Technical University of Munich
TUM School of Engineering and Design

Data-Driven Analysis of Interdependencies in Electrode Manufacturing of Lithium-Ion Battery Cells

Sajedeh Haghi

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Ingenieurwissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitz: Prof. Dr.-Ing. Karsten Stahl

Prüfende der Dissertation:

1. Prof. Dr.-Ing. Rüdiger Daub
2. Prof. James Marco

Die Dissertation wurde am 21.05.2024 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 04.10.2024 angenommen.

Editors' Preface

In times of global challenges, such as climate change, the transformation of mobility, and an ongoing demographic change, production engineering is crucial for the sustainable advancement of our industrial society. The impact of manufacturing companies on the environment and society is highly dependent on the equipment and resources employed, the production processes applied, and the established manufacturing organization. The company's full potential for corporate success can only be taken advantage of by optimizing the interaction between humans, operational structures, and technologies. The greatest attention must be paid to becoming as resource-saving, efficient, and resilient as possible to operate flexibly in the volatile production environment.

Remaining competitive while balancing the varying and often conflicting priorities of sustainability, complexity, cost, time, and quality requires constant thought, adaptation, and the development of new manufacturing structures. Thus, there is an essential need to reduce the complexity of products, manufacturing processes, and systems. Yet, at the same time, it is also vital to gain a better understanding and command of these aspects.

The research activities at the Institute for Machine Tools and Industrial Management (*iwb*) aim to continuously improve product development and manufacturing planning systems, manufacturing processes, and production facilities. A company's organizational, manufacturing, and work structures, as well as the underlying systems for order processing, are developed under strict consideration of employee-related requirements and sustainability issues. However, the use of computer-aided and artificial intelligence-based methods and the necessary increasing degree of automation must not lead to inflexible and rigid work organization structures. Thus, questions concerning the optimal integration of ecological and social aspects in all planning and development processes are of utmost importance.

The volumes published in this book series reflect and report the results from the research conducted at *iwb*. Research areas covered span from the design and development of manufacturing systems to the application of technologies in manufacturing and assembly. The management and operation of manufacturing systems, quality assurance, availability, and autonomy are overarching topics affecting all areas of our research. In this series, the latest results and insights from our application-oriented research are published, and it is intended to improve knowledge transfer between academia and a wide industrial sector.

Rüdiger Daub Gunther Reinhart Michael Zäh

Acknowledgements

This dissertation is the result of my work as a research associate at the Institute for Machine Tools and Industrial Management (*iwb*) at the Technical University of Munich (TUM). First of all, I would like to thank Professor Gunther Reinhart and Professor Michael F. Zäh for providing me with this opportunity. Most importantly, I would like to express my gratitude to Professor Rüdiger Daub for taking over the supervision of my dissertation and for his continuous support and mentorship. I am also deeply thankful to Professor James Marco for hosting me at the University of Warwick, offering invaluable guidance, and evaluating this dissertation. Additionally, I would like to thank Professor Karsten Stahl for coordinating the examination committee.

My gratitude extends to all my former colleagues and fellow researchers at the *iwb* for their continuous support and fruitful discussions. In particular, I would like to thank Josef Keilhofer and Matthias Leeb for their critical review of this dissertation and valuable feedback. I am also thankful to the members of the Battery Systems Research Group at the University of Warwick for the engaging discussions and their collaborative spirit during my research stay. I would like to express my deep appreciation to all my students who have been part of this research journey and contributed to this project. Additionally, I would like to thank Dr. Malte Schönemann, my colleague at PowerCo SE, for his insightful comments during the finalization of this dissertation.

Furthermore, I am thankful to my friends, both near and far, for providing the necessary distraction from work and supporting me during the dissertation, often without even knowing it. Above all, my heartfelt gratitude goes to my family. My parents and sister have always been there for me, offering a constant source of strength. I am truly grateful for your immense love and support throughout this journey and beyond.

Brunswick, November 2024 Sajedeh Haghi

Abstract

Battery technology plays a key role in driving the automotive and energy sectors toward achieving net-zero emissions. However, the production of lithium-ion batteries, which are central to this transformation, poses significant challenges. Electrode manufacturing, as the core of battery cell production, is a complex process chain characterized by numerous interrelated variables. This complexity, combined with high material costs, results in a costly scrap in battery cell production. To achieve a cost-effective, quality-oriented optimization of the process chain, a comprehensive understanding of the quality-relevant variables, the existing interdependencies, and the collective impact on intermediate and final product properties is essential. In light of recent advancements in digitalization and information technology, data-driven approaches have emerged as a promising solution to address these challenges.

Within this context, this dissertation aims to facilitate the development of data-driven models for the analysis of interdependencies in electrode manufacturing. For this purpose, a holistic framework is developed, encompassing crucial aspects that include the identification of quality-relevant parameters, the possibility of inline collection of the parameters, data generation and evaluation, model development, and the derivation of insights. The overarching objective is to enable a holistic, efficient, and quality-oriented analysis of interdependencies in electrode manufacturing. The application of the proposed framework is demonstrated through use cases based on data generated at a research pilot production line.

Contents

List of Abbreviations

AI	Artificial Intelligence
ALE	Accumulated Local Effects
ANOVA	Analysis of Variance
BI	Business Intelligence
CPPS	Cyber-Physical Production System
CRediT	Contributor Roles Taxonomy
CRISP-DM	Cross-Industry Standard Process for Data Mining
DIN	German Institute for Standardization (in German: Deutsches Institut für Normung)
DM	Data Mining
DoE	Design of Experiments
DRM	Design Research Methodology
DSM	Design Structure Matrix
DT	Decision Tree
EIS	Electrochemical Impedance Spectroscopy
ESC	Electrode Seperator Composite
FDS	Fraction of Design Space
FI	Feature Importance
FMEA	Failure Mode and Effect Analysis
GBT	Gradient Boosted Trees
ISO	International Organization for Standardization
iwb	Institute for Machine Tools and Industrial Management (in German: Institut für Werkzeugmaschinen und Betriebswissenschaften)
KDD	Knowledge Discovery in Databases
KPI	Key Performance Indicator
LFP	Lithium Iron Phosphate
LIB	Lithium-Ion Battery
LR	Linear Regression
LTO	Lithium Titanite

MAE	Mean Absolute Error
MDI	Mean Decrease of Impurity
ML	Machine Learning
MLR	Multiple Linear Regression
MSE	Mean Squared Error
NCA	Nickel Cobalt Aluminum
NIST	National Institute of Standards and Technology
NMC	Nickel Manganese Cobalt
NN	Neural Network
OFAT	One-Factor-At-a-Time
PCA	Principle Component Analysis
PolyR	Polynomial Regression
R^2	R-squared value, also known as coefficient of determination
RF	Random Forest
RMSE	Root Mean Squared Error
RSM	Response Surface Methodology
SEI	Solid Electrolyte Interphase
SHAP	SHapley Additive exPlanations
SISSO	Sure Independent Screening and Sparsifying Operator
SO	Sub-Objective
SoH	State of Health
SVM	Support Vector Machine
VIF	Variance Inflation Factor
XAI	eXplainable Artificial Intelligence
XGBoost	eXtreme Gradient Boosting
XML	eXplainable Machine Learning

List of Notations

e	Unobservable random variable, also known as the error component
k	Number of subsets or folds into which the dataset is divided in a cross-validation approach
M	Total number of elements of a system being analyzed
n	Degree of polynomial
N	Number of data points or samples
p	Total number of independent variables
R_i^2	Coefficient of determination when predicting x_i using all other predictors x_j
x	Independent variable, also referred to as predictor variable
x_j	All other predictor variables except x_i
y	Dependent variable, also referred to as response or target variable
$\bar{y}$	Mean of the actual values of the dependent variable
y_i	Actual value of the dependent variable for the i-th data point
$\hat{y}_i$	Predicted value of the dependent variable for the i-th data point
β_0	Intercept term
β_i	Coefficients of the predictor variables

List of Symbols

Greek Letters

α	Bruggeman exponent, unit: -
ϵ	Porosity, unit: %
ρ	Density, unit: $\mathrm{g\,cm^{-3}}$
ρ_{bulk}	Bulk density, unit: $\mathrm{g\,cm^{-3}}$
ρ_{com}	Composite density, unit: $\mathrm{g\,cm^{-3}}$
τ	Tortuosity, unit: -

Latin Letters

C-rate	Charge or discharge capacity, unit: $\mathrm{h^{-1}}$
C_{n}	Nominal capacity, unit: mAh
d_{coating}	Coating thickness, unit: μm
I	Electrical current, unit: A
L_{cv}	Length of the control volume, unit: μm
L_{p}	Ion transport path length, unit: μm
M_{coating}	Coating mass loading, unit: $\mathrm{mg\,cm^{-2}}$
R	Resistance, unit: Ω

Chapter 1

Introduction

1.1 Motivation

With the rise of global warming and natural disasters, the United Nations Framework Convention on Climate Change established the Paris Climate Agreement. This international pact aims to combat climate change and accelerate the necessary actions and investments for a sustainable, low-carbon future. It involves both individual and collective commitments with the overarching goal of limiting global warming to below 2 °C, preferably to 1.5 °C, compared to pre-industrial levels. Achieving this ambitious target necessitates a substantial reduction in greenhouse gas emissions, approximately 45 % below the 2010 levels, by 2030, and ultimately net-zero emissions by 2050. (UNITED NATIONS 2022)

One of the key contributors to a sustainable, low-carbon future is the widespread adoption of electromobility. However, to make electromobility a viable option for the broader population, there are still certain challenges that need to be addressed, including improving the performance and affordability of the Lithium-Ion Battery (LIB) cell technology as the key component (KWADE et al. 2018, p. 290). A recent report published by BloombergNEF, highlighted a disruption in the decreasing trend of battery prices, as the volume-weighted average price of LIB cell reached an 11 % increase in 2022, marking the first rise in a decade, driven primarily by the material costs (BLOOMBERG 2022). This unexpected increase emphasizes the need for efficient and quality-oriented production processes. Effective strategies to reduce costs and improve cell quality are essential to accelerate the global transition toward electromobility and ultimately contribute to achieving the net-zero vision.

Electrode manufacturing is a critical phase in battery cell production, determining the majority of the electrochemical and mechanical properties of the cell. The process chain is inherently complex due to the large number of parameters involved and their interdependencies. This complexity is further compounded by the largely unknown relationships between individual processes and the properties of the intermediate and final products. (GÜNTHER et al. 2016)

Over the last few years, various approaches have been adopted to address this complexity and enhance the process understanding through analysis of the existing cause-and-effect relationships. These approaches range from expert-based methods (WESTERMEIER 2016), analytical (BILLOT 2021) and simulation models (THOMITZEK 2022) to data-driven approaches (TURETSKYY 2022). Among these approaches, the data-driven solutions hold particular significance due to their unique advantages, such as the ability to quantify complex, nonlinear, and potentially unknown interdependencies as well as enabling low-latency inline process monitoring and optimization (MUNIR et al. 2021; SHEN et al. 2007).

The development of data-driven solutions involves several critical steps. These steps typically include identifying relevant parameters, generating and collecting data, selecting appropriate methods for analysis, and ultimately developing models to extract insights. A set of generic guidelines, such as Cross-Industry Standard Process for Data Mining (CRISP-DM) (CHAPMAN et al. 2000) and Knowledge Discovery in Databases (KDD) (FAYYAD et al. 1996), has been proposed to structure the execution of data-driven projects. While such guidelines can support practitioners in developing data-driven solutions, they may encounter certain limitations when dealing with highly complex use cases, such as battery cell production. To address these limitations, practitioners in specialized domains often need to augment the generic approaches with domain-specific knowledge and expertise. These enhancements aim to elaborate on the approaches, taking into account the unique requirements and challenges of the domain, thus ensuring a more effective solution development.

1.2 Objective of the Dissertation

Given the existing challenges and the potential described, the overarching objective of this dissertation is defined as follows:

To facilitate the development of data-driven models for a holistic, efficient, and quality-oriented analysis of interdependencies in electrode manufacturing of LIB cells

In this context, the term holistic refers to a comprehensive analysis encompassing two dimensions. Firstly, it covers the steps required for the development of data-driven models, from the identification of quality-relevant parameters to data generation, model development, evaluation, and derivation of insights. Secondly, the focus of the dissertation extends beyond a single process step or specific aspect, aiming to provide the foundation to explore the cross-process interdependencies along the process chain. The efficiency concentrates on the prioritization of quality-relevant parameters. This emphasis is strengthened by incorporating methods that support an in-depth process understanding while concurrently minimizing the number of experiments required. By providing a holistic framework, the development and implementation of data-driven models can be streamlined, with the ultimate goal of enhancing process understanding and achieving improvements in product quality.

1.3 Research Methodology and Structure of the Dissertation

The research conducted within this dissertation was guided by the Design Research Methodology (DRM), a framework conceived to facilitate the planning and execution of research more effectively and efficiently (BLESSING and CHAKRABARTI 2009, p. 14). In the following, a brief description of the four main stages of DRM—Research Clarification, Descriptive Study I, Prescriptive Study, and Descriptive Study II—along with their context in this dissertation, is provided.

The objective of the *Research Clarification* stage is to identify evidence and indications that support the objectives of the research. Based on the findings, an initial description of both the existing and the desired situation is established. (BLESSING and CHAKRABARTI 2009, p. 15) Within the analyzed context, the complexity of electrode manufacturing, due to the manifold interdependencies, coupled with its relevance and impact on cell quality, underlines the importance of the data-driven approaches and the objective of this dissertation.

In the *Descriptive Study I*, a more comprehensive analysis is conducted to explore the current situation and identify the influencing factors (BLESSING and CHAKRABARTI 2009, p. 16). This phase can be conducted using a review-based or empirical approach. The latter is applied when the literature review reveals a deficit in understanding the subject matter. (BLESSING and CHAKRABARTI 2009, p. 80) The presented research work encompasses both methods and can be classified as comprehensive.

The *Prescriptive Study* builds upon the findings of the previous stage and proceeds to develop *support* in the form of knowledge, guidelines, methods, or models aimed at impacting the influencing factors identified in Descriptive Study I (BLESSING and CHAKRABARTI 2009, p. 141). In this context, the results include a framework outlining the quality-relevant parameters, the possibility of inline collection of the parameters, methods for data generation and evaluation, modeling techniques, and data-driven models, along with relevant methods to derive insights from these models. The application of the framework is demonstrated exemplarily, focusing on the interdependencies that have not been investigated in the literature.

The *Descriptive Study II* involves exploring the impact through application, evaluation, and derivation of suggestions for improvement (BLESSING and CHAKRABARTI 2009, p. 181). The proposed framework has been exemplarily implemented in a research pilot line for battery cell production, the application is evaluated according to the defined requirements.

Consequently, among the seven types of research according to DRM, this dissertation can be categorized as type 5 (BLESSING and CHAKRABARTI 2009, p. 60). Guided by the outlined methodology, Figure 1.1 provides an overview of the structure of the dissertation, considering the main DRM stages, the intended objectives, and the adopted approaches for each stage. Chapter 1 presents the motivation and the overarching objective of this dissertation. Chapters 2 and 3 aim to provide a more in-depth understanding of the current situation and identify the research opportunities. Driven by the identified need, the detailed objectives and requirements are established and presented as part of the conceptual design in Chapter 4. The conceptual design builds the foundation for the Prescriptive Study, with the aim of developing a comprehensive solution that effectively addresses the detailed objectives and complies with the specified requirements. In the context of the publication-based dissertation, Chapter 4 offers an overview of the proposed framework and the relevant publications, while Chapter 5 provides a concise summary of the research findings. The final stage of DRM involves evaluating the application of the proposed solution and identifying improvement potentials, which are addressed in Chapters 6 and 7, respectively.

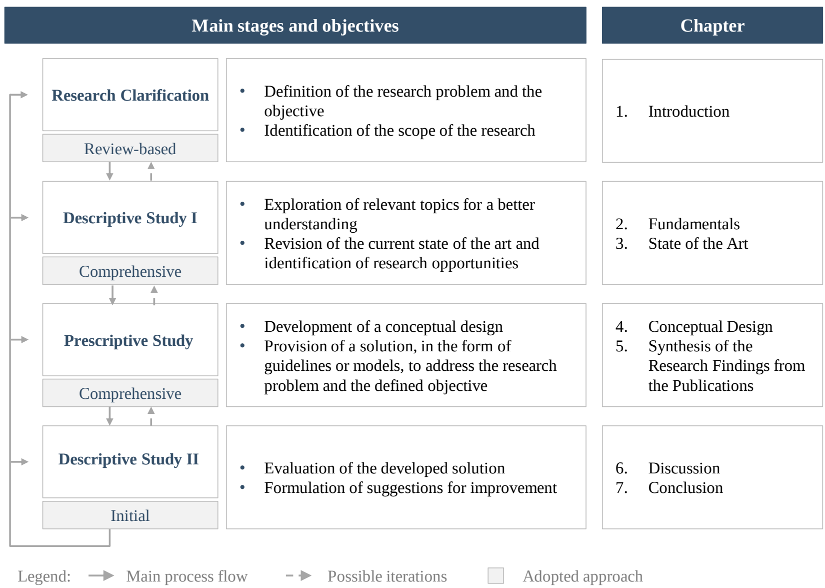

Figure 1.1: Overview of the structure of the dissertation, along with the corresponding main stage in DRM, adapted from BLESSING and CHAKRABARTI (2009)

Chapter 2

Fundamentals

In order to establish a foundational understanding of the subject matter, this chapter begins by clarifying the essential terminology and methods in Section 2.1. Subsequently, Section 2.2 explores the fundamental aspects of LIB as a technology and the production process chain. Additionally, within the scope of this dissertation, a synopsis of the methods used in the fields of Design of Experiments (DoE) and data analytics, particularly Machine Learning (ML), is provided in Sections 2.3 and 2.4, respectively.

2.1 General Terminology and Methods

2.1.1 Approach, Methodology, Framework, and Method

In this subsection, an overview of the general terms used in combination with a problem-solving task is presented. ANDIAPPAN and WAN (2020) elaborated on the different terminologies such as approach, methodology, and method that are used in the process systems engineering field. Given the analyzed context in this dissertation, the definitions are adopted as follows.

Approach

Approach is defined as "[. . .] the basic philosophy concerning a given subject matter. It is a way or direction used to address a problem based on a set of assumptions." (HOFLER 1983, p. 71; ANDIAPPAN and WAN 2020, p. 551)

Methodology or Framework

Methodology refers to the overarching strategy employed to address a particular problem and is seen as a crucial element required to realize an approach. It serves "as a guideline, allowing the practitioner to make choices within a certain set of rules and boundaries." Given this definition, methodology can be equated with the term framework and is understood as a system of methods employed in conjunction with established rules or criteria. (ANDIAPPAN and WAN 2020, p. 551)

Method

Method describes the practical steps required for the implementation of an approach (HOFLER 1983, p. 71). While the terms method and methodology are often used interchangeably, a clear distinction exists between the two. ANDIAPPAN and WAN (2020) clarified this distinction by noting that "if a practitioner engages with a method and follows it like a recipe regardless of the situation, then it remains as a method. If the method is not regarded as a formula, but as a guideline, then it would be clearly classified as a methodology or a framework." (ANDIAPPAN and WAN 2020, p. 552)

2.1.2 Design Structure Matrix

The *Design Structure Matrix (DSM)* is a modeling tool utilized to represent the elements of a system and their corresponding interactions (EPPINGER and BROWNING 2012, p. 2). This method can help to understand a complex system and is based on three main steps: (i) decomposing the system into its constituent elements, (ii) identifying the relationships between these elements, and (iii) defining the boundaries of the system under consideration (BROWNING 2001, p. 292). The DSM is presented as an M×M square matrix with identical labels and order, which maps the interdependencies among the system's M elements. By systematically organizing these interactions, the DSM offers a compact and efficient visualization of the system's structure and interdependencies. (EPPINGER and BROWNING 2012, pp. 2–3)

2.1.3 MoSCoW Analysis

The *MoSCoW analysis* is a prioritization technique first introduced as a component of the Dynamic Systems Development Method, a comprehensive framework for rapid application development (CLEGG and BARKER 1994; STAPLETON 1997). The term MoSCoW is an acronym derived from the four prioritization categories: *Must have*, which are fundamental requirements for the system; *Should have*, important but not mandatory requirements; *Could have*, desirable but not necessary requirements; and *Would have or Won't have this time,* which are requirements of the lowest priority (STAPLETON 1997, p. 29).

2.1.4 Mapping Study

A systematic literature review is conducted with the aim of answering a particular research question that can be addressed empirically. In contrast, a *mapping study* is adopted to explore a more extensive topic and classify the principal findings within a specific domain. (KITCHENHAM et al. 2011, p. 639) The main steps involved in conducting a mapping study include (i) determining the scope of the research, (ii) performing a literature search to identify the primary studies, (iii) screening the studies for inclusion and exclusion, (iv) establishing a classification scheme, and (v) extracting and aggregating data (PETERSEN et al. 2008, pp. 2–5).

2.1.5 Production System and Scalability

A *system* can be defined as a collection of interconnected and purposefully organized elements that continuously interact and collaborate. Similarly, a *production system* is defined as structured assemblies of highly interdependent individual functions, with the overarching objective of producing products. (WESTKÄMPER and ZAHN 2008, p. 28) The primary purpose of a production system is to manufacture the correct products, both in type and quantity, at the right time, with a specified quality, and at an acceptable cost. This underscores the importance of precision in production processes, ensuring that products align with demand specifications, are produced on time to meet market needs, maintain consistent quality standards, and are produced cost-effectively to ensure profitability and value to the business. (WESTKÄMPER 2006, p. 195)

In the context of battery cell production, it is important to recognize the existence of three main production scales, each tailored to specific objectives and distinguished by distinct characteristics. The *lab-scale* focuses primarily on material development and formulation screening. This production system is characterized by discontinuous manual processes and a wide variety of materials and formulations. The *pilot scale* aims to evaluate the scalability of the defined formulations, investigate the suitability of process parameters and their interaction, and optimize the quality control measures. The production system at this level features semi-continuous or continuous processes, often incorporating semi-automated to automated machinery. This stage serves as a vital bridge between the lab-scale and full-scale industrial production. In *industrial production*, the primary focus lies in ensuring process robustness and implementing automated defect detection. (KEPPELER et al. 2021)

2.1.6 Measuring Instrument and Characterization

A *measurement system*, often also referred to as a *measuring instrument*, is defined as "a system that provides information about a variable being measured." (MORRIS and LANGARI 2012, p. 4) In this context, the term *characterization*, according to the National Materials Advisory Board, is defined as "the description of those features of the composition and structure of a material, including defects, that are significant for a particular preparation, study of properties or use, and suffice for the reproduction of the material." (GROVES and WACHTMAN 1986, p. 425) According to OBERKAMPF and ROY (2010, p. 373), characterization involves the measurement of the essential properties required to describe or model a system.

The measuring instruments can be broadly divided into three categories: offline, at-line, and inline. It is important to note that the first two terms are often used interchangeably. An *offline* characterization refers to an analytical method that necessitates the sampling and subsequent transportation of the sample to a specialized device or laboratory. This approach typically demands highly skilled personnel and occurs within centralized laboratory facilities. In offline characterization, the sample is physically separated from the production process and analyzed in a separate location, which may result in a time delay between sampling and analysis, making active process adjustment during the process infeasible. An *at-line* method involves manual or (semi)-automated sampling and measurement conducted in close proximity to the production site. An *inline* system is used to obtain information about process or product characteristics directly and in real-time. In this approach, the measuring instrument is seamlessly integrated into the product flow, allowing immediate and continuous monitoring. (KESSLER 2012, pp. 15–17)

2.2 Lithium-Ion Battery

The LIB is classified as a secondary battery, also known as an accumulator, designed for the generation of electrical energy (WINTER and BRODD 2004, p. 4253). This implies that, in contrast to the primary battery, which is intended for single use and must be discarded upon exhaustion, LIB has the capability to be recharged and reused multiple times (WINTER and BRODD 2004, p. 4247). This feature, combined with a number of advantageous characteristics such as high energy density and long cycle life, turns LIB into a more cost-effective and environmentally sustainable option for various applications ranging from portable electronic devices to electric vehicles and renewable energy storage systems.

2.2.1 Structure and Operating Principles of Battery Cell

The operating principle of the LIB is depicted in Figure 2.1, with a focus on the smallest operational unit of the battery. This essential unit consists of several critical components, each of which plays a significant role in the overall function and performance of the battery. The electrodes consist of a layer of active material coated onto a current collector (VUORILEHTO 2018, p. 23). According to the German Institute for Standardization (DIN), the terminology for the electrodes is based on the discharge process: with the positive electrode referred to as the *cathode*, while the negative one is named the *anode* (DIN40729 1985, p. 3). The two electrodes are electrically isolated from each other through a porous membrane called the *separator*. The separator prevents direct electrical contact between the electrodes while allowing the flow of ions. The ion-conducting *electrolyte* serves as the medium for transporting lithium ions between the electrodes during charge and discharge cycles. (LEUTHNER 2018, p. 14) The process of lithium ions being inserted into the layered structure of the electrode material is called *intercalation*, while the reverse process is referred to as *de-intercalation* (VUORILEHTO 2018, p. 22).

Depending on the application and requirements, various active materials are applied. For the cathodes, the most common materials are Lithium Iron Phosphate (LFP), Nickel Manganese Cobalt (NMC), and Nickel Cobalt Aluminum (NCA) (GRAF 2018). For the anodes, potential materials include graphite, Lithium Titanite (LTO), and graphite-silicon composites. Graphite is widely used due to its prolonged life cycle and safety characteristics. Silicon composites are anticipated to gain significance in the future due to their higher energy density. However, this will only be possible if the existing challenges regarding life cycle and safety are effectively addressed. (WURM et al. 2018)

In addition to the active material, schematically shown in Figure 2.1, conductive additives and binders are the necessary inactive components of the electrode. While the active material is mainly responsible for energy storage, the conductive additives and binders are crucial for electron transport and maintaining the mechanical integrity of the electrode, respectively (LIU et al. 2011, p. 214). The mechanical integrity encompasses the cohesion of the electrode particles and their adhesion to the current collector. In the case of conventional electrode manufacturing, solvents are additionally required to produce a homogeneous, coatable mixture. (VUORILEHTO 2018, p. 23)

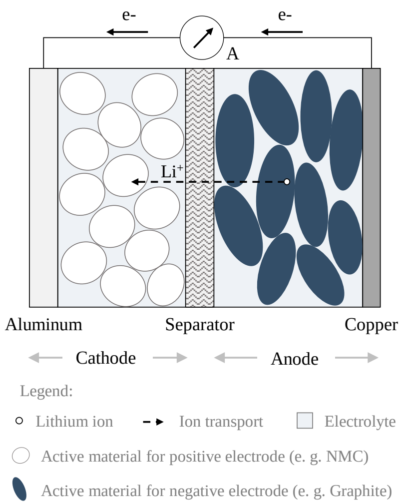

Figure 2.1: Structure and operating principle of a LIB cell, shown during the discharge process, adapted from LEUTHNER (2018, p. 15)

2.2.2 Main Characteristics of Battery Cell and Electrode

The LIB cell, along with its main component, the electrode, can be characterized by several factors. This subsection briefly describes the main characteristics that are relevant to this dissertation.

