Kombination von überwachten und teilüberwachten Methoden des maschinellen Lernens für eine adaptive und robuste Zustandsüberwachung industrieller Applikationen

Dissertation
zur Erlangung des Grades
des Doktors der Ingenieurwissenschaften
der Naturwissenschaftlich-Technischen Fakultät
der Universität des Saarlandes

von

M.Sc. Steffen Klein

Saarbrücken
2024

Tag des Kolloquiums:	07.03.2025
Dekan:	Prof. Dr.-Ing. Dirk Bähre
Berichterstatter:	Prof. Dr. Andreas Schütze
	Prof. Dr.-Ing. Dirk Bähre
Akad. Mitarbeiter:	Dr.-Ing Amine Othmane
Vorsitz:	Prof. Dr. -Ing. Matthias Nienhaus

Zusammenfassung

Industrielle Datenlandschaften werden zunehmend komplexer, wodurch Methoden des maschinellen Lernens immer populärer werden, um die Wertschöpfung auf Basis der gesammelten Daten zu erhöhen. Zentrale Versprechen stellen dabei die Optimierung der Anlagenverfügbarkeit auf Basis moderner Instandhaltungsstrategien sowie eine ressourcensparende und effiziente Prozessführung dar. Viele der vorgeschlagenen Algorithmen sind überwachte Methoden, die versteckte Muster in den Daten lernen und wiedererkennen können. Zum Erlernen der Muster werden jedoch Trainingsdaten in hoher Qualität benötigt, die in einem industriellen Umfeld aufgrund der hohen Kosten nur in begrenztem Umfang generiert werden können. Zudem stellt die Stabilität der Modelle durch variierende Umgebungsbedingungen eine der maßgeblichen Hürden für die breitbandige Nutzung der Methoden dar.

In dieser Arbeit wird ein methodenunabhängiger, übergeordneter Algorithmus zur Kombination teilüberwachter und überwachter Methoden des maschinellen Lernens vorgestellt, welcher eine robuste und adaptive Zustandsüberwachung industrieller Prozesse ermöglicht. Dadurch können die datengetriebenen Modelle in der Anwendung an variierende Umgebungseinflüsse oder bis dato unbekannte Systemzustände angepasst werden. Zur Absicherung der für die Algorithmen benötigten Datenqualität wird zudem eine dem Modellbildungsprozess vorgelagerte Mess- und Datenplanung vorgeschlagen. Um die breite Anwendungsmöglichkeit und Leitungsfähigkeit der Methoden zu demonstrieren, erfolgt die Validierung anhand vier stark unterschiedlicher industrieller Anwendungen.

Abstract

Industrial data landscapes are becoming increasingly complex, leading to a growing popularity of machine learning methods aimed at enhancing the added value of industrial applications based on collected data. Key promises include optimizing equipment availability through modern maintenance strategies and ensuring resource-efficient and effective process management. Many of the proposed algorithms are supervised methods capable of learning and recognizing hidden patterns in the data. However, high-quality training data is required to learn these patterns, which can only be generated to a limited extent in industrial environments due to the high costs involved. Furthermore, the stability of the models under varying environmental conditions in complex industrial settings poses a significant challenge for the widespread adoption of these methods.

This paper presents a method-independent, superordinate algorithm for combining semi-supervised and supervised machine learning methods, enabling robust and adaptive condition monitoring of industrial processes. This allows the data-driven models in the application to be adapted to varying environmental influences or previously unknown system states. To ensure the data quality required for the algorithms, measurement and data planning upstream of the modeling process is proposed. To demonstrate the broad applicability and effectiveness of the methods, validation is conducted using four significantly different industrial applications.

Inhaltsverzeichnis

1 EINLEITUNG 1

2 GRUNDLAGEN UND STAND DER TECHNIK 7

2.1 ZUSTANDSÜBERWACHUNG 7

2.2 MASCHINELLES LERNEN FÜR DIE ZUSTANDSÜBERWACHUNG 9

2.2.1 Datenqualität 13

2.2.2 Domain Shift 19

2.2.3 Design of Experiment 20

2.3 METHODENPOOL FÜR DAS MASCHINELLE LERNEN 22

2.3.1 Signalvorverarbeitung 24

2.3.2 Merkmalsextraktion 25

2.3.3 Merkmalsselektion 31

2.3.4 Dimensionsreduktion 36

2.3.5 Klassifizierung und Regression 39

2.3.6 Novelty Detection 42

2.3.7 Validierung 46

2.3.8 Hyperparameteroptimierung 47

2.3.9 Stacks 49

3 METHODIK FÜR EINE GANZHEITLICHE ZUSTANDSÜBERWACHUNG 51

3.1 MESS- UND DATENPLANUNG 51

3.2 ADAPTIVE ZUSTANDSÜBERWACHUNG 55

3.2.1 Time-Frequency-Extractor 57

3.2.2 Stacks für Novelty Detection, Klassifizierung und Regression 59

3.2.3 Grenzwertbestimmung für die Novelty Detection 61

4 ANWENDUNGSSZENARIEN 65

4.1 ZUSTANDSÜBERWACHUNG EINES HYDRAULISCHEN SYSTEMS 66

4.1.1 Einführung 66

4.1.2 Experimenteller Aufbau 67

4.1.3 Mess- und Datenplanung 70

4.1.4 Modelle für überlagerte Fehler 72

4.1.5 Einfluss von Domain Shifts 74

4.1.6 Adaptive Zustandsüberwachung *78*
4.1.7 Zusammenfassung und Diskussion der Anwendung *92*
4.2 ZUSTANDSÜBERWACHUNG EINER SERVOPNEUMATISCHEN SCHWEIßZANGE 95
4.2.1 Einführung *95*
4.2.2 Experimenteller Aufbau *96*
4.2.3 Mess- und Datenplanung *97*
4.2.4 Adaptive Zustandsüberwachung *99*
4.2.5 Zusammenfassung und Diskussion der Anwendung *113*
4.3 AKUSTISCHE LECKAGEDETEKTION AN PNEUMATIKZYLINDERN 115
4.3.1 Einführung *115*
4.3.2 Experimenteller Aufbau *116*
4.3.3 Mess- und Datenplanung *119*
4.3.4 Auswertung und Modellbildung *125*
4.3.5 Adaptive Zustandsüberwachung *128*
4.3.6 Zusammenfassung und Diskussion der Anwendung *132*
4.4 VERSCHLEIßDETEKTION BEIM WÄLZSCHÄLEN 134
4.4.1 Einführung *134*
4.4.2 Experimenteller Aufbau *135*
4.4.3 Mess- und Datenplanung *137*
4.4.4 Auswertung und Modellbildung *151*
4.4.5 Zusammenfassung und Diskussion der Anwendung *160*

5 DISKUSSION 163

6 ZUSAMMENFASSUNG UND AUSBLICK 167

LITERATURVERZEICHNIS 171

ABBILDUNGSVERZEICHNIS 195

TABELLENVERZEICHNIS 203

EIGENE VERÖFFENTLICHUNGEN 205

DANKSAGUNG 209

1 Einleitung

Mit der Einführung moderner Qualitätsmanagement-Strategien wird die Fertigungsmesstechnik zunehmend näher in die Produktion verlagert. Die Zahl der verwendeten Sensoren und damit auch die Menge und Komplexität der erfassten Daten nimmt seit den 80er Jahren stetig zu [1]. Zentrale Versprechen dieser Entwicklung sind eine umfassende Zustandsüberwachung technischer Systeme sowie eine kontinuierliche Überwachung der Produktqualität. Diese senken Kosten durch unerwartete Maschinenausfälle oder Produktionsausschuss und sind damit in vielen industriellen Bereichen von zentraler Bedeutung [2]. Auch die Steigerung der Betriebs- und Mitarbeitersicherheit sowie ein verantwortungsvollerer Umgang mit den verfügbaren Ressourcen stellen Vorteile der sich weiterentwickelnden Datenstrategien vieler Unternehmen dar [3].

Die Umsetzung dieser Strategien basiert auf der Entwicklung zuverlässiger Diagnosesysteme, die den Betriebszustand einzelner Komponenten komplexer Maschinen präzise erfassen können. Neben der geeigneten Sensorik zur Messung maschinenspezifischer Parameter beinhalten solche Systeme auch Methoden und Algorithmen zur Bewertung von Schädigungs- und Degradationsmechanismen. Dabei kommen bevorzugt signalanalytische oder datenbasierte Zustandsdiagnoseverfahren aus dem Bereich des maschinellen Lernens zum Einsatz. Diese Verfahren sollen Prognosen über die verbleibende Restlebensdauer der Komponenten ermöglichen, wodurch Ausfälle vorhergesagt und Instandhaltungsmaßnahmen in bereits geplante Stillstandszeiten integriert werden können [4, 5].

Der Erfolg zustandsbewertender Verfahren auf Basis maschinellen Lernens hängt dabei maßgeblich von der Qualität und Verfügbarkeit geeigneter Daten ab [6]. Gerade im industriellen Umfeld ist die Erzeugung von Trainingsdaten sehr teuer bzw. häufig nicht möglich, ohne beispielsweise in die Serienproduktion und damit in die Wertschöpfungskette einzugreifen. Des Weiteren ist es bei der Mehrheit der Anwendungen nicht möglich, alle relevanten Systemzustände oder Events in den Trainingsdaten abzubilden, wodurch die Vollständigkeit, aber vor allem auch die Gültigkeit der Modelle im Betrieb über längere Zeit nicht gewährleistet werden kann.

Laboranwendungen dagegen bieten häufig die Möglichkeit, Daten anhand eines Demonstrators oder Prüfstands zu generieren, was das Erzeugen einer umfangreichen

Trainingsdatenbasis erlaubt, in dem eine Anlage oder Prozess gezielt aus dem Normalzustand in relevante Systemfehlerzustände gebracht wird. Viele der auf Basis solcher Daten hergeleiteten Modelle zeigen sich sehr leistungsfähig, was häufig zur Demonstration des Potentials datengetriebener Methoden genutzt wird [7, 8, 9]. Problematisch bei dieser Art der Modellbildung ist häufig die Übertragbarkeit der Anwendungen aus der Laborumgebung in die deutlich komplexe industrielle Umgebung, in der viele Quereinflüsse und instabile Umgebungsbedingungen vorzufinden sind. Diese als Domain Shifts bezeichneten Effekte führen häufig zur Minderung der Performanz oder sogar zur völligen Ungültigkeit der im Labor generierten Modelle [10].

Die dieser Dissertation vorangegangenen Projekte ICM-Hydraulik (2012-2015) und MoSeS-Pro (Modulare Sensorsysteme für die Echtzeit-Prozesssteuerung und smarte Zustandsbewertung, 2015-2018) zeigen die Potentiale datengetriebener Modelle für die Zustandsüberwachung industrieller Applikationen, wie beispielsweise hydraulische Aggregate oder elektromechanische Antriebe [7, 8, 9, 11]. Für die Generierung der Daten im Rahmen dieser Projekte wurden jeweils komplexe Prüfaufbauten mit einer umfangreichen Instrumentierung entwickelt und aufgebaut. Neben den außerordentlichen Potentialen zeigten sich in diesem Rahmen jedoch auch die genannten Hürden, welche den Einsatz der Methoden in realen industriellen Umgebungen erschweren. Einerseits wurden die Modelle massiv von teils nicht kontrollierbaren Umweltbedingungen beeinflusst, andererseits zeigten sich trotz baugleicher technischer Systeme erhebliche Probleme bei der Übertragbarkeit der Modelle, was beispielsweise auf Fertigungstoleranzen oder spezifische Ausfallmechanismen zurückzuführen ist.

Ziel des im Rahmen dieser Dissertation begleiteten Projektes MessMo (Messtechnisch gestützte Montage, 2018–2020) war die Entwicklung bzw. Identifikation geeigneter Algorithmen zur Zustandsüberwachung von Produktqualität und Betriebsmitteln an komplexen Montagelinien, basierend auf ohnehin zu Regelzwecken vorhandenen Prozesssensoren. Hier zeigten sich weitere, größtenteils strukturelle Probleme in realen industriellen Datenlandschaften, welche den Einsatz bzw. die Wertschöpfung maschineller Lernverfahren erschweren [12]. Als Beispiele sind hier eine mangelnde Synchronisierung der Daten zwischen verschiedenen Prozessschritten, eine frühzeitige Reduktion der Signale auf nicht ausreichende Merkmalsdaten oder eine nicht gegebene Rückverfolgbarkeit von Rückläufern zu nennen. Dadurch verlagerte sich der Fokus von der anfänglich geplanten Entwicklung, Identifikation und Erprobung geeigneter Algorithmen hin zur Absicherung und Steigerung der Datenqualität [13].

Kerninhalte des ebenfalls im Rahmen der Dissertation begleiteten Projektes KI-MUSIK4.0 (Mikroelektronik-basierte universelle Sensor-Schnittstelle mit Künstlicher Intelligenz für Industrie 4.0, 2020–2024) war die Entwicklung hochleistungsfähiger Mikroelektronik in Kombination mit Sensorik und eingebetteter Software, welche Prozessdaten von Industrieanlagen sammelt und verarbeitet. Die dabei anfallenden großen Datenmengen, beispielsweise bei der Verwendung akustischer Sensoren, sollen sensornah vorverarbeitet und bewertet werden, wodurch eine erhebliche Entlastung der industriellen IT-Infrastruktur erreicht wird [14]. Ein wesentlicher Teil der Forschungsarbeiten stellt dabei die Identifikation bzw. Entwicklung geeigneter Algorithmen des maschinellen Lernens dar, welche trotz der begrenzten Ressourcen im Rahmen einer dezentralen Datenverarbeitung eingesetzt werden können. Auf Basis der Erkenntnisse der vorangegangenen Projekte erfolgte die Datenaufzeichnung der hier betrachteten industriellen Anwendungen unter Berücksichtigung einer für die maschinellen Lernverfahren optimierten Datenqualität.

In einem 2018 veröffentlichten Bericht der Fraunhofer Gesellschaft werden wichtige zukunftsweisende Forschungsfragen im Zusammenhang mit maschinellem Lernen definiert. Hier spielen unter anderem die Datenqualität, die Nachvollziehbarkeit der Modelle sowie das Lernen auf kleinen Datenmengen eine wichtige Rolle. Ebenfalls wird betont, dass im Vergleich zu Schwerpunktländern der künstlichen Intelligenz wie beispielsweise China oder Amerika, in denen große B2C Konzerne über enorme Datensätze für tiefe maschinelle Lernverfahren verfügen, in Deutschland aufgrund der produktionsnäheren Industriestruktur sowie strenger Datenschutzregelungen erheblich weniger und andersartige Daten zur Verfügung stehen [15].

Das übergeordnete Ziel dieser Dissertation ist es, einen Beitrag zur Lösung der genannten Probleme zu liefern und dadurch die Hürden für den breiten Einsatz datengetriebener Modelle im Kontext einer industriellen Zustandsüberwachung zu senken. Um dies zu erreichen, wird im Rahmen der Dissertation ein methodenunabhängiger, übergeordneter Algorithmus zur Kombination generischer teilüberwachter und überwachter Methoden des maschinellen Lernens vorgestellt, welcher eine robuste und adaptive Zustandsüberwachung industrieller Prozesse ermöglicht. Im Gegensatz zu überwachten Verfahren, welche rein auf Basis annotierter Daten anwendbar sind, kombinieren teilüberwachte Verfahren Information aus annotierten und nicht annotierten Daten und ermöglichen dadurch die Detektion bis dahin unbekannte Zustände. Dadurch kann eine Zustandsüberwachung technischer Systeme ausgehend vom Normalzustand begonnen werden, welcher beispielsweise nach der Inbetriebnahme oder im Zuge einer Wartung bzw. Instandsetzung durch entsprechendes Personal sichergestellt werden kann. Nach Antrainieren des Normalzustandes können auftretende, unbekannte Fehlerzustände durch das

teilüberwachte Modell erkannt und sukzessive in das überwachte Modell übertragen werden. Beim nächsten Auftreten der Fehlerzustände können diese vom überwachten Modell diagnostiziert und ein Hinweis an das entsprechende Personal ausgegeben werden. Die Implementierung des übergeordneten Algorithmus beruht auf einer MATLAB basierten Open-Source-Toolbox für das automatisierte maschinelle Lernen, welche gemeinsam mit den zum Verständnis der Arbeit benötigten Grundlagen in Kapitel 2 vorgestellt wird [8, 9].

Aufgrund der im Rahmen der Dissertation identifizierten Probleme hinsichtlich der Datenqualität realer industrieller Datenlandschaften wird eine der Datengenerierung und Modellbildung vorgelagerte Mess- und Datenplanung vorgeschlagen. Diese basiert auf der Sicherstellung der strukturellen Datenqualität durch ein Szenario-Format, welches ein klar definiertes Lernproblem und alle dazu benötigten Variablen in strukturierter Form enthält, sowie einem iterativen Prozess zur Prüfung der Daten auf Plausibilität und Informationsgehalt mittels dimensionsreduzierender Verfahren und Visualisierung [13]. Sowohl die Vorstellung der Methodik für die adaptive Zustandsüberwachung als auch der vorgeschlagenen Mess- und Datenplanung erfolgt in Kapitel 3. Hier wird zudem eine automatisierte Methode zur Grenzwertbestimmung für die verwendeten teilüberwachten Methoden eingeführt, welche sich an der klassischen Messunsicherheitsbetrachtung orientiert. Diese besitzt keine Hyperparameter und kann automatisiert, basierend nur auf der Verteilung der Daten, angewendet werden, wodurch sich eine gute Integrierbarkeit in die ohnehin hochautomatisierte Toolbox ergibt.

Die Validierung der vorgestellten Methoden erfolgt in Kapitel 4 anhand vier verschiedener Anwendungen. Bei Anwendung 1 handelt es sich dabei um einen öffentlich verfügbaren Datensatz resultierend aus dem Projekt ICM-Hydraulik, welcher die Zustandsüberwachung eines Hydraulikaggregates anhand typischer Fehlermechanismen simuliert. Das Lernproblem ist dabei als Klassifizierung definiert, wobei jeder Fehlermechanismus in verschiedenen Intensitäten vorhanden ist. Die Generierung der Daten erfolgte hier anhand eines realitätsnahen, mit typischer Prozesssensorik ausgestatteten Prüfaufbaus. Der Datensatz von Anwendung 2 wurde im Rahmen des Projektes MessMo vom Projektpartner Festo zur Verfügung gestellt und beinhaltet die Zustandsüberwachung einer servopneumatischen Schweißzange, ebenfalls in Form eines Klassifizierungsproblems. Auch hier konnten mittels eines Prüfaufbaus typische Fehlermechanismen in verschiedenen Schweregraden simuliert und die Daten der installierten Prozesssensoren sowie Regelgrößen aufgezeichnet werden. Anwendung 3 umfasst eine akustische Leckagedetektion an Pneumatikzylindern, welche im Rahmen des Forschungsprojektes KI-MUSIK4.0 gemeinsam mit dem Projektpartner Festo bearbeitet wurde. Ziel ist hierbei die

Quantifizierung der an den Kolbenstangendichtungen auftretenden Leckagen der Zylinder. Auch hier erfolgte die Datengenerierung an einem speziell von Festo entwickeltem Prüfstand. Auch Anwendung 4 stammt aus dem Projekt KI-MUSIK4.0 und wurde hauptsächlich gemeinsam mit dem Anwendungspartner Schaeffler bearbeitet. Ziel ist hier die Quantifizierung des Werkzeugverschleißes bei einem Wälzschälprozess zur Fertigung von Innenverzahnungen. Im Gegensatz zu Anwendung 1-3 erfolgte die Datenaufnahme hier an einer Testmaschine direkt im Produktionsumfeld bei Schaeffler.

