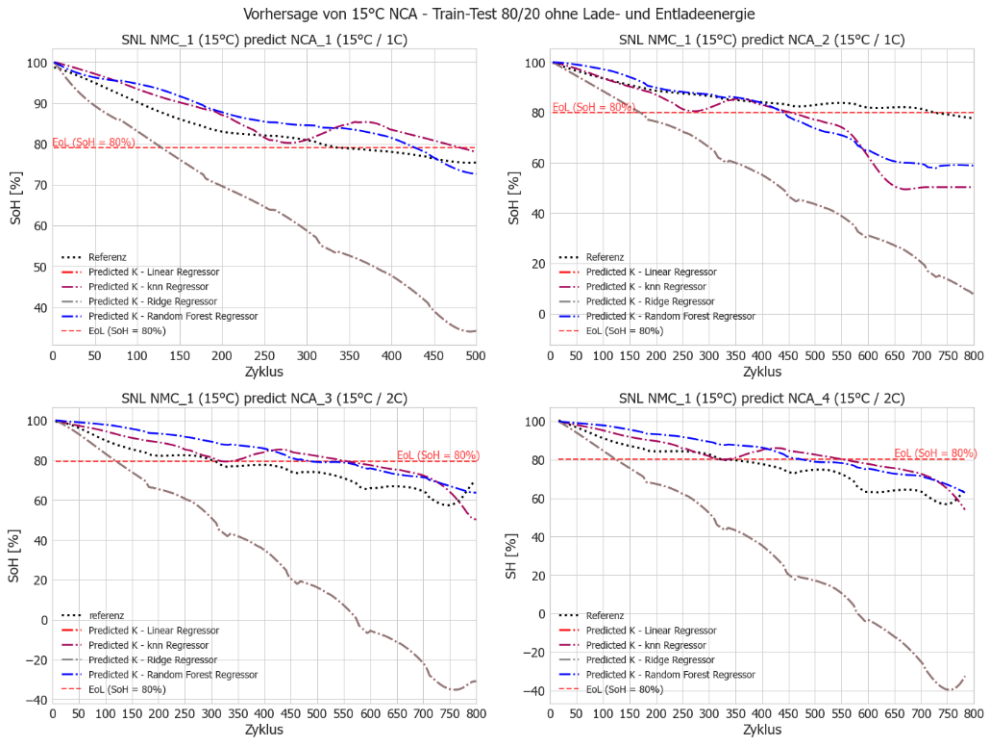
SoH-Kurve 15°C-NCA(2)	95%	90%	85%	80%
Referenzkurve	71	165	339	712
Lineare Regression	70	163	327	715
Ridge-Regression	70	162	326	714
knn-Regression	36	70	108	159
RFR-Regression	50	114	159	305

SoH-Kurve 15°C-NCA(3)	95%	90%	85%	80%
Referenzkurve	56	95	153	294
Lineare Regression	61	102	165	303
Ridge-Regression	61	102	165	303
knn-Regression	40	82	126	188
RFR-Regression	65	117	173	305

SoH-Kurve 15°C-NCA(4)	95%	90%	85%	80%
Referenzkurve	51	98	165	309
Lineare Regression	48	95	165	312
Ridge-Regression	48	95	165	312
knn-Regression	42	83	128	201
RFR-Regression	52	108	163	352