Capacity, Discharge Capacity, and Nominal Capacity

The *capacity* denotes the overall quantity of electric charge that can be exchanged between the electrodes within a battery. This capacity is influenced by several factors, such as the discharge current, the cut-off voltage, the temperature, and both the type and quantity of active material. It is typically expressed in Ah or mAh. (LEUTHNER 2018, p. 16) Accordingly, the *discharge capacity* specifies the total amount of electric charge that a battery can deliver from a fully charged state. Additionally, according to DIN40729 (1985, p. 10), the term *nominal capacity* is used to report the capacity of the battery based on specific conditions defined by the manufacturer, such as the temperature.

Energy

The *energy* of a battery reflects the amount of electric charge that can be stored and is the product of capacity and average discharge voltage (LEUTHNER 2018, p. 16). The *gravimetric energy density* refers to the energy that can be stored per unit weight of a material (LIN et al. 2018, p. 3) and is expressed in $\mathrm{Wh\,kg^{-1}}$. A higher gravimetric energy density means that the material can hold more energy for its weight, which is particularly important for battery applications where weight is a critical factor.

Power

The *power* indicates the speed at which a battery can supply energy. It is calculated by multiplying the current and voltage, for instance, during discharge, and is expressed in W (LEUTHNER 2018, p. 16). The *power density* quantifies the amount of power that a battery can deliver relative to its unit of volume or mass (WU 2015, p. 8). It is labeled as *volumetric power density* when expressed per unit volume, and as *gravimetric power density* when expressed per unit mass.

C-Rate

The *C-rate* represents the rate at which a battery is charged or discharged relative to its nominal capacity. This measure offers a standardized approach to compare the performance of batteries, regardless of their individual capacities. A discharge rate of 1C indicates that the discharge current will deplete the battery in 1 hour, whereas a 2C discharge rate implies the battery will be discharged in half an hour. (WU 2015, p. 6) The C-rate is calculated as the ratio of the current I to the nominal capacity C_n, as shown in Equation 2.1.

$$C\text{-rate} = \frac{I}{C_n} \tag{2.1}$$

Internal Resistance

The *internal resistance* is a comprehensive term used to describe the total resistance within a battery (WU 2015, p. 6). This resistance can be broken down into four components: electric resistance R_c, electrolyte bulk resistance R_{sol}, ionic resistance in pores R_{ion}, and charge transfer resistance for lithium intercalation R_{ct} (OGIHARA et al. 2012, p. 1034).

Mass Loading

The *mass loading* refers to the amount of material applied per unit area of the electrode (LIN et al. 2018, p. 2). The mass loading of the active material is a crucial parameter in battery design as it directly impacts the capacity of the battery. At small production scales, the mass loading is typically expressed in $\mathrm{mg\,cm^{-2}}$, while at the industrial scale, $\mathrm{g\,m^{-2}}$ is commonly adopted.

Density

The *density* of the electrode, ρ, refers to its volumetric mass density, which can be determined by calculating the composite density of the coated materials (SMEKENS et al. 2016, p. 2). The composite density is calculated by considering the density of each material and its respective fraction in the electrode formulation and is usually expressed in $\mathrm{g\,cm^{-3}}$ (SANTEE et al. 2014, p. 66).

Porosity

The *porosity* represents the estimated theoretical void space within the materials applied to the current collector, quantified as a percentage of the total volume of the electrode (SANTEE et al. 2014, p. 66). Various methods, including mercury intrusion porosimetry, can be employed to determine porosity. Alternatively, porosity can also be estimated based on the main physical properties of the electrode (FROBOESE et al. 2017). Specifically, porosity ϵ can be calculated using the bulk density ρ_{bulk} and the composite density of electrode ρ_{com}, as shown in Equation 2.2. The bulk density of the electrode is derived from the mass loading of the coating $M_{coating}$ and its thickness $d_{coating}$.

$$\epsilon = 1 - \frac{\rho_{bulk}}{\rho_{com}} = 1 - \frac{M_{coating}}{d_{coating} \cdot \rho_{com}} \tag{2.2}$$

Tortuosity

The *tortuosity* quantifies the degree of elongation of the transport path, L_{p}, through a porous structure compared to the straight-line distance or the length of the control volume, L_{cv}. It represents the porous structure of the electrode and its impact on ionic diffusion and is expressed as follows: (TJADEN et al. 2016, pp. 45–46)

$$\tau = \frac{L_{\text{p}}}{L_{\text{cv}}} \tag{2.3}$$

In battery production, a common simplification is adopted to describe the correlation between tortuosity and porosity, as shown in Equation 2.4. This approach, with the exponent α equal to 0.5, was originally postulated for spherical particles and is known as the Bruggeman relation. (BRUGGEMAN 1935; LANDESFEIND et al. 2016)

$$\tau = \epsilon^{-\alpha} \tag{2.4}$$

It should be noted that there are several limitations to this simplification, particularly when considering the complex structure of the electrode, which includes a variety of materials with different particle sizes and shapes (TJADEN et al. 2016, p. 49). However, a set of existing studies has often featured modified versions of the Bruggeman relation (MAYILVAHANAN et al. 2021). Several experimental methods, including the use of Electrochemical Impedance Spectroscopy (EIS) (LANDESFEIND et al. 2016), have been introduced in the literature to assess tortuosity (HAWLEY and LI 2019, p. 19).

Adhesion

The term *adhesion* describes the interfacial bonds formed when two distinct solid surfaces are brought into contact. In practical terms, the adhesion strength can be quantified by the mechanical load necessary to break or separate the assembled entities. (SCHULTZ and NARDIN 2003, p. 53) In battery production, the adhesion strength, also known as pull-off strength, is an important mechanical property that represents the maximum force needed to detach the coating from the underlying substrate foil, typically expressed in kPa. This property significantly impacts the performance and durability of the cell. (HASELRIEDER et al. 2015, pp. 1–2) Furthermore, it holds a significant role in ensuring the processability of the electrode throughout the production process chain.

2.2.3 Battery Cell Production Chain

The LIB production process can be broadly divided into three phases: electrode manufacturing, cell assembly, and cell finalization (PETTINGER et al. 2018,

p. 212). Among these, electrode manufacturing is considered the core phase of the process chain, as the majority of the properties that determine the electrochemical performance of a LIB cell are established during this phase (GÜNTHER et al. 2016, p. 307). As this dissertation focuses on electrode manufacturing, an overview of the main process steps in this phase is provided below, followed by a brief introduction to the main steps involved in cell assembly and finalization.

Mixing

The first step in conventional solvent-based electrode manufacturing is to mix the electrode components—active material, conductive and binding agents, additives, and solvent—with the primary objective of producing a homogeneous mixture known as slurry. An important aspect that is established during mixing is the slurry formulation, which indicates the mass ratio of the various components used in the mixture. (PETTINGER et al. 2018, pp. 213–214) The characteristics of the slurry have an irrevocable impact not only on the cell properties, but also on the processability of the slurry in subsequent processes (KAISER et al. 2014, p. 697).

Coating

In the coating process, the produced slurry is applied onto the current collector. Various technologies can be used for this purpose, including doctor blade, slot die, and comma bar coating. The coating may be applied in a continuous form or intermittently. Moreover, both sides of the current collector can be coated either simultaneously or in subsequent steps. Achieving a consistent and uniform coating is a key requirement in this process step, particularly in terms of thickness and mass loading. (PETTINGER et al. 2018, p. 214) Among the existing technologies, slot die coating stands out for its precision and is considered the most relevant technology for large-scale production. However, this precision is accompanied by increased complexity. Within a certain process window for slot die coating, an undesirable quality issue known as edge elevation can occur. (SCHMITT et al. 2013, p. 32)

Drying

Following the coating process, the wet film should be dried immediately. Different technologies are used for this purpose. On pilot and industrial production scales, this process is carried out continuously using a set of dryers in a roll-to-roll machine (KAISER et al. 2014, p. 701). The temperature profile set for the dryers has a critical effect on the adhesion strength of the coating to the substrate foil (PETTINGER et al. 2018, p. 214). Extreme drying conditions can lead to migration of binding agents toward the coating surface, an undesirable phenomenon known as binder migration (HAGIWARA et al. 2014; LIM et al. 2013).

Calendering

The produced coating layer has high porosity. Calendering is the subsequent process used to achieve improved energy density, enhance the contact between the particles, and reduce the electric resistance. (KAISER et al. 2014, p. 703) This process involves passing the electrode between two rollers to decrease its thickness. It is imperative to maintain an optimal load during calendering, as applying excessive pressure might result in failures such as foil tearing or particle cracking. (PETTINGER et al. 2018, p. 215)

The produced coils can be trimmed to a certain width based on the product specification, a process referred to as slitting. Depending on the cell type, the manufactured electrodes undergo further processing during cell assembly. The cell assembly usually takes place in a dry room, which maintains a dew-point temperature ranging from -40 °C to -65 °C. This controlled environment is crucial to prevent any moisture-related issues that could compromise cell performance. As a preparatory step, the electrodes are first dried in a vacuum dryer to ensure the removal of any residual moisture before being transferred to the dry room for further processing. (PETTINGER et al. 2018, p. 215) Based on the final dimensions of the battery cell, the electrodes are precisely cut to match the specific formats. The assembly process to build the Electrode Seperator Composite (ESC) varies depending on the type of the battery cell. For the common large-format cells, the processes are winding, stacking, or z-folding. (KURFER et al. 2012, p. 2) Subsequently, the ESCs are welded to tabs, connecting electrodes of the same type, and placed in a cell housing, which is then sealed. The cell is filled with electrolyte, aiming for a homogeneous wetting of the pores of the electrodes. (PETTINGER et al. 2018, pp. 217–218)

Following the cell assembly, the cells go through formation as the final step in production. During formation, the cell is initially charged, typically at a low current, to properly form the protective Solid Electrolyte Interphase (SEI) layer on the anode side. This procedure serves as a preliminary assessment of the electrochemical performance of the battery. (PETTINGER et al. 2018, p. 219) After formation, a series of charging and discharging cycles can be executed to age the cells. This procedure helps to identify cells that demonstrate reduced performance. (KWADE et al. 2018, p. 294)

2.3 Design of Experiments

This section aims to introduce the fundamental concepts and methods used in the DoE. This includes an overview of the common techniques and their potential. Additionally, strategies for evaluation and enhancement of the design are discussed, ensuring that the most reliable and informative results can be obtained from the experimental procedure.

2.3.1 Objectives, Approach and Common Design Types

In today's highly competitive market landscape, production companies encounter relentless demands to enhance both the quality of their products and the efficiency of their production processes. However, achieving these improvements cannot rely solely on the analysis of historical production data. The limitation arises from the complex and multifaceted interdependencies between product properties, process parameters, and quality management practices. In the dynamic landscape of production, making alterations to a product property or process parameter can lead to significant ramifications. (KLEPPMANN 2016, p. 1) To understand the impact of such modifications and the underlying cause-and-effect relationships in a system, it is essential to conduct systematic and purposefully designed experiments. While the mere observation of a system or process can provide preliminary insights and generate various theories or hypotheses about its operating principle, it is only through rigorous experimentation that these theories can be solidified and validated. (MONTGOMERY 2012, p. 1)

Objectives of DoE

DoE is a robust statistical method that facilitates the systematic examination of the various factors influencing a production system. By adopting DoE, researchers and practitioners can efficiently explore multiple variables simultaneously to understand their interrelationships and their collective impact on a system's response. Beyond its significant advantages in terms of cost and resource savings, this method deepens process understanding and assists in refining the production system parameters for improved outcomes. (KLEPPMANN 2016, pp. 4–5)

Approach

The fundamental phases in the conduction of a DoE include planning, execution, data analysis, and conclusion (ROMÁN-RAMÍREZ and MARCO 2022, p. 4). During the planning phase, the problem statement and the objective of the study are formulated. This is followed by choosing a response variable, selection of factors, and levels. (MONTGOMERY 2012, pp. 14–16) The *response variable* captures the output or result of the experiment and is central to understanding the behavior of the system. The *factors* are defined as the input variables that are deliberately changed or set in the experiment, these are commonly the variables assumed to have the highest impact on the system's response. Once the factors are defined, it is imperative to establish the values or settings that each of these factors will have during the experiment. These predefined values or settings are termed as *levels*. (KLEPPMANN 2016, pp. 12–14) The final aspect of the planning phase of a DoE involves choosing an appropriate design type and subsequently generating a design plan that outlines how the experiments will be conducted. The DoE plan consists of *experimental runs*, also referred to as *configurations*, representing unique combinations of the levels within the experimental space. Based on the generated plan, the experiments are conducted, and the results are evaluated. From the results of the experiments, conclusions are drawn, and any necessary subsequent actions or adjustments to the study are identified. (MONTGOMERY 2012, pp. 19–20)

Common Design Types

The type of experimental design chosen is often directly correlated with the specific objectives of the study. A brief overview of several common design types is provided below, highlighting their unique characteristics and applications. (MONTGOMERY 2012, p. 14)

Factor screening is a widely adopted technique, particularly when dealing with a novel system about which there is minimal prior knowledge. The primary objective of this method is to identify the most influential factors and understand their combined effects on the desired response(s). Unlike the traditional One-Factor-At-a-Time (OFAT) approach, which is not only resource-intensive but also overlooks potential interactions between factors, the screening method ensures a more comprehensive analysis. (MONTGOMERY 2012, p. 14) Fractional factorial design is recognized as a common screening method (ROMÁN-RAMÍREZ and MARCO 2022, p. 2).

Optimal design, which can be categorized under Response Surface Methodology (RSM) (ANDERSON-COOK et al. 2009, p. 630), is a type of design that is used when a fundamental understanding of the system exists. Instead of solely identifying influential factors, the objective shifts toward fine-tuning these factors. The overarching goal of optimization is to pinpoint the settings of the factors that lead to the desired response(s). (MONTGOMERY 2012, pp. 14–15) RSM encompasses various design techniques with their distinct objectives and characteristics (ANDERSON-COOK et al. 2009; CHENG 1996). For instance, two of these techniques are the G-optimal and I-optimal designs. In the G-optimal design, the model developed from the generated data strives to minimize the maximum prediction variance within the design space. On the other hand, the I-optimal design is structured to produce a model that results in the smallest average prediction variance across the design space. Thus, while G-optimal focuses on reducing peak variance, I-optimal seeks to optimize average prediction accuracy. (ROMÁN-RAMÍREZ and MARCO 2022, p. 5)

Robust design, commonly referred to as the Taguchi method, was introduced by Genichi Taguchi in the 1980s (MYERS et al. 2016, p. 619). This type of design aims to identify the optimal settings of controllable factors within a process to enhance its resilience against uncontrollable or noise factors, thereby minimizing variations in the system's response (MONTGOMERY 2012, p. 15).

In the design methods described above, the factors and their levels are independent of each other. However, there is another type of design known as *mixture design*, with the factors defined as components of a mixture. As a result, the levels of the factors are interdependent; changing the proportion of one component inevitably affects the proportions of the others. (MONTGOMERY

2012, p. 530) Such designs are relevant, for instance, when analyzing different formulations in battery production to investigate the impact of variations in components such as active material, binder, and additives.

2.3.2 Common Statistical Methods for Design Evaluation

In this subsection, an overview of commonly used methods for evaluating the design and its predictive capability is provided, with a particular focus on those relevant to this dissertation.

Correlation Matrix

A correlation indicates an association between two variables but does not establish causation or specify the nature of their interdependence. The strength and direction of this relationship can be quantified using the correlation coefficient. The *Pearson correlation coefficient*, often referred to simply as the *correlation coefficient*, is a unitless metric that measures linear dependencies between variables. This coefficient ranges from -1 to 1, with values close to the extremes indicating strong negative and positive correlations, respectively. (SCHIEFER and SCHIEFER 2018, p. 90) An absolute correlation coefficient below 0.3 suggests a weak linear correlation, whereas values between 0.3 and 0.7 can be classified as moderate (RATNER 2009, p. 140). When analyzing numerous factors, the *correlation matrix* is adopted, providing a comprehensive overview of the correlation coefficients for each possible pairing of the analyzed factors (SIEBERTZ et al. 2017, p. 64).

Coefficient of Determination

The R-squared value, often referred to as the *coefficient of determination R^2*, quantifies how much of the variance in the dependent variable y can be accounted for by the independent variables in a model. Essentially, R^2 measures the degree to which the regression predictions align with the actual data points. A perfect fit, where the model predictions precisely match each data point, corresponds to an R^2 value of 1. (MYERS et al. 2016, p. 26) When considering y_i as the observed data point and $\hat{y}_i$ as the corresponding predicted value, the calculation of R^2 follows Equation 2.5, with $\bar{y}$ representing the mean value of the observed data points (SIEBERTZ et al. 2017, p. 237).

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{2.5}$$

Variance Inflation Factor

In regression analysis, it is crucial to understand the nature and significance of the relationship between the independent factors and the response variable.

However, when there is a high degree of correlation among the predictor variables, a phenomenon known as *multicollinearity* arises. This can complicate the analysis, as it becomes challenging to distinguish the individual influence of each factor on the response due to their overlapping effects. (KUTNER et al. 1983, pp. 278–279)

The *Variance Inflation Factor (VIF)* is a metric employed to detect multicollinearity in regression analysis. It quantifies how much the variance of an estimated regression coefficient increases when predictors are correlated. For a given predictor x_i, its VIF is computed based on R_i^2, the coefficient of determination obtained when predicting x_i using all other predictor variables x_j in the model, as shown in Equation 2.6. (MYERS 1990, p. 127)

$$VIF_{x_i} = \frac{1}{1 - R_i^2} \tag{2.6}$$

A VIF value exceeding 10 indicates a problematic level of multicollinearity, suggesting that the predictor variables might be highly correlated. In some conservative analyses, a threshold of 5 is defined as acceptable. (SIEBERTZ et al. 2017, p. 65)

Fraction of Design Criterion

The *Fraction of Design Space (FDS)* is a valuable tool in design evaluation, offering insights into the predictive capability of the design within its allocated space. The FDS plot visually represents the portion of the design space that is captured and provides information regarding the associated variance or prediction error. (ZAHRAN et al. 2003, p. 380)

This visualization can be used to compare different designs, assess their robustness, and pinpoint areas with insufficient experimental data (OZOL-GODFREY et al. 2008, p. 218). Such areas indicate where additional runs might be required for enhanced understanding or optimization.

2.3.3 Design Augmentation

Following the introduction of different evaluation methods, this subsection offers an overview of possible techniques that can be employed to enhance the design. By refining the design space, the results achieved from the evaluation methods can be improved, and the data generated can be used for a more accurate and extensive modeling of the system.

One potential approach to counteract issues such as high VIF is to extend the design plan, a procedure often referred to as *design augmentation*. By conducting additional runs subsequently, more comprehensive coverage of the design space can be achieved. (KLEPPMANN 2016, p. 275) Design augmentation can also be considered as a strategic method when consolidating historical datasets. These datasets, often stemming from diverse experiments or studies, may encompass a broad spectrum of analyzed factors. In such scenarios, the significance of design augmentation becomes more pronounced, as it ensures that the consolidated data is not only cohesive but also remains representative and relevant. There are various methods to achieve this objective. For instance, the *fold over* is a type of design augmentation method, in which the levels of one or all factors are reversed, essentially creating a mirrored version of the original design. Such adjustments allow for the systematic and efficient isolation and examination of the effects of certain variables. (KLEPPMANN 2016, pp. 275–276) Another possibility to enhance the design space is by incorporating *center points*. These points, typically positioned at the midpoint of the design dimensions, enable the detection of curvature or linearity in the response. When significant deviations from the predictions of a linear model are observed at these center points, the presence of curvature is suggested. Consequently, by augmenting the design, models can be refined for a more accurate representation of the system's behavior. (KLEPPMANN 2016, p. 277) Additionally, a set of methods exists under the overarching term *space-filling design*. These methods employ various criteria aimed at ensuring a consistent and uniform distribution of design points throughout the experimental space. (LU and ANDERSON-COOK 2021, p. 1740) One of the methods in this category is the *Maximin distance design*. This method aims to achieve the broadest possible spread of points by identifying sparse regions and maximizing the minimum distance between any two design points, thereby ensuring an optimal coverage of the space. (LU and ANDERSON-COOK 2021, p. 1743)

2.4 Data Analytics

The term *Business Intelligence (BI)* was first introduced in 1958 as a system designed to improve business decision-making by leveraging data-processing machines using fact-based information (LUHN 1958). The primary goal of BI is to transform raw data into actionable insights. This is accomplished using a variety of tools and formats, such as visualizations, reports, and dashboards, all designed to support organizations in understanding their performance and making evidence-based decisions. In the late 2000s, the concept of *business analytics* emerged, capturing the core aspect of BI, the analytical process (DAVENPORT et al. 2006). With the exponential growth of digitalization, the concept of *(big) data analytics* has gained prominence. The term refers to a subset of technologies within BI that relies mainly on advanced statistical analyses and pattern recognition (CHEN et al. 2012, p. 1174). The U.S. Department of Commerce's National Institute of Standards and Technology (NIST) defines *analytics* as "the systematic processing and manipulation of data to uncover patterns, relationship between data, historical trend and attempts at prediction of future behaviors and events." (NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY - BIG DATA PUBLIC WORKING GROUP 2019, p. 6)

Considering the potential value and complexity associated with data analytics, the Gartner Analytics Ascending Model outlines four distinct levels of maturity, as illustrated in Figure 2.2 (LANEY et al. 2012). While slight variations of this model can be found in the literature (DAVENPORT 2013; SCHUH et al. 2017; SIVARAJAH et al. 2017; BELHADI et al. 2019), the fundamental concept remains consistent. The first stage with the lowest level of complexity is *descriptive analytics*. This level focuses on providing a retrospective view of the existing state of the system. This is achieved by employing statistical analysis, generating comprehensive reports, and monitoring Key Performance Indicators (KPIs). Progressing to the next level, *diagnostic analytics* aims to investigate the reasons behind certain trends or observations. An in-depth understanding of the system is generated by *predictive analytics*, wherein ML models, drawing upon historical data, are commonly utilized to forecast the system's potential future states or outcomes. Building upon the solutions from predictive analytics that provide insights, *prescriptive analytics* employ techniques such as simulation or optimization to facilitate the adaptation and fine-tuning of the system. (BELHADI et al. 2019, p. 3; DAVENPORT 2013, p. 13)

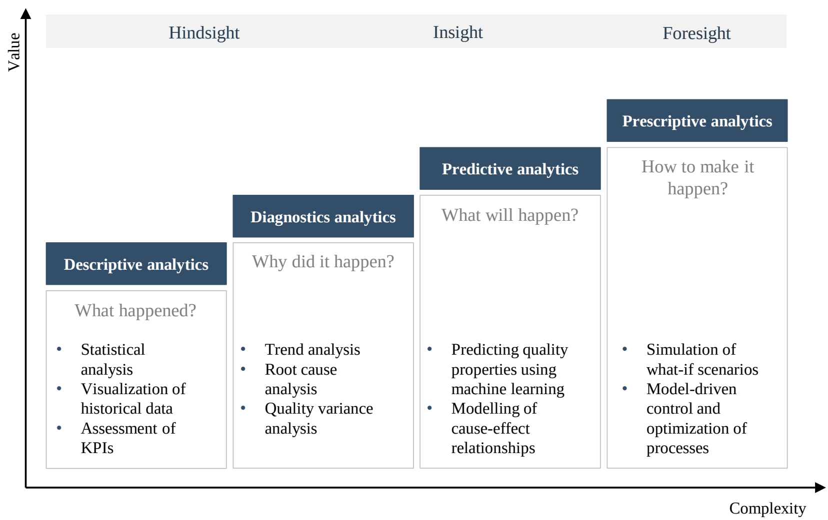

Figure 2.2: The Gartner Analytics Ascending Model, illustrating different stages of data analytics, adapted from LANEY et al. (2012)

In the field of data analytics, the terms *Data Mining (DM)*, *Artificial Intelligence (AI)*, and ML are frequently used interchangeably. In order to provide a solid foundation for this dissertation, a brief description of these terms is provided in the following.

DM is primarily concerned with extraction of meaningful patterns from existing large datasets (FAYYAD et al. 1996, p. 28). DM can be divided into two main categories: descriptive and predictive. Descriptive DM delves into existing datasets to uncover unexpected structures or clusters, while predictive DM, bearing a resemblance to ML, relies on developing models tailored for precise prediction. (IZENMAN 2008, p. 5)

AI, a term first introduced by John McCarthy in 1955, represents a specialized branch of computer science. Its primary objective is to enable machines to imitate human cognitive functions. Although McCarthy is credited with coining the term, the foundational concept can be traced back to Alan Turing (TURING 1950). (BUCHANAN 2005)

While AI aims to equip machines with human-like intelligence, enabling them to think rationally and solve problems, ML as a subset evolved out of AI, allows computers to learn without the need for explicit programming (SAMUEL 1959, p. 535). In a complementary view, MITCHELL (1997, p. 2) states that if a computer program is designed to improve its performance measure for a specific set of tasks based on its experiences, it is said to be *learning* from those experiences. Given the descriptions provided, this dissertation primarily centers on ML. Hence, the relevant methods and concepts in this field are presented in the following subsections.

2.4.1 Terminology and Types of Machine Learning

This subsection serves as a foundational introduction to the key terminology used in the field of ML. Furthermore, an overview of the primary categories of ML is provided.