Basierend auf den Erkenntnissen der vorangegangenen Projekte muss ein Zustandsüberwachungssystem in der Lage sein, sich an komplexe industrielle Umweltbedingungen und bis dato unbekannte Systemzustände anzupassen. Aus diesem Grund untersucht diese Dissertation als zentrale Fragestellung die Machbarkeit einer adaptiven Zustandsüberwachung für industrielle Applikationen. Bei der Mehrheit der veröffentlichten Methoden zur datenbasierten Zustandsüberwachung handelt es sich um überwachte Modelle, welche bekannte Fehler mit hoher Genauigkeit vorhersagen können, bis dato unbekannte Zustände jedoch nicht erkennen [16, 17, 18]. Ein vergleichsweise kleiner Teil der veröffentlichten Ansätze beschäftigt sich dagegen mit teilüberwachten Methoden, welche in der Lage sind, neue Zustände zu detektieren, jedoch nur begrenzt eine Diagnose bzw. Quantifizierung bekannter Fehler ermöglichen [19]. Im Kontext einer adaptiven Zustandsüberwachung muss folglich untersucht werden, inwiefern eine Kombination dieser Methodenfamilien eingesetzt werden kann. Ziel ist es dabei zum einen, die Abhängigkeit einer Zustandsüberwachung von Trainingsdaten zu verringern, indem neue Zustände während des Betriebs erkannt und in das überwachte Modell übertragen werden können. Zum anderen muss eine robuste Zustandsüberwachung in der Lage sein, die Beeinflussung des überwachten Modells beispielsweise durch sich ändernde Umweltbedingungen zu erkennen und idealerweise zu kompensieren. Um eine breite Anwendbarkeit sicherzustellen, sollte das System zudem die Integration anwendungsspezifischer Algorithmen ermöglichen. Dies erlaubt eine schnelle Anpassung an die betrachtete Anwendung in Abhängigkeit des formulierten Lernproblems (Klassifikation / Regression) sowie der verfügbaren Sensorsignale. Darüber hinaus ist zu untersuchen, ob ein solches System vollständig automatisierbar ist oder ob es hauptsächlich als Assistenzsystem für die entsprechenden Prozessingenieure genutzt werden kann.

Aufgrund der zuvor identifizierten Probleme mit der Datenqualität in industriellen Anwendungen widmet sich diese Dissertation als sekundäre Fragestellung dem Thema der Mess- und Datenplanung. Ziel ist es dabei den Prozess der Datengenerierung im industriellen Kontext über ein vereinheitlichtes Verfahren abzubilden, welches die

Qualität der aufgezeichneten Daten für die spätere Modellbildung gewährleistet. Dies beinhaltet einerseits die Bereitstellung der Daten in geeigneter Form, um die Verarbeitung der Daten mit den entsprechenden Algorithmen zu ermöglichen. Andererseits muss nachgewiesen und sichergestellt werden, dass die aufgezeichneten Daten die relevante Information in Bezug auf die zu bestimmende Zielgröße enthalten.

Speziell bei den letzten beiden untersuchten Anwendungen handelt es sich zudem um bis dato noch nicht untersuchte Probleme, weshalb ein weiteres Ziel der Dissertation an dieser Stelle der Aufbau von Prozesswissen und eine generelle Untersuchung der Machbarkeit darstellt. Durch die Breite der vorgestellten Anwendungen, soll einerseits die generelle Anwendbarkeit und Leistungsfähigkeit des übergeordneten Algorithmus sowie der verwendeten Methoden des maschinellen Lernens untersucht werden. Andererseits sollen aber auch die Grenzen und die damit verbundenen Einschränkungen in der industriellen Praxis aufgezeigt werden.

2 Grundlagen und Stand der Technik

In diesem Kapitel werden alle erforderlichen Grundlagen, die zum Verständnis der Arbeit erforderlich sind, vorgestellt. Hierfür werden Begrifflichkeiten und der Stand der Technik aus dem Bereich der Zustandsüberwachung und des maschinellen Lernens eingeführt sowie die spezifischen zur Analyse und Modellbildung verwendeten Methoden erläutert. Auf eine umfassende Vorstellung der technologischen Grundlagen der einzelnen Anwendungsszenarien wird an dieser Stelle verzichtet. Eine kurze Einführung erfolgt stattdessen zu Beginn der Kapitel der jeweiligen Anwendungsszenarien.

2.1 Zustandsüberwachung

Mit den ständig zunehmenden Leistungs- und Funktionsdichte technischen Systemen steigen auch die Anforderungen an deren Verfügbarkeit und Wirtschaftlichkeit. Klassische Instandhaltungsstrategien wie präventive Wartung oder feste Inspektionsintervalle reichen zunehmend nicht mehr aus, um diesen Anforderungen gerecht zu werden [2, 20]. Dies kann zu hohen Kosten aufgrund von Produktionsausfällen oder nicht vollständig ausgeschöpften Abnutzungsvorräten führen. Daher gewinnen Strategien der vorausschauenden und zustandsbasierten Wartung und Instandhaltung für Anlagenbetreiber immer mehr an Bedeutung [21]. Der Begriff der Zustandsüberwachung und -diagnose (engl. Condition Monitoring) ist in der ISO 13372 (Zustandsüberwachung und -diagnostik von Maschinen - Begriffe) als Untersuchung der Symptome zur Feststellung der Natur des Fehlers oder Ausfalls, etwa Typ, Ausmaß und Situation definiert [22]. Ein Fehler bezeichnet demnach einen Zustand einer Maschine, der eintritt, wenn eines ihrer Bauteile oder eine ihrer Baugruppen sich verschlechtert oder ein abnormales Verhalten zeigt, das zum Ausfall der Maschine führen kann. Ein Ausfall liegt dabei vor, wenn die Maschine durch einen fortschreitenden Fehler nicht mehr in der Lage ist, eine erforderliche Funktion zu erfüllen. Die Norm ist Teil einer Serie von Normen, die sich mit der Zustandsüberwachung und -diagnose von Maschinen befassen, wie z.B. die ISO 17359

(Zustandsüberwachung und -diagnose von Maschinen - Allgemeine Leitlinien) oder die ISO 10816 (Mechanische Schwingungen - Bewertung der Schwingungen von Maschinen durch Messungen an nicht rotierenden Teilen).

Im Bereich der sensorgestützten Maschinendiagnose gibt es gerade in Bezug auf die sich entwickelnde Industrie 4.0 eine Vielzahl von Ansätzen zur Zustandsüberwachung. Aus [20] und [21] können zahlreiche Beispiele für Zustandsüberwachungssysteme für verschiedenste Anlagentypen entnommen werden. Diese bestehen zumeist aus speziell zugeschnittener Sensorik in Kombination mit herstellerspezifischen Auswertealgorithmen. Die Art der verwendeten Sensoren ist dabei sehr stark von der jeweiligen Anwendung abhängig. Ziel ist es immer auf ohnehin im Prozess installierte Sensorik zurückzugreifen. Gerade bei komplexeren Prozessen oder in Bereichen, in denen keine Sensorik vorhanden ist, werden aber zusätzliche dedizierte Sensoren installiert. Hier haben sich vor allem im Bereich rotierender Maschinen Luft- und Körperschallsensoren durch ihre hohe Leistungsfähigkeit im Bereich der Zustandsüberwachung und ihre einfache, auch nachträgliche Integrierbarkeit in den Prozess hervorgetan. Abhängig von den Anforderungen des betrachteten Prozesses an die Bandbreite der verwendeten Sensoren, stehen im Bereich des Körperschalls zudem Acoustic Emission Sensoren zur Verfügung. Diese können Oberflächenwellen bis in den MHz Bereich erfassen, wodurch sie sich vor allem im Bereich der Werkzeugüberwachung von Schneide, Fräs- oder Schleifprozessen etabliert haben [23, 24].

Die Auswertung von Luft- und Körperschallsensoren kann in Zeit- und Frequenzbereich erfolgen, wobei sich frequenzselektive Verfahren auf Basis charakteristischer Frequenzen und weiterführender Analyseverfahren wie beispielsweise die Ordnungsanalyse oder Hüllkurvenanalyse bewährt haben. Charakteristische Frequenzen bezeichnen dabei von der grundlegenden Drehfrequenz der Welle abhängende Frequenzen, welche beispielsweise durch Überrollvorgänge von Oberflächenschäden angeregt werden [25]. Die Herleitung dieser Frequenzen basiert rein auf den Geometrien der betrachteten Anwendung, weshalb diese in Kombination mit den zugrundeliegenden Drehzahlen berechnet werden können. Durch die Vielzahl der erfassten Messgrößen mit unterschiedlichen Abtastraten und Datenformaten ergeben sich im heutigen industriellen Umfeld komplexe, hochgradig heterogene Datenlandschaften. In Abbildung 2-1 ist eine Zusammenfassung der im Bereich der Zustandsüberwachung häufig verwendeten Sensoren, Methoden sowie der angestrebten Ziele gezeigt [15, 26].

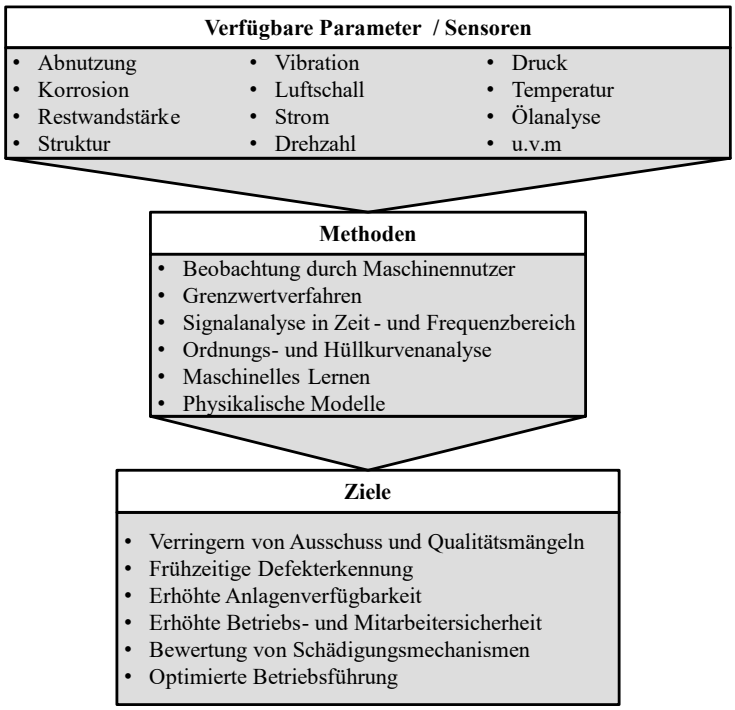


Abbildung 2-1: Übersicht über verwendete Sensoren, Methoden und Ziele der Zustandsüberwachung (angepasst nach [15]).

2.2 Maschinelles Lernen für die Zustandsüberwachung

Durch die komplexen und umfangreichen Datenlandschaften heutiger Produktionsanlagen steigt der Bedarf an Methoden des maschinellen Lernens, um Daten verschiedenster Ursprünge in Modellen zu konzentrieren und damit als Information nutzbar zu machen [27]. In Abbildung 2-2 ist eine Übersicht verwendeter Methoden des maschinellen Lernens im Kontext einer Zustandsüberwachung gegeben. Unterschieden werden grundsätzlich überwachte, teilüberwachte, unüberwachte und verstärkende Lernmethoden. Eine generelle Struktur von überwachten und teilüberwachten Methoden teilt sich in eine Trainings- und eine Testphase. Während in der Testphase Muster aus historischen Daten identifiziert und gelernt werden, erfolgt in der Testphase (auch Inferenz) eine Klassifizierung oder Regression unbekannter Daten [18, 28]. Eine Klassifizierung wird dabei typischerweise bei Vorliegen einer kategorischen Zielgröße mit diskreten Klassen- bzw. Gruppenzugehörigkeiten verwendet, wohingegen eine Regression eine kontinuierliche Zielgröße voraussetzt und damit eine quantitative Vorhersage ermöglicht. Verstärkendes Lernen kann dabei weder als überwacht noch unüberwacht bezeichnet werden, da es ausschließlich auf der Belohnung bzw. Bestrafung von erwünschtem bzw. unerwünschtem Verhalten des Modells basiert [29]. Da es im Rahmen dieser

Dissertation keine Anwendung findet, wird an dieser Stelle auf eine tiefgehende Erläuterung verzichtet.

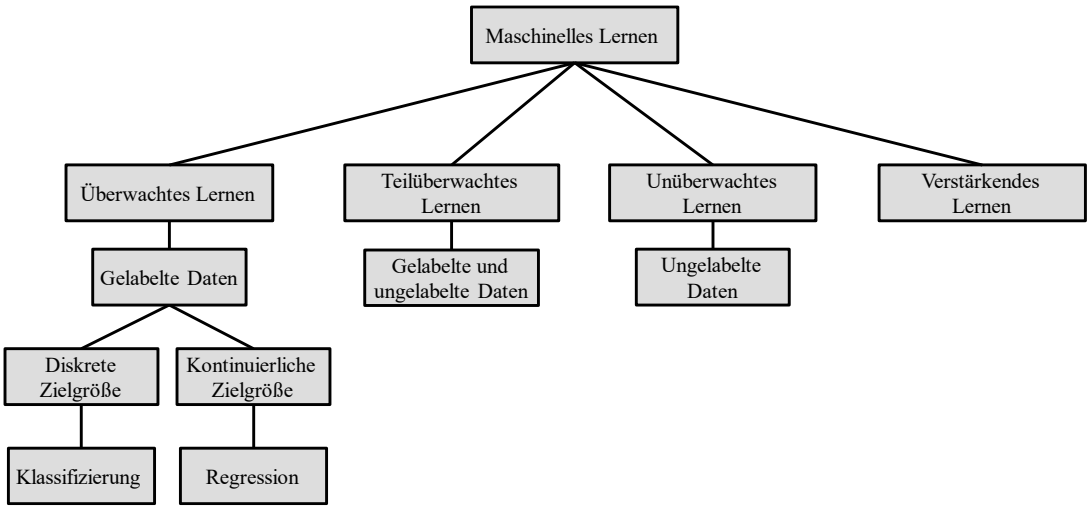


Abbildung 2-2: Übersicht über verfügbare Methodenfamilien im Bereich des maschinellen Lernens (angepasst nach [18]).

Unüberwachte Methoden zeichnen sich durch ihre strukturentdeckenden Eigenschaften aus. Sie benötigen keine gelabelten Daten und eignen sich vor allem zur Analyse und Visualisierung dominanter Effekte. Teilüberwachte Verfahren wie beispielsweise die Anomaliedetektion (auch Neuheitserkennung oder Ausreißererkennung, engl. Novelty Detection) werden auf einem Teil bekannter, gelabelter Daten trainiert und erlauben dadurch die Identifikation neuer unbekannter Zustände. Überwachte Lernverfahren dagegen benötigen für das Training eine vollständige Datenbasis bestehend aus Sensordaten und Zielgröße, mit deren Hilfe sie Muster in den Daten identifizieren und lernen können [16, 17, 18].

Ausgehend von den grundlegenden Funktionsweisen lassen sich je nach verwendeter Methode unterschiedliche Überwachungsszenarien und Instandhaltungsstrategien umsetzen. In Abbildung 2-3 ist dies schematisch anhand des Abnutzungsvorrats einer Maschine dargestellt. Im Allgemeinen stützen sich überwachte Methoden auf Trainingsdaten, die verschiedene Maschinenzustände bzw. Fehler repräsentieren, weshalb sie auf die Erkennung dieser bekannten Zustände beschränkt sind. Allerdings können überwachte Algorithmen dadurch verborgene Muster in den Daten erkennen, die von anderen Effekten wie Rauschen oder wechselnden Umgebungsbedingungen überlagert sind. Daher haben überwachte Modelle eine hohe Empfindlichkeit gegenüber bekannten Fehlern und ermöglichen eine Diagnose und Quantifizierung, auf deren Basis moderne prädiktive Instandhaltungsstrategien aufbauen [30].

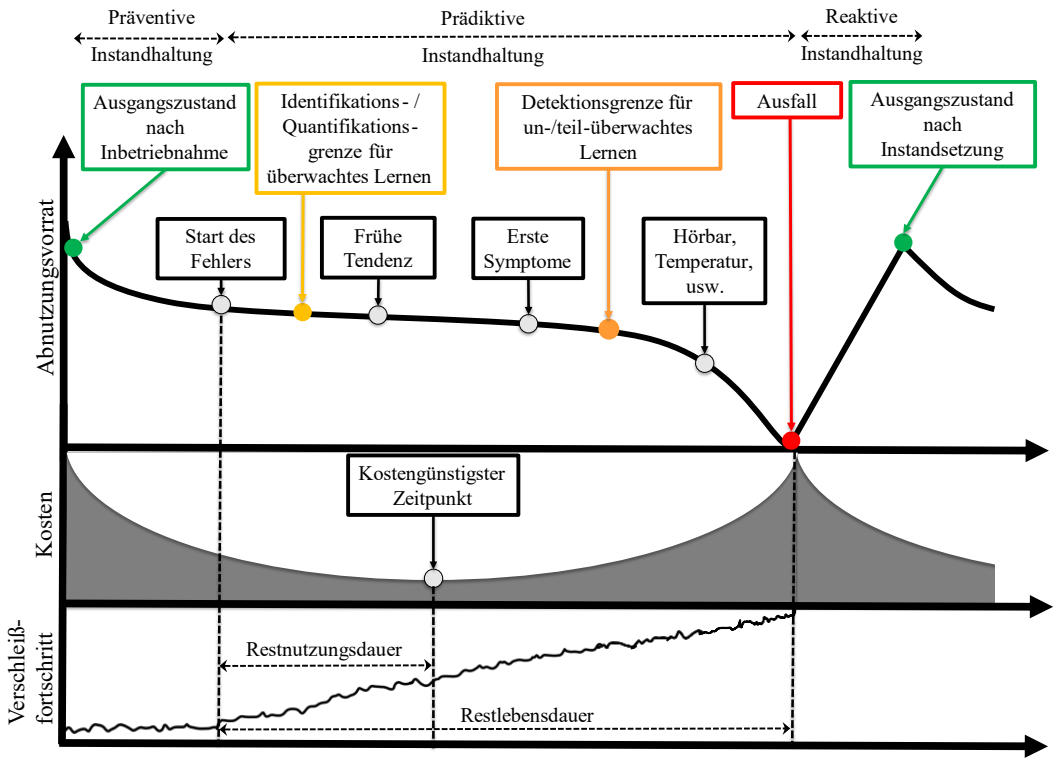


Abbildung 2-3: Entwicklung des Abnutzungsvorrats an einer Maschine mit den zugehörigen Detektionsschwellen für überwachtes und unüberwachtes maschinelles Lernen sowie den Handlungszeiträumen für verschiedene Instandhaltungsstrategien. Bei den Verläufen von Abnutzungsvorrat, Kosten und Verschleißfortschritt handelt es sich um schematische Darstellungen, welche in der Realität stark vom individuellen Prozess abhängig sind (angepasst nach [31]).

Im Gegensatz dazu können un- bzw. teilüberwachte Methoden angewendet werden, wenn keine Trainingsdaten für relevante Fehler verfügbar sind und ermöglichen dadurch die Detektion neuer, unbekannter Zustände oder Fehler. Ein Beispiel hierfür stellt die teilüberwachte Novelty Detection dar, welche auf Basis der Daten der Maschine im Gutzustand trainiert wird und damit in der Lage ist, ein vom Gutzustand abweichendes Systemverhalten zu identifizieren. Ohne eine Kombination mit Prozesswissen in Form von zeitlich lokalisierbaren oder auf charakteristischen Frequenzen basierenden Merkmalen eignet sie sich jedoch nur zur Erkennung dominanter Effekte und liefert dabei auch keine direkte Aussage über die Ursache. Nichtsdestotrotz werden sie benötigt, um unvorhergesehene und bis dahin unbekannte Fehler zu erkennen und einen Ausfall der Maschine zu vermeiden [32].

Im heutigen wissenschaftlichen Umfeld gibt es eine nahezu unüberschaubare Zahl an Veröffentlichungen, die sich mit auf spezifische Datensätze zugeschnittenen Methoden bzw. datensatzspezifischen Problemen beschäftigen [18]. Neben der Unterscheidung der von den Algorithmen benötigten Trainingsdaten wird dabei häufig zwischen konventionellem maschinellem Lernen und tiefem maschinellem Lernen (eng.: Deep Learning) unterschieden [33, 34].

Beim konventionellen maschinellen Lernen werden mehrstufige Signalverarbeitungsketten (auch Pipelines) aus einzelnen Methoden zusammengestellt. Diese arbeiten grundsätzlich in den Schritten der Signalvorverarbeitung, Merkmalsextraktion (engl. Feature Extraktion), Merkmalsselektion (engl. Selektion) mit abschließender Klassifizierung (Classification) oder Regression (engl. Regression) und werden im Folgenden als FESR/C Pipelines oder Stacks bezeichnet. Die Funktion und das Ziel sowie bekannte Vertreter der einzelnen Signalverarbeitungsschritte werden im Kapitel 2.3 vorgestellt.