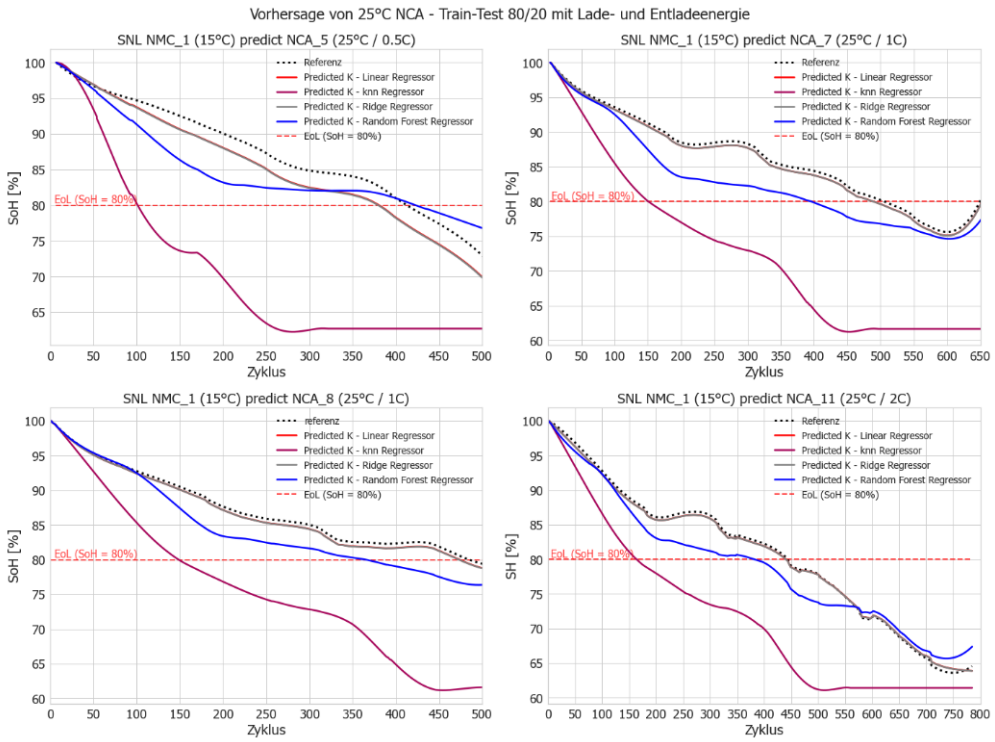
Quelle: eigene Ausarbeitung

SoH 15°C NCA 1, 2, 3, 4 mit Train-Test 80/20 ohne Lade- und Entladeenergie.



Ergebnisse 15°C NCA 1, 2, 3, 4 - Train-Test 80/20 ohne Lade- und Entladeenergie Quelle: eigene Ausarbeitung

SoH 25°C NCA 5, 7, 8, 11 mit Train-Test 80/20 mit Lade- und Entladeenergie.



Ergebnisse 25°C NCA 5, 7, 8, 11 - Train-Test 80/20 mit Lade- und Entladeenergie Quelle: eigene Ausarbeitung

Übersicht der Ergebnisse 25°C NCA 5, 7, 8, 11 - Train-Test 80/20 mit Lade- und Entladeenergie

SoH-Kurve 25°C-NCA(5)	95%	90%	85%	80%
Referenzkurve	85	194	285	406
Lineare Regression	73	156	244	373
Ridge-Regression	72	155	243	372
knn-Regression	38	56	73	94
RFR-Regression	57	105	163	415