For the development of ML models, data plays a key role. The collected data is referred to as a *dataset*, in which each individual *data point*, also termed as *sample*, provides a detailed description of a specific event or object. These descriptions, which are assigned to each data point, are commonly referred to as *features* or *input variables*. (ZHOU 2021, p. 3) An *outlier* within a dataset is a data point or observation that significantly deviates from the rest of the data samples. An outlier can either indicate abnormal system behavior or result from recording errors, especially those caused by faulty sensors. It is essential to identify and eliminate such data points to ensure the reliability of the final result. (ALPAYDIN 2014, p. 199)

An *algorithm* is a systematic and structured set of instructions and computational techniques designed to transform a given set of inputs into a desired output (ALPAYDIN 2014, p. 2). In the context of ML, the algorithm serves as the foundation for developing a mathematical model or approximation of a function, with the core task of drawing valuable conclusions or insights from a dataset (ALPAYDIN 2014, p. 3). The process of employing ML algorithms to construct models from data is often termed *learning* or *training*. In this phase, a particular dataset, referred to as the *training dataset*, is used to instruct the algorithm, allowing it to identify patterns, relationships, and insights within the information provided. (ZHOU 2021, p. 3)

Hyperparameters are adjustable settings that can be used to control the performance and behavior of a complex algorithm (GOODFELLOW et al. 2016, p. 118). Hyperparameters can be configured using various approaches. One possibility involves a trial-and-error process, often necessitating domain-specific expertise. In the case of more complex analyses, this approach is known to be time-consuming and error-prone. (BISCHL et al. 2023, p. 2) An alternative strategy is based on heuristics. In this approach, a range of hyperparameters is systematically tested by defining a grid of values for each hyperparameter and evaluating the performance of the model for each combination. This grid search allows for a thorough exploration of the hyperparameter space. Another effective technique is random search, which is frequently used in scenarios involving higher-dimensional hyperparameter optimization. In contrast to structured grid search, random search is based on the random selection of combinations of hyperparameters for evaluation. (BISCHL et al. 2023, pp. 7–8)

Depending on the nature of the data being analyzed and the objective of the study, ML techniques can be broadly divided into two major categories: supervised and unsupervised learning. In *supervised learning*, the objective is to predict the value of an output, also referred to as the target variable, based on a set of input variables. The term supervised captures the core of this approach, as it implies that the learning process is guided and supervised by the presence of the target variable, also referred to as the labeled output. In the case of *unsupervised learning*, there is no target variable available, but the objective is to describe the associations, patterns, and clusters among the input variables. (HASTIE et al. 2009, p. 2) As the field of ML has evolved over the years, additional branches have emerged alongside these two main categories, including semi-supervised learning (VAN ENGELEN and HOOS 2020, p. 374). The objective of this dissertation is to analyze and understand the interdependencies in electrode manufacturing and their influence on intermediate and final product quality. Given that the product properties are treated as labeled target variables, the scope of this dissertation falls under the category of supervised ML.

If the target variable in a supervised learning scenario consists of numerical values, the problem is typically referred to as *regression*. In a regression task, the goal is to predict a continuous and quantitative output. Conversely, when the output variable is characterized by discrete categories or classes, supervised ML falls under the domain of *classification*, with the aim of predicting a predefined set of categories. (HASTIE et al. 2009, p. 10)

2.4.2 Machine Learning Regression Algorithms

In the context of this dissertation, the problem under investigation is treated as a regression task, primarily concerned with predicting a continuous numerical target variable. Hence, in this subsection, a brief description of the supervised ML regression algorithms relevant to this dissertation and their working principle is presented.

Linear Regression

The *Linear Regression (LR)* algorithm is regarded as one of the most fundamental techniques in the field of ML, given that numerous powerful nonlinear models are derived from LR by incorporating multi-layer structures or high-dimensional mappings (ZHOU 2021, p. 58). The basic concept of this algorithm is to establish a linear relationship between input and output variables. Depending on the number of these variables, LR can be further categorized into distinct types.

The *Multiple Linear Regression (MLR)* is defined as an LR model based on various input variables, as shown in Equation 2.7. While y stands for the output, also termed as the dependent variable, x_i represents the independent variables, where i is the number of independent variables, ranging from 1 to p, each contributing to the prediction. The term e is an unobservable random variable, also referred to as the error component. The coefficients for each independent variable are shown by β_i, indicating the strength and direction of their influence on y, while β_0 represents the intercept, serving as the baseline of the regression model. (IZENMAN 2008, p. 108)

$$y = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \mathrm{e} \tag{2.7}$$

Polynomial Regression

The Polynomial Regression (PolyR) is a type of regression model in which the relationship between the independent variables and the dependent variable is represented by an n^{th}-degree polynomial. In the case of two independent variables, Equation 2.8 shows a second-degree polynomial, providing an example that includes two-factor interaction. (WALL and AMEMIYA 2007, p. 326)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \mathrm{e} \tag{2.8}$$

Decision Tree

The Decision Tree (DT) is a versatile algorithm used for both classification and regression purposes (BREIMAN et al. 1984). Building a DT involves repeatedly splitting the dataset, where at each split, the data is divided into two groups. This division aims to maximize the consistency of each group with respect to the target variable. To predict values, a regression model is applied to each node during this process, resulting in a tree structure that helps make predictions based on the input data. (De'ath and Fabricius 2000, p. 3178) A DT consists of *internal decision nodes* and *terminal or leaf nodes*. A decision node is a point where the dataset is divided into branches based on certain conditions or criteria related to the independent variables. A leaf or terminal node represents the ultimate outcome for a specific path of the DT. (Alpaydin 2014, pp. 213–214)

Random Forest

The *Random Forest (RF)*, introduced by Breiman (2001), is a tree-based ensemble learning algorithm. It is constructed from a substantial ensemble of decorrelated DTs. The ensemble approach combines the predictions of these trees to achieve robust and accurate results through averaging. (Hastie et al. 2009, p. 587)

eXtreme Gradient Boosting

Based on the foundational gradient tree boosting technique, which seeks to synthesize predictions from a collection of models to create a more robust and precise model (Friedman 2001), Chen and Guestrin (2016) introduced eXtreme Gradient Boosting (XGBoost) as a scalable ML algorithm for tree boosting. Both RF and XGBoost are considered methods designed to enhance the performance of the model by leveraging simple base models. Nevertheless, they function on distinct principles, as illustrated in Figure 2.3. While RF relies on the construction of multiple independent DTs, the XGBoost builds trees subsequently, with each new tree correcting its predictions based on the residuals from the previous iteration. (Natras et al. 2022, p. 9)

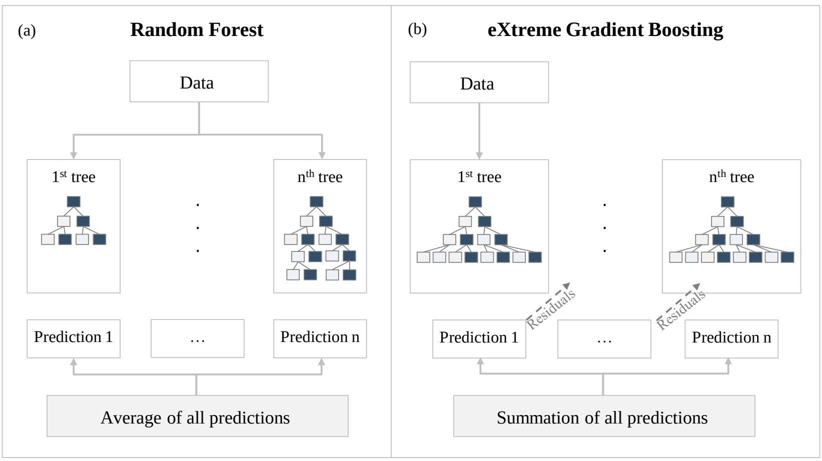

Figure 2.3: Overview of the working principles of (a) RF and (b) XGBoost, adapted from Natras et al. (2022)

Support Vector Machine

The Support Vector Machine (SVM), introduced by BOSER et al. (1992) with an initial focus on classification tasks, is a powerful algorithm, available in both linear and nonlinear versions (IZENMAN 2008, p. 369). The SVM regression algorithm transforms the input variables into a high-dimensional feature space and aims to find an optimal separating hyperplane that effectively describes the newly established feature space (VAPNIK 1999, p. 133).

2.4.3 Metrics and Methods for Performance Evaluation

There is a diverse range of mathematical expressions, referred to as *metrics*, that can be used to evaluate the performance of regressive models in ML. These metrics provide quantifiable measures reflecting the effectiveness, accuracy, and generalization capabilities of the ML models. (JOSHI 2020, p. 169) Aside from R^2, which was introduced in Subsection 2.3.2, there are additional metrics relevant to this dissertation, which are presented in the following.

The *Mean Absolute Error (MAE)* is a metric used in statistics and ML to measure the average magnitude of errors between predicted values $\hat{y}_i$ and actual observed values y_i, according to the following Equation (JOSHI 2020, p. 170).

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{2.9}$$

The *Mean Squared Error (MSE)* measures the average of the squared differences between predicted values and actual observed values. The inclusion of the squared term in the MSE, as shown in Equation 2.10, leads to a more pronounced penalty for larger errors, which are often associated with outliers. (JOSHI 2020, p. 170)

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{2.10}$$

Another common evaluation metric is *Root Mean Squared Error (RMSE)*, which calculates the error by taking the square root of the MSE, as shown in Equation 2.11. The results have the same unit as the target variable and are less sensitive to outliers. (JOSHI 2020, p. 171)

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{2.11}$$

The range for MAE, MSE, and RMSE generally begins at a minimum value of 0 in an ideal scenario, where the model's predictions precisely align with the actual values. There is no upper limit for these metrics, but lower values are indicative of better model performance.

The primary goal of ML is not simply to reproduce the training data, but rather to make accurate predictions for new cases. The measure of how effectively a model trained on the training dataset, can accurately predict the target variable for new, previously unseen instances is referred to as *generalization*. In essence, generalization assesses the ability of the model to apply its learned patterns and knowledge to new data points, which is a crucial aspect of evaluating its performance and practical application. (ALPAYDIN 2014, pp. 38–39)

When dealing with a large dataset, a common approach to assess a model's generalization is to evaluate its performance on independent data, often referred to as a *validation set*. Nonetheless, there are numerous scenarios in which data availability is restricted. In such cases, a small validation set can result in an imprecise assessment of the model's predictive performance due to increased noise. One practical solution to address this issue is the *cross-validation* approach. This approach divides the available dataset into two subsets: a training set, and a *test set* reserved for assessing the model's performance. Cross-validation involves iteratively partitioning the data into separate training and test subsets. This process is conducted k times, where k represents the number of iterations or partitions. Each of these subsets, often referred to as a *fold*, is assessed once as the test set while the remaining k-1 folds are used for training, as shown in Figure 2.4. This iterative process offers a more robust estimation of the model's generalization performance. In the case of data scarcity, a simplified alternative involves employing a *random train-test split*. While this method can be resource-efficient, it is important to note that it introduces a certain degree of randomness. Consequently, the effectiveness of the model's evaluation may vary depending on the specific data split employed. (BISHOP and NASRABADI 2006, pp. 32–33)

2.4.4 Key Challenges in Machine Learning

The potential of ML, particularly when dealing with complex systems such as LIB production, is evident (LOMBARDO et al. 2021, p. 10953). However, there are certain challenges that need to be considered throughout the development and prior to deployment of data-driven models.

Before developing ML models, it is imperative to pay substantial attention to the quality and the veracity of the available dataset. A misdefinition of the problem or use of an inaccurate or biased dataset can easily lead to incorrect ML predictions and misleading insights. DoE serves as a powerful statistical tool that can effectively address this challenge by maximizing the statistical significance of the data to be obtained and minimizing the number of experiments required to analyze and understand the system. (LOMBARDO et al. 2021, p. 10918)

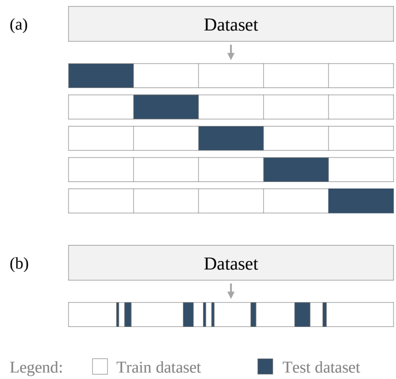

Figure 2.4: Possible methods for evaluating generalization in a supervised ML model with limited data: (a) k-fold cross-validation, with k=5 and (b) random train-test split, adapted from BISHOP and NASRABADI (2006, p. 33)

In terms of model generalization, there is often a trade-off between bias and variance (HASTIE et al. 2009, p. 38). *Variance* describes the degree to which the predictions made by a model would vary when it is trained using a different training dataset. In other words, it measures how sensitive the model's performance is to the particular dataset on which it was trained. A high variance implies that the model is overly influenced by the training data, potentially making it less reliable when faced with new, unseen data. *Bias*, on the other hand, is denoted as the error introduced when a complex system or problem is approximated by a considerably simpler model. These simpler models are typically characterized by limited flexibility, resulting in a higher level of bias. (JAMES et al. 2013, pp. 34–35) A complex model characterized by low bias and high variance indicates that the training dataset has been learned in great detail, including the noise. This phenomenon is commonly referred to as *overfitting*. Conversely, *underfitting* is observed when the variability of the data cannot be captured by a model with high bias. (ALPAYDIN 2014, p. 82) Figure 2.5 illustrates the concept of generalization, based on the example of regression models, and highlights the two common issues of underfitting and overfitting. To achieve a balance between a model's complexity and its ability to generalize effectively and reduce the risk of overfitting, several strategies can be adopted, including the implementation of techniques such as regularization or the incorporation of a cross-validation approach. (HASTIE et al. 2009, p. 400)

Another conflict arises when considering the trade-off between the flexibility and complexity of an ML model and its potential for explainability and interpretability (JAMES et al. 2013, p. 25). Models with higher flexibility often possess the capacity to capture the complexity of a system more effectively. However, gaining insight into the decision-making processes of such complex models is challenging. Such models are often referred to as *black boxes* as it is difficult to understand the reasons behind specific predictions or identify the

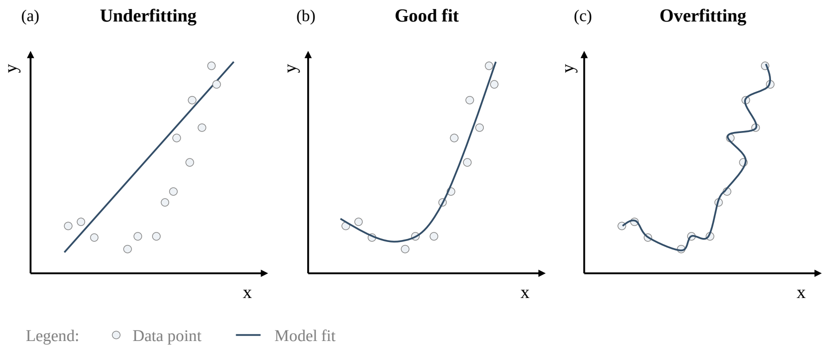

Figure 2.5: Examples of regression model generalization, with (a) underfitting occurred by a linear model, (b) good fit, and (c) overfitting by a high-degree polynomial model, adapted from BADILLO et al. (2020, p. 876)

features influencing the outcomes. (LINARDATOS et al. 2020, pp. 1–2) Figure 2.6 illustrates the trade-off outlined for a set of common ML algorithms.

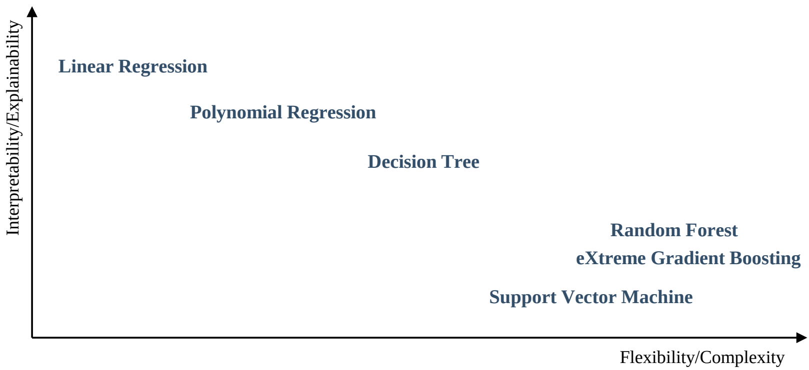

Figure 2.6: Trade-off between the complexity of the model and its interpretability across a set of common ML algorithms, adapted from JAMES et al. (2013, p. 25)

2.4.5 Explainable Machine Learning Methods

The ever-increasing demand for ML models, which not only offer high performance and robustness but also provide trust and transparency for real-world applications, has led to the revival of the field of eXplainable Artificial Intelligence (XAI), alternatively referred to as eXplainable Machine Learning (XML) (LINARDATOS et al. 2020, p. 2). The XML methods are designed to offer insight into the decision-making and prediction processes of a model, effectively transforming what is often referred to as a black box model into a *glass box* model (Rai 2020, pp. 137–138). Within the domain of XML, there is a set of methods tailored to a particular family of algorithms, known as *model-specific* methods. In contrast, there are also methods that are universally applicable to any algorithm, and these are referred to as *model-agnostic* methods. While certain methods offer explanations that encompass the entire model, unveiling global effects, there are other methods that focus on revealing local, instance-specific impacts. (Linardatos et al. 2020, p. 5) In the following, a brief description of the XML methods relevant to this dissertation is presented.

Mean Decrease of Impurity

The *Feature Importance (FI)* based on *Mean Decrease of Impurity (MDI)* is an XML method that evaluates the significance of features in the context of tree-based algorithms. In DTs, as the main components of tree-based ensemble models, a key aspect involves partitioning the dataset into subsets. The overarching goal is to have the subsets that foster greater homogeneity and minimize impurity with respect to the target variable. To assess the relevance of an input variable, the decrease of impurity in the nodes is used as a measure. This calculation involves the accumulation of impurity decrease at each node across the entire forest, where a split based on the considered input variable has been executed. In essence, it quantifies how much the input variable contributes to making the subsets purer and in line with the target variable. (Nembrini et al. 2018, pp. 3711–3712)

Permutation Feature Importance

The *Permutation FI* is a heuristic approach used to address bias in conventional FI. This method is based on the assumption that the random importance of a feature follows a specific probability distribution. The relevance of a feature is evaluated by measuring how much the model's prediction error increases after permuting or shuffling the feature value. (Linardatos et al. 2020, p. 13)

SHapley Additive exPlanations

The *SHapley Additive exPlanations (SHAP)*, an XML method inspired by the Shapely values from the game theory, was initially introduced for tree-based models (Lundberg et al. 2020, p. 56). This method provides insight into both the global feature relevance and the local instance-specific contributions (Lundberg et al. 2020, p. 61). For this purpose, the contribution of each feature to the prediction of the target variable is calculated, considering all possible combinations of features (Faraji Niri et al. 2023, p. 11).

Accumulated Local Effects

The *Accumulated Local Effects (ALE)* plot is a visualization method used to interpret the impact of features on the model's prediction (Apley and Zhu

2020, p. 1061). Taking into account the conditional distribution of features, ALE plots are capable of dealing with correlated features. The method is based on calculating the average impact of a specific feature of interest while mitigating the influence of correlated features. (LINARDATOS et al. 2020, p. 14). ALE is categorized as a model-agnostic method.

Chapter 3

State of the Art

This chapter summarizes the relevant literature contributing to the objective of this dissertation. The approach adopted, including the main research areas identified as relevant for the literature review, is outlined in Section 3.1. The subsequent sections (3.2 to 3.4) present a detailed analysis of the existing studies, with each section focusing on a particular research area, accompanied by a brief summary. Section 3.5 provides concluding remarks and delves into possible research perspectives.

3.1 Literature Review Approach

In order to obtain a thorough overview of the relevant concepts in the subject matter, a systematic concept-centric literature review was conducted. As suggested by WEBSTER and WATSON (2002), this approach offers a set of advantages when compared with a publication-centric one, including a more comprehensive analysis and a holistic perspective on the relevant aspects within the domain being analyzed. A systematic literature review can be broadly divided into three major phases: review planning, search execution, and, analysis and evaluation (KITCHENHAM et al. 2009, pp. 8–9). In the planning phase, the objective and the scope of the literature review are defined. Within the context of this dissertation, the scope was restricted to LIB electrode manufacturing.

In line with the motivation and the overarching objective established in Chapter 1, and guided by the adopted concept-centric approach, three main research areas have been identified as relevant to enable a holistic, efficient, and quality-oriented data-driven analysis of interdependencies in the electrode manufacturing of LIB cells. These areas are briefly described below.

- Parameters and measurement solutions: Data is the backbone of any data-driven solution. Therefore, this research area explores the literature on quality-relevant parameters and available measuring systems for the characterization of (intermediate) products and data collection, which is a prerequisite for the development of ML models.

- Data generation and evaluation: As outlined in Subsection 2.4.4, data quality is a critical aspect in ML models. In this research area, the extent to which data generation and evaluation methodologies have been utilized in the existing literature related to data-driven studies in LIB electrode manufacturing is explored.
- Data-driven models and derivation of insights: This research area is centered on the investigation of the existing data-driven studies in electrode manufacturing. The existing literature is analyzed considering various aspects such as processes and parameters, materials, and the ML techniques employed. With the overarching goal of comprehending the existing interdependencies, enhancing process understanding, and consequently product quality, there is a deliberate emphasis on extracting insights from the developed models. Therefore, the use of XML methods is additionally considered.

In all of the defined research areas, the methodological aspect of the existing solutions is also taken into account. This emphasis on method-based solutions not only fosters a profound understanding of the proposed solutions, but also facilitates their adaptation and transfer as necessary. Figure 3.1 presents an overview of the relevant research areas and their key elements, which have been considered during the conduction of the literature review, with a focus on LIB electrode manufacturing.

Figure 3.1: Overview of the relevant research areas identified and their key elements for the concept-centric literature review focused on LIB electrode manufacturing

Based on the identified research areas, an initial literature review was conducted in January 2021, using the *Scopus* and *Google Scholar* databases. This was followed by extensive mapping studies carried out as part of the solution development, which were subsequently published. Detailed information on the search strategies and the keywords used in the queries can be found in Haghi et al. (2022, p. 9) and Haghi, Hidalgo, et al. (2023, p. 3). In addition, there is a set of extensive literature reviews available for each of the research areas, which were adopted as reference points and are outlined in the respective sections. To ensure a comprehensive literature review, a thorough examination of the existing literature was conducted in March 2023. The results are presented in the following sections.

3.2 Parameters and Measurement Solutions

In a comprehensive review, AYERBE et al. (2022) summarized the current status, challenges, and future perspectives in the field of digitalization for battery cell production. Within this context, data acquisition is identified as a key component, given the high number of parameters required to monitor the quality of both intermediate and final products (AYERBE et al. 2022, p. 3).

An early approach to address the challenge of managing a multitude of parameters in battery cell production was introduced by WESTERMEIER et al. (2013). Elaborating on the complexity of the process chain, a five-step methodology for quality planning in LIB production chain was developed, consisting of parameter identification, parameter classification, parameter selection, process chain DoE, and process chain optimization. In the first step for parameter identification, the methodology involved an expert-based evaluation of potential correlations between parameters, facilitated by the application of Failure Mode and Effect Analysis (FMEA). (WESTERMEIER et al. 2013) In the subsequent publication, based on multiple domain matrices, parameters were classified into those with direct and indirect impact on the quality of the final product (WESTERMEIER et al. 2014). The proposed methodology was detailed and refined in the dissertation of WESTERMEIER (2016). The applicability of the proposed solution was demonstrated on the LIB battery cell production adopted as a use case, with a primary focus on the analysis of the cell assembly and finalization processes. The results were obtained through interviews conducted with 12 experts from research projects in the field of battery production; further details on the profile of the experts were not provided. Although the proposed solution meets its methodological requirements, there are still certain limitations. One constraint arises from the necessity for experts to possess a comprehensive cross-process understanding of the complex interdependencies along the entire production process and the cumulative effects, particularly when focusing solely on the quality of the final product. Additionally, one key aspect that remains unaddressed throughout the methodology is the feasibility of collecting the identified relevant parameters, ideally in real-time, through appropriate measurement solutions. The framework does not take into account the practical considerations of implementing measurement solutions, which are essential to ensure the efficient application of the proposed methodology.

TURETSKYY, THIEDE, et al. (2020) introduced a data-driven concept for the acquisition of relevant data, data management, and data analytics in battery cell production. Building on this concept, the dissertation of TURETSKYY

(2022) presented a holistic framework for data analytics in battery cell production that integrated the required elements from *CRISP-DM* (Chapman et al. 2000) and incorporated them into the *Cyber-Physical Production System (CPPS)*, a methodology introduced by Thiede (2018). The work presented by Turetskyy (2022) can be regarded as a significant contribution to the current state of the art in this field. In this section, only the aspects related to parameters and measurability are discussed, while the data analytics aspects are explored in Section 3.4. The proposed framework was developed based on the example of the *Battery LabFactory Braunschweig*. For this purpose, a primarily expert-based analysis was conducted with the technical and scientific staff of the facility to identify different parameters, data sources, and data types available. (Turetskyy, Thiede, et al. 2020) The data management aspect, including the definition of the data acquisition strategy, data consolidation, and data storage, has been extensively elaborated (Turetskyy 2022). However, the underlying foundation, which relates to the relevance of the parameters, has not been discussed.

In an extensive literature review, Zanotto et al. (2022) provided data specifications for the development of computational models that can be used as a digital twin of a battery production pilot line. The review presented an overview of parameters classified into three categories based on their criticality for model development. The models considered in this step included empirical, physics-based, discrete, and ML models. Approximately 65 parameters were identified for the electrode manufacturing. The reported parameters were likely based on doctor blade coating, given that crucial details specific to slot die coating—an industrially relevant technology—such as slot die angle or pump flow rate, were not considered. Furthermore, the review included an assessment of the measurability and accuracy of the sensor technology for the parameters. It is important to highlight that the work presented by Zanotto et al. (2022) is closely aligned with the solution developed in this dissertation and presented in Haghi et al. (2022), with both articles published simultaneously in July 2022. However, it is worth noting that the methodological aspect of the solution presented by Zanotto et al. (2022) could benefit from additional clarity, as the foundational details regarding the evaluation of the parameters, their measurability, and accuracy are not explicitly provided.

In a recent publication, Kampker et al. (2023) highlighted the relevance of parameters in the development of a digital twin in battery cell production. Based on a literature review, an overview of process and product parameters was provided. The findings were used to develop a digital battery product twin concept using an information model. The relevance of the parameters or their measurability was not discussed.