Ein großer Teil der Publikationen beschäftigt sich mit künstlichen neuronalen Netzen (engl. Artificial Neural Network, ANN), welche angelehnt an die Funktionsweise des menschlichen Gehirns durch das Verschalten einzelner Neuronen gebildet werden [35]. Tiefe neuronale Netze stellen dabei End-To-End Learning Pipelines dar, welche ausgehen von den Rohdaten eine Merkmalsextraktion, -selektion bis hin zur abschließenden Klassifizierung oder Regression durchführen. Der Aufbau oder die Architektur der ANNs ist dabei sehr stark von der Lernaufgabe und der Natur der zu verarbeitenden Daten abhängig. Besonders hervorzuheben sind an dieser Stelle faltende neuronale Netze (engl. Convolutional Neural Network), welche ihren Ursprung im Bereich des maschinellen Sehens und der Sprachverarbeitung haben und mittlerweile auch im Forschungsumfeld der industriellen Zustandsüberwachung vermehrt Anwendung finden [36, 37]. Eine Übersicht über gängige Architekturen, deren Funktion sowie deren Anwendungsfelder können [20] entnommen werden. Die unterschiedlichen Herangehensweisen des konventionellen maschinellen Lernens sowie des Deep Learnings sind in Abbildung 2-4 dargestellt.

Die benötigte Komplexität der Modelle wird unter dem bekannten Aspekt des Bias-Variance-Trade-off zusammengefasst [38]. Bei hohem Bias (Verzerrung) neigt ein Modell zum Underfitting, d.h. die Komplexität des Modells ist nicht hoch genug, wodurch sowohl ein großer Fehler auf Trainings- als auch auf Testdaten zu beobachten ist. Ist die Variance (Varianz) hingegen hoch, neigt das Modell zum Overfitting, wodurch die Generalisierbarkeit des Modells ungenügend ist, was sich in einem niedrigen Fehler im Training und einem hohen Fehler auf Testdaten äußert.

Speziell im Bereich der Zustandsüberwachung bestehen neben der reinen Leistungsfähigkeit in Bezug auf die Komplexität der abbildbaren Lernprobleme zusätzliche Anforderungen an die Methoden. In diesem Zusammengang spielen die Transparenz und die Plausibilisierbarkeit der erzielten Ergebnisse, gerade im industriellen Umfeld, eine immer wichtigere Rolle [15, 39]. Dadurch entsteht das Vertrauen der Anlagenbetreiber in die abgeleiteten Modelle und neben der reinen Anwendung dieser Modelle kann zusätzliches Prozesswissen generiert werden.

Gerade in diesem Punkt zeigen sich aktuell noch Vorteile des klassischen maschinellen Lernens im Vergleich zum Deep Learning. Weitere wichtige Punkte stellen die Übertragbarkeit sowie die Menge der benötigten Trainingsdaten dar. Während komplexe Deep Learning Methoden bei großen Datensätzen häufig eine gesteigerte Performance zeigen, bietet das klassische maschinelle Lernen einen Vorteil bei einer geringeren Anzahl im Training verfügbarer Stichproben, gerade im Bereich der Übertragbarkeit der Modelle [40, 41].

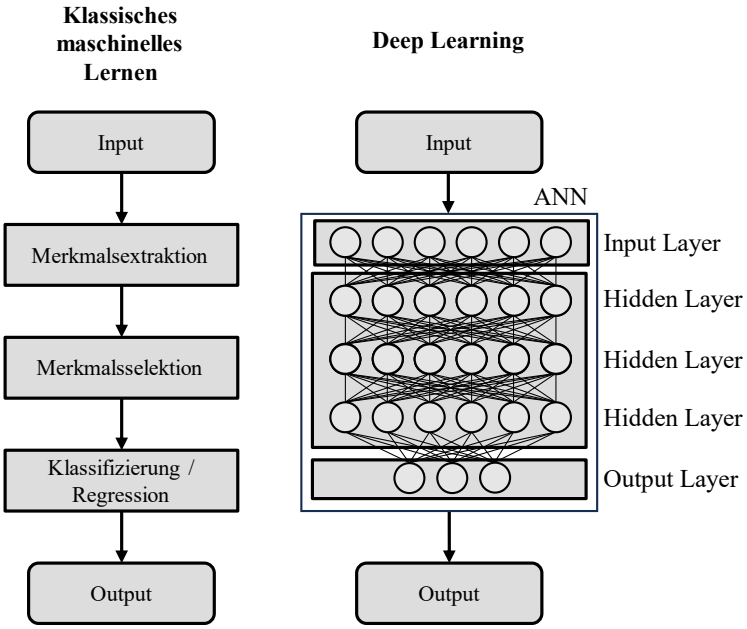


Abbildung 2-4: Unterschiedliche Herangehensweisen von klassischem maschinellem Lernen und Deep Learning (angepasst nach [42]).

2.2.1 Datenqualität

2.2.1.1 Messunsicherheit

Die Ergebnisse datengetriebener Analysen hängen stark von der Qualität der zugrunde liegenden Daten ab [43]. Aus diesem Grund nimmt die Abschätzung der Datenqualität sowohl in der Messtechnik als auch beim maschinellen Lernen eine wichtige Rolle ein [44, 45, 46]. Zugrunde liegen häufig eine statistische Betrachtung der Daten und Methoden, wie beispielsweise die Messunsicherheitsbetrachtungen [47]. Diese erlaubt eine Abschätzung der Abweichung einer realen Messung vom wahren zu messenden Wert. Die Ursachen für die Messabweichungen sind vielfältig und können beispielsweise in Umwelteinflüssen, dem Messsystem oder dem Menschen liegen. Die auftretenden Messabweichungen werden dabei in grobe, systematische und zufällige Fehler unterteilt. Während grobe Fehler, wie beispielsweise ein falsches Ablesen des Messwerts, durch eine sorgfältige Arbeitsweise ausgeschlossen werden können,

werden systematische Fehler maßgeblich durch die Kalibrierung des Messsystems beseitigt. Zufällige Fehler hingegen sind nicht erfassbar oder beeinflussbar und werden durch die Angabe des Ergebnisses mit Messunsicherheit einschließlich Fehlerfortpflanzung berücksichtigt [47]. Die Angabe der Messunsicherheit erfolgt meist in Form einer Gaußschen Normalverteilung der Form:

$$f(x) = \frac{1}{\sqrt{2\,\pi} \cdot \sigma} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2} \tag{2.1}$$

Dabei steht μ für den Erwartungswert, auch bekannt als das erste statistische Moment, während σ^2 die Varianz der Verteilung repräsentiert und als zweites zentrales statistisches Moment bezeichnet wird. Da nie die gesamte zugrundeliegende Verteilung bekannt ist, wird als Schätzwert für den Erwartungswert der Mittelwert $\bar{x}$ nach Formel (2.2) berechnet, welcher für $n \rightarrow \infty$ gegen den Erwartungswert der zugrundeliegenden Wahrscheinlichkeitsverteilung konvergiert [47]. Werden alle systematischen Fehler korrigiert, entspricht der Erwartungswert μ dem wahren Messwert.

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{2.2}$$

Die Varianz σ^2 ist ein Maß für die Verteilung zufälliger Messfehler um den Erwartungswert und wird als zweites zentrales Moment bezeichnet. Als bestmögliche Abschätzung der positiven Wurzel der Varianz wird die Standardabweichung s wie folgt berechnet:

$$s = \sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^{N} (x_i - \bar{x})^2} \tag{2.3}$$

Die Standardabweichung beschreibt ein Konfidenzintervall, in dem sich 68,3 % aller Messwerte befinden. Im Folgenden wird $\bar{x}$ und μ sowie s und σ als äquivalent betrachtet.

Die Messunsicherheit $\Delta\bar{y}$ für mehrere statistisch unabhängige Messgrößen x_j kann dann nach der klassischen Gaußschen Fehlerfortpflanzung (auch quadratische Fehlerfortpflanzung) für den allgemeinen Fall $y = f(x_1, \dots, x_n)$ wie folgt berechnet werden:

$$\Delta\bar{y} = \sqrt{\sum_{j=1}^{N} \left(\frac{\delta f(x_1, \dots, x_n)}{\delta x_j} \cdot \Delta x_j \right)^2} \tag{2.4}$$

mit $\Delta x_j =$ Messunsicherheiten der einzelnen Messgrößen

Eine Verallgemeinerung der Messunsicherheitsbetrachtung ist im „Guide to the Expression of Uncertainty in Measurement“ (GUM) und ihre Erweiterungen GUM-S1 und GUM-S2 definiert [48, 49, 50]. Diese stellt eine vereinheitlichte und international anerkannte Vorgehensweise zur Ermittlung, Bewertung und Angabe der Messunsicherheit dar. Konkret fordert eine Messunsicherheitsbetrachtung nach GUM dabei die Beschreibung der Messaufgabe, die Einschätzung der beteiligten Größen mithilfe von Wahrscheinlichkeitsverteilungen sowie das Aufstellen der Modellgleichung. Für das in der Praxis häufig als Problem angesehene Aufstellen der Modellgleichungen bestehen Verfahren, die eine einfache Anpassung an die konkrete Messaufgabe zulassen, wodurch sich das GUM-Verfahren als praxisgerechte, schrittweise abarbeitbare Prozedur anwenden lässt [51].

In aktuellen Forschungsarbeiten wurde der Einfluss der Messunsicherheit auf die Leistungsfähigkeit von maschinellen Lernalgorithmen untersucht, indem die Fortpflanzung der Messunsicherheit über den gesamten Modellbildungsprozess betrachtet wurde. Dafür wurden die Methoden der Merkmalsextraktion, -selektion, Klassifizierung und Regression, der in Kapitel 2.3 beschriebenen Toolbox für das automatisierte maschinellen Lernen um eine Messunsicherheitsbetrachtung nach GUM erweitert [52, 53]. Dadurch können die modellbasierten Vorhersagen unter Berücksichtigung der Messunsicherheit der Sensor-Rohdaten angegeben und hinsichtlich ihrer Qualität und Vertrauenswürdigkeit bewertet und interpretiert werden [45].

2.2.1.2 Test auf Normalverteilung

Der zentrale Grenzwertsatz besagt, dass der Mittelwert einer großen Anzahl von unabhängigen und identisch verteilten Zufallsvariablen, unabhängig von der ursprünglichen Verteilung, näherungsweise normalverteilt ist [54]. Das bedeutet, dass viele reale Prozesse, die das Ergebnis von vielen unabhängigen Einflüssen sind, tendenziell eine Normalverteilung aufweisen. Ausgehen von dieser Annahme legen viele Verfahren aus den Bereichen der Messtechnik sowie des maschinellen Lernens eine Normalverteilung der Daten zugrunde [55, 56]. Um festzustellen, ob eine Normalverteilung vorliegt, wird üblicherweise eine Kombination aus grafischen und statistischen Prüfverfahren eingesetzt [57]. Ein verbreiteter grafischer Ansatz ist das Quantil-Quantil-Diagramm (Q-Q-Plot) [58]. Hierbei werden die Quantile der unbekannten Verteilung gegen die Quantile einer bekannten Verteilung aufgetragen. Liegen die Punkte nahe an einer Geraden, kann angenommen werden, dass die Daten aus einer ähnlichen Verteilung stammen. In Abbildung 2-5 sind exemplarisch die Normalverteilung sowie der Einfluss von Schiefe (3. statistisches Moment) und

Kurtosis (4. statistisches Moment) auf die Verteilungsfunktionen dargestellt. Die Q-Q-Plots der Normalverteilungen bilden dabei eine Gerade, wohingegen sich für die anderen Verteilungen klare Abweichungen von der Geraden ergeben. Somit erlauben grafische Verfahren die Identifikation der Lage bzw. Skalierungsunterschiede der Verteilungen sowie das Detektieren von Ausreißern [58].

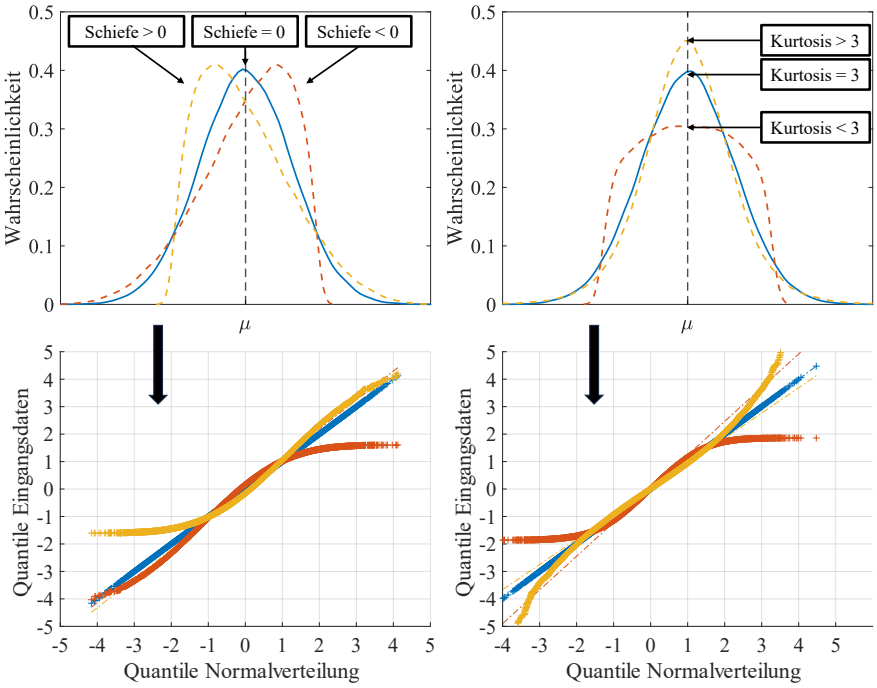


Abbildung 2-5: Einfluss von Schiefe (3. statistisches Moment) und Kurtosis (4. Statistisches Moment) auf die Dichtefunktion einer Normalverteilung sowie die zugehörigen Q-Q-Plots zum Überprüfen der Daten auf Normalverteilung.

Obwohl es sich bei den grafischen Methoden um nützliche Instrumente handelt, werden diese für eine objektive und automatisierbare Aussage häufig mit statistischen Tests kombiniert. In der Literatur sind mehr als 40 verschiedene Tests auf Normalverteilung bekannt, wobei es sich üblicherweise um Hypothesentests handelt. Diese stellen der Nullhypothese, welche eine Normalverteilung der Daten annimmt, einer Alternativhypothese gegenüber [56, 59]. Typische Vertreter solcher Tests sind beispielsweise der Kolmogorov-Smirnov-Test (in verschiedenen Varianten und Modifikationen) [60], der Shapiro-Wilk-Test [59], der Jarque-Bera-Test [59] und der Anderson-Darling-Test [59]. Im Vergleich zu den grafischen Methoden weist die Leistungsfähigkeit dieser Tests eine starke Abhängigkeit von dem Stichprobenumfang auf. In mehreren Untersuchungen wurde festgestellt, dass der Shapiro-Wilk-Test in vielen Situationen leistungsfähiger als andere Methoden ist, weshalb er an dieser Stelle genauer vorgestellt wird [56, 57, 61].

Die Teststatistik W des Shapiro-Wilk-Tests wird nach Formel (2.5) berechnet. Diese berechnet im Zähler einen Schätzwert für die Varianz, der die erwartete Varianz einer

Stichprobe unter der Annahme einer normalverteilten Grundgesamtheit darstellt und vergleicht diesen mit einem Schätzwert für die tatsächliche Varianz der Stichprobe im Nenner.

$$W = \frac{b^2}{(n-1) \cdot s^2} = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{2.5}$$

Hierbei beschreiben $x_{(i)}$ die nach aufsteigender Größe geordneten Stichprobenwerte und $\bar{x}$ den Stichprobenmittelwert. Die Konstanten a_i berechnen sich wie folgt:

$$(a_1, \dots a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{\frac{1}{2}}} \tag{2.6}$$

Dabei ist $m = (m_1, m_2, \dots, m_n)^T$ der Vektor der erwarteten Werte der Ordnungsstatistik, die unabhängige und identisch verteilte Zufallsvariablen darstellen, die der Normalverteilung folgen. V ist die Kovarianzmatrix dieser Ordnungsstatistiken.

Ist die Grundgesamtheit normalverteilt, führen beide Varianzschätzungen unabhängig voneinander zu einem ähnlichen Ergebnis. Je geringer die Abweichung zwischen den geschätzten Varianzen ist, desto wahrscheinlicher ist es, dass die Grundgesamtheit der Stichprobe tatsächlich normalverteilt ist. Basierend auf der Teststatistik W wird ein p-Wert bestimmt, der angibt, wie wahrscheinlich es ist, dass die beobachtete Teststatistik unter der Nullhypothese auftritt. Ein kleiner p-Wert (typischerweise < 0.05) führt zur Ablehnung der Nullhypothese, was darauf hinweist, dass die Daten nicht normalverteilt sind [62].

2.2.1.3 Erweiterte Datenqualität für das maschinelle Lernen

Neben den klassischen Betrachtungen der Datenqualität in Form der Messunsicherheit rückt im Bereich des maschinellen Lernens zunehmend eine erweiterte Betrachtung der Datenqualität in den Fokus aktueller Forschungsarbeiten [13, 63, 64, 65, 66, 67]. Hierbei werden neben der Qualität der zugrundeliegenden Messungen vor allem auch die Struktur und der Informationsgehalt der Daten untersucht. Die Struktur bezieht sich in diesem Fall auf die Form der abgelegten bzw. verarbeiteten Daten und stellt damit die grundsätzliche Anwendbarkeit der Algorithmen sicher. Der Informationsgehalt beinhaltet einerseits die Informationen bezüglich der zu bestimmenden Zielgröße und andererseits die in den Daten enthaltene Varianz relevanter Umgebungsbedingungen, welche im realen Betrieb einen Einfluss auf die Performanz des Modells haben.

Ein Beispiel zur Sicherung bzw. Erhöhung der strukturellen und informationsbezogenen Datenqualität an Bestandsanlagen stellt die in [68] vorgestellte Checkliste dar. Diese beinhaltet einen an kleine und mittelständische Unternehmen adressierten Leitfaden zur Durchführung datengetriebener Projekte, wodurch Fehler sowohl bei der Datenaufzeichnung als auch bei der Herleitung und Interpretation der Modelle verhindert werden sollen.

Eine objektive Beurteilung der Datenqualität kann anhand der in [69] beschriebenen konkreten Kriterien bzw. Dimensionen erfolgen. Die 15 Dimensionen adressieren ein allgemeines Qualitätswesen für Daten und stehen damit nicht direkt im Kontext des maschinellen Lernens. Die Dimensionen lauten wie folgt:

1. Zugänglichkeit
2. Angemessener Umfang
3. Glaubwürdigkeit
4. Vollständigkeit
5. Übersichtlichkeit
6. Einheitliche Darstellung
7. Bearbeitbarkeit
8. Fehlerfreiheit
9. Eindeutige Auslegbarkeit
10. Objektivität
11. Relevanz
12. Hohes Ansehen
13. Aktualität
14. Verständlichkeit
15. Wertschöpfung

Für die Dimensionen Vollständigkeit, Fehlerfreiheit, Konsistenz und Aktualität werden zudem konkrete Metriken vorgestellt. Eine Erweiterung der Dimensionen sowie entsprechende Metriken speziell im Kontext des maschinellen Lernens werden in [64] vorgestellt. Bei den erweiterten Dimensionen handelt es sich um die Genauigkeit der Merkmale, die Genauigkeit der Zielgröße, die Einzigartigkeit und die Klassenbalance. Durch eine gezielte Verschlechterung einzelner Dimensionen werden hier zudem die Auswirkungen auf die Performanz datengetriebener Modelle zur Klassifizierung, Clustering und Regression untersucht. Die detaillierte Beschreibung der Merkmale sowie die Berechnungsvorschriften können [64] entnommen werden. Da die verwendeten Datensätze nicht im industriellen Kontext, sondern hauptsächlich im Finanzwesen angesiedelt sind, müsste eine Übertragbarkeit der Dimensionen auf industrielle Daten zunächst noch untersucht werden.