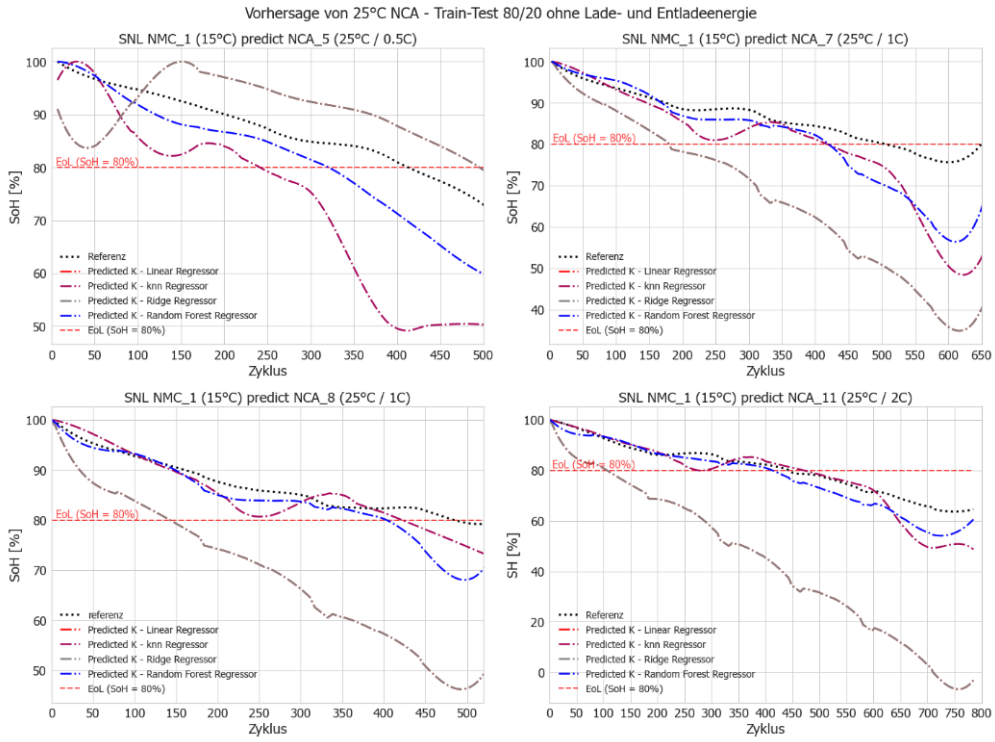
SoH-Kurve 25°C-NCA(7)	95%	90%	85%	80%
Referenzkurve	64	172	356	497
Lineare Regression	59	163	334	481
Ridge-Regression	59	162	334	480
knn-Regression	34	66	101	148
RFR-Regression	54	123	172	387

SoH-Kurve 25°C-NCA(8)	95%	90%	85%	80%
Referenzkurve	54	159	296	478
Lineare Regression	50	151	264	467
Ridge-Regression	50	150	262	466
knn-Regression	34	66	101	147
RFR-Regression	55	125	173	360

SoH-Kurve 25°C-NCA(11)	95%	90%	85%	80%
Referenzkurve	72	132	312	437
Lineare Regression	65	126	309	434
Ridge-Regression	65	126	309	434
knn-Regression	37	73	112	163
RFR-Regression	57	122	172	376