In two exclusive reviews, Reynolds et al. (2021) and Zhang et al. (2022) explored the possible measuring solutions in the coating and drying processes, respectively. Underlying the need for a comprehensive overview of the existing measurement options as a prerequisite for the development of data-driven models, Reynolds et al. (2021) presented a summary of the slurry properties and the product and process parameters during the coating process. The review provided an overview of possible measuring instruments, including exemplary comparisons with the advantages and the disadvantages for two product parameters, specifically the coating thickness and the quality of the coating surface in the context of defect detection solutions (REYNOLDS et al. 2021). However, the comparisons lacked quantitative data as a benchmark; instead, qualitative aspects such as affordability and precision were mentioned as advantages. ZHANG et al. (2022) presented a summary of the relevant parameters and elaborated on their impact in the context of possible defects arising during the drying process. The review included an overview of characterization methods, their scale, and the possibility of inline or offline applications. Additionally, the methods that are already in use in the drying process of LIB production were highlighted.

In conclusion, the existing literature has addressed the topic of LIB electrode manufacturing parameters to a certain extent. Some authors, such as WESTERMEIER (2016), have underscored the importance of evaluation of the parameters based on their relevance. This is necessitated by the large number of parameters involved and the inherent complexity of the process chain. It is worth noting that the discussion around the measurability of these parameters has predominantly taken place in isolation within the current literature. The existing analyses lack a certain degree of transparency, and the comprehensive literature reviews fall short of addressing the entirety of the electrode manufacturing process chain and the relevant parameters. These aspects underscore the need for a more holistic and systematic examination of this research area.

3.3 Data Generation and Evaluation

The necessity of data quality in data-driven models and the advantages of DoE methods in this context were elaborated in Chapter 2. In a comprehensive review, ROMÁN-RAMÍREZ and MARCO (2022) analyzed the application of DoE methods in various domains of battery research. While the review did not focus exclusively on ML studies, the results were adopted as the foundation for the literature review in this section. The review revealed that the application of DoE methods, particularly in the context of slurry formulation, electrode manufacturing, and cell production, has been rather strictly constrained, with only 12 related studies identified (ROMÁN-RAMÍREZ and MARCO 2022, pp. 11, 13). From a methodological perspective, despite the higher efficiency offered by optimal designs, in comparison with other DoE design types, these methods have also received limited attention.

For a comprehensive analysis, an in-depth investigation of the data generation and evaluation steps for the development of ML models in electrode manufacturing was additionally conducted. In the following, only the articles providing details on these two steps are discussed, while the aspects relevant to model development are analyzed in Section 3.4.

Faraji Niri et al. (2021) investigated the impact of the coating parameters on the physical and electrochemical properties of the electrode. The study was based on DoE methods, conducted in two parts. In the first part, a screening design was used to identify the main influencing parameters. The target variables were electrode mass loading, thickness, and porosity, while the analyzed factors included comma bar gap, web speed, drying temperature, air speed, and coating ratio. The first phase consisted of 12 cathode configurations, with a correlation matrix used to evaluate the data generated by the screening design. Due to the strong correlation observed between the physical properties of the electrode and the comma bar gap, web speed, and coating ratio, the second DoE phase focused exclusively on these parameters. In this phase, the Box-Wilson composite design was utilized as a method within the framework of RSM, with five levels for each factor. This approach resulted in a total of 20 experimental runs. This dataset was subsequently adopted in a following article by Faraji Niri, Liu, Apachitei, et al. (2022) and supplemented by a series of experiments carried out on the anode. In the case of the anode, a correlation matrix between the input variables was used as a first step to restrict the number of factors. Afterward, the RSM for the anode was conducted based on only the comma bar gap and the coating ratio.

Duquesnoy et al. (2021) examined the influence of the slurry formulation, solid content, and coating gap on the heterogeneity of the electrode, taking into account its mass loading and thickness. The reported parameters indicate that a comprehensive experiment was carried out using a full factorial design. This resulted in a total of 144 electrode configurations. Subsequently, a Principle Component Analysis (PCA) was conducted to identify the interdependencies among the input variables, and, if applicable, reduce the number of relevant variables. This analytical method serves as a valuable tool to reduce the dimensionality of the dataset while retaining the crucial information.

In a study focused on the calendering process, Faraji Niri, Apachitei, et al. (2022) analyzed the effects of mass loading, target porosity, calendering gap, and rollers temperature on the electrochemical properties of the electrode. Using a DoE method, 18 experimental runs were defined and conducted. Given that the pressure applied during the calendering was kept constant, it was expected that the analyzed factors, in particular the calendering gap and the target porosity, would show a high correlation when all other factors, such as mass loading, remained unchanged. However, the study does not provide further details on the DoE method adopted, the rationale behind the selection of the factors, or the approach taken to address this aspect.

Wang et al. (2022) presented an XML-based study, investigating the impact of slurry formulation and different binder types on the discharge capacity at different C-rates. The data utilized in this study was sourced from the dataset originally published by RYNNE et al. (2019, 2020), which was generated using a DoE method.

In a recent publication, DUQUESNOY et al. (2023) introduced an approach for generation of a synthetic dataset from physics-based simulation models, which was then used to train an ML model and make predictions concerning the mesoscale properties of the electrode. While the study mentioned the use of a DoE for data generation, it did not provide further details on this aspect.

A number of studies, such as SCHNELL et al. (2019), KORNAS et al. (2019), and TURETSKYY, THIEDE, et al. (2020), have demonstrated the application of DM methods using available historical datasets collected throughout the production chain. Data evaluation was carried out as part of the data preparation process, resulting in a reduction of the datasets. For instance, in the case of the article presented by SCHNELL et al. (2019), out of 714 available data points, only 113 were identified as suitable for model development after data preparation. However, these studies did not provide further details on the methods applied to evaluate the data or the background of the historical data. Furthermore, the use cases were presented with a certain degree of anonymity concerning the parameters. This lack of transparency makes it difficult to gain a complete understanding of the interdependencies between different factors, the range of variations considered, and the statistical significance of the results.

In summary, it is evident that the topic of data generation and evaluation has not received extensive attention in the existing data-driven literature. While there are noteworthy examples in the field of formulation studies, such as the data generated by RYNNE et al. (2019) and adopted by WANG et al. (2022) for ML analysis, the application of process-based DoE has been limited. The only examples in this field are presented by FARAJI NIRI et al. (2021) and FARAJI NIRI, APACHITEI, et al. (2022), which predominantly focus on single process steps. Remarkably, a comprehensive cross-process analysis of the electrode manufacturing process chain, carried out efficiently through systematic data generation, and potentially incorporating the use of DoE techniques, remains a relatively unexplored area of research. Furthermore, in studies utilizing historical data, an in-depth analysis of data quality, particularly with respect to issues such as multicollinearity and possible measures to improve the dataset, has been largely overlooked.

3.4 Data-Driven Models and Derivation of Insights

In a critical review, LOMBARDO et al. (2021) explored the ML applications in battery research, covering areas such as material design, cell production, as well as cell diagnosis and prognosis. Among the 200 articles analyzed, only 6 % were dedicated to electrode manufacturing and cell production (LOMBARDO et al. 2021, p. 10908). Similarly, in a comprehensive review, FARAJI NIRI et al. (2023) investigated the application of XML methods in battery research, encompassing aspects from production to state and performance estimation. The findings revealed that XML has emerged as a prominent trend across all research domains in recent years (FARAJI NIRI et al. 2023, p. 30). However, it is important to note that a significant proportion of the existing studies, particularly in the battery production field, relied predominantly on the application of FI method, analyzing only the global relevance of the parameters (FARAJI NIRI et al. 2023, p. 27).

The findings presented by LOMBARDO et al. (2021) and FARAJI NIRI et al. (2023) served as the foundation upon which a comprehensive literature review was conducted. This section provides a summary of the relevant studies analyzing the interdependencies in LIB electrode manufacturing. Some of the studies have been consolidated, as they represent a cohesive investigation presented across multiple consecutive publications. It is important to note that certain studies, such as those focusing on the use of ML models for faster parameterization of simulation models or those concentrating exclusively on the interdependencies in cell assembly processes, were excluded from this analysis. In the conducted mapping study, which is a part of the solution development in this dissertation (see Chapter 5), an overview of ML-based studies along the entire process chain is provided. Further details in this regard can be found in HAGHI, HIDALGO, et al. (2023). The relevant studies in this section were analyzed from two perspectives: production and data analytics. While the former investigated the process steps analyzed, the target variables, and the production scale to assess the potential for achieving a comprehensive understanding within electrode manufacturing, the latter delved into the methodological considerations, including aspects such as interpretability, to extract valuable insights from the developed models.

THIEDE et al. (2019) demonstrated the application of a DM approach, followed by the CPPS framework introduced in an earlier article (THIEDE 2018), to predict the quality properties of a battery cell based on the parameters collected along the process chain at pilot scale. The study utilized a dataset obtained from 172 LIB pouch cells, considering parameter variations during calendering, cutting, and z-folding. A multivariate regression model was used to predict maximal capacity, capacity loss after 400 cycles, and formation loss after the first cycle. The coefficients of the developed regression model were used as indicators to identify the influencing factors, which were unfortunately reported anonymously with only reference numbers.

CUNHA et al. (2020) analyzed the impact of slurry properties such as solid content, on electrode characteristics, specifically porosity after drying and mass loading. The study was based on a classification analysis, using DT, SVM, and deep Neural Network (NN). With 82 different NMC cathode configurations, the dataset generated at a pilot line included more than 600 data points. Despite the use of complex models such as NN, the study did not delve into the interpretability aspects. Based on the extensive dataset provided by CUNHA et al. (2020), several subsequent studies have been published with a similar objective, aiming to evaluate the performance of different models (LIU, YANG, et al. 2021; Liu, Hu, et al. 2021; Liu, Wei, et al. 2021; Liu, Li, et al. 2021; Chen et al. 2021; Liu, Peng, et al. 2022). The majority of these studies included a set of XML methods such as FI to quantify the impact of input variables.

Duquesnoy et al. (2020) introduced a hybrid methodology based on the experimental data used with physics-based models to generate mesoscale properties of the electrode such as tortuosity. Subsequently, ML models were employed to make predictions using the generated dataset. The experimentally obtained dataset at pilot scale consisted of 54 data points based on 14 distinctive cathode configurations. The analyzed parameters included active material content in the slurry, electrode thickness, and calendering pressure. The Sure Independent Screening and Sparsifying Operator (SISSO) was chosen as the ML algorithm for the model development. The study did not include further analysis of the impact of the parameters.

In a comprehensive study, Drakopoulos et al. (2021) conducted an extensive analysis of various parameters, including slurry formulation, mixing protocol, coating gap, coating speed, drying temperature, and porosity. They investigated the impact of these factors on the rheological properties of the slurry, the adhesion strength of the electrode, and the electrochemical performance using half-cells. The experimental dataset was based on 27 different anode configurations produced at laboratory scale. This study stands out as one of the most comprehensive cross-process analyses in the literature. However, there are some limitations to consider. While it was acknowledged that no systematic DoE approach was used for data generation, the article did not provide further details on how a mixture design with variations in slurry formulation was incorporated into process-focused experiments, such as those involving alterations in the coating gap. The study adopted exclusively Alchemite™, as the ML algorithm for the model development, without applying any other ML algorithms or using XML methods. Given the large number of factors under analysis, coupled with limitations in dataset size and transparency concerning data generation, it remains challenging to draw robust and generalizable conclusions from the findings. Additionally, it is worth noting that the scale-up of the effect of the parameters from the isolated drying step at the lab-scale to roll-to-roll drying, where a combination of web speed and temperatures can affect electrode quality, is another crucial aspect that was not addressed. This transition to a larger production scale may introduce complexities and variations that were not captured in the lab-scale experiments. Hence, it is essential to consider potential differences in outcomes between these two scales.

Duquesnoy et al. (2021) investigated how variations in slurry formulation, solid content, and coating gap impact electrode heterogeneity, taking into account its mass loading and thickness. With a dataset consisting of 144 different cathode configurations produced at a pilot line, the study adopted Gaussian Naive Bayes as the classifier. Probability plots were employed to examine the classification results and to explore the analyzed interdependencies and the associated results.

Using a two-step DoE-based approach, FARAJI NIRI et al. (2021) analyzed the impact of comma bar gap, web speed, and coating ratio on the relevant physical properties of the electrode, specifically mass loading, thickness, and porosity after the drying process. The produced cathodes were additionally characterized electrochemically through rate capability tests performed on half-cells. Based on a dataset derived from 32 electrode configurations and 115 coin cells, the study conducted a comparative analysis of various ML models including SVM, LR, NN, DT and Gradient Boosted Trees (GBT). Furthermore, the study examined how varying the number of folds in cross-validation affected the performance of the SVM models. While this study did not utilize XML methods to explore the significance of the parameters analyzed, in a subsequent publication (FARAJI NIRI, LIU, APACHITEI, et al. 2022), the same dataset was employed in conjunction with RF and GBT along with XML techniques. The factor contribution analysis and FI were used for this purpose. The study included additionally a dataset for anode production, analyzing the impact of comma bar gap and coating ratio on the predefined target variables. This work stands out as one of the few ML studies that provides data generated from pilot-scale anode production, encompassing 25 distinct configurations. The dataset from cathode production was also utilized in an XML study presented by LIU, FARAJI NIRI, et al. (2022). This study aimed to explore the influence of electrode mass loading, thickness, and porosity on discharge capacity, gravimetric, and volumetric capacity. The analysis was based on RF as the primary model, and the results were interpreted using feature interactions and ALE plots. Additionally, the developed model was compared with LR, SVM, and AdaBoost. Among these models, the ensemble methods, AdaBoost and RF, exhibited the highest performance, achieving an R^2 value of 0.98 in predicting the volumetric capacity.

In an extensive lab-scale study, FARAJI NIRI, REYNOLDS, et al. (2022) investigated the impact of various parameters from mixing and coating processes on cathode physical properties and areal capacity using half-cells. The study encompassed variations in slurry properties such as density, viscosity, solid content, and surface tension, coupled with adjustments in coating speed, coating gap, and coating dry density, resulting in a total of 67 electrode configurations. For model development, RF and GBT were employed. The significance of the analyzed parameters were evaluated using MDI and ALE techniques. The target variables analyzed included electrode wet thickness, dry thickness, mass loading, coating density, porosity, and cell capacity. The models exhibited varying performance for these target variables, with porosity demonstrating the lowest R^2 of approximately 0.58, while for wet thickness a high R^2 value of 0.94 was reported.

ROHKOHL et al. (2022) introduced a DM approach to streamline the setup and optimization of continuous processes in battery cell production, using the extrusion process as an example. The approach was based on three steps, with each step built upon an AI model. In the first stage, based on the domain knowledge, 15 parameters were identified that can affect the quality of the produced slurry. A DT model was developed to map the target product characteristics to the corresponding process parameters to be set. In the second step, a digital twin of the process was employed based on the set process parameters and their expected distribution. This digital twin served as a virtual representation of the process to identify the optimal process parameters considering economic and environmental aspects. The last step involved the process modeling and optimization. This step relied on the data collected during production to continuously monitor the product quality while taking into account certain process fluctuations. The last step was realized by developing an NN model using a 10-fold cross-validation approach to predict the shear viscosity at specific shear rates. This work stands out as the only study in the field of ML-based battery research to analyze the extrusion process, utilizing anode production data collected at a pilot line.

Using a dataset generated through DoE, WANG et al. (2022) investigated the effect of slurry formulation and different binder types on the rate capability of the battery cell at different C-rates. The study included the analysis of both LFP and LTO electrodes, produced at a lab-scale. For the electrochemical characterization half-cells were used. From the data analytics perspective, XGBoost was adopted for the model development in combination with XML methods. Furthermore, models were created based on SVM, NN, and DT to evaluate their performance in comparison to XGBoost. Both NN and XGBoost exhibited the highest performance in predicting the discharge capacity at a high C-rate, achieving an impressive R^2 value of 0.92. It is worth noting that the RMSE for NN was slightly larger than that for XGBoost.

FARAJI NIRI, APACHITEI, et al. (2022) analyzed the impact of product and process parameters in the calendering process on the cell impedance and capacity. Utilizing a dataset consisting of 18 different cathode configurations, resulting in a total of 54 half-cells, the discharge capacity at different cycles within a 50-cycle range and the area-specific impedance were predicted using the Extra Trees algorithm. The study encompassed a comprehensive analysis of feature contributions through various XML methods, including FI, SHAP, and ALE plots.

In a recent publication, DUQUESNOY et al. (2023) presented an ML-based optimization approach facilitated by synthetic data generated by physics-based simulation models. The simulation models were based on a number of parameters along the electrode manufacturing, including slurry formulation, solid content and compaction rate during the calendering process, enabling the estimation of the mesoscale properties of the electrode such as tortuosity. It should be noted that certain factors, such as drying temperature, were not considered in these physics-based models. The synthetic data generated by the simulation models for cathodes was utilized to train an ML model. For this purpose, the SISSO algorithm was adopted using a random test-train split. Based on the developed ML model, a Bayesian multi-objective optimization was conducted to determine the input space for the optimal mesoscale target variables. The results were visualized using a partial dependence plot. To validate the developed approach, the proposed parameters were adopted to produce electrodes, which were subsequently characterized. Since the model initially lacked certain key parameters, such as drying temperature, these were adjusted by domain experts. The results demonstrated a reasonable agreement between the predicted and the measured values.

The dissertation of TURETSKYY (2022), which was based on a number of relevant articles including TURETSKYY, WESSEL, et al. (2020), TURETSKYY, THIEDE, et al. (2020) and TURETSKYY et al. (2021), aimed to develop a concept for data analytics in battery production systems. As outlined in Section 3.2, data acquisition and management, focusing on different data sources, formats, and types, was defined as one of the key aspects that were addressed in the dissertation. Concerning data analytics, three use cases were introduced to demonstrate the capabilities of data-driven models. These included production quality planning, implementation of quality gates in the production process, and optimization of processes for energy efficiency. The first two defined use cases were considered relevant in this section and were analyzed further. As noted by TURETSKYY (2022), these two use cases addressed the same question from a data analytics perspective, aiming to identify the factors influencing battery cell performance. The primary distinction between the use cases lay solely in their industrial deployment (TURETSKYY 2022, p. 155). The first use case was presented in detail, using data derived from 191 LIB pouch cells, encompassing 1029 intermediate product features prior to data preprocessing and cleaning. Further details on the number of different configurations included in the dataset were not disclosed. An Analysis of Variance (ANOVA) was conducted, leading to the selection of maximum capacity, self-discharge, and State of Health (SoH) after 400 cycles as the target variables. Subsequently, a Pearson correlation analysis was carried out, using a threshold of 0.64, to identify highly correlated features, which were then removed. This process resulted in a total of 35 intermediate product features. Following this, a feature selection process, informed by domain know-how, resulted in choosing 10 intermediate product parameters for model development. These features included cathode solid content, cathode electric resistance, remaining electrolyte per thickness, cathode chamfer width and heat-affected zone, electrode porosity volume ratio, electrode overlapping rate, anode delamination area, cathode chamfer angle, amount of active material in the cathode, and the ratio of theoretical capacity of cathode to anode. The algorithms employed included LR, NN, SVM, RF, and XGboost. Among these, the last two models demonstrated the highest performance, yielding an R^2 value of approximately 0.86 when employing a cross-validation approach. However, despite a slightly lower R^2, NN was chosen over these models for further analysis. This choice was rationalized by the lower interpretability associated with the tree-based models. To enhance understanding of the developed model and provide decision support, a variance-based sensitivity analysis was performed as a statistical technique. This method provides insight into the global impact of the input variables. While the work presented by TURETSKYY (2022) has made a significant contribution to the field of data analytics in battery production, particularly through the systematic application of DM methods, it is important to highlight some remaining open aspects. The use case presented has shown the complexity involved in tackling high-dimensional problems and the necessity of feature reduction to develop an appropriate model using historical data. However, it is crucial to point out that the chosen threshold for the Pearson correlation value can be considered relatively critical, as it suggests the presence of highly correlated features within the dataset. Additionally, considering the final parameters selected for model development, it can be noted that the complexity associated with digitalization or the possibility of inline collection of the parameters was not taken into account. Moreover, the utilization of XML methods, particularly at the local level for deriving insights, was not considered.

To sum up, it is evident that data-driven approaches have received increasing attention over the last few years. From the production perspective, individual process steps have been the primary focus of most studies, while the investigation of interdependencies along the process chain, as demonstrated by the research works of DRAKOPOULOS et al. (2021) and TURETSKYY (2022), represents a distinct subset. The drying process, particularly at the pilot scale, has received limited attention, despite its status as an energy-intensive and quality-critical process step. The majority of studies have centered solely on the physical properties of the electrode, such as mass loading, thickness, and porosity. Subsequent to these, there are studies that have also included electrochemical characterization, enabling analysis of the impact of product and process parameters on cell performance. The mechanical property of the electrode, which is a key factor not only influencing the final cell performance but also the processability of the electrode throughout the process chain, has been the subject of investigation in only one study. From the data analytics perspective, the majority of studies have focused on evaluating the suitability of different algorithms for modeling purposes. However, the application of XML methods to extract insights from data-driven studies, particularly in the context of cross-process analyses with complex models, has remained limited.

3.5 Concluding Remarks and Research Opportunities

The literature review presented in Sections 3.2, 3.3, and 3.4 has reaffirmed the complexities involved in the field of data-driven analysis of battery cell production and summarized the existing contributions toward efficient, intelligent, and quality-oriented production. However, considering the objective of this dissertation, there are still certain research opportunities that should be explored and addressed.

Most of the previous research has primarily concentrated on individual steps involved in the data-driven analysis of the interdependencies in electrode manufacturing. These studies often centered on either the data generation and model development phases or, alternatively, just the development of models using historical datasets. A significant research opportunity lies in the development of a holistic framework that can provide guidelines to understand the process chain as a fundamental prerequisite for adopting data-driven models. This framework should extend beyond a basic understanding of the various process steps in the electrode manufacturing. It should also include the identification of parameters that are critical for maintaining product quality, explore methods for data generation and evaluation, and enable a comprehensive cross-process analysis. From an applicability perspective, an essential consideration involves the possible measurement solutions for the parameters identified as relevant. The inline collection of these parameters is vital to ensure the long-term viability and sustainability of ML applications in industrial production settings.

The majority of the existing data-driven studies have included a comparison of different modeling techniques. Hence, from the modeling perspective, a practical guide covering different algorithms, dataset sizes, and analyzed aspects can be highly beneficial. Such a guideline can serve as a valuable reference point for researchers and practitioners, offering insights into data-driven techniques, streamlining the modeling process, and ensuring best practices.

Furthermore, with the overall objective of extracting valuable insights from data-driven analysis, two critical dimensions should be taken into account. The first dimension involves a comprehensive assessment of data quality and addressing potential correlations between different features within the dataset. Neglecting this aspect can introduce inaccuracies and misinterpretations into the analysis, potentially leading to misleading results. Therefore, a research opportunity is to provide a blueprint for data generation and evaluation and to demonstrate the possible measures for improving data quality, especially when multicollinearity is present in the dataset. A statistically reliable and insightful dataset establishes the foundation for further in-depth analysis. The second dimension focuses on deriving insights from data-driven models. Considering the complexity of the process chain and the models developed in the existing literature, there is an inevitable need for interpretability of the ML models. The XML methods offer a pathway to navigate the complexity of modern industrial processes, data-driven decision-making, and process optimization by providing insights at both global and local levels.

Chapter 4

Conceptual Design

Following the overarching objective formulated in Chapter 1, along with the current state of the art explored in Chapter 3, the following phase according to DRM is the Prescriptive Study, starting with the conceptual design. Hence, this chapter begins by establishing the detailed Sub-Objectives (SOs) and the requirements necessary to meet these SOs in Section 4.1. Subsequently, Section 4.2 provides an overview of the proposed framework, which aims to facilitate holistic, efficient, and quality-oriented data-driven analysis of interdependencies in LIB electrode manufacturing.

4.1 Sub-Objectives and Requirements

Based on the research opportunities summarized in Section 3.5, four key SOs were defined using a deductive approach. The SOs aim to guide and facilitate the realization of the primary research goal of this dissertation, providing a structured and systematic framework for comprehensive data-driven analysis in electrode manufacturing.

SO1. *Identification of quality-relevant parameters and measuring instruments*

To establish a holistic framework applicable across different production lines, the first SO involves systematically identifying the quality-relevant parameters. This is followed by an overview of possible measuring instruments that can be used to characterize intermediate products as quality indicators along the process chain.

SO2. *Mapping of potential modeling techniques*

To provide practitioners with a comprehensive reference for data-driven modeling techniques and their associated facets, this SO aims to outline the modeling techniques employed and the aspects analyzed in the existing literature for LIB production.

SO3. *Methods for data generation and evaluation*

Following the identification of quality-relevant parameters, two use cases can be considered. The first one is based on the generation of a dataset, while the second use case evaluates the potential for consolidating and, if necessary, enhancing historical data.

SO4. *Development of data-driven models and derivation of insights*

Building on the results of the first three SOs, this step focuses on the exemplary development and evaluation of data-driven models using both newly generated and historical datasets. Additionally, it aims to provide comprehensive insights into the interdependencies within electrode manufacturing through the application of XML methods.

Aligned with the defined SOs, a set of requirements was formulated. Originating from the field of software engineering, according to the International Organization for Standardization (ISO), a *requirement* is described as "a statement which translates or expresses a need and its associated constraints and conditions" (IEEE/ISO/IEC 29148 2018, p. 4). Requirement definition is a precise and careful examination of the specific needs that a system or solution is expected to meet, its attributes or constraints. This step plays a key role in establishing the foundation for the subsequent phase of solution development. (ROSS and SCHOMAN 1977, p. 6) Various approaches can be employed to derive requirements. The requirements presented below, denoted as R1, R2, and so forth, were formulated using the conventional techniques according to NUSEIBEH and EASTERBROOK (2000, p. 40), relying primarily on the analysis of the existing literature and the identification of potential use cases. While certain requirements are specific to individual SOs, some should be considered throughout the solution development phase and across multiple SOs. The relevant SOs are additionally noted for each requirement.

R1. Definition of the characteristics of the production system, the process technologies considered, and the boundaries | SO1

R2. Integration of domain know-how as part of a holistic methodology | SO1 and SO2

R3. Provision of an overview of product and process parameters in electrode manufacturing | SO1

R4. Consideration of the relevance of the parameters from the quality management perspective | SO1

R5. Consideration of the digitalization aspect and the complexity involved in collecting the parameters for data-driven analysis | SO1

R6. Enabling a holistic cross-process analysis along the process chain | SO1, SO3, and SO4

R7. Consideration of the size of dataset when selecting potential ML techniques for analysis of interdependencies | SO2

R8. Consideration of possible use cases concerning data availability and quality | SO3

R9. Visualization of data-driven models in a comprehensive and interpretable form | SO4

R10. Quantification of the analyzed interdependencies and their impact on (intermediate) product properties | SO4

R11. Realization of a modular methodology, adjustable based on the objective of the analysis | SO1, SO2, SO3, and SO4

R12. Ensuring transparency and transferability by elaborating on methods and techniques adopted in each phase | SO1, SO2, SO3, and SO4

4.2 Overview of the Proposed Framework

Based on the formulated SOs and requirements, this section provides an overview of the framework proposed to facilitate the development of data-driven models for a holistic, efficient, and quality-oriented analysis of interdependencies in electrode manufacturing. The framework is built on the foundation of five distinct publications listed below, each of which addresses specific SOs.