Gerade im Bereich der wissenschaftlichen und akademischen Communities gewinnt das Konzept der FAIR Data immer mehr an Bedeutung, welches die Verwaltung und Nutzung von Datensätzen vereinfachen und verbessern soll. FAIR bedeutet in diesem Kontext Findable, Accessible, Interoperable und Reusable (Auffindbar, Zugänglich, Interoperabel und Wiederverwendbar) sowie 15 weitere untergeordnete Prinzipien [70]. Ihr Hauptaugenmerk liegt auf der Maschinenlesbarkeit, d. h. der Fähigkeit von Maschinen, z. B. Computern, digitale Ressourcen ohne menschliches Zutun oder nur

mit minimaler menschlicher Unterstützung zu finden, darauf zuzugreifen, zu interagieren und wiederzuverwenden. Um dies zu erreichen, werden beispielsweise neben den eigentlichen Sensordaten definierte Metadaten abgelegt, welche eine Beschreibung des Datensatzes beinhalten. Auch in industriellen Bereichen wie beispielsweise der Pharmaindustrie erhält das Konzept stetig mehr Zustimmung und hat damit das Potential, sich zu einem internationalen Standard zu entwickeln [71].

Ein ganzheitlicher Ablaufplan zur Durchführung datengetriebener Projekte im Kontext industrieller Applikationen ist in [13] gegeben. Hier werden Herausforderungen für die Durchführung solcher Projekte, wie beispielsweise unzureichende bzw. nicht im notwendigen Detaillierungsgrad aufgezeichnete Messdaten oder die fehlende interdisziplinäre Zusammenarbeit von Expertenteams genannt. Um diesen Herausforderungen zu entgegnen, soll der Modellbildungsprozess durch wissensgetriebene Ansätze gestützt und die Datenqualität mittels einer der Modellbildung vorgelagerten Mess- und Datenplanung abgesichert werden. Diese greift auf diverse Methoden zur Wissensmodellierung wie Kausalitätsgraphen, Signalflussgraphen oder Merkmalentstehungsbäume zurück und sieht eine iterative Prüfung der Datenqualität in Form von Plausibilitätsprüfungen und Visualisierungen vor.

2.2.2 Domain Shift

Eine grundlegende Voraussetzung von Methoden des maschinellen Lernens ist, dass die Trainingsdaten und die Testdaten aus derselben zugrundeliegenden Verteilung stammen [72, 73]. Gerade im industriellen Umfeld ist diese Annahme durch intrinsische Einflüsse wie variierende Maschinen und Prozesskonfigurationen oder extrinsische Einflüsse, wie sich ändernde Umweltbedingungen häufig nichtzutreffend, wodurch der breitbandige Einsatz datengetriebener Methoden massiv gehemmt wird [63, 74]. Als Beispiele für nicht bzw. nur teilweise beeinflussbare Umwelteinflüsse können hier unter anderem variierende Temperaturen, Materialchargen oder benachbarte Prozesse aufgeführt werden. Dies führt häufig dazu, dass ein im Labor entwickeltes Modell nicht erfolgreich auf Anlagen im Feld übertragen werden kann [75]. Aber auch prozessinterne Faktoren wie die konfigurierte Bearbeitungsgeschwindigkeit, der Arbeitsdruck oder die Geometrie gefertigter Produkte können Domain Shifts und damit eine Verschlechterung der Modellperformanz bis hin zur völligen Modellungültigkeit zur Folge haben [10]. Die Ursachen für Domain Shifts sind dabei nicht nur vom Prozess und den Umweltfaktoren, sondern auch von der verwendeten Sensorik abhängig. Ein Mikrofon, welches breitbandig Geräusche in einer industriellen Umgebung erfasst, ist

beispielsweise stärker von benachbarten Fertigungsprozessen betroffen als ein Drucksensor, welcher stärker auf systeminterne oder temperaturbedingte Veränderungen reagiert.

Aus diesem Grund gibt es eine Vielzahl an Arbeiten, welche sich der Lösung dieses Problems widmen und unter dem Begriff Domain Adaptation zusammengefasst werden können [72, 74, 75, 76, 77]. Die Art der verfügbaren Domain Adaptation Methoden ist dabei wiederum stark von den verwendeten Modellbildungsmethoden abhängig. In [73] ist eine Übersicht der verfügbaren Domain Adaptation Methoden gegeben. Der Fokus der Veröffentlichungen liegt insgesamt sehr stark auf ANNs vor allem aus dem Bereich des maschinellen Sehens. Ansätze beschäftigen sich unter anderem mit der Extraktion Domänen-invarianter Merkmale, der künstlichen Erweiterung des Trainingsdatensatzes (Data Augmentation) oder der Kalibrierung der Modelle [40, 77]. Viele dieser Methoden stellen allerdings wiederum überwachte Verfahren dar, welche Daten der verschiedenen Verteilungen erfordern. Bevor eine Anpassung des Modells erfolgen kann, muss also in erster Linie festgestellt werden, dass neu eintreffende Daten von einem Domain-Shift betroffen sind. Zudem ergeben sich im industriellen Umfeld eine unüberschaubare Anzahl an intrinsischen und extrinsischen Ursachen für Domain Shifts, welche nicht alle über eine künstliche Erweiterung der Trainingsdaten abdeckbar sind. Zudem konnte in [40] ein Zusammenhang zwischen der Modellkomplexität und dem Einfluss von Domain Shifts abgeleitet werden, wonach komplexe Modelle stärker unter unbekannten Einflüssen leiden. Um die Robustheit der Modelle zu erhöhen, sollte die Komplexität der verwendeten Modelle also an die Komplexität des entsprechenden Lernproblems angepasst werden.

Trotz der Entwicklung in diesem Bereich ist es teilweise nicht möglich, bestehende Modelle an die veränderten intrinsischen oder extrinsischen Einflüsse anzupassen bzw. ein ausreichend generalisiertes und robustes Modell herzuleiten, weshalb das bestehende Modell verworfen und Daten zum erneuten Training des Modells gesammelt werden müssen.

2.2.3 Design of Experiment

Besteht die Möglichkeit im Rahmen von Versuchen, Daten für das überwachte Lernen zu sammeln, ist dies speziell im Bereich der industriellen Anwendungen häufig mit hohen Kosten verbunden. Hierzu müssen entweder spezielle Versuchsanlagen aufgebaut oder Anlagen aus dem Fertigungsbereich modifiziert werden. Beim Sammeln von Daten an Versuchsaufbauten muss sichergestellt werden, dass die Umweltbedingungen identisch zu den realen Fertigungsbedingungen sind, um ein

reibungsloses Überführen des Modells ins Feld zu ermöglichen. Werden Anlagen aus der Serienfertigung für die Durchführung der Versuche verwendet, müssen die Anlagen häufig gezielt aus den optimalen Betriebsbereichen ausgelenkt werden und stehen dadurch nur eingeschränkt für die Fertigung zur Verfügung. Um Informationen über spezifische Fehlerbilder zu erhalten, werden zudem zielgerichtet Produkte mangelnder Qualität gefertigt, welche dann in der Wertschöpfungskette nicht mehr zur Verfügung stehen. Den entstehenden Kosten steht die Forderung nach einer möglichst vollständigen Abdeckung des Parameterraums aller bekannten und relevanten Einflüsse entgegen, welche im Betrieb des Modells zu erwarten sind, wodurch die Robustheit und Generalisierbarkeit des abgeleiteten Modells erhöht wird. Aus diesem Grund werden statistische Versuchspläne (engl. Design of Experiments, DoE) verwendet, um die Datengenerierung so effizient wie möglich zu gestalten [78].

In [79] ist eine Übersicht an gängigen Methoden zur Generierung eines DoEs im industriellen Kontext gegeben. Typischerweise werden bei der Versuchsplanung die zu berücksichtigen Einflüsse als Faktoren und deren Einstellung als Stufen bezeichnet. Als grundlegende Methode ist der vollfaktorielle Versuchsplan zu nennen, welcher alle möglichen Kombinationen testet. Dadurch ergibt sich der Versuchsaufwand zu [80]:

$$n_r = n_s^{n_f} \tag{2.7}$$

mit

n_r: Anzahl Versuche

n_s: Anzahl Stufen

n_f: Anzahl Faktoren

Durch den hohen Versuchsaufwand eignet sich diese Art von Versuchsplan nur bei einer geringen Anzahl an Faktoren und Stufen oder falls eine hohe Automatisierung für die Datengenerierung erreicht werden kann. Im Gegenzug eignen sich die Daten eines vollfaktoriellen Versuchsplans ausgezeichnet, um Systemwissen aufzubauen und den Effekt einzelner Faktoren sichtbar zu machen. Um die Wirkung der Faktoren getrennt voneinander untersuchen zu können, wird möglichst versucht, orthogonale Versuchspläne zu erstellen, sodass keine signifikante Korrelation zwischen zwei Fakturen und deren Stufen existiert [80]. Da sich in realen Versuchen die Flexibilität der einzelnen Faktoren stark unterscheiden kann, muss diese Information als Randbedingung für die Versuchsplanung berücksichtigt werden. Zu nennen sind hier beispielsweise aufwändige Rüstvorgänge bei Werkzeugmaschinen oder langsam einlaufende Temperatureffekte, deren häufiges Verändern die Dauer zwischen den Versuchen enorm erhöhen würde.

Ist ein vollfaktorielles DoE aufgrund des hohen Versuchsaufwands nicht möglich, bietet das Latin Hypercube Sampling (LHS) eine effiziente Alternative, um eine möglichst vollständige und homogene Abdeckung des Parameterraums zu erreichen [81]. Zur Erstellung eines LHS werden für alle Faktoren die verfügbaren Parameterbereiche definiert und diese in n gleich große Intervalle unterteilt, wobei n die Anzahl der durchzuführenden Parameterkombinationen darstellt. Im Anschluss werden aus allen Intervallen der Faktoren Stichproben gezogen, wobei es sich dabei um Zufallsstichproben oder systematische Stichproben wie beispielsweise den Intervall Mittelpunkt handeln kann. Auch die finalen Parametergruppierungen werden randomisiert bestimmt, wobei Erweiterungen des ursprünglichen LHS eine iterative Optimierung der Versuchspläne hinsichtlich reduzierter Korrelationen oder optimaler Parameterraumabdeckung ermöglichen [82, 83].

2.3 Methodenpool für das maschinelle Lernen

Die im Rahmen dieser Arbeit verwendeten Methoden sind Teil der MATLAB-basierten, open-source Toolbox zum maschinellen Lernen für zyklische Sensordaten entwickelt am Lehrstuhl für Messtechnik der Universität des Saarlandes [8, 84]. Die Toolbox stellt einen umfangreichen Methodenpool dar und arbeitet dabei in den für das maschinelle Lernen typischen Schritten der Merkmalsextraktion, Merkmalsselektion, Dimensionsreduktion und anschließender Klassifizierung bzw. Regression. In Abbildung 2-6 ist die ursprüngliche Struktur der Toolbox für die überwachte Klassifizierung dargestellt. Diese kombiniert fünf Methoden der Merkmalsextraktion (vgl. Abschnitt 2.3.2) mit drei Ansätzen zur Merkmalsselektion (vgl. Abschnitt 2.3.3). Mittels Dimensionsreduktion (vgl. Abschnitt 2.3.4) und Klassifizierung (vgl. Abschnitt 2.3.5) wird über eine 10-fache Kreuzvalidierung (vgl. Abschnitt 2.3.7) die optimale Methodenkombination für den jeweiligen Datensatz bestimmt. Die enthaltenen Methoden wurden im Rahmen einer umfassenden Untersuchung von insgesamt 14 Algorithmen zur Merkmalsextraktion und 66 Algorithmen zur Merkmalsselektion identifiziert [85]. Zugrunde liegt hier das „No Free Lunch Theorem“, welches im Kontext des maschinellen Lernens besagt, dass es keinen universellen Algorithmus gibt, der für alle möglichen Datensätze oder Probleme die beste Lösung liefert [86]. Stattdessen muss der Algorithmus immer an das spezifische Problem und die zugrunde liegende Struktur der Daten angepasst werden, um optimale Ergebnisse zu erzielen. Um ein nahezu ideales Klassifizierungsergebnis unabhängig von der Herkunft des Datensatzes zu erreichen, wurden deshalb aus den untersuchten Algorithmen sich auf theoretischer Seite ergänzende Methoden für die Merkmalsextraktion und -selektion ausgewählt.

Neben der möglichst generalisierbaren Anwendbarkeit wurde bei der Auswahl der Methoden für die Merkmalsextraktion vor allem darauf geachtet, dass sie später sensornah auf einem Mikroprozessor bzw. einer FPGA implementiert werden können. Dadurch entfällt bei der Anwendung der Algorithmen in der Zustandsüberwachung die Notwendigkeit, umfangreiche Rohdaten auf höhere Systemebenen der Prozesssteuerung transferieren zu müssen. Stattdessen werden nur noch weniger charakteristische, in der Trainingsphase identifizierte Merkmale übertragen [87].

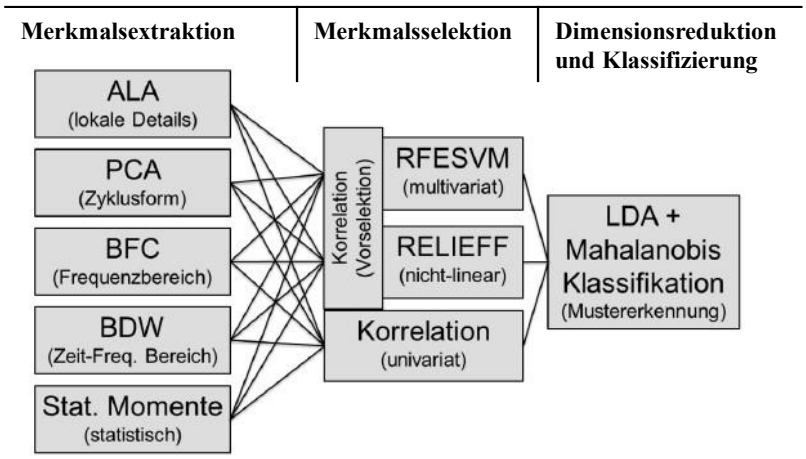


Abbildung 2-6: Vollautomatisierte Toolbox für die überwachte Klassifizierung (angepasst nach [9]).

Die Anwendung des Methodenbaukastens setzt eine bestimmte Datenstruktur voraus, wie sie typisch für Methoden des maschinellen Lernens ist. Die Sensordaten müssen in einer zyklischen Form, etwa einer Matrix

$$\boldsymbol{X} = \left(x_{i,j}\right) = \begin{pmatrix} x_{1,1} & \cdots & x_{1,N} \\ \vdots & \ddots & \vdots \\ x_{M,1} & \cdots & x_{M,N} \end{pmatrix} \tag{2.8}$$

vorliegen. Hierbei beschreibt der Laufindex i die Nummer des betrachteten Zyklus und j den Zeitpunkt bzw. die Nummer des Samples innerhalb dieses Zyklus. Liegen die Daten in Form kontinuierlicher Zeitreihen vor, kann die benötigte zyklische Struktur über eine Segmentierung der Zeitreihen erreicht werden. Hier wird deutlich, dass die Sensordaten mit einer ausreichenden Synchronisierung aufgezeichnet werden müssen, damit sich relevante Muster immer im selben Zyklusbereich befinden. Eine mangelnde Synchronisierung der Daten kann in bestimmten Fällen zu einer erheblichen Verschlechterung der Performanz des Modells führen [88]. Da es sich im Bereich der Zustandsüberwachung typischerweise um Multisensorprobleme handelt, stellt ein zentraler Punkt die Sensordatenfusion, also die Kombination der Informationen der vorhandenen Sensorsignale dar. Erreicht wird dies im Anschluss and die Merkmalsextraktion durch die Verkettung der aus einem Zyklus (also einer Zeile der Sensor-Rohdaten) extrahierten Merkmale zu Merkmalvektoren, welche dadurch

wiederum die in Formel (2.9) dargestellte Struktur aufweisen, wobei die Zeilen jetzt die sensorübergreifenden, verketteten Merkmale enthalten.

Falls teilüberwachte oder überwachte Verfahren eingesetzt werden sollen, muss ebenfalls eine Zielgröße der Form

$$\boldsymbol{y} = (y_i) = \begin{pmatrix} y_1 \\ \vdots \\ y_M \end{pmatrix} \tag{2.9}$$

vorhanden sein. Im Falle einer Klassifizierung enthält $\boldsymbol{y}$ die Klasse, welche beispielsweise den Maschinenzustand abbildet. Im Falle einer Regression beinhaltet $\boldsymbol{y}$ kontinuierliche Daten, wie beispielsweise eine Leckagerate oder den Abnutzungsvorrat.

Im Rahmen der kontinuierlichen Weiterentwicklung der Toolbox wurden zahlreiche Methoden hinzugefügt und die Funktionalität der Toolbox beispielsweise auf Problemstellungen wie Regression oder Anomaliedetektion erweitert. Durch ein einheitliches Interface lassen sich die Methoden der einzelnen Ebenen modular zusammensetzen, was eine einfache Anpassung an verschiedene Use Cases erlaubt. Die Verkettung der Methoden und eine daraus folgende schrittweisen Modellbildung bzw. Signalverarbeitung, wird im Folgenden als Stack bezeichnet. Des Weiteren bestehen umfangreiche Visualisierungsmöglichkeiten auf Basis dimensionsreduzierender Transformationen, welche im Hinblick auf eine explorative Datenanalyse eingesetzt werden können. Im Folgenden wird eine Übersicht aller zum Zeitpunkt der Anfertigung dieser Arbeit verfügbaren Methoden mit einer funktionalen Einordnung gegeben. Zudem werden die grundsätzlichen Schritte des maschinellen Lernens anhand ausgewählter und für das Verständnis dieser Arbeit relevanter Elemente des Methodenpools erläutert.

2.3.1 Signalvorverarbeitung

Eine dem maschinellen Lernen vorgelagerte Signalvorverarbeitung gilt als ein fundamentaler Bestandteil des maschinellen Lernens, der massiv die Qualität und Robustheit der Modellbildung beeinflusst [89]. Typische Methoden zur Signalvorverarbeitung umfassen unter anderem die Datenbereinigung, die das Entfernen unvollständiger Zyklen und das Herausfiltern von Ausreißern beinhaltet, sowie die Datenanpassung, wie beispielsweise die Fensterung im Falle nichtzyklischer Daten. Auch Verfahren zur Signalglättung wie beispielsweise der gleitende Mittelwert- oder der Savitzky-Golay-Filter finden häufig Anwendung [90]. Zudem kommt im Bereich der industriellen Zustandsüberwachung Standardisierungsverfahren eine erhebliche Bedeutung zu, da die Sensorsignale häufig

von verschiedenen Sensortypen stammen, deren Signale über sehr unterschiedliche Skalen verfügen [91]. Die verwendeten Methoden sind dabei abhängig vom Typ der verarbeiteten Daten, aber auch von den nachgelagerten Methoden. Gerade bei der Verwendung von Klassifikatoren basierend auf Abstandsmaßen, wie beispielsweise k-Nearest Neighbors (kNN, vgl. Kap. 2.4.6), können Merkmale mit höheren Absolutwerten die Berechnung der Abstände und somit die Klassifizierung dominieren [92]. Je nach Art der verwendeten Methoden variiert die Tiefe, mit der die Methoden in die Signalverarbeitungskette eindringen, sehr stark [93]. Eine Standardisierung kann beispielsweise auf Rohdaten oder auf bereits extrahierte Merkmale angewendet werden. Abbildung 2-7 zeigt die in der Toolbox verfügbaren Methoden zur Signalvorverarbeitung.

Abbildung 2-7: In der Toolbox verfügbare Methoden zur Signalvorverarbeitung (Standardisierung).

Im Rahmen dieser Arbeit erfolgt hauptsächlich eine Standardisierung der extrahierten Merkmale mittels Z-Score (auch Z-Transformation genannt), welche die Stichprobe um einen Mittelwert 0 zentriert und auf eine Standardabweichung von 1 skaliert.

$$x_{Z|ij} = Z\left(x_{ij}\right) = \frac{x_{ij} - \overline{x_i}}{\sigma_i} \tag{2.10}$$

x_{ij} beschreibt hierbei das j-te Sample im i-ten Zyklus, wobei $\overline{x_i}$ den Mittelwert und σ_i die Standardabweichung des gesamten Zyklus beschriebt. Zur Driftkompensation, beispielsweise im Bereich der Gassensorik, wird zudem häufig eine Mittelwertskalierung verwendet [94].

$$x_{mittel} = \frac{x_{ij}}{\overline{x_i}} \tag{2.11}$$

2.3.2 Merkmalsextraktion

Industrielle Anwendungsfälle wie die Zustandsüberwachung oder die Prozessoptimierung basieren häufig auf Zeitreihendaten, die von einer Vielzahl von Sensoren mit unterschiedlichen Abtastraten stammen [95]. Abhängig von der Abtastrate und der Anzahl der Sensoren, die für die Analyse verwendet werden, ist die Dimensionalität der Daten deutlich zu hoch, um sie direkt mit herkömmlichen

Klassifizierungs- oder Regressionsalgorithmen zu verarbeiten. Daher werden Algorithmen zur Merkmalsextraktion eingesetzt, um die Informationen aus den verfügbaren Signalen hervorzuheben und zu komprimieren [96]. Je nach Anwendungsfall und Art der verfügbaren Signale können die Informationen im Zeit-, Frequenz- oder Zeitfrequenzbereich erwartet werden. So werden beispielsweise Schwingungssensoren, die häufig bei der Zustandsüberwachung rotierender Maschinen eingesetzt werden, vorzugsweise im Frequenzbereich ausgewertet, während nichtperiodische Signale hauptsächlich im Zeitbereich ausgewertet werden [97]. Abbildung 2-8 zeigt die in der Toolbox verfügbaren Merkmalsextraktionsmethoden. Im Folgenden werden die einzelnen Merkmalsextraktionsmethoden vorgestellt.