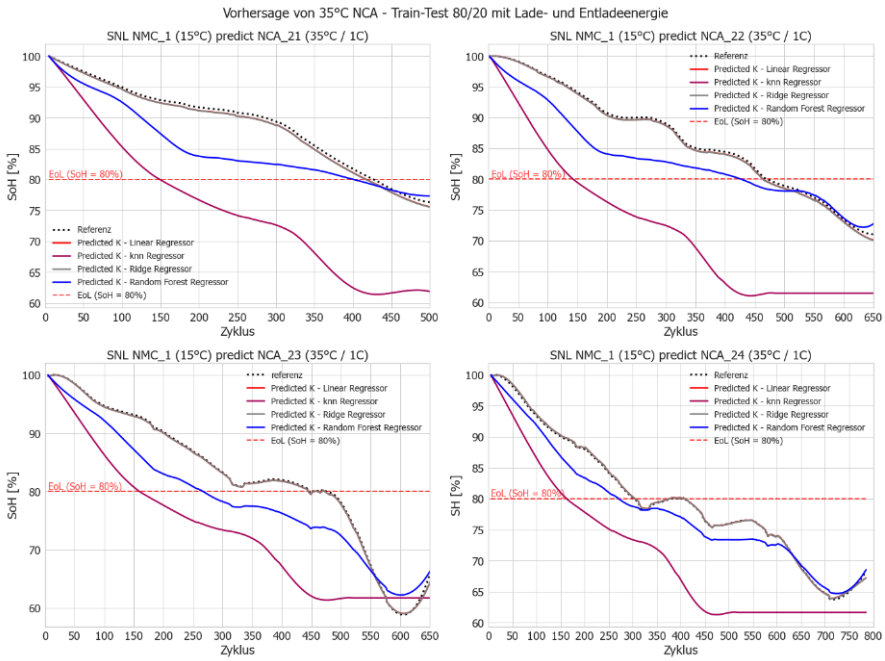
Quelle: eigene Ausarbeitung

SoH 25°C NCA 5, 7, 8, 11 mit Train-Test 80/20 ohne Lade- und Entladeenergie.



Ergebnisse 25°C NCA 5, 7, 8, 11 - Train-Test 80/20 ohne Lade- und Entladeenergie Quelle: eigene Ausarbeitung

SoH 35°C NCA 21, 22, 23, 24 mit Train-Test 80/20 mit Lade- und Entladeenergie.



Ergebnisse 35°C NCA 21, 22, 23, 24 - Train-Test 80/20 mit Lade- und Entladeenergie
Quelle: eigene Ausarbeitung

Übersicht der Ergebnisse 35°C NCA 21, 22, 23, 24 - Train-Test 80/20 mit Lade- und Entladeenergie

SoH-Kurve 35°C-NCA(21)	95%	90%	85%	80%
Referenzkurve	96	278	349	420
Lineare Regression	88	259	342	410
Ridge-Regression	88	257	341	410
knn-Regression	33	64	98	145
RFR-Regression	54	121	171	393

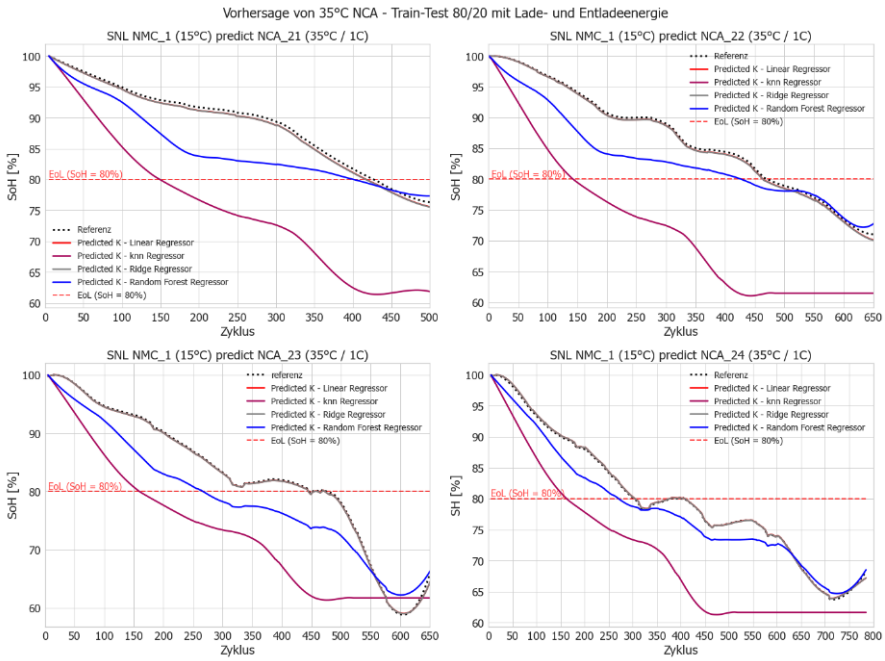
SoH-Kurve 35°C-NCA(22)	95%	90%	85%	80%
Referenzkurve	135	223	347	465
Lineare Regression	131	208	337	458
Ridge-Regression	131	207	336	457
knn-Regression	33	64	98	142
RFR-Regression	66	128	178	420

SoH-Kurve 35°C-NCA(23)	95%	90%	85%	80%
Referenzkurve	91	200	270	437
Lineare Regression	87	197	268	436
Ridge-Regression	87	197	268	436
knn-Regression	36	69	106	155
RFR-Regression	57	119	168	260