I. Haghi, S., Summer, A., Bauerschmidt, P., Daub, R. "Tailored Digitalization in Electrode Manufacturing: The Backbone of Smart Lithium-Ion Battery Cell Production". In: Energy Technology 10 (2022) 10, 2200657 pp. 1-19. DOI: 10.1002/ente.20220065

II. Haghi, S., Leeb, M., Molzberger, A., Daub, R. "Measuring Instruments for Characterization of Intermediate Products in Electrode Manufacturing of Lithium-Ion Batteries". In: Energy Technology 11 (2023) 9, 2300364 pp. 1-13. DOI: 10.1002/ente.202300364

III. Haghi, S., Hidalgo, M. F., Faraji Niri, M., Daub, R., Marco, J. "Machine Learning in Lithium-Ion Battery Cell Production: A Comprehensive Mapping Study". In: Batteries & Supercaps 6 (2023) 7, e202300046 pp. 1-14. DOI: 10.1002/batt.202300046

IV. Haghi, S., Keilhofer, J., Schwarz, N., He, P., Daub, R. "Efficient Analysis of Interdependencies in Electrode Manufacturing Through Joint Application of Design of Experiments and Explainable Machine Learning". In: Batteries & Supercaps 7 (2024) 2, e202300457 pp. 1-18 DOI: 10.1002/batt.202300457

V. Haghi, S., Chen, Y., Molzberger, A., Daub, R. "Interdependencies in Electrode Manufacturing: A Comprehensive Study Based on Design Augmentation and Explainable Machine Learning". In: Batteries & Supercaps 7 (2024) 5, e202300556 pp. 1-11 DOI: 10.1002/batt.202300556

Figure 4.1 provides an overview of the proposed solution from the research perspective, structured according to the defined SOs and the corresponding publications. This structured representation is intended to facilitate a clear understanding of the alignment of the solution with the defined SOs and the contributions of each publication to the overarching objective.

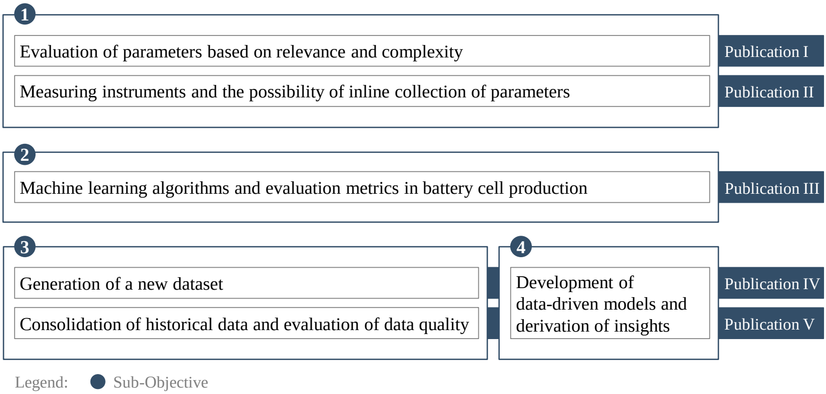

Figure 4.1: Overview of the proposed solution from the research perspective, structured according to the defined SOs and the corresponding publications

To address the initial SO and the corresponding requirements, in *Publication I* (HAGHI et al. 2022) a reference process encompassing the selected process technologies was defined, followed by an overview of product and process parameters in electrode manufacturing. Utilizing a thorough two-step approach, involving both literature review and expert interviews, these parameters were evaluated concerning their relevance from the quality management perspective and the level of complexity associated with the digitalization.

Expanding on the findings of the initial publication and addressing the first SO, *Publication II* (HAGHI, LEEB, et al. 2023) offered a detailed comprehensive assessment of potential measuring instruments for the characterization of intermediate products in the electrode manufacturing. For this purpose, a market analysis was conducted to gain insights into the available measuring instruments and pinpoint the parameters that are currently exclusively measurable using offline methods.

In alignment with the second SO, which aims to provide practitioners with an overview of ML techniques used in battery cell production along with best practices, a comprehensive mapping study was conducted as part of *Publication III* (HAGHI, HIDALGO, et al. 2023). The results not only highlighted the most commonly used algorithms and evaluation metrics, but also offered an overview of aspects that have received limited attention, including specific process steps and parameters.

To address the third and fourth SOs, which encompass topics such as data generation, evaluation, and model development, two use cases were defined. The first use case, outlined in *Publication IV* (HAGHI, KEILHOFER, et al. 2024), demonstrated an efficient analysis of interdependencies using a limited set of newly generated data. The second use case was based on the assumption of having access to historical data and showcased the methods used to assess data quality prior to model development. Design augmentation was adopted to enhance the data quality, ultimately enabling a comprehensive analysis of interdependencies in electrode manufacturing, achieved through the utilization of XML methods. The approach and findings were documented in *Publication V* (HAGHI, CHEN, et al. 2024).

While the first three publications were methodologically defined and developed based on a reference process and established criteria, the last two were demonstrated through experimental analysis conducted at the battery pilot production line available at the Institute for Machine Tools and Industrial Management (*iwb*) at the Technical University of Munich. The use cases exemplified the required methods and approach, utilizing the technology available at the research production line of the *iwb*, specifically a roll-to-roll doctor blade coating machine with three infrared dryers. The research gaps identified in the conducted mapping study (HAGHI, HIDALGO, et al. 2023) were also considered by the demonstrated use cases, ensuring that not only the methodology but also the research opportunities from a production perspective were addressed.

Figure 4.2 offers an overview of the proposed framework from the user's perspective, accompanied by respective sections summarizing the key findings for each step. The proposed framework can support users in selecting the relevant parameters depending on the objective of the analysis, while also considering the possibility of inline data collection. This is followed by the selection of potential data-driven techniques, including the ML algorithms. The framework covers two use cases for model development based on data availability. While the initial steps have a certain level of abstraction, the modeling step and associated methods were concretely demonstrated through empirical analysis, utilizing data generated at a research pilot line.

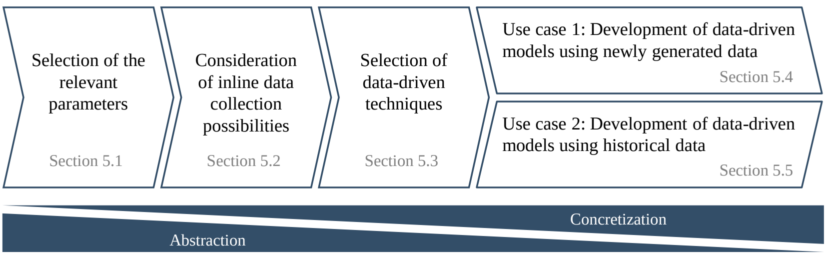

Figure 4.2: Overview of the proposed framework from the user's perspective

Chapter 5

Synthesis of the Research Findings from the Publications

This chapter provides a brief overview of the methods used and the key findings of the publications that address the SOs of this publication-based dissertation. Each publication is summarized in a dedicated section (Sections 5.1 to 5.5). A detailed account of the author's contributions to each publication can be found in Appendix B.

5.1 Publication I: Evaluation of Parameters Based on Relevance and Complexity

The overall objective of Publication I was to develop a tailored digitalization concept in electrode manufacturing. Following the introduction of key terms, including digitalization and traceability, the article emphasized that the realization of smart data-driven production with a high degree of data granularity is associated with substantial costs. This is particularly significant in electrode manufacturing, where processes are predominantly continuous and large volumes of time-series data are generated. This consideration served as motivation for the evaluation and prioritization of the parameters.

Following establishment of the requirements and definition of the scope and scale of the production system, including the considered process technologies such as slot die coating (HAGHI et al. 2022, pp. 2–3), a two-step approach was introduced. In the first step, a comprehensive mapping study was conducted to prioritize the parameters based on the existing interdependencies described in the literature. For this purpose, the DSM was used. The first step in the development of the DSM involved the literature-based identification of the product and process parameters as the matrix elements. Subsequently, the matrix was populated with data obtained from the conducted mapping study, which examined the literature over a ten-year period. In total, approximately 200 relevant articles were identified, analyzed, and incorporated into the matrix. The development of the DSM was based solely on the interrelationships between the parameters, without necessarily suggesting a cause-and-effect relationship. This approach assumed that a parameter with a high number of interdependencies can be considered relevant from the quality management perspective, highlighting the need for monitoring. The parameters within each process step were classified into two categories based on the number of interdependencies: highly relevant and less relevant.

In the second step of the proposed approach, a problem-centered semi-structured interview method was adopted. A total of 12 interviews were conducted with experts coming from prominent research institutes, each with substantial experience in a particular process of LIB electrode manufacturing. The workshops were tailored to each specific process and conducted individually. Detailed information on the expert profiles, the structure of the expert interviews, and the methods employed, can be found in HAGHI et al. (2022, pp. 11–12). The MoSCoW method, as introduced in Section 2.1, was adopted to evaluate the parameters. The experts categorized the parameters into four groups by considering two dimensions: parameter importance as well as the complexity and effort involved in digitalization. When assessing the importance of the parameters, aspects such as the influence of a parameter on the intermediate or final product, as well as on subsequent process steps and their associated parameters, were considered. Concerning the complexity involved in the digitalization, several factors, including the possibility of inline measurement, calibration effort, accuracy, and associated cost, were taken into account.

The final assessment of the parameters was based on the results of the conducted mapping study and the expert interviews. For this purpose, the importance of the parameters was evaluated using the results of DSM and MoSCoW analyses. In instances where discrepancies arose between the literature-based and expert-based evaluations, priority was given to the latter, as there are certain parameters that are considered relevant but have not yet been extensively analyzed in the literature. Based on the complexity aspect evaluated by the experts, a final evaluation and prioritization of the parameters was followed, resulting in the assessment of over 100 parameters. The coating and drying processes accounted for the majority of parameters categorized as *must* in the MoSCoW analysis.

In summary, Publication I addressed SO1 by providing a novel systematic approach for the identification of quality-relevant parameters in electrode manufacturing. The complexity associated with the collection of the parameters was qualitatively assessed through expert interviews. This publication established a comprehensive methodology that incorporated both expert knowledge and a systematic approach to extract insights from the existing literature. The result can serve as a guideline for the selection and prioritization of critical parameters, thereby contributing to more efficient and quality-oriented electrode manufacturing.

5.2 Publication II: Measuring Instruments and the Possibility of Inline Collection of Parameters

While the first publication addressed the topic of digitalization and the possible measurement strategies primarily through a qualitative, expert-based approach, Publication II aimed to conduct a comprehensive market analysis of potential measuring instruments. This analysis focused specifically on the characterization of intermediate products in electrode manufacturing. Intermediate products can serve as quality indicators for both the setup process parameters and the final product properties. By characterizing and monitoring intermediate product parameters, valuable insights can be gained from the production processes, revealing the complex interdependencies.

For a systematic evaluation of possible measuring instruments, a three-step approach, according to CAULFIELD et al. (2019, p. 7), was adopted. In the first step, the evaluation criteria and the boundary conditions were defined. The evaluation criteria included aspects such as measurement range, investment cost, and measurement strategy, indicating the possibility of inline or offline data collection. The boundary conditions and the considered product parameters were adopted from the previous publication (HAGHI et al. 2022). In the second step, a market analysis was conducted to identify the potential measuring instruments available on the market. This led to the identification of more than 40 suppliers, who were subsequently contacted. In the final step, a desk-based evaluation of the measuring instruments was performed based on the defined evaluation criteria, using the product catalogs and data sheets provided by the suppliers. Six distinct categories were established for the assessment of the investment cost as one of the evaluation criteria. Further detailed information in this regard can be found in HAGHI, LEEB, et al. (2023, p. 4). The results of the comprehensive market analysis revealed that certain product parameters in electrode manufacturing can currently only be acquired using offline characterization methods. For example, adhesion strength, a mechanical property of the electrode considered highly relevant from the quality management perspective, can only be measured through destructive offline techniques. To address such cases, the research community is currently exploring two strategies: the indirect assessment of such properties using innovative inline characterization methods such as spectrophotometry (WEBER et al. 2023), and the application of data-driven models (DRAKOPOULOS et al. 2021). By developing a comprehensive ML model based on the key product and process parameters that can be collected inline, there is potential to reduce the frequency and sample size needed to apply offline characterization methods. Ultimately, this paves the way for quality-oriented and efficient production.

Publication II extended the findings of the first publication by providing a more in-depth quantitative analysis of the measuring instruments available on the market for the characterization of intermediate products in electrode manufacturing. Based on the findings of the first two publications, SO1 was fully addressed. By presenting a comprehensive overview of both product and process parameters, highlighting their relevance from the quality management perspective, and exploring the possibility of collecting these parameters inline, the first two publications established a solid foundation for the development of data-driven solutions.

5.3 Publication III: Machine Learning Algorithms and Evaluation Metrics in Battery Cell Production

Publication III was dedicated to conducting a mapping study focused on the ML methods employed in the context of battery cell production. Based on a thorough examination of the ML application landscape in battery cell production, the study not only provided insights into potential use cases, but also extracted and synthesized the critical information such as the commonly adopted algorithms and evaluation metrics. The overarching objective was to derive best practices and assist practitioners with effective knowledge transfer.

For this purpose, a comprehensive mapping study was carried out based on the main steps introduced in Section 2.1. In total, 215 publications were retrieved, examined, and shortlisted, resulting in 38 articles that were identified as relevant and subsequently subjected to comprehensive analysis. Further details in this regard can be found in HAGHI, HIDALGO, et al. (2023, p. 3). The relevant articles were analyzed from two perspectives. The first perspective delved into the production aspect, exploring process steps, product and process parameters used as input variables, target variables, the materials analyzed, and the production scale. When interdependencies were examined in conjunction with electrochemical characterization at the cell level, the type of cell—coin, pouch, or prismatic—was also incorporated into the analysis. The second perspective focused on the ML aspects and included considerations such as the algorithms employed, the evaluation metrics, and the size of the dataset.

The results indicated that a significant portion of the studies primarily utilized supervised ML to assess how product and process parameters affected cell characteristics. Following this, some articles delved into analyzing the impact of parameters on intermediate product properties. Approximately 60 % of the studies concentrated on electrode manufacturing, with the majority of them focusing on cathodes. The datasets used in the ML studies were predominantly generated at pilot production lines. However, for the characterization of the produced electrodes, coin cells, both in half-cell and full-cell formats, were mostly employed. Among the cell characteristics, the discharge capacity at various C-rates and the capacity loss after a certain number of cycles have been the most extensively studied parameters. In these investigations, common input variables included intermediate product parameters such as active material weight, electrode thickness, and porosity. As elaborated in Section 3.4, the findings revealed an uneven distribution of research attention among different process steps in data-driven studies. Notably, the drying process has been less explored, specifically at the pilot scale.

Regarding modeling techniques, the most frequently used algorithms were NN and RF, with SVM being the next most popular choice. In terms of evaluating the developed models, common metrics included R^2, RMSE, and MAE. The mapping study also aimed to investigate the interplay between the number of input variables, the size of the dataset, and the algorithms adopted. For this purpose, only studies that offered detailed information on dataset configurations and unique instances, including replicates, were considered. Nevertheless, the outcome did not reveal any discernible pattern or overarching principle, suggesting that a trial-and-error approach is frequently employed in algorithm selection.

The mapping study also highlighted the aspects that demand greater attention. These aspects encompass the adoption of best practices for documenting and publishing ML models, along with the incorporation of XML methods. The latter is particularly relevant due to the findings of the study, which revealed that the prevailing algorithms commonly employed in battery cell production are often quite complex. Consequently, this complexity may impact their trustworthiness and hinder their successful deployment in large-scale production environments.

The conducted mapping study addressed SO2 by offering a comprehensive overview of the most commonly adopted algorithms and techniques, serving as valuable reference points. Furthermore, the study unveiled research gaps from both production and ML perspectives. While the primary objective of the subsequent publications was to showcase the required methods for data-driven analysis of interdependencies in electrode manufacturing, they also committed to actively address and bridge the identified research gaps.

5.4 Publication IV: Data-Driven Analysis of Interdependencies Using Newly Generated Data

Building on the results of the initial three publications, Publication IV aimed to address the first use case of SO3 and develop data-driven models within the context of SO4. Taking into account the existing literature on formulation and mixing (see Section 3.4), this study narrowed its scope to a specific single slurry formulation, with the objective of providing a systematic approach for analyzing the impact of quality-relevant parameters in electrode manufacturing.

The use case involved employing the I-optimal method to generate a new dataset. The choice of this method was driven by two main considerations. Firstly, the I-optimal design enables a comprehensive evaluation of the impact of input variables on the response, encompassing both quantitative and qualitative aspects. It allows for the identification of optimal regions while simultaneously minimizing the average prediction variance across the entire experimental space. Secondly, the application of the I-optimal method addressed a research gap highlighted in Chapter 3, concerning the currently limited utilization of optimal design methods in battery cell production. As outlined in Section 4.2, for the experimental analysis, the process technology available at the pilot production line at the *iwb*, had to be taken into consideration. Accordingly, the results of Publication I were utilized and adapted to determine the quality-relevant parameters to be incorporated into the DoE. One of the ultimate objectives of the study was to determine the impact of the quality-relevant parameters on the cell properties. Intermediate product parameters can be considered as the main quality benchmarks at different stages of the process chain, ultimately influencing the final cell properties in large-scale production. To ensure a consistent foundation for the analysis of multiple aspects, the intermediate product parameters were considered as factors for DoE. It should be noted that, depending on the objective of the analysis, the results of Publication I can guide the selection of suitable parameters. For instance, in a Taguchi design, the quality-relevant process parameters could be considered as factors. Table 5.1 offers an overview of the product parameters that were rated as highly relevant in both the literature-based approach and the expert interviews, as documented in the results of HAGHI et al. (2022), along with their complexity with respect to digitalization.

While defects in both coating process, mainly caused by slot die coating or inhomogeneous slurry, and subsequent calendering process are relevant topics, a conscious decision was made to exclude these aspects from the scope of analysis within the DoE. Several considerations and contextual factors influenced this choice. The primary goal of the DoE-based study was to provide a focused and precise examination of the key factors influencing product quality throughout the electrode manufacturing. To achieve this, it was imperative to restrict the number of variables under investigation. Furthermore, there are extensive studies in the literature dedicated exclusively to the examination of defects in both coating process (BARET DE LIMÉ et al. 2022) and calendering process (GÜNTHER et al. 2020).

As outlined in SO4, the objective of the study was to gain insight into existing cross-process interdependencies through the application of XML methods. Consequently, it was crucial to ensure the independence of the input variables. As a solution, the input variables were narrowed down to the electrode porosity after calendering and mass loading, effectively eliminating the third dependent variable, the electrode thickness. Similarly, due to the existing linear correlation between wet film thickness and mass loading, only the latter was included in the analysis. Due to the considerable complexity associated with the digitization of adhesion (HAGHI, LEEB, et al. 2023) and the challenges in determining appropriate levels to be included in the DoE, the temperature of the second dryer was chosen as an influential factor affecting adhesion. This choice was guided by insights derived from the literature concerning different mechanisms during the drying process (KUMBERG et al. 2019, pp. 1–2).

Table 5.1: Overview of the intermediate product parameters evaluated as highly important and their complexity in terms of digitalization, based on the results of HAGHI et al. (2022), adapted from HAGHI, KEILHOFER, et al. (2024, p. 2)

Process	**Intermediate product parameter**	**Importance**	**Complexity**
Coating	Wet film thickness	High	Low
	Quality of wet film	High	High
Drying	Adhesion	High	High
	Electrode thickness	High	Low
	Mass loading	High	Low
	Porosity	High	High
	Quality of electrode (defects)	High	High
Calendering	Electrode thickness	High	Low
	Porosity	High	High
	Adhesion	High	High
	Quality of electrode (defects)	High	High

Consequently, the primary factors selected for analysis included mass loading, second dryer temperature, and porosity after calendering. These factors formed the basis for generating an I-optimal design. The DoE plan, consisting of 17 different electrode configurations, was generated using Design-Expert® software. To guarantee the absence of significant multicollinearity within the dataset, the VIF was employed as an evaluation metric. The produced anodes were characterized mechanically, and electrochemically. The mechanical characterization included adhesion measurements, while the electrochemical characterization was based on EIS using symmetric cells and the rate capability test performed on full coin cells. For detailed information on the analyzed ranges, the configurations, the experimental setup, and the characterization methods, please refer to HAGHI, KEILHOFER, et al. (2024, pp. 3–5).

Following a detailed description of the data preprocessing and cleaning phase, the study introduced three scenarios for the development of data-driven models. The first scenario was based on sample-specific product parameters considered in the DoE and drying temperature. The second scenario utilized averaged corresponding process parameters, while the last scenario incorporated both process parameters and intermediate product parameters collected during the process. The primary aim of the study was to employ XML methods to uncover interdependencies in electrode manufacturing and assess the influence of various parameters on intermediate and final product properties. Hence, it was essential to ensure the independence of input variables. Notably, the last scenario did not meet this requirement and was not considered in conjunction with XML methods. This becomes apparent when considering factors such as coating gap as a process parameter, which correlates with intermediate product parameters such as electrode wet mass loading and dry mass loading, which were collected inline. However, this scenario was included to investigate the significance of inline measuring instruments and their contribution to the development of high-performing ML models. In the first two scenarios, the suitability of potential input variables was assessed using a correlation matrix with a predefined threshold of 0.3 for the Pearson correlation value.

Guided by the findings of the conducted mapping study (HAGHI, HIDALGO, et al. 2023) and given the size of the dataset, which ranged from approximately 50 to 85 data points, with a minimum of three samples considered for each configuration, RF and SVM were chosen for model development. Furthermore, the study incorporated simpler algorithms, including MLR and PolyR. The model development was based on an 80-20 split, with the test data points being randomly selected. Therefore, to mitigate potential overfitting issues associated with RF, XGBoost was employed in specific instances. A number of target variables were explored, including physical properties such as electrode thickness and mass loading after drying, as well as tortuosity, adhesion strength, ionic resistance, discharge capacity, and gravimetric capacity at different C-rates. For the evaluation of the developed models, R^2 and RMSE were employed. These evaluation metrics were reported for both the training and test datasets, serving as benchmarks for assessing the generalization of the developed models. Additionally, the ratio of MSE for the training dataset to MSE for the test dataset was taken into consideration as an indicator of generalization.

The study included an additional verification step to evaluate the methodology employed and the models developed. This verification approach involved comparing the predictions made by the model with the tortuosity estimates derived from the adjusted Bruggeman relationship (see Section 2.2). The results demonstrated strong overall agreement. For further details, please refer to HAGHI, KEILHOFER, et al. (2024, pp. 11–12).

In terms of interpretability, an assessment of global FI was conducted using the MDI method. To gain a more profound understanding of the developed models and the analyzed interdependencies, SHAP plots were used to provide instance-level explanations. Additionally, the results of the XML methods were compared with the normalized coefficients derived from the developed MLR models. This comparative analysis aimed to investigate the level of agreement between these methods, considering that they were derived from models with varying degrees of complexity and performance. As an illustrative example, Figure 5.1 offers an overview of the results for ionic resistance and tortuosity as target variables, derived from the developed XGBoost and MLR models. It should be noted that in terms of R^2 values, the XGBoost models outperformed the MLR models for both target variables. While the developed MLR model demonstrated high predictive performance for ionic resistance, achieving an average R^2 value of approximately 0.86 between the training and test datasets, it is noteworthy that the results derived from the normalized coefficients (cf. 5.1 (b)) diverged from those obtained by the MDI method (cf. 5.1 (a)). Specifically, mass loading was identified as the least influential parameter impacting ionic resistance based on the coefficients of the developed MLR model. In contrast, for tortuosity, both methods yielded consistent parameter rankings, identifying porosity as the most significant parameter and drying temperature as the least influential one (cf. 5.1 (c), (d)). The observed discrepancy in the case of ionic resistance underscored the possible limitations of MLR and the importance of applying multiple techniques, both from a modeling and interpretation perspective, to achieve a holistic understanding of the analyzed interdependencies.

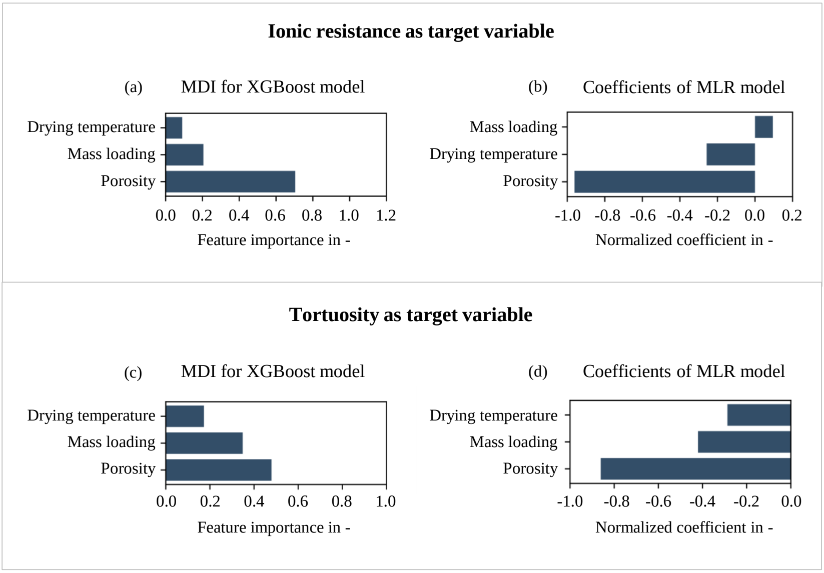

Figure 5.1: Exemplary results of the applied XML method and the MLR coefficients, revealing the influence of the analyzed features on the ionic resistance (top) and the tortuosity (bottom), adapted from Haghi, Keilhofer, et al. (2024, p. 12)

Publication IV addressed SO3 and SO4, based on a use case involving the systematic, efficient generation of a new dataset. Such use cases become particularly relevant when introducing new materials or undertaking specific optimization studies. The results can be summarized from three main perspectives. In terms of the production aspect, the possibility of predicting physical parameters such as mass loading has been investigated. Such ML models can be used to support practitioners in finding the appropriate process parameters in the ramp-up phase for a predefined intermediate product parameter, or can be adopted in combination with an inline control system to dynamically adjust the process parameters as part of a prescriptive analytics. Additionally, the study explored the potential for predicting parameters that, according to the current state of the art (HAGHI, LEEB, et al. 2023), cannot be measured in real-time during the production, showcasing the ability to estimate these parameters based on data collected inline. Moreover, the impact of quality-relevant parameters on cell performance was investigated. Through the definition of three distinct scenarios and a comparison of model performance, the research work explored the relevance of sensor data and sample-specific parameters, thus highlighting the valuable contributions of digitalization and tracking and tracing systems in the field of battery cell production. From a methodological perspective, the study can be considered as one of the early endeavors to provide a novel systematic framework for conducting efficient, quality-oriented cross-process analyses throughout the electrode manufacturing process chain. Addressing the research gaps identified in the conducted mapping study (HAGHI, HIDALGO, et al. 2023), the article aimed to offer detailed insights into the steps required to perform such analyses effectively. From the ML perspective, the study explored the capabilities of various algorithms, ranging from complex models such as SVM and RF to simpler approaches such as MLR. The results revealed that for a set of target variables, such as mass loading or discharge capacity at lower C-rates, simple algorithms yielded commendable performance. On the contrary, when dealing with intricate and multifaceted phenomena such as adhesion or discharge capacity at higher C-rates, complex models proved to be more effective. With the overall objective of providing insight into the existing interdependencies, XML methods were adopted, offering both global and local explanations for the analyzed aspects.