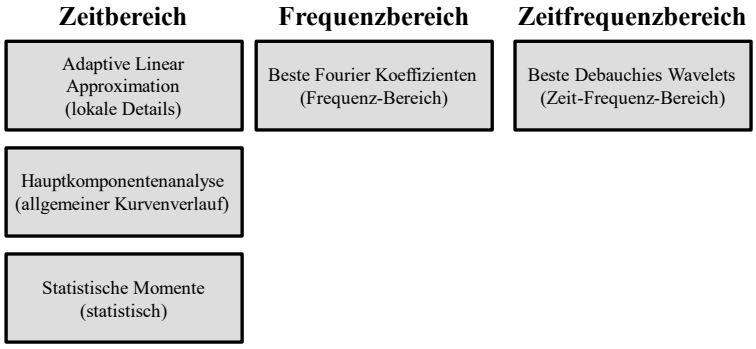


Abbildung 2-8: In der Toolbox verfügbare Methoden zur Merkmalsextraktion.

2.3.2.1 Merkmale im Zeitbereich

Statistische und formbeschreibende Merkmale haben sich als wirkungsvolle Merkmale im Bereich der Zustandsüberwachung erwiesen [8, 11, 98, 99]. Ein verbreitetes Merkmal stellt dabei der Median dar. Der Median ist ein statistisches Maß, das die Mitte einer nach der Größe geordneten Stichprobe mit n Elementen repräsentiert, wobei 50 % der Werte unterhalb und 50 % oberhalb des Medians liegen. Bei einer ungeraden Anzahl von Datenpunkten ist der Median der mittlere Wert, bei einer geraden Anzahl wird er als Mittelwert der beiden mittleren Werte berechnet.

$$Median(x) = \begin{cases} x_{\frac{n+1}{2}} \ \text{für } n \ ungerade \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) \ \text{für } n \ gerade \end{cases} \tag{2.12}$$

Im Fall der Schwingungsanalyse mittels piezoelektrischer Beschleunigungssensoren stellen die Signale mittelwertfreie Oszillationen dar, weshalb der Median an dieser Stelle nicht als aussagekräftiges Merkmal eingestuft werden kann. Ein vergleichbares Merkmal, welches auch im Fall von Schwingungen verwendet werden kann, ist der nach (2.13) berechnete Effektivwert oder auch Root Mean Square (RMS).

$$x_{RMS} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x_n^2} \tag{2.13}$$

Die Toolbox nutzt zudem die höheren zentralen statistischen Momente Varianz (2. Ordnung), Schiefe (3. Ordnung) und Kurtosis (4. Ordnung) als einfache, aber effektive Merkmale, welche je nach Konfiguration des Extraktors aus dem gesamten oder aus Teilbereichen des Zyklus berechnet werden. Die zentralen statistischen Momente berechnen sich wie folgt:

$$Varianz(x) = \sigma^2(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{2.14}$$

$$Schiefe(x) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{\sigma}\right)^3 \tag{2.15}$$

$$Kurtosis(x) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{\sigma}\right)^4 \tag{2.16}$$

Bei den zentralen Momenten beschreibt x_i den i-ten Datenpunkt, $\overline{x}$ das arithmetische Mittel und σ die Standardabweichung der betrachteten Stichprobe. Speziell die Kurtosis stellt eine bekannte Kennzahl für die Wälzlagerdiagnose dar. Veranschaulicht beschreibt sie in dem Fall eine spitzer werdende Verteilung bei einer Zunahme von schadensbedingten Stoßimpulsen, beispielsweise in dem Signal eines Beschleunigungssensors. Eine ähnliche Kennzahl stellt der sogenannte Crest-Faktor da. Er beschreibt entsprechend Gleichung (2.17) das Verhältnis der maximal auftretenden Amplitude $\hat{a}$ (auch Spitzenwert genannt) zum RMS a_{RMS} und findet ebenfalls bei der Überwachung von Wälzlagern Anwendung.

$$C_{rest} = \frac{\hat{a}}{a_{RMS}} \tag{2.17}$$

Die hier zugrundeliegende Theorie ist ebenfalls die Erzeugung von Stoßimpulsen beim Überrollen einer schadhaften Stelle oder die Änderung der Signalcharakteristik durch spielbedingte Schläge von Maschinenkomponenten [100].

Neben statistischen Kennzahlen finden ebenfalls formbeschreibende Merkmale Anwendung. Als einfache Vertreter sind hier z.B. die auftretenden Extremwerte und deren Positionen im Zyklus bzw. Spektrum zu nennen. Gerade im Bereich von transienten Vorgängen können weitere formbeschreibende Merkmale, welche die Steigung eines bestimmten Signalabschnitts widerspiegeln, von Interesse sein. Hierzu wird ein bestimmter Signalabschnitt mittels eines Regressionsmodells 1-ter Ordnung abgebildet. Die Steigung der bestimmten Gerade fungiert als Merkmal.

Eine ähnliche Herangehensweise stellt die Merkmalsextraktion mittels adaptiver linearer Approximation dar. Diese unterteilt die Zyklen in diskrete Segmente variabler

Länge und approximiert jedes resultierende Segment mit einer linearen Funktion, die durch Mittelwert und Steigung dargestellt wird. Die beste Segmentierung wird mit dem in [101] beschriebenen Algorithmus zur Minimierung des quadratischen Approximationsfehlers zwischen Originalsignal und segmentweiser Approximation für eine bestimmte Anzahl von Segmenten gefunden. Durch die abschnittsweise Approximation des Signals ermöglicht diese Methode eine hervorragende Abbildung lokaler Details [9].

Im Gegensatz dazu eignet sich die Hauptkomponentenanalyse (engl. principal component analysis, PCA) zur Extraktion von Merkmalen, falls die Information in der allgemeinen Zyklusform enthalten ist. Die PCA stellt die beste lineare Transformation der Daten in Bezug auf den Approximationsfehler sowie eine Zerlegung des Signals in lineare unabhängige und orthogonale Komponenten dar. Dabei erklären die ersten Komponenten die meiste Varianz, d. h. sie erfassen die dominantesten Einflüsse, während höhere Komponenten oft nur zufällige Schwankungen, d. h. Rauschen, enthalten [102]. Im Rahmen dieser Arbeit wird die PCA hauptsächlich zur Dimensionsreduktion eingesetzt, weshalb sie in Abschnitt 2.3.4.1 noch einmal detailliert erläutert wird.

2.3.2.2 Merkmale im Frequenzbereich

Im Bereich der Zustandsüberwachung etablierte Messgrößen wie Beschleunigung, Luftschall und Motorstrom stellen Schwingungen dar. Folglich erfassen die entsprechenden Sensoren periodische Signale, welche sich hervorragend für eine Auswertung im Frequenzbereich eignen. Grundlage dafür stellt die Fourier-Transformation dar, welche die Signale vom Zeit- in den Frequenzbereich überführt. Vibrationen an Maschinen setzen sich aus einer Überlagerung vieler Schwingungen verschiedener Amplituden und Frequenzen zusammen. Um die verschiedenen Ursachen von Schwingungen im Rahmen der Maschinendiagnose bewerten und identifizieren zu können, ist es notwendig, den Frequenzinhalt der Signale zu analysieren. Dazu wird das Zeitsignal relevanter Sensoren mithilfe der Fourier-Transformation in den Frequenzbereich transformiert. Für kontinuierliche Signale sind die Transformation und die inverse Transformation wie folgt definiert [103]:

$$F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-j\omega t} dt \tag{2.18}$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(t) e^{j\omega t} d\omega \tag{2.19}$$

j: komplexe Einheit

ω: Kreisfrequenz

Das Ergebnis der Fourier-Transformation einer kontinuierlichen Zeitfunktion $f(t)$ ist die komplexe Funktion $F(\omega)$, welche in Betrags- und Phasendarstellung vorliegt. In der Realität werden Signale mithilfe eines Analog-Digital-Wandlers digitalisiert und liegen demnach nie in kontinuierlicher Form vor. Für die durch Digitalisierung entstehenden zeit- und wertdiskreten Signale muss die diskrete Fourier-Transformation (DFT) angewendet werden. Diese ist folgendermaßen definiert [104]:

$$F_j = \frac{1}{N} \sum_{k=0}^{N-1} f_k \, {W_N}^{-kj} \tag{2.20}$$

mit

$W_N = e^{\frac{2\pi i}{N}}$

N: Anzahl der Abtastwerte

Der Berechnungsaufwand für diese Transformation beträgt aufgrund der Vektor-Matrix-Multiplikation N^2. Dies kann im Kontext der Zustandsüberwachung, bei der tendenziell große Datenmengen verarbeitet werden müssen, zu sehr langen Berechnungszeiten führen. Um dieses Problem zu lösen, entwickelten J.W. Cooley und John Tukey die schnelle Fourier-Transformation (Fast Fourier-Transformation, FFT), einen Algorithmus zur effizienteren Berechnung der DFT. Dieser reduziert den Berechnungsaufwand auf $N\ log(N)$ und hat sich somit als Standard in der modernen Signalverarbeitung etabliert. Ursprüngliche Voraussetzungen für die Berechnung der FFT, wie etwa die Notwendigkeit, dass die Anzahl der verwendeten Stützstellen eine Zweierpotenz ist, werden von modernen Algorithmen, wie sie beispielsweise in Matlab implementiert sind, umgangen und können daher vernachlässigt werden.

Durch die Transformation diskreter Zeitsignale in den Frequenzbereich liegen die resultierenden Spektren ebenfalls als eine Folge diskreter Werte vor. Um bei den im Verlauf dieser Arbeit untersuchten Spektren benachbarte Frequenzen ausreichend gut separieren zu können, spielt die Auflösung der Frequenzen eine wichtige Rolle. Diese ist nach [104] wie folgt definiert:

$$\triangle f = \frac{f_s}{N_s} \tag{2.21}$$

mit

f_s: Abtastrate

N_s: Anzahl Samples

Neben der erzielten Frequenzauflösung ist die maximal zu beobachtende Frequenz im Spektrum von entscheidender Bedeutung. Diese ergibt sich aus dem Nyquist-Shannon-Abtasttheorem als die Hälfte der Abtastfrequenz. Die Ergebnisse der FFT werden in Form von Frequenzspektren dargestellt. Hierbei werden in einem Koordinatensystem der Amplitudengang auf einer Ordinate und die Frequenz auf der Abszisse aufgetragen. Analog gilt dies für die Phasenspektren.

Merkmale, die aus den Spektren berechnet werden, reichen von Kennzahlen wie beispielsweise der Schwingungsenergie, welche breitbandig aus dem Spektrum berechnet wird, bis hin zu frequenzselektiven Kennzahlen, welche auf der vorherigen Bestimmung prozesscharakteristischer Frequenzen beruhen [105]. Im Falle der vorgestellten Toolbox erfolgt die Extraktion der Fourier-Koeffizienten des Amplituden- und Phasenspektrums, wobei 10% der Koeffizienten mit dem höchsten durchschnittlichen Betrag als Merkmale verwendet werden.

Ändert sich die zugrundeliegende Kinematik des zu bewertenden Prozesses beispielsweise durch drehzahlvariable Antriebe, verlagern sich auch die informationstragenden, drehsynchronen Frequenzen im Spektrum, was eine Auswertung auf Basis herkömmlicher Spektren erschwert. In solchen Fällen hat sich die Verwendung von Ordnungsspektren als zielführend erwiesen [106, 107]. Diese werden erzeugt, indem die Frequenzachse der Spektren auf die aktuelle Drehzahl der antreibenden Welle normiert wird, was also die Erfassung der Drehzahl voraussetzt. Dadurch bleiben die Ordnungen auch bei variabler Drehzahl im Ordnungsspektrum entlang der Ordinate fixiert.

2.3.2.3 Merkmale im Zeitfrequenzbereich

Gerade bei nichtstationären Prozessen ist der Verlauf von mehreren zeitlich aufeinanderfolgenden Spektren von Interesse, um Drehzahlveränderungen bzw. Änderungen des Frequenzinhalts über die Zeit zu beobachten. Eine Methode, welche dies ermöglicht, ist die Kurzeit-Fourier-Transformation. Diese multipliziert das Zeitsignal $x(t)$ mit einem um τ verschiebbaren Analysefenster $\gamma(t-\tau)$ und berechnet die Fourier-Transformation für das gefensterte Signal [108]:

$$F(\tau,\omega) = \int_{-\infty}^{+\infty} x(t)\,\gamma(t-\tau)e^{-j\omega t}dt \tag{2.22}$$

mit:

j : komplexe Einheit

ω: Kreisfrequenz

Außerhalb des gefensterten Bereichs wird das Zeitsignal unterdrückt, wodurch ein lokales Spektrum entsteht. Wurde das Fenster über das gesamte Zeitsignal verschoben, erfolgt die Darstellung häufig in Form des Betragsquadrates, welches auch als Spektrogramm bezeichnet wird. Hierbei werden in einem Koordinatensystem der Amplitudengang auf einer Ordinate und die Frequenz auf der Abszisse und die Zeitachse in einer dritten Dimension aufgetragen. Ein Problem der STFT stellt der Trade-off zwischen Zeit- und Frequenzauflösung dar. Bei einer Wahl großer zeitlicher Fenster ergibt sich nach Formel (2.21) eine gute Auflösung im Frequenzbereich, jedoch eine schlechte zeitliche Auflösung von Frequenzänderungen und umgekehrt. Die definierte Frequenzauflösung bleibt dabei über die gesamte Bandbreite konstant.

Ein mittlerweile weit verbreiteter Ansatz zur Merkmalsextraktion und Signalanalyse im Zeitfrequenzbereich ist die Wavelet-Transformation [109, 110]. Im Gegensatz zur STFT erlaubt sie eine hohe Auflösung im Frequenzbereich und eine niedrige Auflösung im Zeitbereich für niederfrequente Signale und umgekehrt für hochfrequente Signale. Damit eignet sich die Wavelet-Transformation besser für die Analyse von Signalen mit dynamischen Änderungen des Frequenzinhalts. Die kontinuierliche Wavelet-Transformation einer Funktion $f(t)$ ist wie folgt definiert:

$$W_\psi f(a,b) := |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(t)\ \psi\left(\frac{t-b}{a}\right) dt \tag{2.23}$$

Hierbei beschreibt

$$\psi_{ab}(t) = |a|^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right) \tag{2.24}$$

das zugrundeliegende Wavelet, welches mittels a skaliert und b verschoben wird. Die Merkmalsextraktionsmethode in der Toolbox basiert auf der diskreten Wavelet-Transformation, was eine effiziente Implementierung in Form einer Filterbank ermöglicht [111]. Bei dem verwendeten Wavelet handelt es sich um das Daubechies-4-Wavelet, wobei wiederum die 10% der Koeffizienten, welche den höchsten durchschnittlichen Betrag aufweisen, als Merkmale verwendet werden [9].

2.3.3 Merkmalsselektion

Abhängig von der verwendeten Merkmalsextraktionsmethode kann eine sehr große Anzahl von Merkmalen entstehen, welche immer noch eine zu hohe Dimensionalität für die direkte Verarbeitung durch einen Klassifizierungs- oder Regressionsalgorithmus aufweist. Zudem beinhaltet nicht jedes Merkmal relevante Informationen bzw. beinhalten mehrere Merkmale dieselben oder zumindest redundante Informationen. Aus diesem Grund werden

Merkmalsselektionsalgorithmen verwendet, um ein Merkmalsranking, basierend auf einem Score zu erstellen, welcher eine möglichst optimale Merkmalszusammenstellung in Bezug auf Anzahl und Interaktion der Merkmale ermöglicht [112]. Grundsätzlich kann unterschieden werden zwischen einer manuellen Merkmalsselektion, welche auf Expertenwissen oder einer vorhergehenden Signalanalyse beruht, und einer automatisierten Merkmalsselektion [113]. Die in der Literatur bekannten Methoden zur automatisierten Merkmalsextraktion werden typischerweise in Filter-, Embedded- und Wrapper-Methoden unterteilt [112, 114], [115]. Das grundsätzliche Vorgehen der Methoden ist in Abbildung 2-9 dargestellt.

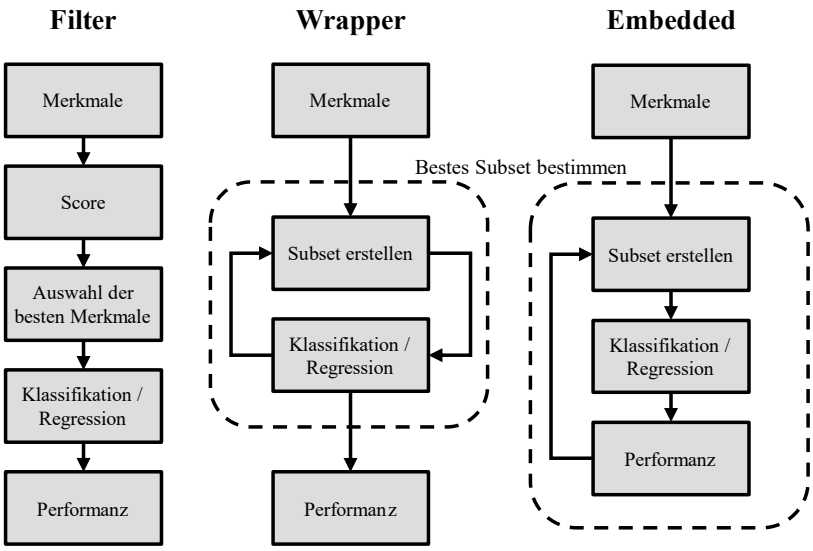


Abbildung 2-9: Grundsätzlicher Ablauf der Merkmalsselektion für die verfügbaren Typen Filter, Wrapper und Embedded, angepasst nach [116].

Filtermethoden bewerten die Relevanz von Merkmalen unabhängig vom nachfolgenden Modell anhand statistischer Metriken bzw. Scores wie beispielsweise Korrelation, Distanz oder Information. Sie bieten eine schnelle und sehr gut skalierbare Implementierung, wobei die Merkmalsinteraktion ungeachtet der nachfolgenden Modellbildungsmethoden beurteilt wird. Univariate Merkmalsfilter bewerten (und ordnen) in der Regel ein einzelnes Merkmal, während multivariate Filter eine ganze Merkmalsteilmenge auswerten und somit zumindest einfache Merkmalsinteraktionen berücksichtigt werden [117].

Wrapper-Methoden hingegen stellen iterative Verfahren dar, bei denen ein Klassifikator oder Regressor auf Merkmalssubsets trainiert und für die Auswahl eines optimalen Merkmalssubsets verwendet wird. Durch das wiederholte Trainieren eines Modells ergibt sich ein hoher Berechnungsaufwand, wobei hier wiederum eine starke Abhängigkeit vom verwendeten Klassifizierungs- oder Regressionsalgorithmus besteht. Zudem sind die gefundenen Merkmale auf die zur Selektion verwendeten Algorithmen optimiert. Die Anpassung des Subsets nach jedem Trainingsvorgang

erfolgt anhand spezifischer Suchstrategien, wie beispielsweise der Forward Selection oder der Backward Elimination [117]. Während bei der Forward Selection mit einem anfangs leeren Merkmalspool gestartet wird, welcher dann über die Iterationen mit immer mehr Merkmalen befüllt wird, startet die Backward Elimination mit allen Merkmalen und verwirft iterativ Merkmale, welche keinen nennenswerten Beitrag zur Performanz des Modells beitragen [118]. Obwohl Wrapper-Methoden hinsichtlich der Qualität der selektierten Merkmale und der berücksichtigten Merkmalsinteraktion anderen Methoden überlegen sind, finden sie aufgrund des hohen Berechnungsaufwands in der Praxis nur bedingt Anwendung.