SoH-Kurve 35°C-NCA(24)	95%	90%	85%	80%
Referenzkurve	78	147	238	295
Lineare Regression	80	151	240	298
Ridge-Regression	80	151	239	298
knn-Regression	36	71	108	159
RFR-Regression	59	115	168	264

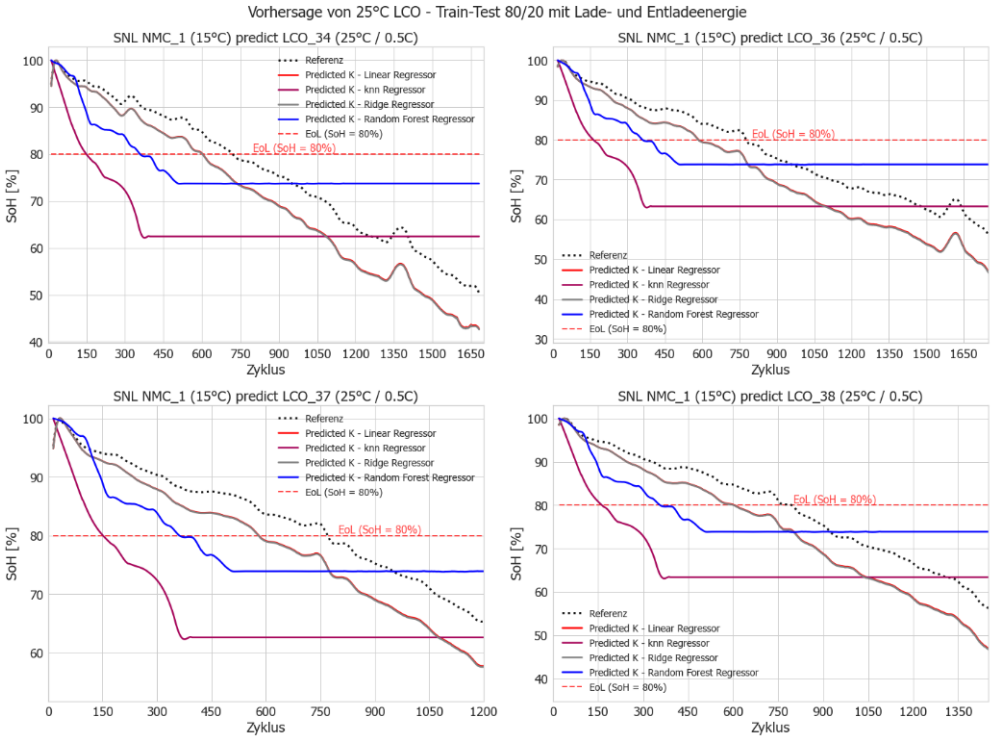
Quelle: eigene Ausarbeitung

SoH 35°C NCA 21, 22, 23, 24 mit Train-Test 80/20 ohne Lade- und Entladeenergie.



Ergebnisse 35°C NCA 21, 22, 23, 24 - Train-Test 80/20 ohne Lade- und Entladeenergie
Quelle: eigene Ausarbeitung

SoH 25°C LCO 34, 36, 37, 38 mit Train-Test 80/20 mit Lade- und Entladeenergie.