5.5 Publication V: Data-Driven Analysis of Interdependencies Using Historical Data

Publication V aimed to address SO3 and SO4 by focusing on a use case that involved the consolidation of historical data. The overarching objective was to demonstrate the essential methods required to assess data quality when dealing with historical data, and the possible measures to enhance data quality for comprehensive analysis prior to model development.

The article emphasized that historical data, particularly when originating from pilot line production, stems from a collection of studies, each designed with specific objectives and analyzed in a particular context. Therefore, historical data of this nature may encompass biases or discrepancies that can have a considerable impact on the development of the model and the subsequent insights derived from it. Consolidation of such data extends beyond mere aggregation; it necessitates a comprehensive understanding and evaluation of the dataset. When deemed appropriate, adjustments or refinements to the experimental space may be essential to ensure the reliability of the data and its relevance for model development. The use case presented aimed to simultaneously address the research gap identified in the mapping study from a production perspective, enabling a comprehensive analysis of interdependencies in the drying process, along with the closely related process steps of coating and calendering.

The historical data consisted of two datasets generated at the pilot production line at the *iwb*, which employed a roll-to-roll coating machine with three infrared dryers. The first dataset resulted from the analysis presented in Publication IV, while the second dataset featured variations in mass loading, porosity, and temperature of the first dryer. In each dataset, the remaining drying conditions, including the web speed and temperature of the other two dryers, were held constant. However, it is worth noting that these conditions differed between the two datasets. This range of variations within the historical dataset was expected to allow a holistic analysis of the influence of various parameters, including first and second dryer temperatures and drying web speed. To verify this potential, an initial assessment of data quality was conducted using the Pearson correlation analysis of the potential input parameters. The result revealed a strong correlation between the drying web speed and the temperature of the first dryer, with an absolute coefficient value of 0.73. This suggests that the historical data, in its existing state, lacks the capability to represent these parameters independently. To address the existing multicollinearity in the dataset and enable a comprehensive analysis of the considered parameters, design augmentation was adopted. By conducting additional runs using space-filling design augmentation approach, a more comprehensive coverage of the design space was achieved, resulting in the generation of an enriched dataset for in-depth investigation of the effects of individual parameters. The design augmentation was conducted using Design-Expert® software, resulting in a final dataset of 40 distinct electrode configurations. For further details on the analyzed ranges, the historical and augmented datasets, please refer to HAGHI, CHEN, et al. (2024). Figure 5.2 presents the correlation matrices before and after the design augmentation. This visual comparison highlights the impact of the additional runs, which effectively balanced the dataset and eliminated critical strong correlations present in the historical data. In addition to the correlation matrix, VIF and FDS were adopted as quality indicators to evaluate the prediction capability of the dataset. Through a comparative analysis of these measures, the study highlighted the improvements achieved, particularly in the case of the correlation coefficients and VIF values, as a direct outcome of the design augmentation techniques applied.

Following the comprehensive data quality assessment and improvement, this study aimed to explore the existing interdependencies between key product and process parameters in electrode manufacturing, with electrode adhesion and cell characteristics, adopted as target variables. Building upon the insights derived from the preceding study and the outlined scenarios (HAGHI, KEILHOFER,

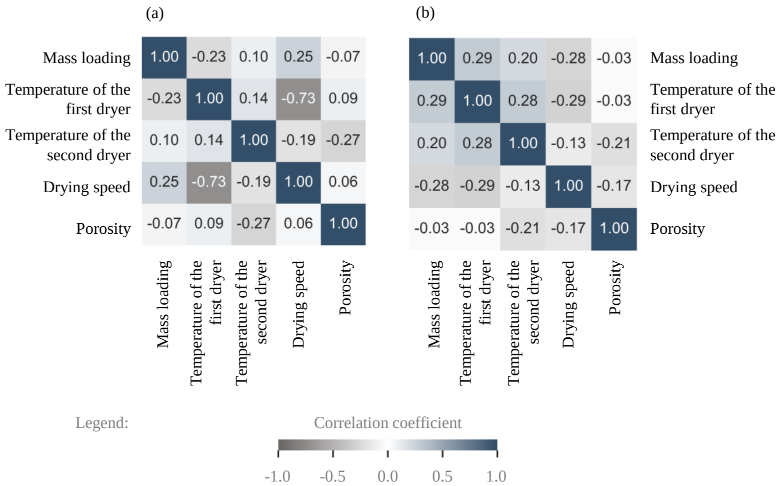

Figure 5.2: Correlation matrices based on (a) historical dataset, and (b) augmented dataset for potential input parameters considered for model development, adapted from HAGHI, CHEN, et al. (2024, pp. 3–4)

et al. 2024), this article integrated sample-specific product parameters, including mass loading and porosity, in conjunction with process parameters, such as drying temperatures and drying web speed, as input variables for model development. ML models were developed and evaluated using a 5-fold cross-validation approach.

The final dataset consisted of approximately 125 data points. Given the size of the dataset and the relatively large number of input variables, in conjunction with the chosen cross-validation approach, RF and SVM were selected for model development. The developed models were evaluated using R^2, RMSE, and MAE. The evaluation metrics were reported based on the average values computed across the folds.

In terms of interpretability, three methods were adopted. The permutation FI method, known for its robustness in handling correlated features, was used to provide a global understanding of the analyzed parameters. In addition, SHAP and ALE plots were employed to offer detailed, instance-level explanations. Figure 5.3 provides a representative illustration of the results obtained from the adopted permutation FI and SHAP analyses for the developed RF model predicting adhesion strength. The permutation FI highlights the relative significance of each parameter and their contributions to the adhesion strength. Whereas the SHAP plot not only quantifies the importance of each parameter, with the most influential parameter listed at the top of the plot, but also provides insight into the direction in which these parameters impact the adhesion strength (cf. 5.3 (b)). Both methods identified porosity as the most influential parameter when analyzing adhesion strength. A low porosity, indicating a high compaction rate, was found to positively impact adhesion strength. Mass loading was identified as the second influential parameter. Concerning the drying process, within the analyzed ranges, the temperature of the first dryer demonstrated the highest impact on adhesion strength, followed by the drying web speed and the temperature of the second dryer. It is worth noting that since the presented global ranking does not exist in the literature, only certain aspects, such as the correlation between mass loading and adhesion, could be verified based on the existing experimental-based studies, often conducted using the conventional OFAT approach. Notably, the results for these specific aspects were consistent with those reported in the literature. For detailed analysis and interpretation of the findings, please refer to HAGHI, CHEN, et al. (2024).

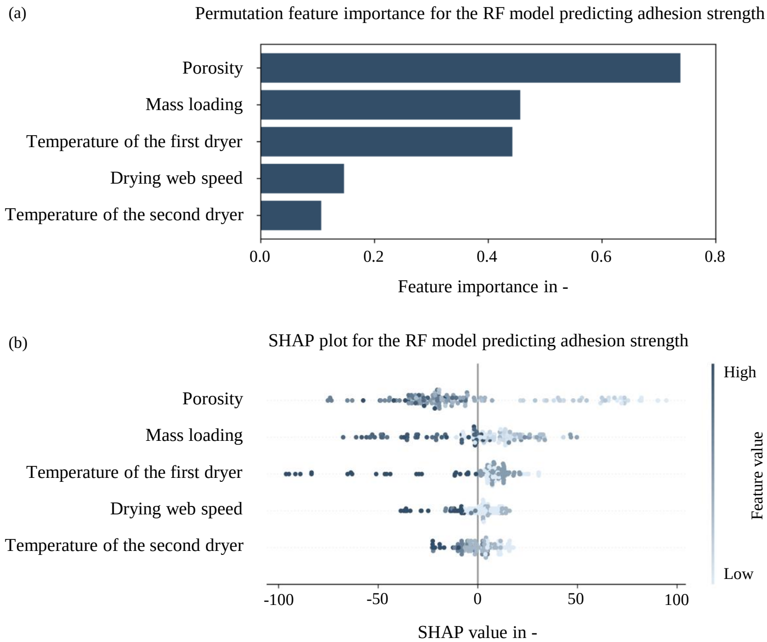

Figure 5.3: Exemplary results of the applied XML methods, illustrating the impact of input features on adhesion as a target variable in the developed RF model, using (a) permutation FI and (b) SHAP plot, adapted from HAGHI, CHEN, et al. (2024, p. 7)

Publication V addressed SO3 and SO4 by examining a use case involving the utilization of historical data for a comprehensive analysis of interdependencies in electrode manufacturing. This demonstrated use case encompassed various aspects, spanning from methods for assessing data quality and enhancing predictive capabilities through design augmentation, to model development, and derivation of valuable insights. The enriched dataset obtained through the design augmentation, in conjunction with the use of XML methods, facilitated a comprehensive, in-depth analysis of various critical parameters within the drying process, a field that has previously remained largely unexplored in data-driven battery production studies. The developed models, particularly the RF, demonstrated good performance, with an average R^2 value of 0.72, 0.97, and 0.87 for adhesion strength, discharge capacity at 0.1C, and discharge capacity at 5C, respectively. However, it should be noted that this performance was slightly lower than the results presented in Publication IV. This difference can be attributed to the chosen cross-validation approach and the comprehensive analysis that involved various input variables in Publication V. Through the use cases presented in Publications IV and V, the last two SOs of this dissertation were effectively and comprehensively addressed.

Chapter 6

Discussion

This chapter focuses on the evaluation of the proposed framework. Section 6.1 assesses the practical application of the framework and its contribution to the current state of the art. This evaluation is considered as part of the Descriptive Study II within the context of DRM. For this purpose, the exemplary application of the framework, which was based on the data generated at the LIB pilot battery production line of the *iwb*, was taken into account. Furthermore, Section 6.2 delves into a discussion concerning the transferability and limitations of the framework.

6.1 Evaluation of the Proposed Framework

Emerging from the domain of software engineering, verification and validation are considered as pivotal elements in the system development process. Verification primarily assesses whether the system has been constructed correctly throughout the development process, while validation encompasses a range of activities to ensure that the correct system has been built. (IEEE/ISO/IEC 29148 2018, p. 7) The exemplary ML models developed were subjected to verification through formal analysis and a number of methods, including comparison with the modified Bruggeman estimation, existing literature, and cross-validation approach, as detailed in Sections 5.4 and 5.5. Nonetheless, it is essential to carry out a thorough and holistic evaluation of the proposed framework, considering the defined objective and requirements.

6.1.1 Evaluation of Application

In a series of publications, PEDERSEN et al. (2000) and SEEPERSAD et al. (2006) proposed a framework for the validation of engineering design methods. This framework encompasses two primary aspects: structural validity and performance validity, both to be evaluated from theoretical and empirical perspectives. *Structural validity* assesses the internal consistency of the methods and their underlying assumptions within the intended domain of application. From an empirical perspective, it is important to consider the relevance of the exemplary use case adopted to evaluate the framework. *Performance validity* involves a quantitative evaluation based on an exemplary use case. Furthermore, the theoretical perspective as part of the performance validity considers evaluating the framework beyond the exemplary use case. Through these two aspects, the efficiency and effectiveness of the proposed solution, based on the defined objectives, can be evaluated. (PEDERSEN et al. 2000) Figure 6.1 provides an overview of the adopted approach and the key aspects assessed, serving as a guideline for the evaluation of the proposed framework in this dissertation.

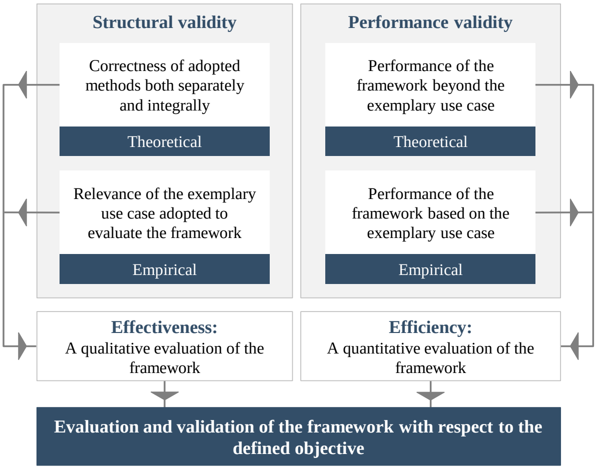

Figure 6.1: Overview of the adopted validation approach, adapted from PEDERSEN et al. (2000, p. 6)

From the structural perspective, the methods adopted within the proposed framework, including mapping studies combined with DSM, MoSCoW analysis, and the DoE techniques, are widely recognized and accepted in fields such as complexity management and process analysis. To underscore their effectiveness as an integrated system, the structural validity of these adopted methods was demonstrated through a practical use case presented in Publication IV. This use case was built upon the foundational insights and research findings from the initial three publications. Considering the empirical aspect of the structural validity, in line with the defined requirements outlined in Chapter 4, the first SO was addressed using a broadly applicable reference process chain. This ensured the relevance and effectiveness of the proposed solution across a range of scenarios and contexts in electrode manufacturing. For the model development phase, two distinct use cases were defined to represent possible relevant scenarios. By providing a detailed description throughout the development process, transparency was ensured so that the proposed framework and the adopted methods could be adjusted as necessary.

From the performance perspective, the proposed framework addressed the overall objective of the dissertation, enabling a holistic, efficient, and quality-oriented analysis of interdependencies in electrode manufacturing. Based on the results of the first two publications, the cross-process analysis conducted in the first use case was restricted to three quality-relevant parameters. By employing an optimal DoE method, the number of required experiments was minimized to 17 runs. It should be noted that within the analyzed context, considering the factors and their defined levels, a total of 64 experimental runs would have been needed for a conventional full factorial design. The proposed framework streamlined the investigation of existing interdependencies in electrode manufacturing by identifying quality-relevant parameters, integrating DoE methods to generate insightful data efficiently, and incorporating XML methods. This comprehensive methodology enhanced the analysis of interdependencies, leading to the development of high-performing data-driven models, while at the same time reducing the experimental cost and effort. Through the integration of existing knowledge, collected from the conducted mapping studies and expert interviews, along with the inclusion of data generated at a pilot production line, the domain-specific theoretical performance validity of the proposed framework was addressed, extending its application beyond a specific use case and ensuring a certain degree of generality and transferability. Nevertheless, it is important to acknowledge that the performance evaluation was restricted to one specific pilot line. Hence, from the DRM perspective, the approach adopted for this stage represents an initial assessment rather than a comprehensive one.

6.1.2 Fulfillment of Requirements

Following the evaluation of the application of the proposed framework, the extent to which the requirements have been fulfilled is discussed in this subsection.

R1. Definition of the characteristics of the production system, the process technologies considered, and the boundaries: In compliance with this requirement, a reference process chain was defined as part of Publication I. This process chain included common technologies, such as slot die coating, and provided a detailed description of the characteristics and boundary conditions of the production system. This initial step was crucial to establish a solid foundation for the first two publications and to ensure transparency throughout the developed framework (cf. R12).

R2. Integration of domain know-how as part of a holistic methodology: The proposed framework comprehensively addressed various crucial aspects involved in the development of data-driven models, ranging from the identification of relevant parameters and the possibility of collecting these parameters inline, to the selection of appropriate algorithms, data generation, model development, and the derivation of insights. Through mapping studies, expert interviews, and market analysis, domain know-how was systematically extracted, structured, and subsequently used to provide guidelines within the holistic framework.

R3. Provision of an overview of product and process parameters in electrode manufacturing: An overview of product and process parameters was presented as part of Publication I. For this purpose, a literature-based approach was adopted, guided by the defined reference process chain (cf. R1).

R4. Consideration of the relevance of the parameters from the quality management perspective: This requirement was fully addressed in Publication I using a two-step approach. In the first step, a DSM was developed based on the existing interdependencies identified through an extensive mapping study. In the second step, valuable insights from expert interviews were integrated into the assessment. Collectively, these steps ensured a comprehensive evaluation of the relevance of the parameters.

R5. Consideration of the digitalization aspect and the complexity involved in collecting the parameters for data-driven analysis: Initially, this requirement was addressed qualitatively during the expert interviews as part of Publication I. In Publication II, the emphasis was shifted toward providing quantitative results, which were obtained through market analysis and a desk-based evaluation of possible measuring instruments. The results of the evaluation included factors such as investment cost and the possibility of collecting the parameters inline.

R6. Enabling a holistic cross-process analysis along the process chain: This requirement was addressed through a holistic and comprehensive assessment of product and process parameters in electrode manufacturing in Publication I, along with the presentation of a demonstrated use case with a specific objective and a detailed description of the parameter selection approach in Publication IV. Both data-driven use cases presented were based on cross-process analyses, encompassing a set of quality-relevant parameters from coating, drying, and calendering processes.

R7. Consideration of the size of dataset when selecting potential ML techniques for analysis of interdependencies: Publication III addressed this requirement by conducting a comprehensive mapping study that involved the extraction and aggregation of relevant information. It is important to note that in certain aspects, such as the size of the dataset and the selected algorithm in combination with the number of features, no overarching pattern could be identified in the existing studies. Nevertheless, the multi-perspective comparison provided a solid starting point that can be utilized by practitioners in this field.

R8. Consideration of possible use cases concerning data availability and quality: To address this requirement, the development of data-driven models was presented in two use cases, with the first one involving the generation of a new dataset and the second one relying on the utilization of historical data. To provide a comprehensive approach, the second demonstrated use case highlighted poor data quality with interrelated input parameters. This issue was effectively addressed through design augmentation, resulting in an enriched dataset. The improvement in data quality was evaluated using three methods, as discussed in Publication V.

R9. Visualization of data-driven models in a comprehensive and interpretable form: The data-driven analyses presented in Publications IV and V were complemented with XML methods to address this requirement. The results included various visualizations such as SHAP plots.

R10. Quantification of the analyzed interdependencies and their impact on (intermediate) product properties: The adopted XML methods provided detailed insights into the impact of the analyzed parameters. The results revealed not only the global ranking of the parameters, but also the local impact and direction of their influence on the considered target variables.

R11. Realization of a modular methodology, adjustable based on the objective of the analysis: According to BALDWIN and CLARK (2003, p. 86), *modularity* is a strategy to efficiently organize complex systems or processes. In a modular approach, independent units or modules are developed separately and can function cohesively as an integrated whole. The modularity of the proposed framework was implicitly achieved as a result of addressing SOs individually in separate publications. Each publication focused on specific SOs, allowing for a modular approach to the overall objective. The cohesiveness of the framework was successfully demonstrated in Publication IV for a specific use case with a defined objective. However, it is important to note that no additional evaluation, specifically with respect to modularity, was conducted beyond this demonstration.

R12. Ensuring transparency and transferability by elaborating on methods and techniques adopted in each stage: This requirement was consistently considered and addressed throughout the solution development phase. This involved the adoption of recognized methods, elaboration on the approaches and criteria used to conduct mapping studies and extract information, outlining the profiles of the experts, and addressing discrepancies when they occurred. From a data analytics perspective, a detailed description of the selection of parameters to be considered in the DoE method was disclosed, accompanied by information on the distinct configurations and variations included in the dataset, the data preprocessing and cleaning steps, the size of the dataset, the model selection, the specific hyperparameters chosen, and the evaluation metrics applied. All relevant software tools and methods, as well as the materials and characterization techniques employed, have been disclosed in the respective publications. These efforts were aimed at creating transparent documentation that ensures the adaptability and transferability of the framework and facilitates an in-depth understanding of the methods and techniques used at each stage.

6.1.3 Contribution to the State of the Art

As the requirements discussed in the previous subsection were mainly derived from the existing literature, this dissertation is expected to make distinct contributions to the state of the art. These contributions are briefly discussed below.

From the methodological perspective, the proposed framework covered the main steps involved in analyzing interdependencies based on data-driven models. These steps included identification of relevant parameters considering the quality management and digitalization perspective, data generation, selection of algorithms, development of models, and derivation of insights. Notably, this dissertation did not delve into the topic of data acquisition and management, as a comprehensive work on this subject was presented by TURETSKYY (2022). Concerning the first SO, as outlined in Section 3.2, a study closely aligned with the results of Publication I was simultaneously presented by ZANOTTO et al. (2022). However, in terms of transparency and comprehensiveness, there are certain aspects where the study could potentially benefit from further exploration. These aspects were fully addressed in Publications I and II. The topic of measuring instruments has been investigated so far in isolated studies, covering individual process steps, such as the review articles presented by REYNOLDS et al. (2021) and ZHANG et al. (2022). Publication II effectively addressed this gap by providing a quantitative evaluation of the existing measuring instruments available on the market for the entire electrode manufacturing process chain.

While there have been comprehensive literature reviews that have explored the application of ML in the battery field, such as the work presented by LOMBARDO et al. (2021), Publication III went beyond a literature review by extracting and synthesizing the results of the conducted mapping study. The article provided an extensive, multi-perspective comparison, highlighting the most commonly adopted algorithms, evaluation metrics, and use cases. This comprehensive mapping study contributed to the enrichment and consolidation of the collective expertise in this domain. Furthermore, the article highlighted both overarching and production-specific aspects that had remained largely unexplored. These aspects have been considered and subsequently partially addressed in the following use cases within the proposed framework.

As discussed in Section 3.3, there were only few data-driven articles in the literature that relied on DoE methods for data generation. These articles focused predominantly on individual process steps, as exemplified by the work presented by FARAJI NIRI, APACHITEI, et al. (2022). Building upon the findings of the first publication, Publication IV introduced a systematic approach for analyzing quality-relevant parameters. These parameters were then incorporated into a process-focused DoE method, enabling an efficient data generation and cross-process analysis. While a number of studies have demonstrated the application of DM methods based on historical data, such as the articles presented by SCHNELL et al. (2019) and THIEDE et al. (2019), they have often lacked transparency in their data handling practices, including the measures employed to assess data quality. In the use case presented in the dissertation of TURETSKYY (2022), a correlation analysis was carried out, using a threshold of 0.64 to identify highly correlated features, which were then removed prior to model development. This threshold can still be considered relatively high, indicating the presence of multicollinearity in the dataset. To the best of the author's knowledge, the existing literature on battery production has largely overlooked the topic of common methods for evaluating data quality and potential measures such as data enhancement through design augmentation. Publication V addressed this aspect by exploring three evaluation methods: Pearson correlation matrix, FDS, and VIF. The article demonstrated the benefits of design augmentation and presented a comprehensive analysis of quality-relevant parameters in electrode manufacturing. Leveraging an enriched dataset and employing XML methods, Publication V simultaneously addressed the research gap identified in the conducted mapping study from a production perspective. It revealed the impact of various parameters in electrode manufacturing, including the drying process, on less explored target variables such as adhesion strength as well as cell properties.

6.2 Transferability and Limitations

The evaluation of the proposed framework was primarily conducted based on an empirical-inductive approach. Therefore, it is important to discuss its transferability and limitations.

General approaches, such as CRISP-DM, offer a foundation for the development of data-driven models, focusing mainly on the basic philosophy and generic steps, regardless of the application field and domain-specific challenges. This dissertation aimed to introduce a framework that goes beyond generic principles by integrating domain-specific know-how, best practices, methods, and guidelines that can be used to facilitate the analysis of interdependencies in LIB electrode manufacturing. Consequently, the discussion on transferability will be limited to the context of battery cell production. While the applied methods, such as DSM to extract insights from the conducted mapping study or the MoSCoW analysis, can be employed when considering other novel technologies, such as extrusion or dry coating in battery production, it is important to acknowledge that the list of parameters and the evaluation may need to be adapted accordingly. In cases where limited knowledge is available in the literature, the expert-based evaluation can be extended by the Delphi method (LINSTONE, TUROFF, et al. 1975).

As part of Publication IV, a systematic approach was introduced, based on a predefined objective, to identify the quality-relevant parameters considered in the DoE for data generation. However, the definition of the levels of certain factors, particularly those related to the drying process, was mainly based on domain know-how. As a result, a certain degree of process knowledge is assumed for the empirical analysis. In cases where there is no knowledge of the respective ranges, an initial trial-and-error approach may be necessary. This need becomes even more pronounced when analyzing a large number of interrelated parameters, as was the case in the conducted design augmentation. The framework was based on the assumption that the process constraints are known. These are, for example, the combination of the largest mass loading with the lowest drying temperatures and the highest drying web speed possible, while still ensuring the successful production of a thoroughly dried electrode. In case of a lack of domain expertise, the efficiency of the proposed framework may be compromised, necessitating a more extensive trial-and-error approach. Furthermore, it is worth noting that the data generation step, guided by DoE and design augmentation, was founded on the assumption of a single slurry formulation, which can be considered a relevant assumption from an industrial perspective. Consequently, aspects such as systematic integration of a mixture design (cf. Section 2.3) with process-focused DOE to streamline data generation were not included in the framework.

The demonstrated use cases were based on selective data generated at the LIB pilot production line for analyzing interdependencies. The framework did not address aspects such as the computational resources required, best practices for managing larger datasets, and algorithm scalability in the context of ML for mass production applications. Furthermore, it is essential to acknowledge that, given the wide variety of algorithms available, the conducted mapping study on ML applications in battery cell production should be considered more as a starting reference point. It may not provide a comprehensive overview of the optimal choices for all types of analyses and scenarios, as the selection of algorithms may vary depending on the specific objectives and data characteristics.