Bei Embedded Methoden ist der Selektionsprozess direkt in die Modellbildung integriert und die Auswertung der Modellperformanz wird zur Auswahl des optimalen Merkmalssubsets verwendet. Embedded Methoden sind nicht in der Toolbox implementiert, weshalb an dieser Stelle auf eine ausführlichere Beschreibung verzichtet wird. Als Beispiele für Embedded Methoden können aber Least Absolute Shrinkage and Selection Operator (LASSO) [119] oder Ridge [120] genannt werden.

In Abbildung 2-10 sind die in der Toolbox verfügbaren Methoden zur Merkmalsselektion abgebildet. Bei den implementierten Selektionsalgorithmen handelt es sich um Ranking Algorithmen, welche die Merkmale unter Verwendung entsprechender Metriken bezüglich ihrer Relevanz bewerten und sortieren. Die Bestimmung der optimalen Anzahl an Merkmalen erfolgt in Kombination mit den in Abschnitt 2.3.8 vorgestellten Verfahren zur Hyperparameteroptimierung.

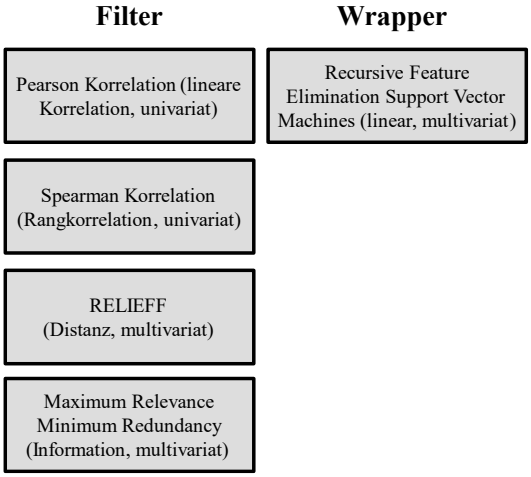


Abbildung 2-10: In der Toolbox verfügbare Methoden zur Merkmalsselektion.

Im Rahmen dieser Arbeit finden vor allem die Merkmalsselektion mittels Pearson Korrelationskoeffizient und Maximum Relevance Minimum Redundancy (MRMR) Anwendung, weshalb diese im Folgenden detaillierter erläutert werden. Letztere wurde der Toolbox im Rahmen dieser Arbeit hinzugefügt. Informationen über

RELIEFF und Recursive Feature Elimination Support Vector Machines können [121] und [122] entnommen werden.

2.3.3.1 Pearson Korrelationskoeffizient

Der Pearson Korrelationskoeffizient r beschreibt die Stärke und die Richtung der linearen Beziehung zwischen zwei Variablen X und Y und berechnet sich über die Kovarianz und die Standardabweichung beider Variablen.

$$r = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}} = \frac{\sum_{i=1}^{M}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{M}(x_i - \bar{x})^2 \sum_{i=1}^{M}(y_i - \bar{y})^2}} \tag{2.25}$$

Ein positiver Wert von r bedeutet eine positive Korrelation, ein negativer Wert eine negative Korrelation (gegenläufig). Ein Wert nahe Null bedeutet, dass keine lineare Beziehung zwischen den beiden Variablen vorliegt. Da lediglich die Stärke der Korrelation für eine Merkmalsselektion entscheidend ist, wird bei der Implementierung der Betrag von r als Kriterium für die Relevanz des Merkmals verwendet. Da ausschließlich die Korrelation zur Zielgröße betrachtet wird, handelt es sich um eine univariate Merkmalsselektionsmethode. Aufgrund der geringen Berechnungskomplexität wird die Pearson Selektion, wie in Abbildung 2-6 dargestellt, für eine Vorselektion, also eine Begrenzung des Merkmalspools vor der eigentlichen Merkmalsselektion, eingesetzt. Als Standard werden hierbei die 500 höchstkorrelierten Merkmale vorselektiert, um den Berechnungsaufwand der folgenden rechenintensiven Merkmalsselektionsmethoden einzugrenzen. Speziell in Verbindung mit Merkmalsextraktoren, die eine große Zahl an redundanten Merkmalen extrahieren, wie es bei dem in Abschnitt 3.2.1 vorgestellten Time-Frequency-Extractor der Fall sein kann, führt eine zu starke Begrenzung der Merkmalspools mittels Pearson Korrelation gegebenenfalls zu wenig komplementärer Information. Aus diesem Grund wird die Pearson Korrelation im Rahmen dieser Arbeit lediglich als Vorselektion verwendet, gefolgt von der multivariaten Filter Methode MRMR, welche im Folgenden vorgestellt wird.

2.3.3.2 Maximum Relevance Minimum Redundancy

MRMR-Methoden bestehen im Allgemeinen aus einer Metrik zur Beurteilung der Beziehung einer Variable x zu der Zielgröße y (Relevanz, engl. Relevance) sowie einer Metrik zur Beurteilung der Beziehungen zwischen Paaren von Variablen (Redundanz, engl. Redundancy). Für die Berechnung von Redundanz und Relevanz existieren mittlerweile umfangreiche Sammlungen von Metriken [123]. Im Rahmen dieser Arbeit werden die ursprünglich in [124] veröffentlichte und in MATLAB

implementierte Variante, basierend auf der Transinformation (auch gegenseitige Information, engl. Mutual Information) verwendet, welche im Folgenden vorgestellt wird.

Für diskrete/kategoriale Variablen wird die Transinformation von zwei Variablen x und y auf der Grundlage ihrer gemeinsamen Wahrscheinlichkeitsverteilung $p(x, y_j)$ und der jeweiligen Randwahrscheinlichkeiten $p(x_i)$ und $p(y_j)$ wie folgt berechnet:

$$I(x,y) = \sum_{i,j} p(x_i, y_j) \log\left(\frac{p(x_i, y_j)}{p(x_i)p(y_j)}\right) \tag{2.26}$$

Wenn x und y unabhängig sind, wird I zu 0. Wenn x und y die gleiche Zufallsvariable sind, dann entspricht I der Entropie von x [125]. Um die Berechnung der Transinformation auf kontinuierliche Methoden auszuweiten, diskretisiert die in MATLAB implementierte Variante die kontinuierlichen Variablen in 256 Bins. Der in [126] vorgestellte adaptive Algorithmus findet dabei optimale bivariate Bins für jedes Variablenpaar [127].

Das Ziel der MRMR-Methode ist also, ein optimales Subset S an Merkmalen zu identifizieren, welches die Transinformation V_s zwischen Merkmalen und Zielgröße im Subset maximiert (Relevanz) und gleichzeitig die Transinformation W_s zwischen den Merkmalen im Subset minimiert (Redundanz). Diese sind definiert durch:

$$V_s = \frac{1}{|S|} \sum_{x \in S} I(x,y) \tag{2.27}$$

$$W_S = \frac{1}{|S|^2} \sum_{x,z \in S} I(x,z) \tag{2.28}$$

Hier beschreibt $|S|$ die Anzahl der Merkmale im Subset. Die Auswahl der Merkmale basiert dann auf einer Forward Selection, in dem im ersten Schritt das Merkmal mit der höchsten Relevanz V_x zur Zielgröße aus dem gesamten Merkmalspool Ω ausgewählt und in das Subset S eingefügt wird.

$$\max_{x \in \Omega} V_x = I(x,y) \tag{2.29}$$

Im Anschluss werden Merkmale basierend auf dem in Formel (2.30) dargestellten Mutual Information Quotient (MIQ_x) zum Subset S hinzugefügt.

$$MIQ_x = \frac{V_x}{W_x} \tag{2.30}$$

Hierbei beschreibt W_x die Redundanz des betrachteten Merkmals zum bereits ausgewählten Merkmalssubset S und ist wie folgt definiert:

$$W_x = \frac{1}{|S|} \sum_{z \in S} I(x, z) \tag{2.31}$$

Je höher also der MIQ eines Merkmals, desto höher ist die Qualität des Merkmals hinsichtlich der enthaltenen, komplementären Information. Der Selektionsprozess ist beendet, wenn alle Merkmale entsprechend des MIQ sortiert und dem Merkmalssubset S hinzugefügt wurden.

2.3.4 Dimensionsreduktion

Auch nach der Merkmalsselektion ist die Dimensionalität des Merkmalsraums häufig noch zu hoch, um die Merkmale direkt mittels eines Klassifizierungs- oder Regressionsalgorithmus zu verarbeiten. Aus diesem Grund werden Methoden der Dimensionsreduktion verwendet, um die Dimensionalität des Merkmalsraums weiter zu verringern. Zusätzlich erlauben dimensionsreduzierende Verfahren eine kompakte Visualisierung hochdimensionaler Daten, verringern die Komplexität und erhöhen die Interpretierbarkeit der folgenden Modellbildungsschritte [128, 129]. In Abbildung 2-11 sind die in der Toolbox verfügbaren Methoden zur Dimensionsreduktion dargestellt und werden im Folgenden erläutert.

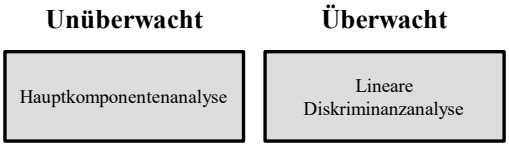


Abbildung 2-11: In der Toolbox verfügbare Methoden zur Dimensionsreduktion.

2.3.4.1 Hauptkomponentenanalyse

Ein Standardverfahren zur unüberwachten Dimensionsreduktion stellt die Hauptkomponentenanalyse (eng. principal component analysis, PCA) dar [130]. Ihr Ziel ist es, die Dimensionen der untersuchten Daten zu verringern, während ein Großteil der in den Daten enthaltenen Informationen, welche in diesem Fall mit der Varianz gleichgesetzt ist, erhalten bleibt. Dies geschieht durch das Überführen der Daten in ein neues Koordinatensystem, welches mittels sogenannter Hauptkomponenten (engl. principal component, PC) gebildet wird [131]. Die Hauptkomponenten stellen die Eigenvektoren der Kovarianzmatrix der Daten dar, während die Eigenwerte die Varianz der Daten entlang der neuen Hauptachsen beschreiben. Die Hauptkomponenten sind orthogonal zueinander und werden sinkend nach der erklärten Varianz sortiert. Eine detaillierte mathematische Beschreibung der PCA kann [132], [133] entnommen werden.

Die eigentliche Dimensionsreduktion entsteht durch das Verwerfen höherer Komponenten und kann damit als verlustbehaftet bezeichnet werden. Durch die Konzentration der erklärten Varianz in den ersten Hauptkomponenten eignet sich die PCA hervorragend zur Visualisierung hochdimensionaler Daten in Form von 2- oder 3-dimensionalen Streudiagrammen. Durch das Gleichsetzen von Information mit der in den Daten enthaltenen Varianz besteht allerdings die Gefahr, Informationen, die in höheren Hauptkomponenten enthalten sind, zu verwerfen bzw. nicht zu entdecken [134]. Die PCA eignet sich aus diesem Grund hauptsächlich zur Analyse dominanter Effekte in Datensätzen. Um zu verhindern, dass Variablen mit großen Skalen die PCA dominieren, sollte vor der PCA eine Standardisierung beispielsweise mittels des in Abschnitt 2.3.1 vorgestellten Z-Scores erfolgen.

2.3.4.2 Lineare Diskriminanzanalyse

Die Lineare Diskriminanzanalyse (LDA) ist eine multivariate Methode zur Dimensionsreduktion, welche erstmals 1936 von Fisher vorgestellt wurde [135]. Im Gegensatz zu der gerade vorgestellten Hauptkomponentenanalyse handelt es sich bei der LDA um eine überwachte Methode, weshalb für die Anwendung der LDA ein Vektor $\boldsymbol{y}$ mit den bekannten Klassenzugehörigkeiten der Daten vorhanden sein muss. Dazu werden die Daten von dem ursprünglich hochdimensionalen Merkmalsraum in einen niederdimensionalen Diskriminanzraum auf Basis orthogonaler Diskriminanzfunktionen (DF) überführt [136]. Die zu bestimmenden Diskriminanzfunktionen sind Linearkombinationen der Merkmale und haben allgemein die Form:

$$\boldsymbol{y} = \boldsymbol{W} * \boldsymbol{x} \qquad 2.32$$

bzw.:

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} w_{10} & w_{11} & \cdots & \cdots & w_{1n} \\ w_{20} & w_{21} & & & \vdots \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ w_{m0} & w_{m1} & \dots & \dots & w_{mn} \end{pmatrix} * \begin{pmatrix} 1 \\ x_1 \\ \dots \\ \dots \\ x_n \end{pmatrix} \qquad 2.33$$

Hierbei ist $\boldsymbol{y}$ der Diskriminanzvektor, welcher die Koordinaten der projizierten Daten im Diskriminanzraum enthält, während $\boldsymbol{x}$ ist der Merkmalsvektor darstellt. Die Matrix $\boldsymbol{W}$ enthält die Diskriminanzkoeffizienten, welche so zu optimieren sind, dass die Streuung innerhalb einer Klasse minimiert und die Streuung zwischen den Klassen maximiert wird. Dies erfolgt über die Maximierung des Optimierungskriteriums $J(\boldsymbol{W})$, welches auch als Diskriminanz- oder Fisher-Kriterium bezeichnet wird [137].

$$J(\boldsymbol{W}) = \frac{\boldsymbol{W}^T \boldsymbol{S}_b \boldsymbol{W}}{\boldsymbol{W}^T \boldsymbol{S}_w \boldsymbol{W}} \tag{2.34}$$

Die Streumatrizen S_b (zwischen, engl. between) und S_w (innerhalb, engl. within) sind dabei wie folgt definiert:

$$\boldsymbol{S}_b = \sum_{g=1}^{G} N_g (\overline{\boldsymbol{x}}_g - \overline{\boldsymbol{x}})(\overline{\boldsymbol{x}}_g - \overline{\boldsymbol{x}})^{\boldsymbol{T}} \tag{2.35}$$

$$\boldsymbol{S}_w = \sum_{g=1}^{G} \sum_{n=1}^{N_g} (\overline{\boldsymbol{x}}_{gn} - \overline{\boldsymbol{x}}_g)(\overline{\boldsymbol{x}}_{gn} - \overline{\boldsymbol{x}}_g)^{\boldsymbol{T}} \tag{2.36}$$

mit

G: Anzahl der Klassen

N_g: Anzahl der Stichproben der Klasse g

$\overline{\boldsymbol{x}}_g$: Mittelwertvektor der Klasse g

$\overline{\boldsymbol{x}}$: Globaler Mittelwertvektor

$\overline{\boldsymbol{x}}_{gn}$: Merkmalsvektor der Klasse g

Die Optimierung der Projektionsmatrix $\boldsymbol{W}$ kann als Eigenwertproblem formuliert und dadurch effizient gelöst werden [138]. Die Dimensionalität des Diskriminanzraums ergibt sich dabei zu G-1, wodurch im Gegensatz zur PCA in der Regel eine tatsächliche Dimensionsreduktion erreicht wird. Analog zur PCA werden die Ergebnisse der LDA typischerweise in 2- oder 3-dimensionalen Streudiagrammen dargestellt, was eine Validierung und Plausibilisierung der hergeleiteten Modelle ermöglicht [8].

Bei der LDA handelt es sich um eine parametrische Methode, weshalb bei der Anwendung gewisse Annahmen bezüglich der zugrundeliegenden Daten getroffen werden. Vorausgesetzt ist beispielsweise eine multivariate Normalverteilung der Daten, für deren Nachweis nur bedingt praktikable Methoden existieren [139]. Empirische Untersuchungen zeigen jedoch, dass eine Abweichung von dieser Annahme nur geringen Einfluss auf die Ergebnisse der LDA haben [140]. Zudem bestehen Anforderungen bezüglich der Unkorreliertheit der zugrundeliegenden Variablen, was bei Nichtbeachtung zu numerischer Instabilität aufgrund einer Singularität bei der Berechnung der inversen Kovarianzmatrix führen kann [140]. Speziell in Kombination mit der in Abschnitt 3.2.1 vorgestellten Merkmalsextraktion mittels TFEx kann dies bei stationären Prozessen zu Komplikationen führen, weshalb dieser in Kombination mit der LDA immer mit einer multivariaten Merkmalsselektion verwendet werden sollte.

Die unterschiedlichen Ziele der beiden vorgestellten dimensionsreduzierenden Verfahren, also die Projektion der Richtung größter Varianz für die PCA und die Projektion in Richtung bestmöglicher Diskriminanz für die LDA, sind in Abbildung 2-12 dargestellt.

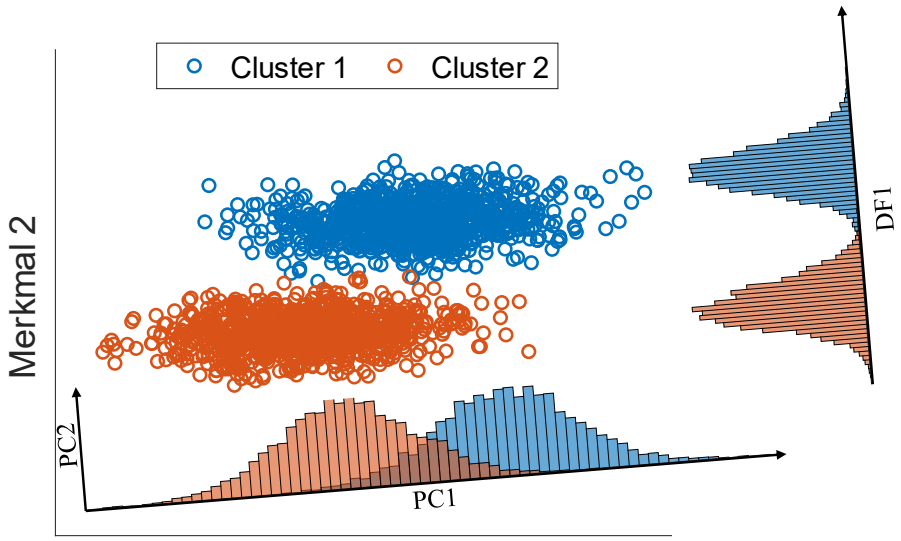


Abbildung 2-12: Gegenüberstellung der Projektionsrichtungen der beiden vorgestellten dimensionsreduzierenden Verfahren PCA und LDA anhand eines Testdatensatzes mit zwei Clustern bzw. Gruppen.

2.3.5 Klassifizierung und Regression

Beim überwachten Lernen stellen Klassifizierungs- oder Regressionsalgorithmen das Ende der Signalverarbeitungskette dar. Unabhängig, ob das betrachtete Lernproblem als Klassifizierung oder Regression formuliert wird, bestehen eine Vielzahl an Methoden, welche abhängig vom Typ und der Komplexität des betrachteten Datensatzes, den verfügbaren Rechenressourcen sowie weiteren Randbedingungen wie Transparenz, Verfügbarkeit von Expertenwissen oder Echtzeitfähigkeit ausgewählt werden [141]. Zu gängigen Algorithmen im Bereich der Zustandsüberwachung zählen dabei Abstandsmaße, Entscheidungsbäume, Support Vector Machines und künstliche neuronale Netze [142, 143, 144, 145, 146, 147, 148]. In der Toolbox sind, wie in Abbildung 2-13 dargestellt, jeweils zwei Methoden für die Klassifizierung und Regression integriert.

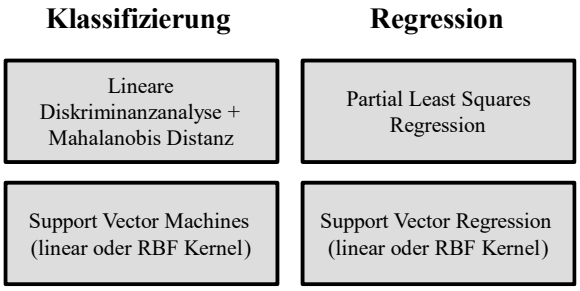


Abbildung 2-13: In der Toolbox verfügbare Methoden zur Klassifizierung und Regression.