Ergebnisse 25°C LCO 34, 36, 37, 38 - Train-Test 80/20 mit Lade- und Entladeenergie Quelle: eigene Ausarbeitung

Übersicht der Ergebnisse 25°C LCO 34, 36, 37, 38 - Train-Test 80/20 mit Lade- und Entladeenergie

SoH-Kurve 25°C-LCO(34)	95%	90%	85%	80%
Referenzkurve	148	335	581	704
Lineare Regression	90	240	418	589
Ridge-Regression	89	239	416	587
knn-Regression	30	59	89	136
RFR-Regression	100	129	234	345

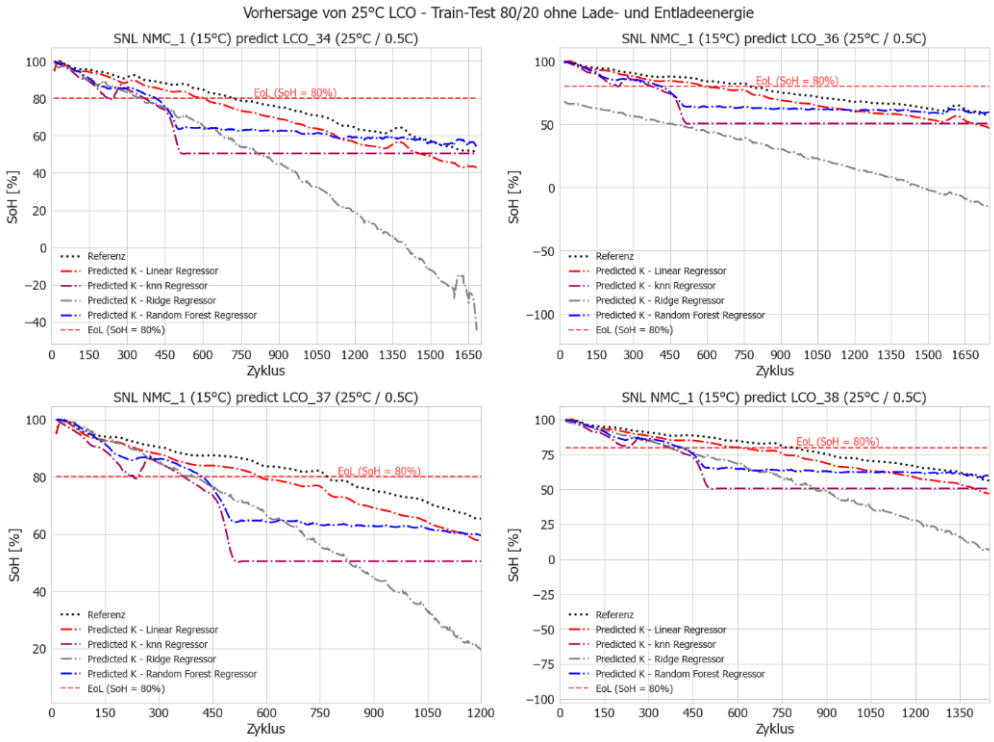
SoH-Kurve 25°C-LCO(36)	95%	90%	85%	80%
Referenzkurve	109	295	552	753
Lineare Regression	82	221	349	563
Ridge-Regression	82	220	348	561
knn-Regression	30	58	89	144
RFR-Regression	93	125	231	339

SoH-Kurve 25°C-LCO(37)	95%	90%	85%	80%
Referenzkurve	88	303	554	751
Lineare Regression	76	221	354	563
Ridge-Regression	76	220	352	562
knn-Regression	30	58	89	137
RFR-Regression	99	130	240	345

SoH-Kurve 25°C-LCO(38)	95%	90%	85%	80%
Referenzkurve	117	307	540	770
Lineare Regression	83	226	379	581
Ridge-Regression	83	225	371	576
knn-Regression	30	57	88	143
RFR-Regression	91	124	232	334

Quelle: eigene Ausarbeitung

SoH 25°C LCO 34, 36, 37, 38 mit Train-Test 80/20 ohne Lade- und Entladeenergie.



Ergebnisse 25°C LCO 34, 36, 37, 38 - Train-Test 80/20 ohne Lade- und Entladeenergie Quelle: eigene Ausarbeitung

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Masterarbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Stellen sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder in gleicher noch in ähnlicher Form einer anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Dornbirn, am 10 Januar 2023 Rafael Pires Mimoso