It should also be noted that the insights derived from the production perspective through XML methods, particularly the global parameter rankings in combination with different sets of target variables, should be contextualized within the boundaries of the examined ranges. For example, a broader range of drying web speed or mass loading may result in different parameter rankings concerning adhesion strength. However, in order to facilitate the development of data-driven models and ensure the transferability of the generated results to other pilot production lines using methods such as transfer learning, the collected data, combined with the selected hyperparameters, were made publicly available (Haghi, Keilhofer, et al. 2024; Haghi, Chen, et al. 2024). This was intended to support and encourage the battery production community in their efforts toward advancing the application of data-driven models, aiming for sustainable, quality-oriented battery cell production.

Chapter 7

Summary and Outlook

In this chapter, a summary of the work presented, including the key findings, is provided in Section 7.1. Additionally, an outlook on potential further research is offered in Section 7.2.

7.1 Summary

This dissertation aimed to contribute to the state of the art in battery cell production, with a particular focus on the development of data-driven models for analyzing interdependencies in electrode manufacturing. Chapter 1 provided a brief motivation of the topic, including the overarching objective. Chapter 2 covered the fundamental concepts and background information necessary to understand battery cell production. Additionally, the relevant methods in the domains of DoE and ML were briefly introduced.

Following a systematic approach, Chapter 3 discussed the existing contributions to the relevant research areas for the development of data-driven models in battery cell production. These research areas included topics such as parameters and measurement solutions, data generation and evaluation, and data-driven models, along with the derivation of insights. In this chapter, each research area was briefly summarized, and the research opportunities were outlined.

Building on the identified research opportunities, Chapter 4 provided an overview of SOs and the requirements necessary to address the central objective of this dissertation. Consequently, an outline of the proposed solution was provided, structured according to the defined SOs and the corresponding publications. In total, five publications formed the foundation of this publication-based dissertation, which were listed in Section 4.2 and summarized in Chapter 5. The summary included the primary methods applied, the targeted SO, and the key findings. Publication I addressed SO1 by introducing a systematic two-step approach based on a structured mapping study and expert interviews. The results included the evaluation of parameters based on two dimensions: their relevance from the quality management perspective and the level of complexity associated with their digitalization. Publication II further explored the topic of digitalization, focusing specifically on measuring instruments and the possibility of inline collection of intermediate product parameters. Through market analysis, a desk-based evaluation of possible measuring instruments was provided, considering factors such as measurement strategy, accuracy, and capital cost. SO2 was addressed by conducting a mapping study that explored the adopted ML methods in battery cell production, as detailed in Publication III. By extracting and synthesizing the relevant information, from both data analytics and production perspectives, the mapping study provided a starting reference point for the selection of ML algorithms and evaluation metrics. To address the research areas of data generation, evaluation, development of data-driven models, and derivation of insights, two distinct use cases were defined, each approaching SO3 and SO4 from different initial conditions and perspectives. In Publication IV, a use case was presented for the efficient generation of a new dataset to facilitate cross-process analysis. Publication V was based on the assumption of the availability of historical data and explored the possibility of consolidating the data, combined with relevant methods for evaluation of data quality, as well as enhancing it through design augmentation. The two demonstrated use cases were based on the data generated at the pilot production line and included variations in the coating, drying, and calendering processes. By employing XML methods, the studies investigated the significance of the analyzed parameters and their influence on the target physical and mechanical properties of the electrode, as well as the cell properties.

Chapter 6 was dedicated to the discussion of the proposed framework. For this purpose, the framework was evaluated based on its exemplary applications. Furthermore, the extent to which the defined requirements were met and the contribution of the framework to the current state of the art were elaborated. The chapter concluded by discussing the transferability of the framework and its limitations.

7.2 Outlook on Further Research

The proposed framework effectively addressed a number of research perspectives outlined in HAGHI, HIDALGO, et al. (2023). Nonetheless, there are additional research opportunities that can leverage the presented framework to further enhance the integration of ML technology into LIB production.

One potential research avenue involves delving into the adaptation of ML models that were initially developed using data from a particular production line. This research area could explore the essential steps required to implement transfer learning and use these initial models as a foundational starting point to guide the development of ML models intended for deployment in a different production line.

Considering the multi-scale nature of LIB, spanning from micro to meso and macro levels, an additional promising future research field lies in the integration of ML models with in silico approaches. This integration can bridge the different scales within LIB systems and facilitate comprehensive optimizations across these scales.

From the data analytics perspective, the presented use cases fall primarily within the domain of predictive analytics. They involved modeling the essential cause-and-effect relationships throughout the process chain and contributed to a deeper understanding of the electrode manufacturing process chain and the collective impact of various parameters on the intermediate and final product properties. In this context, the logical next step is to progress toward prescriptive analytics. This involves the development of data-driven inline control and multivariable optimization systems. These systems cannot only predict potential issues, but also suggest precise interventions to achieve desired outcomes, ultimately enhancing process efficiency. By transitioning from predictive to prescriptive analytics, it becomes possible to implement more proactive and dynamic approaches to manage and improve the electrode manufacturing process chain.

Bibliography

ALPAYDIN, E. (2014). *Introduction to machine learning*. MIT press. ISBN: 9780262043793.

ANDERSON-COOK, C. M., BORROR, C. M., and MONTGOMERY, D. C. (2009). "Response surface design evaluation and comparison". In: *Journal of Statistical Planning and Inference* 139.2, pp. 629–641. DOI: 10.1016/j.jspi.2008.04.004.

ANDIAPPAN, V. and WAN, Y. K. (2020). "Distinguishing approach, methodology, method, procedure and technique in process systems engineering". In: *Clean technologies and environmental policy* 22, pp. 547–555. DOI: 10.1007/s10098-020-01819-w.

APLEY, D. W. and ZHU, J. (2020). "Visualizing the effects of predictor variables in black box supervised learning models". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82.4, pp. 1059–1086. DOI: 10.1111/rssb.12377.

AYERBE, E., BERECIBAR, M., CLARK, S., FRANCO, A. A., and RUHLAND, J. (2022). "Digitalization of battery manufacturing: current status, challenges, and opportunities". In: *Advanced Energy Materials* 12.17, p. 2102696. DOI: 10.1002/aenm.202102696.

BADILLO, S., BANFAI, B., BIRZELE, F., DAVYDOV, I. I., HUTCHINSON, L., KAM-THONG, T., SIEBOURG-POLSTER, J., STEIERT, B., and ZHANG, J. D. (2020). "An introduction to machine learning". In: *Clinical pharmacology & therapeutics* 107.4, pp. 871–885. DOI: 10.1002/cpt.1796.

BALDWIN, C. Y. and CLARK, K. B. (2003). "Managing in an age of modularity". In: *Managing in the modular age: Architectures, networks, and organizations* 149, pp. 84–93.

BARET DE LIMÉ, A. du, LEIN, T., MALETTI, S., SCHMAL, K., REUBER, S., HEUBNER, C., and MICHAELIS, A. (2022). "Impact of Electrode Defects on Battery Cell Performance: A Review". In: *Batteries & Supercaps* 5.10, e202200239. DOI: 10.1002/batt.202200239.

BELHADI, A., ZKIK, K., CHERRAFI, A., SHA'RI, M. Y., et al. (2019). "Understanding big data analytics for manufacturing processes: insights from literature review and multiple case studies". In: *Computers & Industrial Engineering* 137, p. 106099. DOI: 10.1016/j.cie.2019.106099.

Billot, N. (2021). “Modeling of Adhesion Mechanisms of Graphite-based Anodes for Lithium-ion Batteries”. PhD thesis. Technical University of Munich. isbn: 978-3-8316-4976-1.

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A.-L., et al. (2023). “Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges”. In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 13.2, e1484. doi: 10.1002/widm.1484.

Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*. Vol. 4. 4. Springer. isbn: 978-0-387-31073-2.

Blessing, L. T. M. and Chakrabarti, A. (2009). *DRM, a Design Research Methodology*. Springer London Limited. isbn: 1848825870.

Bloomberg (2022). *Lithium-ion Battery Pack Prices Rise for First Time to an Average of* $151/kWh$. url: https://about.bnef.com/blog/lithium-ion-battery-pack-prices-rise-for-first-time-to-an-average-of-151-kwh/ (visited on 09/01/2023).

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). “A training algorithm for optimal margin classifiers”. In: *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144–152. doi: 10.1145/130385.130401.

Breiman, L. (2001). “Random forests”. In: *Machine learning* 45, pp. 5–32. doi: 10.1023/A:1010933404324.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. CRC press. isbn: 9781315139470.

Browning, T. R. (2001). “Applying the design structure matrix to system decomposition and integration problems: a review and new directions”. In: *IEEE Transactions on Engineering management* 48.3, pp. 292–306. doi: 10.1109/17.946528.

Bruggeman, V. D. (1935). “Berechnung verschiedener physikalischer Konstanten von heterogenen Substanzen. I. Dielektrizitätskonstanten und Leitfähigkeiten der Mischkörper aus isotropen Substanzen”. In: *Annalen der physik* 416.7, pp. 636–664. doi: 10.1002/andp.19354160705.

Buchanan, B. G. (2005). “A (very) brief history of artificial intelligence”. In: *Ai Magazine* 26.4, pp. 53–60. doi: 10.1609/aimag.v26i4.1848.

Caulfield, B., Reginatto, B., and Slevin, P. (2019). “Not all sensors are created equal: a framework for evaluating human performance measurement technologies”. In: *NPJ digital medicine* 2.1, p. 7. doi: 10.1038/s41746-019-0082-4.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R., et al. (2000). “CRISP-DM 1.0: Step-by-step data mining guide”. In: *SPSS inc* 9.13, pp. 1–73.

CHEN, H., CHIANG, R. H., and STOREY, V. C. (2012). "Business intelligence and analytics: From big data to big impact". In: *MIS quarterly*, pp. 1165–1188. DOI: 10.2307/41703503.

CHEN, T., SONG, M., HUI, H., and LONG, H. (2021). "Battery electrode mass loading prognostics and analysis for lithium-ion battery–based energy storage systems". In: *Frontiers in Energy Research* 9, p. 754317. DOI: 10.3389/fenrg.2021.754317.

CHEN, T. and GUESTRIN, C. (2016). "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794. DOI: 10.1145/2939672.2939785.

CHENG, C.-S. (1996). "Optimal design: Exact theory". In: *Handbook of statistics* 13, pp. 977–1006. DOI: 10.1016/S0169-7161(96)13028-5.

CLEGG, D. and BARKER, R. (1994). *Case method fast-track: a RAD approach*. Addison-Wesley Longman Publishing Co., Inc. ISBN: 9780201624328.

CUNHA, R. P., LOMBARDO, T., PRIMO, E. N., and FRANCO, A. A. (2020). "Artificial intelligence investigation of NMC cathode manufacturing parameters interdependencies". In: *Batteries & Supercaps* 3.1, pp. 60–67. DOI: 10.1002/batt.201900135.

DAVENPORT, T. H. et al. (2006). "Competing on analytics". In: *Harvard business review* 84.1, pp. 1–11.

DAVENPORT, T. H. (2013). *Enterprise analytics: Optimize performance, process, and decisions through big data*. Pearson Education. ISBN: 9780133039498.

DE'ATH, G. and FABRICIUS, K. E. (2000). "Classification and regression trees: a powerful yet simple technique for ecological data analysis". In: *Ecology* 81.11, pp. 3178–3192. DOI: http://dx.doi.org/10.2307/177409.

DIN40729 (1985). *Accumulators; galvanic secondary cells; general terms and definitions*. Beuth Verlag GmbH.

DRAKOPOULOS, S. X., GHOLAMIPOUR-SHIRAZI, A., MACDONALD, P., PARINI, R. C., REYNOLDS, C. D., BURNETT, D. L., PYE, B., O'REGAN, K. B., WANG, G., WHITEHEAD, T. M., et al. (2021). "Formulation and manufacturing optimization of lithium-ion graphite-based electrodes via machine learning". In: *Cell Reports Physical Science* 2.12. DOI: 10.1016/j.xcrp.2021.100683.

DUQUESNOY, M., BOYANO, I., GANBORENA, L., CEREIJO, P., AYERBE, E., and FRANCO, A. A. (2021). "Machine learning-based assessment of the impact of the manufacturing process on battery electrode heterogeneity". In: *Energy and AI* 5, p. 100090. DOI: 10.1016/j.egyai.2021.100090.

DUQUESNOY, M., LIU, C., DOMINGUEZ, D. Z., KUMAR, V., AYERBE, E., and FRANCO, A. A. (2023). "Machine learning-assisted multi-objective optimization of battery manufacturing from synthetic data generated by physics-based simulations". In: *Energy Storage Materials* 56, pp. 50–61. DOI: 10.1016/j.ensm.2022.12.040.

DUQUESNOY, M., LOMBARDO, T., CHOUCHANE, M., PRIMO, E. N., and FRANCO, A. A. (2020). "Data-driven assessment of electrode calendering process by combining experimental results, in silico mesostructures generation and machine learning". In: *Journal of Power Sources* 480, p. 229103. DOI: 10.1016/j.jpowsour.2020.229103.

EPPINGER, S. D. and BROWNING, T. R. (2012). *Design structure matrix methods and applications*. MIT press. ISBN: 9780262528887.

FARAJI NIRI, M., APACHITEI, G., LAIN, M., COPLEY, M., and MARCO, J. (2022). "The Impact of Calendering Process Variables on the Impedance and Capacity Fade of Lithium-Ion Cells: An Explainable Machine Learning Approach". In: *Energy Technology* 10.12, p. 2200893. DOI: 10.1002/ente.202200893.

FARAJI NIRI, M., ASLANSEFAT, K., HAGHI, S., HASHEMIAN, M., DAUB, R., and MARCO, J. (2023). "A Review of the Applications of Explainable Machine Learning for Lithium–Ion Batteries: From Production to State and Performance Estimation". In: *Energies* 16.17, p. 6360. DOI: 10.3390/en16176360.

FARAJI NIRI, M., LIU, K., APACHITEI, G., RAMIREZ, L. R., LAIN, M., WIDANAGE, D., and MARCO, J. (2021). "Machine learning for optimised and clean Li-ion battery manufacturing: Revealing the dependency between electrode and cell characteristics". In: *Journal of Cleaner Production* 324, p. 129272. DOI: 10.1016/j.jclepro.2021.129272.

FARAJI NIRI, M., LIU, K., APACHITEI, G., ROMÁN-RAMÍREZ, L. A., LAIN, M., WIDANAGE, D., and MARCO, J. (2022). "Quantifying key factors for optimised manufacturing of Li-ion battery anode and cathode via artificial intelligence". In: *Energy and AI* 7, p. 100129. DOI: 10.1016/j.egyai.2021.100129.

FARAJI NIRI, M., REYNOLDS, C., RAMÍREZ, L. A. R., KENDRICK, E., and MARCO, J. (2022). "Systematic analysis of the impact of slurry coating on manufacture of Li-ion battery electrodes via explainable machine learning". In: *Energy Storage Materials* 51, pp. 223–238. DOI: 10.1016/j.ensm.2022.06.036.

FAYYAD, U., PIATETSKY-SHAPIRO, G., and SMYTH, P. (1996). "From data mining to knowledge discovery in databases". In: *AI magazine* 17.3, pp. 37–54. DOI: 10.1609/aimag.v17i3.1230.

FRIEDMAN, J. H. (2001). "Greedy function approximation: a gradient boosting machine". In: *Annals of statistics*, pp. 1189–1232. DOI: 10.1214/aos/1013203451.

FROBOESE, L., TITSCHER, P., WESTPHAL, B., HASELRIEDER, W., and KWADE, A. (2017). "Mercury intrusion for ion-and conversion-based battery electrodes–structure and diffusion coefficient determination". In: *Materials Characterization* 133, pp. 102–111. DOI: 10.1016/j.matchar.2017.09.002.

GOODFELLOW, I., BENGIO, Y., and COURVILLE, A. (2016). *Deep Learning*. MIT Press. ISBN: 0262035618.

GRAF, C. (2018). "Cathode materials for lithium-ion batteries". In: *Korthauer (Ed.) - Lithium-Ion Batteries: Basics and Applications*. Springer, pp. 29–41.

GROVES, D. and WACHTMAN, J. (1986). "Materials characterization–vital and often successful, yet still a critical problem". In: *Materials Characterization* *for Systems Performance and Reliability*. Springer, pp. 425–443. DOI: 10.1007/978-1-4613-2119-4_22.

GÜNTHER, T., BILLOT, N., SCHUSTER, J., SCHNELL, J., SPINGLER, F. B., and GASTEIGER, H. A. (2016). "The manufacturing of electrodes: key process for the future success of lithium-ion batteries". In: *Advanced Materials Research* 1140, pp. 304–311. DOI: 10.4028/www.scientific.net/AMR.1140.304.

GÜNTHER, T., SCHREINER, D., METKAR, A., MEYER, C., KWADE, A., and REINHART, G. (2020). "Classification of calendering-induced electrode defects and their influence on subsequent processes of lithium-ion battery production". In: *Energy Technology* 8.2, p. 1900026. DOI: 10.1002/ente.201900026.

HAGHI, S., CHEN, Y., MOLZBERGER, A., and DAUB, R. (2024). "Interdependencies in Electrode Manufacturing: A Comprehensive Study Based on Design Augmentation and Explainable Machine Learning". In: *Batteries & Supercaps* 7.5, e202300556. DOI: 10.1002/batt.202300556.

HAGHI, S., HIDALGO, M. F. V., NIRI, M. F., DAUB, R., and MARCO, J. (2023). "Machine Learning in Lithium-Ion Battery Cell Production: A Comprehensive Mapping Study". In: *Batteries & Supercaps* 6.7, e202300046. DOI: 10.1002/batt.202300046.

HAGHI, S., KEILHOFER, J., SCHWARZ, N., HE, P., and DAUB, R. (2024). "Efficient Analysis of Interdependencies in Electrode Manufacturing Through Joint Application of Design of Experiments and Explainable Machine Learning". In: *Batteries & Supercaps* 7.2, e202300457. DOI: 10.1002/batt.202300457.

HAGHI, S., LEEB, M., MOLZBERGER, A., and DAUB, R. (2023). "Measuring Instruments for Characterization of Intermediate Products in Electrode Manufacturing of Lithium-Ion Batteries". In: *Energy Technology* 11.9, p. 2300364. DOI: 10.1002/ente.202300364.

HAGHI, S., SUMMER, A., BAUERSCHMIDT, P., and DAUB, R. (2022). "Tailored Digitalization in Electrode Manufacturing: The Backbone of Smart Lithium-Ion Battery Cell Production". In: *Energy Technology* 10.10, p. 2200657. DOI: 10.1002/ente.202200657.

HAGIWARA, H., SUSZYNSKI, W. J., and FRANCIS, L. F. (2014). "A Raman spectroscopic method to find binder distribution in electrodes during drying". In: *Journal of Coatings Technology and Research* 11, pp. 11–17. DOI: 10.1007/s11998-013-9509-z.

HASELRIEDER, W., WESTPHAL, B., BOCKHOLT, H., DIENER, A., HÖFT, S., and KWADE, A. (2015). "Measuring the coating adhesion strength of electrodes for lithium-ion batteries". In: *International Journal of Adhesion and Adhesives* 60, pp. 1–8. DOI: 10.1016/j.ijadhadh.2015.03.002.

HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. H., and FRIEDMAN, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer. ISBN: 978-0-387-84858-7.

Hawley, W. B. and Li, J. (2019). “Electrode manufacturing for lithium-ion batteries—Analysis of current and next generation processing”. In: *Journal of Energy Storage* 25, p. 100862. doi: 10.1016/j.est.2019.100862.

Hofler, D. B. (1983). “Approach, method, technique a clarification”. In: *Reading World*, pp. 71–72. doi: 10.1080/19388078309557742.

IEEE/ISO/IEC 29148 (2018). *Systems and software engineering — Life cycle processes — Requirements engineering*. International Organization for Standardization.

Izenman, A. J. (2008). *Modern multivariate statistical techniques*. Vol. 1. Springer. isbn: 978-1-4939-3832-2.

James, G., Witten, D., Hastie, T., Tibshirani, R., et al. (2013). *An introduction to statistical learning*. Vol. 112. Springer. isbn: 978-1-4614-7138-7.

Joshi, A. V. (2020). *Machine learning and artificial intelligence*. Springer. isbn: 978-3-030-26624-0.

Kaiser, J., Wenzel, V., Nirschl, H., Bitsch, B., Willenbacher, N., Baunach, M., Schmitt, M., Jaiser, S., Scharfer, P., and Schabel, W. (2014). “Prozess-und produktentwicklung von elektroden für li-ionen-zellen”. In: *Chemie Ingenieur Technik* 86.5, pp. 695–706. doi: 10.1002/CITE.201300085.

Kampker, A., Heimes, H. H., Dorn, B., Clever, H., Ludwigs, R., Li, R., and Drescher, M. (2023). “Concept for Digital Product Twins in Battery Cell Production”. In: *World Electric Vehicle Journal* 14.4, p. 108. doi: 10.3390/wevj14040108.

Keppeler, M., Tran, H.-Y., and Braunwarth, W. (2021). “The Role of Pilot Lines in Bridging the Gap Between Fundamental Research and Industrial Production for Lithium-Ion Battery Cells Relevant to Sustainable Electromobility: A Review”. In: *Energy Technology* 9.8, p. 2100132. doi: 10.1002/ente.202100132.

Kessler, R. W. (2012). *Prozessanalytik: Strategien und Fallbeispiele aus der industriellen Praxis*. John Wiley & Sons. isbn: 9783527608997.

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., and Linkman, S. (2009). “Systematic literature reviews in software engineering–a systematic literature review”. In: *Information and software technology* 51.1, pp. 7–15. doi: 10.1016/j.infsof.2008.09.009.

Kitchenham, B. A., Budgen, D., and Brereton, O. P. (2011). “Using mapping studies as the basis for further research–a participant-observer case study”. In: *Information and Software Technology* 53.6, pp. 638–651. doi: 10.1016/j.infsof.2010.12.011.

Kleppmann, W. (2016). *Versuchsplanung: Produkte und Prozesse optimieren*. Carl Hanser Verlag GmbH Co KG. isbn: 978-3-446-44717-2.

Kornas, T., Daub, R., Karamat, M. Z., Thiede, S., and Herrmann, C. (2019). “Data-and expert-driven analysis of cause-effect relationships in the production of lithium-ion batteries”. In: *2019 IEEE 15th International Conference* *on Automation Science and Engineering (CASE)*. IEEE, pp. 380–385. DOI: 10.1109/COASE.2019.8843185.

KUMBERG, J., MÜLLER, M., DIEHM, R., SPIEGEL, S., WACHSMANN, C., BAUER, W., SCHARFER, P., and SCHABEL, W. (2019). "Drying of lithium-ion battery anodes for use in high-energy cells: influence of electrode thickness on drying time, adhesion, and crack formation". In: *Energy Technology* 7.11, p. 1900722. DOI: 10.1002/ente.201900722.

KURFER, J., WESTERMEIER, M., TAMMER, C., and REINHART, G. (2012). "Production of large-area lithium-ion cells–Preconditioning, cell stacking and quality assurance". In: *CIRP annals* 61.1, pp. 1–4. DOI: 10.1016/J.CIRP.2012.03.101.

KUTNER, M. H., NACHTSHEIM, C. J., and NETER, J. (1983). *Applied linear regression models*. McGraw-Hill Irwin. ISBN: 0-07-238688-6.

KWADE, A., HASELRIEDER, W., LEITHOFF, R., MODLINGER, A., DIETRICH, F., and DROEDER, K. (2018). "Current status and challenges for automotive battery production technologies". In: *Nature Energy* 3.4, pp. 290–300. DOI: 10.1038/s41560-018-0130-3.

LANDESFEIND, J., HATTENDORFF, J., EHRL, A., WALL, W. A., and GASTEIGER, H. A. (2016). "Tortuosity determination of battery electrodes and separators by impedance spectroscopy". In: *Journal of The Electrochemical Society* 163.7, A1373. DOI: 10.1149/2.1141607jes.

LANEY, D., BITTERER, A., SALLAM, R. L., and KART, A. (2012). *Predicts 2013: Information Innovation*. Tech. rep. Gartner.

LEUTHNER, S. (2018). "Lithium-ion battery overview". In: *Korthauer (Ed.) - Lithium-Ion Batteries: Basics and Applications*. Springer, pp. 13–19. DOI: 10.1007/978-3-662-53071-9_2.

LIM, S., AHN, K. H., and YAMAMURA, M. (2013). "Latex migration in battery slurries during drying". In: *Langmuir* 29.26, pp. 8233–8244. DOI: 10.1021/la4013685.

LIN, Z., LIU, T., AI, X., and LIANG, C. (2018). "Aligning academia and industry for unified battery performance metrics". In: *Nature communications* 9.1, p. 5262. DOI: 10.1038/s41467-018-07599-8.

LINARDATOS, P., PAPASTEFANOPOULOS, V., and KOTSIANTIS, S. (2020). "Explainable ai: A review of machine learning interpretability methods". In: *Entropy* 23.1, p. 18. DOI: 10.3390/e23010018.

LINSTONE, H. A., TUROFF, M., et al. (1975). *The delphi Method: Techniques and Applications*. Addison-Wesley Reading, MA. ISBN: 0-201-04294-0.

LIU, G., ZHENG, H., SONG, X., and BATTAGLIA, V. S. (2011). "Particles and polymer binder interaction: a controlling factor in lithium-ion electrode performance". In: *Journal of The Electrochemical Society* 159.3, A214. DOI: 10.1149/2.024203jes].

LIU, K., FARAJI NIRI, M., APACHITEI, G., LAIN, M., GREENWOOD, D., and MARCO, J. (2022). "Interpretable machine learning for battery capacities prediction and coating parameters analysis". In: *Control Engineering Practice* 124, p. 105202. DOI: 10.1016/j.conengprac.2022.105202.

LIU, K., HU, X., ZHOU, H., TONG, L., WIDANAGE, W. D., and MARCO, J. (2021). "Feature analyses and modeling of lithium-ion battery manufacturing based on random forest classification". In: *IEEE/ASME Transactions on Mechatronics* 26.6, pp. 2944–2955. DOI: 10.1109/TMECH.2020.3049046.

LIU, K., LI, K., and CHEN, T. (2021). "Interpretable Sensitivity Analysis and Electrode Porosity Classification for Li-ion Battery Smart Manufacturing". In: *2021 IEEE Sustainable Power and Energy Conference (iSPEC)*. IEEE, pp. 3653–3658. DOI: 10.1109/iSPEC53008.2021.9735647.