Die LDA in Kombination mit der Mahalanobis-Distanz als auch die Partial Least Squares Regression heben sich vor allem durch eine geringe Komplexität sowie einer daraus resultierenden kurzen Berechnungszeit in Kombination mit einer guten Interpretierbarkeit hervor. Die Anwendung der in Abschnitt 3.2 vorgestellten Methodik zur adaptiven Zustandsüberwachung beinhaltet einen iterativen Modellbildungsprozess, weshalb eine geringe Berechnungszeit, gerade bei der Erprobung und Entwicklung der Methodik als notwendig angesehen wird. Im Falle der LDA zeigt sich zudem durch die dimensionsreduzierte Darstellung eine überlegene Transparenz der Modelle, wodurch die erzielten Ergebnisse nachvollzogen und zusätzliches Prozesswissen aufgebaut werden kann.

Aus diesen Gründen erfolgt im Rahmen dieser Arbeit eine Klassifizierung ausschließlich mittels der gerade vorgestellte LDA in Kombination mit einem Mahalanobis Distanz Klassifikator und die Regression auf Basis der Partial Least Squares Regression, weshalb diese im Folgenden detaillierter vorgestellt werden. Informationen über die zudem implementierten Support Vector Machines für Klassifizierung und Regression können [149, 150, 151, 152] entnommen werden.

2.3.5.1 Mahalanobis Distanz

Gerade in Kombination mit vorgelagerten dimensionsreduzierenden Verfahren zeichnen sich Abstandsmaße durch eine gute Nachvollziehbarkeit aus, was sie gerade im Bereich der industriellen Zustandsüberwachung interessant macht. Als einfaches Maß kann hier die quadrierte euklidische Distanz genannt werden. Die euklidische Distanz zwischen zwei Punkten $\boldsymbol{x}$ und $\boldsymbol{y}$ in einem n-dimensionalen Raum berechnet sich wie folgt:

$$d(\boldsymbol{x},\boldsymbol{y}) = \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2} \tag{2.37}$$

Im Gegensatz zur euklidischen Distanz, die die direkte lineare Entfernung zwischen zwei Punkten im Raum misst, berücksichtigt die Mahalanobis-Distanz die Struktur der

Daten, indem sie die Varianz und Kovarianz der Daten in die Berechnung einbezieht [153]. Veranschaulicht bedeutet dies, dass der Abstand eines unbekannten Datenpunktes zu einer Gruppe mittels der Standardabweichung in der entsprechenden Richtung skaliert wird [154]. Abbildung 2-14 zeigt schematisch die Unterschiede zwischen der euklidischen und der Mahalanobis Distanz. Aus Gründen der Effizienz erfolgt die Berechnung der Mahalanobis-Distanz im Rahmen der Toolbox zu den Gruppenmittelpunkten. Die Mahalanobis-Distanz zwischen einem Vektor **x** und dem Mittelpunkt einer Gruppe berechnet sich dabei wie folgt:

$$d_{mahal}(\boldsymbol{x}) = \sqrt{\left(\boldsymbol{x} - \overline{\boldsymbol{x}}_g\right)^T \boldsymbol{S}^{-1}(x - \overline{\boldsymbol{x}}_g)} \tag{2.38}$$

mit

$\boldsymbol{x}$: Merkmalsvektor

$\overline{\boldsymbol{x}}_g$: Mittelpunkt der Gruppe g

$\boldsymbol{S}$: Kovarianzmatrix der Gruppe g

Die Qualität der Klassifizierung wird mittels des Klassifizierungsfehlers bzw. der Klassifizierungsrate als prozentualer Anteil der falsch bzw. richtig klassifizierten Stichproben angegeben [155].

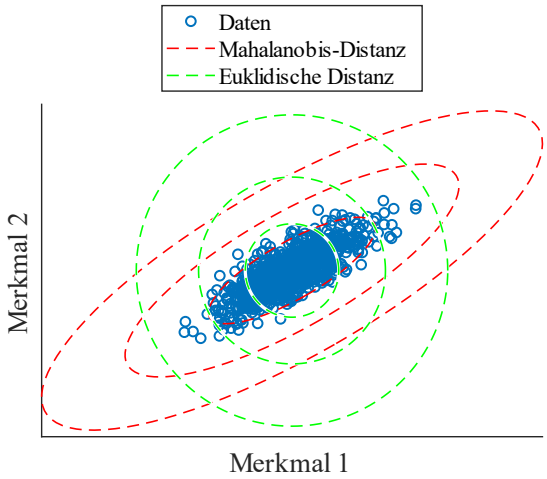


Abbildung 2-14: Gegenüberstellung der euklidischen und der Mahalanobis-Distanz.

2.3.5.2 Partial Least Squares Regression

Die Partial Least Squares Regression hat ihren Ursprung 1984 in der Datenanalyse chemischer Prozesse und stellt in speziellen Bereichen, wie beispielsweise der Auswertung von Halbleitergassensoren, eine der meistverwendeten Methoden dar [156, 157]. Sie wird verwendet, um die Beziehung zwischen einem Satz unabhängiger Variablen, den Prädiktoren, und einem Satz abhängiger Variablen, den

Antwortvariablen, zu modellieren. Im Gegensatz zur klassischen linearen Regression, die darauf abzielt, die Summe der quadrierten Abweichungen der Vorhersagen zu minimieren, konzentriert sich die PLS-Regression darauf, die Kovarianz zwischen den unabhängigen und abhängigen Variablen zu maximieren. Dadurch ist sie besonders nützlich bei multikollinearen und hochdimensionalen Datensätzen [158]. Dies erreicht sie durch eine vorgelagerte PCA-ähnliche Vorverarbeitung, welche unkorrelierte, unabhängige Variablen für die nachfolgende lineare Regression bildet. Vergleichbar zu der in Abschnitt 2.3.4.1 vorgestellten PCA ergibt sich dadurch ein Hyperparameter N_{PLSR}, welcher beschreibt, wie viele Komponenten für die weitere Modellbildung verwendet werden sollen. An dieser Stelle ist zu erwähnen, dass eine hohe Anzahl an verwendeten Komponenten im Vergleich zur Anzahl der im Training vorhandenen Stichproben ein verstärktes Risiko für eine Überanpassung des Modells mit sich bringt, was eine ausreichende Validierung des Modells unabdingbar macht. Eine detaillierte mathematische Beschreibung der PLSR kann [159] entnommen werden. Die finale Qualität der Regression wird mittels des Root Mean Square Errors (RMSE) bestimmt, welcher wie folgt berechnet wird:

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{2.39}$$

mit

y: Zielgröße

$\hat{y}$: Vorhersage des Modells

m: Anzahl Stichproben

2.3.6 Novelty Detection

Aufgrund der Komplexität technischer Systeme sowie der meist komplexen industriellen Umgebungsbedingungen existiert eine Vielzahl möglicher Fehlerarten und variierender Umgebungseinflüsse, für die es nahezu unmöglich ist, ein vollständiges überwachtes Modell zu trainieren. Zudem stehen meist für fehlerhafte Systemzustände nur geringe Datenmengen zur Verfügung, wohingegen Daten aus dem Normalzustand in großer Menge verfügbar sind [160]. In diesen Fällen bieten sich teilüberwachte Methoden wie beispielsweise eine Novelty Detection zur Identifikation eines abweichenden Systemverhaltens an. Methoden der Novelty Detection arbeiten dabei typischerweise in zwei Schritten. Im ersten Schritt wird ein Modell des Normalzustands gebildet, was die Definition eines Novelty Scores sowie die

Festlegung eines Grenzwertes zur Abgrenzung von Normaldaten und Neuheiten (Novelty) einschließt. Im zweiten Schritt erfolgt dann die Bewertung neu eintreffender Datenpunkte hinsichtlich ihres Novelty Scores. Auch im Fall der Novelty Detection existiert keine generell funktionierende Methode, weshalb hier ebenfalls mittlerweile eine Vielzahl an Methoden und Algorithmen existiert, welche zugeschnitten auf spezielle Datentypen oder zugrundeliegende Problemstellungen verwendet werden [161]. Abbildung 2-15 gibt eine Übersicht über die in der Toolbox verfügbaren Methoden für eine Novelty Detection.

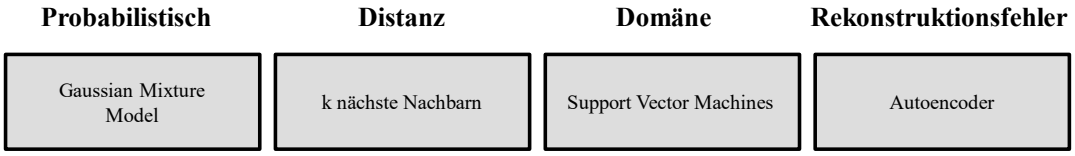


Abbildung 2-15: In der Toolbox verfügbare Methoden zur Novelty Detection.

Grundsätzlich kann hier zwischen probabilistischen Distanz-, Domänen- und Rekonstruktionsfehler-basierten Algorithmen unterschieden werden. Distanz-basierte Verfahren nutzen dabei unterschiedliche Distanzmaße, wie beispielsweise die euklidische oder die Manhattan Distanz zwischen einzelnen Datenpunkten oder Klassenmittelpunkten [162]. Probabilistische Verfahren wie beispielsweise die Gaussian Mixture Models nähern die Wahrscheinlichkeitsverteilung der Trainingsdaten mittels mehrerer Gauß-Verteilungen an und schätzen dadurch die Wahrscheinlichkeit, ob ein unbekannter Datenpunkt zu einer dieser Verteilungen gehört [163]. Domänen basierte Verfahren, wie die One Class Support Vector Machines versuchen eine Hyperebene zwischen Trainingsdaten und neuartigen Daten zu modellieren und nutzen die Distanz zu dieser Ebene als Novelty Score [164, 165]. Autoencoder basieren auf künstlichen neuronalen Netzen und lernen mittels eines Encoders eine niederdimensionale Darstellung der Daten und werden dann über einen Decoder rekonstruiert. Für neuartige Datenpunkte versagt diese Rekonstruktion, wodurch sich ein erhöhter Rekonstruktionsfehler ergibt [166]. Eine umfangreiche Übersicht über verfügbare Methoden sowie deren Vor- und Nachteile können [161], [167] entnommen werden.

Auch im Kontext der industriellen Zustandsüberwachung ist eine nachvollziehbare Entscheidung über die Neuheit eines Datenpunkts bei der Novelty Detection wünschenswert. Hierbei sind insbesondere distanzbasierte Verfahren hervorzuheben, die in Kombination mit einer dimensionsreduzierenden Datenvorverarbeitung, wie sie beispielsweise durch die Hauptkomponentenanalyse (PCA) realisiert wird, einen robusten und vor allem leicht nachvollziehbaren Novelty Score bereitstellen [84], [168]. Als bekanntes Beispiel ist hier das k Nächste Nachbarn (kNN) Verfahren basierend auf der euklidischen Distanz zu nennen. Der Novelty Score ist dabei als die

Summe der Distanzen zu den k nächsten Nachbarn definiert. Wie alle Distanz basierten Verfahren liegt dabei die Annahme zugrunde, dass normale Datenpunkte im betrachteten Merkmalsraum nahe an den im Training vorhandenen Daten aus dem Gutzustand liegen, wohingegen Anomalien größere Abstände aufweisen [161]. Die kNN basierte Novelty Detection ist problemlos im Stande, verteilte Cluster im Normalzustand abzubilden [167]. Probleme durch hohe Dimensionalitäten, wie sie in der Literatur häufig beschrieben werden, können mittels vorgelagerter dimensionsreduzierender Verfahren vermieden werden. Weitere Probleme der kNN basierten Novelty Detection stellen ein hoher Speicherbedarf und Berechnungsaufwand für große Datenmengen dar, was aufgrund der meist geringen Stichprobenanzahlen im Kontext der industriellen Zustandsüberwachung zumindest für eine klassische zentrale Datenverarbeitung als unkritisch angesehen werden kann. Gerade bei einer kleinen Anzahl an Stichproben zeigt sich das kNN-Verfahren dabei robuster als andere Methoden [161]. Aufgrund der überlegenen Nachvollziehbarkeit sowie Robustheit erfolgt die Novelty Detection im Rahmen dieser Arbeit mittels des kNN-Verfahrens in Kombination mit der euklidischen Distanz.

Eine generelle Herausforderung bei der Anwendung teilüberwachter Verfahren, stellt die Festlegung des Grenzwertes dar, ab wann ein Punkt als Novelty eingestuft wird. Für das kNN-Verfahren werden hier beispielsweise feste Grenzwerte auf Basis der Standardabweichung der Gutdaten wie etwa 2σ oder 3σ vorgeschlagen [169]. Sind bereits Daten von fehlerhaften Systemzuständen vorhanden, besteht zudem die Möglichkeit, die Grenzwertbestimmung mittels Kreuzvalidierung oder manuell auf Basis von Expertenwissen durchzuführen [167].

Der Parameter k stellt einen Hyperparameter da, welcher massiven Einfluss auf die Leistungsfähigkeit des Modells hat. Bei kleinen Werten für k (k = 1 oder k = 2) reagiert das Modell sehr empfindlich auf einzelne Datenpunkte. Die Entscheidungsgrenze ist dabei sehr flexibel und folgt stark den lokalen Mustern der Daten, wodurch das Modell zum Overfitting tendiert. Große Werte für k (k > 20) hingegen ergibt sich eine geglättete Entscheidungsgrenze. Dadurch ist das Modell im Stande besser zu Generalisieren, wobei jedoch die Gefahr des Underfitting besteht. Eine pauschale Aussage zur Auswahl von k kann nicht getroffen werden, wobei sich Werte von k = 3 bis k = 10 als zweckmäßig erwiesen haben. Sind bereits Daten unnormaler Zustände vorhanden, können wiederum allgemeine Methoden zur Hyperparameteroptimierung wie beispielsweise eine x-fache Kreuzvalidierung oder die so genannte ROC-Kurven verwendet werden [167]. Für die Beurteilung der Qualität der Novelty Detection stehen verschiedene Kennzahlen wie beispielsweise die False Negative Rate FN_r, die False Positive Rate FP_r oder die Accuracy ACC zur Verfügung:

$$FN_r = \frac{FN}{FP + FN} \tag{2.40}$$

$$FP_r = \frac{FP}{TN + FP} \tag{2.41}$$

$$ACC = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{2.42}$$

mit

FN: False Negatives

FP: False Positives

TN: True Negatives

TP: True Positives

Da viele der in der Literatur beschriebenen Metriken Probleme mit ungleich verteilten Klassen oder einer intuitiven Interpretierbarkeit haben, erfolgt die Angabe der Performanz der Novelty Detection im Rahmen dieser Arbeit mittels der FN, FP, TN und TP [167].

Neben der Erkennung neuer Zustände gibt es gerade in Verbindung mit einer überwachten Modellbildung weitere Anwendungen für die Novelty Detection. In Abbildung 2-16 sind die Eingriffsmöglichkeiten der Novelty Detection in Verbindung mit einem typischen Modellbildungsprozess für eine überwachte Klassifizierung dargestellt.

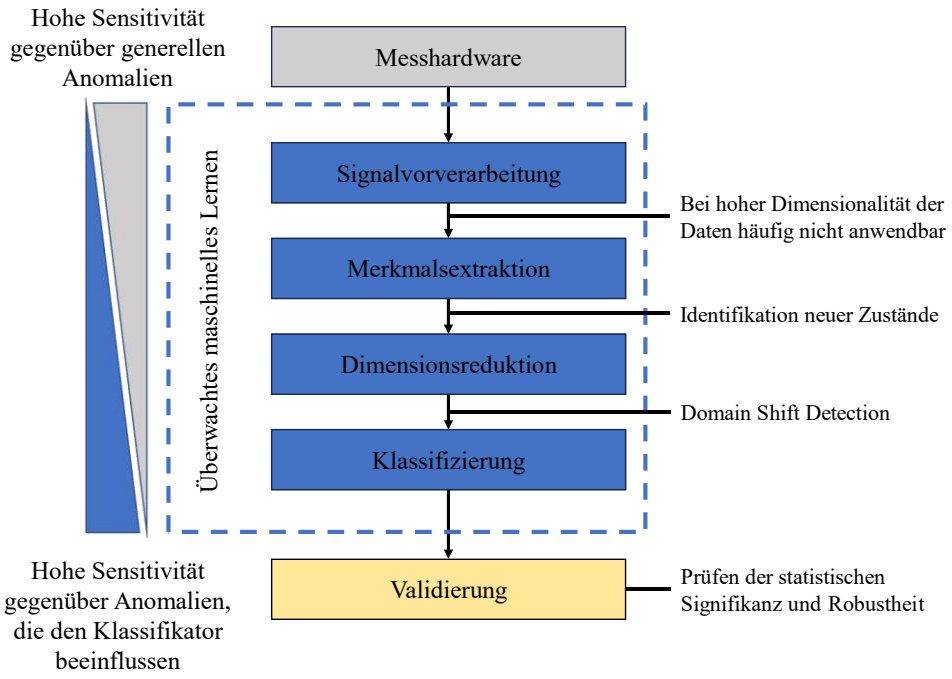


Abbildung 2-16: Eingriffspunkte der Novelty Detection in Kombination mit überwachten maschinellen Lernverfahren (angepasst nach [84]).

Je später im Prozess die Novelty Detection eingesetzt wird, desto geringer ist die Sensitivität gegenüber allgemeinen Novelties und desto höher wird die Sensitivität gegenüber Effekten, welche sich auf den Klassifikator auswirken. Eine spät im Modellbildungsprozess trainierte Novelty Detection kann somit zur Prüfung der Gruppenzugehörigkeit der im überwachten Modell enthaltenen Gruppen neuer Datenpunkte und damit beispielsweise zur Erkennung von Domain Shifts, also als Domain Shift Detection (vgl. Abschnitt 2.2.2), verwendet werden [84].

2.3.7 Validierung

Um die Leistungsfähigkeit und die Generalisierbarkeit der hergeleiteten Modelle sicherzustellen, werden systematische Methoden zur Validierung, wie beispielsweise die x-fache Kreuzvalidierung (engl. x-fold Cross-Validation), verwendet [170]. Bei dieser werden die Daten randomisiert in x Bereiche eingeteilt und anschließend iterativ x-1 Bereiche für das Training und ein Bereich für das Testen des Modells verwendet. Dies wird x-fach wiederholt, bis alle Daten für Training und Test verwendet wurden. Das Vorgehen ist in Abbildung 2-17 a) noch einmal zur Veranschaulichung dargestellt. Um repräsentative Stichproben für die Gesamtheit der Daten zu erhalten, erfolgt die Auswahl der Stichproben stratifiziert [171]. Die Auswahl von x ist dabei stark von dem jeweiligen Datensatz abhängig, weshalb eine pauschale Aussage nur schwer möglich ist [172]. Werte von x = 5 bis x = 10 werden in der Literatur als optimal angesehen [173, 174, 175]. Der in der Toolbox verwendete Standardwert beträgt x = 10, wobei dieser im Rahmen dieser Arbeit abhängig von der betrachteten Anwendung und dem Ziel der Validierung angepasst wird. Die Angabe des Validierungsfehlers erfolgt dann in Form des gemittelten Fehlers über alle Testbereiche.

Eine Abwandlung der randomisierten x-fachen Kreuzvalidierung ist die Leave-One-Group-Out Validierung, bei der die Auswahl der Trainings- und Testsamples nicht randomisiert stattfindet, sondern auf Basis annotierter Quereinflüsse. Dadurch kann beispielsweise die Übertragbarkeit der Modelle zwischen verschiedenen Maschinen oder das Verhalten unter wechselhaften Umweltbedingungen untersucht werden, wodurch sich eine deutlich realitätsnähere Beurteilung der Modellperformanz erzielen lässt [41]. x ergibt sich hierbei automatisch durch die Anzahl der verfügbaren Variationen des untersuchten Quereinflusses.

Gerade bei der Verwendung dimensionsreduzierender Verfahren, wie beispielsweise der PCA oder LDA, lässt sich zudem eine visuelle Validierung der Modelle durchführen. Durch die Projektion der hochdimensionalen Merkmalsvektoren in eine niederdimensionale Darstellung können Ursache-Wirk-Zusammenhänge identifiziert und das Vertrauen in die Modellbildung gesteigert werden. Ordnen sich beispielsweise

Klassen nach quantitativen Abstufungen der Zielgröße, wie in Abbildung 2-17 b) dargestellt, obwohl diese Information bei der Modellbildung nicht verwendet wurde, kann angenommen werden, dass das Modell einen physikalischen Zusammenhang in den Daten identifiziert hat. Zudem können gezielt Zwischenstufen aus dem Training entfernt und anschließend zur Validierung der gefundenen Zusammenhänge in den niederdimensionalen Raum projiziert werden.