LIU, K., PENG, Q., LI, K., and CHEN, T. (2022). "Data-based interpretable modeling for property forecasting and sensitivity analysis of Li-ion battery electrode". In: *Automotive Innovation* 5.2, pp. 121–133. DOI: 10.1007/s42154-021-00169-7.

LIU, K., WEI, Z., YANG, Z., and LI, K. (2021). "Mass load prediction for lithium-ion battery electrode clean production: A machine learning approach". In: *Journal of Cleaner Production* 289, p. 125159. DOI: 10.1016/j.jclepro.2020.125159.

LIU, K., YANG, Z., WANG, H., and LI, K. (2021). "Classifications of lithium-ion battery electrode property based on support vector machine with various kernels". In: *Recent Advances in Sustainable Energy and Intelligent Systems: LSMS 2021/ICSEE 2021*. Springer, pp. 23–34. DOI: 10.1007/978-981-16-7210-1_3.

LOMBARDO, T., DUQUESNOY, M., EL-BOUYSIDY, H., ÅRÉN, F., GALLO-BUENO, A., JØRGENSEN, P. B., BHOWMIK, A., DEMORTIÈRE, A., AYERBE, E., ALCAIDE, F., et al. (2021). "Artificial intelligence applied to battery research: hype or reality?" In: *Chemical reviews* 122.12, pp. 10899–10969. DOI: 10.1021/acs.chemrev.1c00108.

LU, L. and ANDERSON-COOK, C. M. (2021). "Strategies for sequential design of experiments and augmentation". In: *Quality and Reliability Engineering International* 37.5, pp. 1740–1757. DOI: 10.1002/qre.2823.

LUHN, H. P. (1958). "A business intelligence system". In: *IBM Journal of research and development* 2.4, pp. 314–319. DOI: 10.1147/rd.24.0314.

LUNDBERG, S. M., ERION, G., CHEN, H., DEGRAVE, A., PRUTKIN, J. M., NAIR, B., KATZ, R., HIMMELFARB, J., BANSAL, N., and LEE, S.-I. (2020). "From local explanations to global understanding with explainable AI for trees". In: *Nature machine intelligence* 2.1, pp. 56–67. DOI: 10.1038/s42256-019-0138-9.

MAYILVAHANAN, K. S., HUI, Z., HU, K., KUANG, J., MCCARTHY, A. H., BERNARD, J., WANG, L., TAKEUCHI, K. J., MARSCHILOK, A. C., TAKEUCHI, E. S., et al. (2021). "Quantifying uncertainty in tortuosity estimates for porous electrodes". In: *Journal of The Electrochemical Society* 168.7, p. 070537. DOI: 10.1149/1945-7111/ac1316.

MITCHELL, T. M. (1997). *Machine learning*. McGraw-Hill Series in Computer Science. ISBN: 0070428077.

MONTGOMERY, D. (2012). *Design analysis of experiments, Wiley: New York, NY*. John Wiley Sons, Inc. ISBN: 978-1-118-14692-7.

MORRIS, A. S. and LANGARI, R. (2012). *Measurement and instrumentation: theory and application*. Academic Press. ISBN: 978-0-12-800884-3.

MUNIR, N., NUGENT, M., WHITAKER, D., and MCAFEE, M. (2021). "Machine learning for process monitoring and control of hot-melt extrusion: current state of the art and future directions". In: *Pharmaceutics* 13.9, p. 1432. DOI: 10.3390/pharmaceutics13091432.

MYERS, R. H. (1990). *Classical and modern regression with applications*. Vol. 2. Duxbury press Belmont, CA. ISBN: 978-0534380168.

MYERS, R. H., MONTGOMERY, D. C., and ANDERSON-COOK, C. M. (2016). *Response surface methodology: process and product optimization using designed experiments*. John Wiley & Sons. ISBN: 978-1-118-91601-8.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY - BIG DATA PUBLIC WORKING GROUP (2019). *Big Data Interoperability Framework: Volume 1, Definitions*. Tech. rep. National Institute of Standards and Technology U. S. Department of Commerce. DOI: 10.6028/NIST.SP.1500-1r2.

NATRAS, R., SOJA, B., and SCHMIDT, M. (2022). "Ensemble Machine Learning of Random Forest, AdaBoost and XGBoost for Vertical Total Electron Content Forecasting". In: *Remote Sensing* 14.15, p. 3547. DOI: 10.3390/rs14153547.

NEMBRINI, S., KÖNIG, I. R., and WRIGHT, M. N. (2018). "The revival of the Gini importance?" In: *Bioinformatics* 34.21, pp. 3711–3718. DOI: 10.1093/bioinformatics/bty373.

NUSEIBEH, B. and EASTERBROOK, S. (2000). "Requirements engineering: a roadmap". In: *Proceedings of the Conference on the Future of Software Engineering*, pp. 35–46. DOI: 10.1145/336512.336523.

OBERKAMPF, W. L. and ROY, C. J. (2010). *Verification and validation in scientific computing*. Cambridge university press. ISBN: 9780511760396.

OGIHARA, N., KAWAUCHI, S., OKUDA, C., ITOU, Y., TAKEUCHI, Y., and UKYO, Y. (2012). "Theoretical and experimental analysis of porous electrodes for lithium-ion batteries by electrochemical impedance spectroscopy using a symmetric cell". In: *Journal of The Electrochemical Society* 159.7, A1034. DOI: 10.1149/2.057207jes.

OZOL-GODFREY, A., ANDERSON-COOK, C., and ROBINSON, T. J. (2008). "Fraction of design space plots for generalized linear models". In: *Journal of Statistical Planning and Inference* 138.1, pp. 203–219. DOI: 10.1016/j.jspi.2007.05.011.

PEDERSEN, K., EMBLEMSVAG, J., BAILEY, R., ALLEN, J., and MISTREE, F. (2000). "Validating Design Methods and Research: The Validation Square". In: *ASME Design Theory and Methodology Conference*. DOI: 10.1115/DETC2000%2FDTM-14579.

PETERSEN, K., FELDT, R., MUJTABA, S., and MATTSSON, M. (2008). "Systematic mapping studies in software engineering". In: *12th International Conference on Evaluation and Assessment in Software Engineering (EASE) 12*, pp. 1–10. DOI: 10.14236/ewic/EASE2008.8.

PETTINGER, K.-H., KAMPKER, A., HOHENTHANNER, C.-R., DEUTSKENS, C., HEIMES, H., and HEMDT, A. vom (2018). "Lithium-ion cell and battery production processes". In: *Korthauer (Ed.) - Lithium-ion batteries: Basics and applications*. Springer, pp. 211–226. DOI: 10.1007/978-3-662-53071-9_17.

RAI, A. (2020). "Explainable AI: From black box to glass box". In: *Journal of the Academy of Marketing Science* 48, pp. 137–141. DOI: 10.1007/s11747-019-00710-5.

RATNER, B. (2009). "The correlation coefficient: Its values range between+ 1/- 1, or do they?" In: *Journal of targeting, measurement and analysis for marketing* 17.2, pp. 139–142. DOI: 10.1057/jt.2009.5.

REYNOLDS, C. D., SLATER, P. R., HARE, S. D., SIMMONS, M. J., and KENDRICK, E. (2021). "A review of metrology in lithium-ion electrode coating processes". In: *Materials & Design* 209, p. 109971. DOI: 10.1016/j.matdes.2021.109971.

ROHKOHL, E., SCHÖNEMANN, M., BODROV, Y., and HERRMANN, C. (2022). "A data mining approach for continuous battery cell manufacturing processes from development towards production". In: *Advances in Industrial and Manufacturing Engineering* 4, p. 100078. DOI: 10.1016/j.aime.2022.100078.

ROMÁN-RAMÍREZ, L. A. and MARCO, J. (2022). "Design of experiments applied to lithium-ion batteries: A literature review". In: *Applied Energy* 320, p. 119305. DOI: 10.1016/j.apenergy.2022.119305.

ROSS, D. T. and SCHOMAN, K. E. (1977). "Structured analysis for requirements definition". In: *IEEE transactions on Software Engineering* 1, pp. 6–15. DOI: 10.1109/TSE.1977.229899.

RYNNE, O., DUBARRY, M., MOLSON, C., LEPAGE, D., PRÉBÉ, A., AYMÉ-PERROT, D., ROCHEFORT, D., and DOLLÉ, M. (2019). "Designs of experiments for beginners—A quick start guide for application to electrode formulation". In: *Batteries* 5.4, p. 72. DOI: 10.3390/batteries5040072.

RYNNE, O., DUBARRY, M., MOLSON, C., NICOLAS, E., LEPAGE, D., PRÉBÉ, A., AYMÉ-PERROT, D., ROCHEFORT, D., and DOLLÉ, M. (2020). "Exploiting materials to their full potential, a Li-ion battery electrode formulation optimization study". In: *ACS Applied Energy Materials* 3.3, pp. 2935–2948. DOI: 10.1021/acsaem.0c00015.

SAMUEL, A. L. (1959). "Some studies in machine learning using the game of checkers". In: *IBM Journal of research and development* 3.3, pp. 210–229. DOI: 10.1147/rd.33.0210.

SANTEE, S. G., RAVDEL, B., GULBINSKA, M. K., GNANARAJ, J. S., and DICARLO, J. F. (2014). "Optimizing Electrodes for Lithium-ion Cells". In: *Lithium-*

ion Battery Materials and Engineering: Current Topics and Problems from the Manufacturing Perspective. Springer, pp. 63–88. DOI: 10.1007/978-1-4471-6548-4_3.

SCHIEFER, H. and SCHIEFER, F. (2018). *Statistik für Ingenieure*. Springer. ISBN: 978-3-658-20640-6.

SCHMITT, M., BAUNACH, M., WENGELER, L., PETERS, K., JUNGES, P., SCHARFER, P., and SCHABEL, W. (2013). "Slot-die processing of lithium-ion battery electrodes—Coating window characterization". In: *Chemical Engineering and Processing: Process Intensification* 68, pp. 32–37. DOI: 10.1016/j.cep.2012.10.011.

SCHNELL, J., NENTWICH, C., ENDRES, F., KOLLENDA, A., DISTEL, F., KNOCHE, T., and REINHART, G. (2019). "Data mining in lithium-ion battery cell production". In: *Journal of Power Sources* 413, pp. 360–366. DOI: 10.1016/j.jpowsour.2018.12.062.

SCHUH, G., ANDERL, R., GAUSEMEIER, J., TEN HOMPEL, M., and WAHLSTER, W. (2017). *Industrie 4.0 maturity index: die digitale transformation von unternehmen gestalten*. Tech. rep. acatech.

SCHULTZ, J. and NARDIN, M. (2003). "Theories and mechanisms of adhesion". In: *Handbook of Adhesive Technology, eds. A. Pizzi, and KL Mittal*, pp. 53–67.

SEEPERSAD, C. C., PEDERSEN, K., EMBLEMSVÅG, J., BAILEY, R., ALLEN, J. K., and MISTREE, F. (2006). "The validation square: how does one verify and validate a design method". In: *Decision making in engineering design*, pp. 303–314. DOI: 10.1115/1.802469.ch25.

SHEN, C., WANG, L., and LI, Q. (2007). "Optimization of injection molding process parameters using combination of artificial neural network and genetic algorithm method". In: *Journal of materials processing technology* 183.2-3, pp. 412–418. DOI: 10.1016/j.jmatprotec.2006.10.036.

SIEBERTZ, K., HOCHKIRCHEN, T., and BEBBER, D. van (2017). *Statistische versuchsplanung*. Springer. ISBN: 978-3-662-55742-6.

SIVARAJAH, U., KAMAL, M. M., IRANI, Z., and WEERAKKODY, V. (2017). "Critical analysis of Big Data challenges and analytical methods". In: *Journal of business research* 70, pp. 263–286. DOI: 10.1016/j.jbusres.2016.08.001.

SMEKENS, J., GOPALAKRISHNAN, R., VAN DEN STEEN, N., OMAR, N., HEGAZY, O., HUBIN, A., and VAN MIERLO, J. (2016). "Influence of electrode density on the performance of Li-ion batteries: Experimental and simulation results". In: *Energies* 9.2, p. 104. DOI: 10.3390/en9020104.

STAPLETON, J. (1997). *DSDM, dynamic systems development method: the method in practice*. Cambridge University Press. ISBN: 978-0201178890.

THIEDE, S. (2018). "Environmental sustainability of cyber physical production systems". In: *Procedia CIRP* 69, pp. 644–649. DOI: 10.1016/j.procir.2017.11.124.

THIEDE, S., TURETSKYY, A., KWADE, A., KARA, S., and HERRMANN, C. (2019). "Data mining in battery production chains towards multi-criterial quality prediction". In: *CIRP Annals* 68.1, pp. 463–466. DOI: 10.1016/j.cirp.2019.04.066.

THOMITZEK, M. (2022). "Modeling Process-Product Interdependencies in Battery Cell Production". PhD thesis. Technische Universität Braunschweig. ISBN: 9783802783739.

TJADEN, B., COOPER, S. J., BRETT, D. J., KRAMER, D., and SHEARING, P. R. (2016). "On the origin and application of the Bruggeman correlation for analysing transport phenomena in electrochemical systems". In: *Current opinion in chemical engineering* 12, pp. 44–51. DOI: 10.1016/j.coche.2016.02.006.

TURETSKYY, A. (2022). "Data Analytics in Battery Production Systems". PhD thesis. Technische Universität Braunschweig. ISBN: 9783802783685.

TURETSKYY, A., THIEDE, S., THOMITZEK, M., DRACHENFELS, N. von, PAPE, T., and HERRMANN, C. (2020). "Toward data-driven applications in lithium-ion battery cell manufacturing". In: *Energy Technology* 8.2, p. 1900136. DOI: 10.1002/ente.201900136.

TURETSKYY, A., WESSEL, J., HERRMANN, C., and THIEDE, S. (2020). "Data-driven cyber-physical system for quality gates in lithium-ion battery cell manufacturing". In: *Procedia CIRP* 93, pp. 168–173. DOI: 10.1016/j.procir.2020.03.077.

— (2021). "Battery production design using multi-output machine learning models". In: *Energy Storage Materials* 38, pp. 93–112. DOI: 10.1016/j.ensm.2021.03.002.

TURING, A. (1950). "Computing Machinery and Intelligence". In: *Mind* LIX.236, pp. 433–460.

UNITED NATIONS (2022). *Emissions Gap Report 2022: The Closing Window - Climate Crisis Calls for Rapid Transformation of Societies*. Tech. rep. United Nations Environment Programme.

VAN ENGELEN, J. E. and HOOS, H. H. (2020). "A survey on semi-supervised learning". In: *Machine learning* 109.2, pp. 373–440. DOI: 10.1007/s10994-019-05855-6.

VAPNIK, V. (1999). *The nature of statistical learning theory*. Springer science & business media. ISBN: 978-0-387-98780-4.

VUORILEHTO, K. (2018). "Materials and function". In: *Korthauer (Ed.) - Lithium-ion batteries: Basics and applications*. Springer, pp. 21–28. DOI: 10.1007/978-3-662-53071-9_3.

WALL, M. M. and AMEMIYA, Y. (2007). "Nonlinear structural equation modeling as a statistical method". In: *Handbook of latent variable and related models*. Elsevier, pp. 321–343. DOI: 10.1016/B978-044452044-9/50018-5.

WANG, Y., LI, W., FANG, R., ZHU, H., and PENG, Q. (2022). "Capacities prediction and correlation analysis for lithium-ion battery-based energy storage system". In: *Control Engineering Practice* 125, p. 105224. DOI: 10.1016/j.conengprac.2022.105224.

WEBER, M., SCHOO, A., SANDER, M., MAYER, J. K., and KWADE, A. (2023). "Introducing Spectrophotometry for Quality Control in Lithium-Ion-Battery Electrode Manufacturing". In: *Energy Technology* 11.5, p. 2201083. DOI: 10.1002/ente.202201083.

WEBSTER, J. and WATSON, R. T. (2002). "Analyzing the past to prepare for the future: Writing a literature review". In: *MIS quarterly*, pp. xiii–xxiii.

WESTERMEIER, M., REINHART, G., and ZEILINGER, T. (2013). "Method for quality parameter identification and classification in battery cell production quality planning of complex production chains for battery cells". In: *3rd international electric drives production conference (EDPC)*. IEEE, pp. 1–10. DOI: 10.1109/EDPC.2013.6689742.

WESTERMEIER, M. (2016). "Qualitätsorientierte Analyse komplexer Prozessketten am Beispiel der Herstellung von Batteriezellen". PhD thesis. Technical University of Munich. ISBN: 978-3-8316-7243-1.

WESTERMEIER, M., REINHART, G., and STEBER, M. (2014). "Complexity management for the start-up in lithium-ion cell production". In: *Procedia CIRP* 20, pp. 13–19. DOI: 10.1016/j.procir.2014.05.026.

WESTKÄMPER, E. (2006). *Einführung in die Organisation der Produktion*. Springer-Verlag. ISBN: 978-3540260394.

WESTKÄMPER, E. and ZAHN, E. (2008). *Wandlungsfähige Produktionsunternehmen: Das Stuttgarter Unternehmensmodell*. Springer-Verlag. ISBN: 978-3540218890.

WINTER, M. and BRODD, R. J. (2004). "What are batteries, fuel cells, and supercapacitors?" In: *Chemical reviews* 104.10, pp. 4245–4270. DOI: 10.1021/cr020730k.

WU, Y. (2015). *Lithium-ion batteries: Fundamentals and Applications*. Vol. 4. CRC press. ISBN: 978-1466557338.

WURM, C., OETTINGER, O., WITTKAEMPER, S., ZAUTER, R., and VUORILEHTO, K. (2018). "Anode materials for lithium-ion batteries". In: *Korthauer (Ed.) - Lithium-Ion Batteries: Basics and Applications*. Springer, pp. 43–58. DOI: 10.1007/978-3-662-53071-9_5.

ZAHRAN, A., ANDERSON-COOK, C. M., and MYERS, R. H. (2003). "Fraction of design space to assess prediction capability of response surface designs". In: *Journal of Quality Technology* 35.4, pp. 377–386. DOI: 10.1080/00224065.2003.11980235.

ZANOTTO, F. M., DOMINGUEZ, D. Z., AYERBE, E., BOYANO, I., BURMEISTER, C., DUQUESNOY, M., EISENTRAEGER, M., MONTAÑO, J. F., GALLO-BUENO, A., GOLD, L., et al. (2022). "Data specifications for battery manufacturing digitalization: Current status, challenges, and opportunities". In: *Batteries & Supercaps* 5.9, e202200224. DOI: 10.1002/batt.202200224.

ZHANG, Y. S., COURTIER, N. E., ZHANG, Z., LIU, K., BAILEY, J. J., BOYCE, A. M., RICHARDSON, G., SHEARING, P. R., KENDRICK, E., and BRETT, D. J. (2022). "A review of lithium-ion battery electrode drying: mechanisms and metrology". In: *Advanced Energy Materials* 12.2, p. 2102233. DOI: 10.1002/aenm.202102233.

ZHOU, Z.-H. (2021). *Machine learning*. Springer Nature. ISBN: 978-981-15-1966-6.

Appendix A

List of Supervised Student Assistants and Student Theses

In the course of the author's research activities at the *iwb* at the Technical University of Munich, several student assistants and student theses were intensively supervised. The author would like to thank the student assistants who have been part of the research journey over the years, specifically, Armin Summer, Philipp Bauerschmidt, Nico Schwarz, and Roman Mazur. Their assistance, particularly in the electrode manufacturing and characterization, deserves special recognition and appreciation. Table A.1 presents a chronological list of the student theses supervised by the author, focusing on those relevant to this dissertation. These theses were carried out under the author's substantial scientific, technical, and content-related supervision. The author would like to thank the students for their dedication and the valuable discussions during the collaboration.

Table A.1 List of supervised student theses relevant to this dissertation

Time frame	Name	Thesis type	Title of thesis (original)
11.2020–05.2021	Felix Jochimsthal	Semester thesis	Entwicklung eines Digitalisierungskonzeptes für die Batteriezellproduktion
02.2021–08.2021	Simon Otte	Master thesis	Konzept zur Digitalisierung der Batteriezellproduktion
05.2021–11.2021	Sandra Bernhardt	Bachelor thesis	Maschinelles Lernen für die Batteriezellproduktion
09.2021–04.2022	Armin Summer	Master thesis	Entwicklung eines ganzheitlichen Anwendungskonzepts für datengetriebene Ansätze in der Elektrodenherstellung

01.2022–07.2022	Markus Ritzer	Semester thesis	Befähigung einer Beschichtungsanlage zur Erfassung von Inline-Messgrößen bei der Herstellung von Lithium-Ionen-Batterien
04.2022–11.2022	Korbinian Heiss	Master thesis	Untersuchung der qualitätsrelevanten Parameter bei der Herstellung von Lithium-Ionen-Batteriezellen
06.2022–09.2022	Karl Schmidt	Bachelor thesis	Analyse des Trocknungsprozesses und seine Auswirkungen in der Elektrodenherstellung von Lithium-Ionen-Batterien
06.2022–12.2022	Liangtao Jin	Semester thesis	Analysis of Quality-Relevant Parameters in the Drying and Calendering Process of Lithium-Ion Battery Cell Production
07.2022–01.2023	Maximilian Fuchs	Semester thesis	Systematische Untersuchung der qualitätsrelevanten Parameter bei der Herstellung von Lithium-Ionen-Batteriezellen
07.2022–01.2023	Annika Molzberger	Semester thesis	Systematische Bewertung der Messlösungen in der Elektrodenherstellung der Lithium-Ionen-Batterieproduktion
11.2022–04.2023	Nico Schwarz	Semester thesis	Data-Driven Methods for the Analysis of Quality-Relevant Parameters in the Electrode Production and Their Impact on Cell Characteristics
02.2023–05.2023	Pengdan He	Bachelor thesis	Analysis of Quality-Relevant Parameters in the Lithium-Ion Battery Cell Production
03.2023–09.2023	Wenhuang Yao	Semester thesis	Analysis of Quality-Relevant Parameters in the Lithium-Ion Battery Cell Production
03.2023–10.2023	Annika Molzberger	Master thesis	Entwicklung eines Konzepts basierend auf einer Versuchsplanerweiterung zur Analyse der qualitätsrelevanten Parameter in der Batteriezellproduktion

03.2023–10.2023	Yao Chen	Master thesis	Machine Learning Models to Analyze the Quality-Relevant Parameters and Their Influence on Cell Characteristics in Battery Cell Production
05.2023–09.2023	Hye Min Hong	Bachelor thesis	Analyse der Zusammenhänge in der Elektrodenherstellung mit dem Schwerpunkt auf den mechanischen Eigenschaften der Elektrode

Appendix B

Overview of the Publications and Contributions of the Author

In this section, an overview of the contributions made by the author to the scholarly discourse, upon which the foundation of this publication-based dissertation was built, is provided. Below is the list of the five articles, as presented in Chapter 4.

I. Haghi, S., Summer, A., Bauerschmidt, P., Daub, R. "Tailored Digitalization in Electrode Manufacturing: The Backbone of Smart Lithium-Ion Battery Cell Production". In: Energy Technology 10 (2022) 10, 2200657 pp. 1-19. DOI: 10.1002/ente.20220065

II. Haghi, S., Leeb, M., Molzberger, A., Daub, R. "Measuring Instruments for Characterization of Intermediate Products in Electrode Manufacturing of Lithium-Ion Batteries". In: Energy Technology 11 (2023) 9, 2300364 pp. 1-13. DOI: 10.1002/ente.202300364

III. Haghi, S., Hidalgo, M. F., Faraji Niri, M., Daub, R., Marco, J. "Machine Learning in Lithium-Ion Battery Cell Production: A Comprehensive Mapping Study". In: Batteries & Supercaps 6 (2023) 7, e202300046 pp. 1-14. DOI: 10.1002/batt.202300046

IV. Haghi, S., Keilhofer, J., Schwarz, N., He, P., Daub, R. "Efficient Analysis of Interdependencies in Electrode Manufacturing Through Joint Application of Design of Experiments and Explainable Machine Learning". In: Batteries & Supercaps 7 (2024) 2, e202300457 pp. 1-18 DOI: 10.1002/batt.202300457

V. Haghi, S., Chen, Y., Molzberger, A., Daub, R. "Interdependencies in Electrode Manufacturing: A Comprehensive Study Based on Design Augmentation and Explainable Machine Learning". In: Batteries & Supercaps 7 (2024) 5, e202300556 pp. 1-11 DOI: 10.1002/batt.202300556

The author of this dissertation was the main contributor to all of the articles listed above, taking a leading role in developing the concepts, selecting the methods, visualizing the results, and drafting the manuscripts. Co-authors Matthias Leeb and Josef Keilhofer, as colleagues and research associates at the Department of Battery Production at the *iwb*, contributed respectively to Publications II and IV, primarily by discussing and reviewing the results. Publication III was the result of a collaborative effort during the author's research stay at the University of Warwick. Dr. Marc Francis

V. Hidalgo, research fellow at the University of Warwick, contributed to the conceptualization, formal analysis, and the initial drafting of the manuscript. Prof. James Marco, in his role as the Energy Directorate of the Warwick Manufacturing Group, supported the development of the concept and provided valuable guidance through a detailed review of the manuscript. Prof. Mona Faraji Niri, assistant professor at the University of Warwick, contributed to the manuscript review process. Armin Summer, Philipp Bauerschmidt, Nico Schwarz, Annika Molzberger, Pengdan He, and Yao Chen contributed in their roles as supervised students, when named as co-authors in the respective publications. Their contribution mainly involved investigation and data curation. Prof. Dr.-Ing. Rüdiger Daub made significant contributions to all of the publications as the head of the *iwb* and the supervisor of this dissertation.

Table B.1 provides a summary of the author's contributions to the publications, based on the Contributor Roles Taxonomy (CRediT). *Concept* summarizes the conceptualization and methodology aspects, according to CRediT. *Realization* includes investigation, formal analysis, data curation, software, and validation, where applicable. *Documentation* covers aspects such as visualization, and writing, including both the original draft as well as review and editing.

Table B.1 Summary of the contributions of the author of this dissertation to the five publications

Publication	Concept	Realization	Documentation	Averaged contribution
I	90 %	40 %	90 %	73 %
II	95 %	45 %	80 %	73 %
III	80 %	50 %	65 %	65 %
IV	90 %	70 %	85 %	82 %
V	100 %	70 %	95 %	88 %