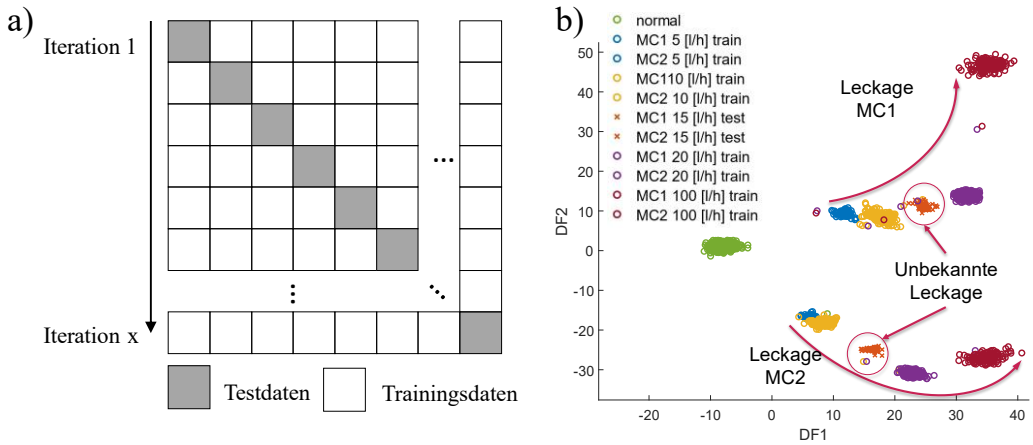


Abbildung 2-17: a) Vorgehen bei der systematischen x-fachen Kreuzvalidierung. b) Visuelle Validierung durch die Projektion von Trainingsdaten, wie es dimensionsreduzierende Verfahren wie in diesem Fall die LDA ermöglichen[168].

2.3.8 Hyperparameteroptimierung

Neben der Entwicklung bzw. Identifikation geeigneter Algorithmen und Methoden spielt vor allem die problemspezifische Optimierung der Parameter eine entscheidende Rolle. Unterschieden wird typischerweise zwischen Modellparametern und Hyperparametern. Während Modellparameter während des Trainingsprozesses angepasst und optimiert werden, müssen Hyperparameter bereits vor der Trainingsphase definiert werden. Als Beispiel für Modellparameter sind die Gewichte von neuronalen Netzen zu nennen, wohingegen die Architektur des Netzes einen Hyperparameter darstellt [176]. Die Anzahl der anpassbaren Parameter ist dabei stark von der Art der verwendeten Methoden abhängig. Da sich die Wahl der Hyperparameter massiv auf die Leistungsfähigkeit des Modells in der Anwendung auswirkt, werden diese häufig mittels entsprechender automatisierter Optimierungsmethoden bestimmt. Diese arbeiten typischerweise mit einer äußeren Validierungsschleife, wie beispielsweise der in Abschnitt 2.3.7 vorgestellten x-fachen Kreuzvalidierung, und prüfen die Leistungsfähigkeit des Modells bei Verwendung eines bestimmten Hyperparametersatzes. Je nach Anzahl der gesuchten Hyperparameter und der damit verbundenen Dimensionalität des Suchraums können verschiedene Optimierungsverfahren verwendet werden. Sind beispielsweise

Erfahrungen oder Prozesswissen bezüglich der verwendeten Methoden oder der spezifischen Problemstellung vorhanden, können diese eingesetzt werden, um bestimmte Hyperparameter vorab festzulegen oder den Suchraum manuell einzugrenzen [177]. Als einfachste Methode ist beispielsweise die erschöpfende Suche (engl. Brute Force) zu nennen, welche alle möglichen Kombinationen an verfügbaren Hyperparametern untersucht und dadurch gesichert die beste Kombination identifiziert. Problematisch ist hierbei die sehr schnell wachsende Komplexität und damit die benötigte Rechenleistung, weshalb diese in den meisten Fällen kein adäquates Mittel zur Hyperparameteroptimierung darstellt [178].

In Abbildung 2-18 sind die in der Toolbox enthaltenen Methoden zur Hyperparameteroptimierung dargestellt. Speziell für die Bestimmung der Anzahl der zur Modellbildung verwendeten Merkmale kann eine Brute Force Suche (Num Feat Ranking) durchgeführt werden. Beinhalten die verwendeten Methoden zusätzliche Hyperparameter, wie es beispielsweise beim TFEx der Fall ist, erweist sich eine Optimierung mittels erschöpfender Suche nicht als zielführend, weshalb die Toolbox um eine Bayes Optimierung erweitert wurde. Beide Methoden werden im Folgenden vorgestellt.

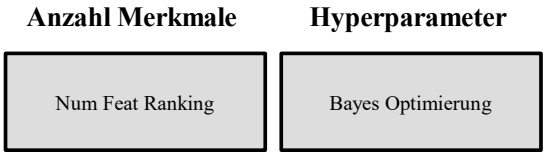


Abbildung 2-18: In der Toolbox verfügbare Methoden zur Hyperparameteroptimierung.

2.3.8.1 Num Feat Ranking

Die Anzahl der für die Modellbildung verwendeten Merkmale stellt einen methodenunabhängigen Hyperparameter der Toolbox dar. Um diesen zu optimieren, bietet die Toolbox eine Optimierung auf Basis einer x-fachen Kreuzvalidierung an. Hierbei erfolgt die Modellbildung mit einer steigenden Anzahl an Merkmalen. Betrachtet wird im Anschluss die Modellperformanz über die Anzahl der verwendeten Merkmale, wobei die Anzahl mit dem kleinsten verbleibenden Testfehler für die finale Modellbildung verwendet wird.

2.3.8.2 Bayes Optimierung

Sind bei den verwendeten Methoden mehrere Hyperparameter zu optimieren, wie es beispielsweise beim TFEx oder der PLSR der Fall ist, kann dies mittels Bayes Optimierung geschehen [179, 180]. Diese wählt zunächst einige gleichmäßig verteilte Sätze für die verfügbaren Hyperparameter aus und evaluiert das Modell mit diesen

(Matlab verwendet vier Stützstellen [181]). Diese Stützstellen werden zum Trainieren eines Regressionsmodells, basierend auf Gaußschen Prozessen, verwendet. Dieses erlaubt ein Abschätzen der Modellperformanz sowie ein Abschätzen der Unsicherheit für alle verfügbaren Hyperparameterkombinationen. Basierend auf diesen Vorhersagen wird iterativ die vielversprechendste Hyperparameterkombination ausgewählt und evaluiert. Zur Identifikation der nächsten Hyperparameterkombination stehen mehrere Methoden mit unterschiedlichen Optimierungszielen zur Verfügung, wie beispielsweise Expected Improvement oder Probability of Improvement [181]. Als Abbruchkriterium für die Optimierung der Hyperparameter kann eine maximale Anzahl an Iterationen, eine maximale Laufzeit, ein Konvergenzkriterium oder das Erreichen eines Zielwertes definiert werden. Eine detaillierte Beschreibung der Bayes Optimierung sowie die konkret verwendete Implementierung kann [182] und [181] entnommen werden.

2.3.9 Stacks

In Abbildung 2-19 sind exemplarische Stacks für eine Novelty Detection sowie eine überwachte Klassifizierung mit Domain Shift Detection dargestellt. Gerade im industriellen Umfeld spielen die Transparenz und Nachvollziehbarkeit der Modelle eine wichtige Rolle. Eine einfache Möglichkeit zur Plausibilisierung der Modelle stellt die Analyse und Visualisierung der für die Modellbildung selektierten Merkmale dar. Der Methodenpool erlaubt dazu die Rückverfolgbarkeit der Merkmale über den gesamten Stack. Dadurch lassen sich Ursache-Wirkzusammenhänge analysieren, wodurch einerseits eine Plausibilisierung der Modellbildung erfolgen, andererseits aber auch zusätzliches Prozesswissen generiert werden kann.

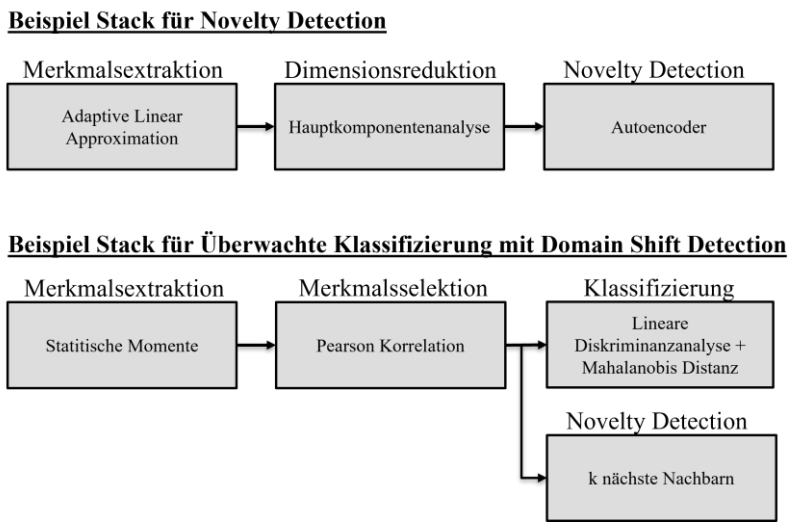


Abbildung 2-19: Exemplarische Stacks für eine ND und überwachte Klassifizierung mit Domain Shift Detection.

3 Methodik für eine ganzheitliche Zustandsüberwachung

Am Anfang des Kapitels erfolgt die Vorstellung einer Mess- und Datenplanung, welche speziell bei Anwendung 3 und 4 zur Sicherstellung der Datenqualität verwendet wird. Zudem erfolgt die generelle Vorstellung der adaptiven Zustandsüberwachung auf Basis des übergeordneten Algorithmus zur Kombination generischer teilüberwachter und überwachter Methoden des maschinellen Lernens. Im Anschluss werden die dafür verwendete, breitbandige Merkmalsextraktion sowie die Stacks für Klassifizierung, Regression und Novelty Detection vorgestellt, welche im Rahmen der adaptiven Zustandsüberwachung für die Anwendungen 1-4 verwendet werden. Ebenfalls wird das für die Novelty Detection verwendete Verfahren zur Grenzwertbestimmung präsentiert.

3.1 Mess- und Datenplanung

Um die für die Algorithmen benötigte strukturelle Datenqualität, aber auch den für die Modellbildung benötigten Informationsgehalt sicherzustellen, wird eine dem Modellbildungsprozess vorgelagerte Mess- und Datenplanung nach dem in Abbildung 3-2 abgebildeten iterativen Prozess durchgeführt. Der Prozess untergliedert sich dabei in ein Prüfen der Datenqualität auf Basis kleinerer Mengen an Testdaten sowie die Generierung umfangreicher Datensätze im Rahmen größerer Messkampagnen. In beiden Fällen handelt es sich um iterative Prozesse, welche abhängig von den erzielten Ergebnissen Anpassung am Prozess, also beispielsweise der verwendeten Sensorik, Datenerfassung, Prozessführung oder den kontrollierbaren Umwelteinflüssen bewirken können. Im Folgenden wird der gesamte Prozess erläutert.

Zu Beginn werden basierend auf der zu untersuchenden technischen Anwendung Testdaten generiert. Bereits hier sollten mögliche Quereinflüsse zumindest bedacht werden. Über die verschiedenen Iterationen der Prüfung der Datenqualität kann bereits eine systematische Variation bekannter und kontrollierbarer Quereinflüsse erfolgen, um die Auswirkungen auf die Daten und die Zielgröße zu analysieren. Die generierten Testdaten werden im ersten Schritt strukturiert. Daten industrieller Anwendungen

stammen dabei aus sich sehr stark unterscheidenden Datenlandschaften, wobei die zugehörigen Daten zu Beginn häufig in unterschiedlichen, teils proprietären, Datenformaten vorliegen. Um diese mittels Methoden des maschinellen Lernens auswerten zu können, muss im ersten Schritt eine Strukturierung der Daten erfolgen. Da die Arbeiten im Rahmen dieser Arbeit auf der in Kapitel 2.3 vorgestellten Toolbox beruhen, wird bei der Sicherstellung der strukturellen Datenqualität im Speziellen auf die Anforderungen dieser Methoden geachtet.

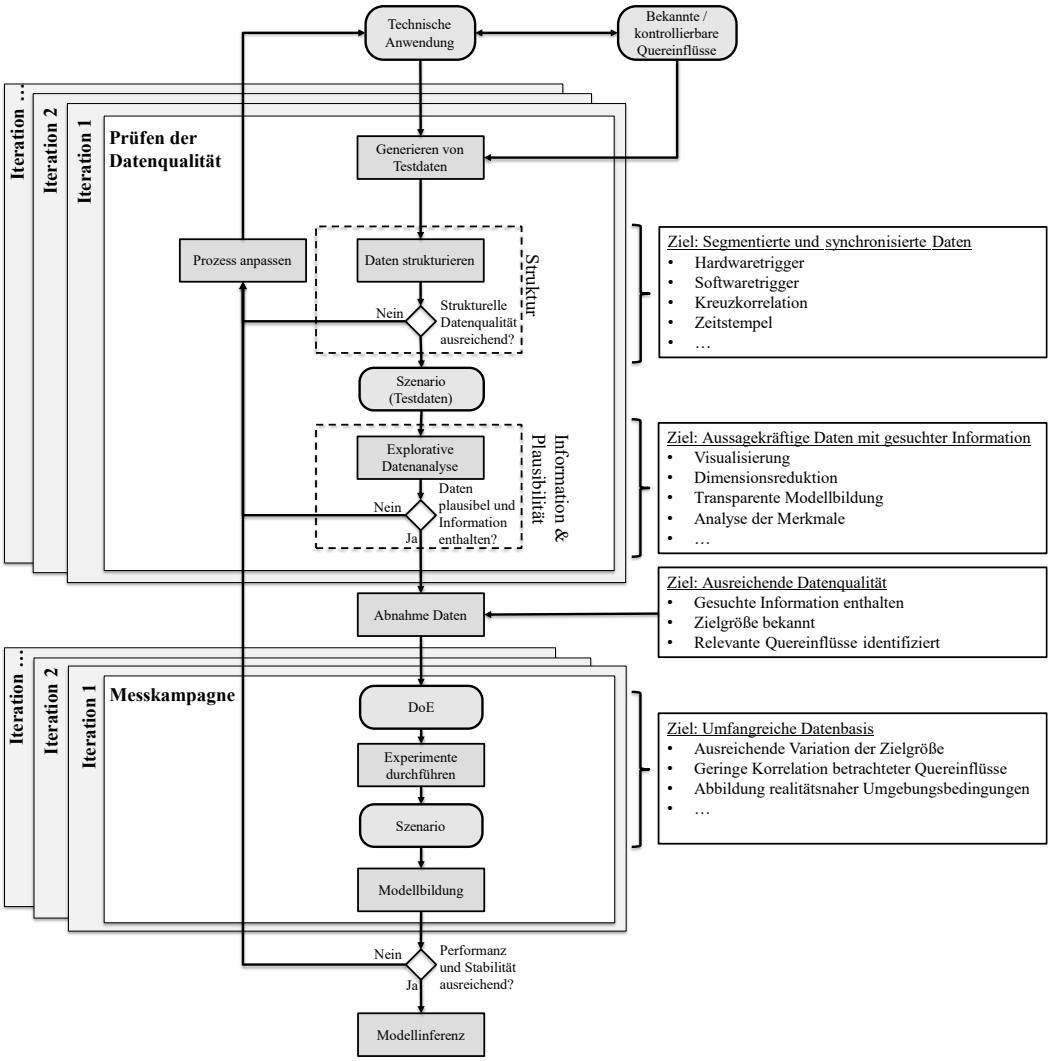


Abbildung 3-1: Ablauf der Mess- und Datenplanung mit anschließender Datengenerierung nach DoE.

Ziel der strukturellen Mess- und Datenplanung ist, die Überführung der Rohdaten in ein klar definiertes Lernproblem (Szenario) in einem standardisierten Format, welches alle benötigten, strukturierten Daten unter Berücksichtigung der in Abschnitt 2.3 beschriebenen Eigenschaften in Form dedizierter Variablen enthält. Neben der leichteren Handhabbarkeit wird dadurch das automatisierbare Benchmarking von Algorithmen ermöglicht. Die Methoden zum Erreichen dieser Struktur sind stark von der Datenlandschaft und der Zielstellung in den einzelnen Anwendungen abhängig,

was eine vollständige Automatisierung, gerade im Hinblick auf die Robustheit, erschwert. Kann die Struktur softwareseitig nicht hergestellt werden, müssen Änderungen am Prozess erfolgen. Auch im Hinblick auf die Inferenz der Modelle stellt die online Bereitstellung der strukturellen Datenqualität den ersten Schritt der Pipeline dar. Die Zielstruktur der Szenarien mit den zugehörigen Variablen ist in Abbildung 3-2 dargestellt und wird im Folgenden kurz vorgestellt. Bei der ersten Untersuchung eines Prozesses wird zudem eine Überinstrumentierung hinsichtlich der verwendeten Sensoren und Abtastraten empfohlen. Dadurch kann sichergestellt werden, dass keine Informationen durch den von Beginn an begrenzten Suchraum verloren gehen.

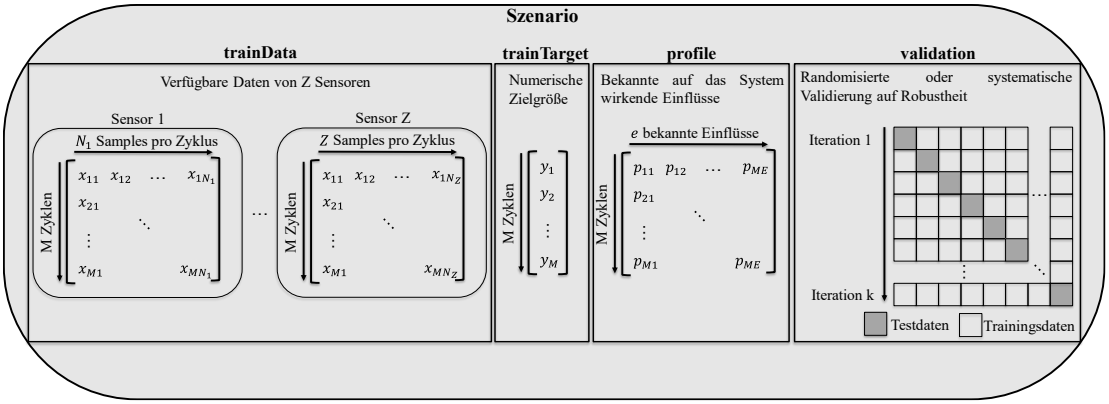


Abbildung 3-2: Szenario-Format mit klar definiertem Lernproblem.

trainData: Enthält die in Matrizen strukturierten Sensordaten. Angelehnt an die in Kapitel 2.3 beschriebene Struktur ist eine zyklische Datenstruktur erforderlich. Im Rahmen dieser Arbeit werden hierzu hauptsächlich prozessinterne Triggersignale verwendet. Diese können einerseits direkt zur Steuerung der Datenerfassung oder als zusätzlicher Sensor erfasst und nachträglich zur softwareseitigen Segmentierung der Daten verwendet werden. Nach der Segmentierung ist bei variierender Prozesslänge gegebenenfalls noch eine Vereinheitlichung der Zyklenlänge erforderlich.

trainTarget: Enthält die für das überwachte Lernen benötigte Zielgröße, welche in Form eines numerischen Vektors kodiert ist. Die Erfassung kann dabei durch eine nachgelagerte Messung mittels eines Referenzmesssystems oder durch Steuerung bzw. Modifikation des Prozesses erfolgen. Enthalten sein können diskrete Zustände im Falle eines Klassifizierungsproblems oder kontinuierliche Werte für eine Regression. Ist es nicht möglich, eine Zielgröße zu erfassen, können synthetische Zielgrößen auf Basis von Expertenwissen oder physikalischen Modellen verwendet werden.

profile: Enthält alle bekannten auf den Prozess wirkenden Einflüsse. Diese können einerseits über Sensorik erfasst oder durch die Prozesssteuerung bzw. Parametrierung vorgegeben werden. Der Parameterraum ist hierbei mit der im realen Modelleinsatz zu erwartenden Varianz abzudecken.

validation: Für eine eindeutige Definition des Lernproblems wird ein Validierungsobjekt erzeugt, welches die Trennung der Daten in Trainings-, Validierungs- und Testdaten beinhaltet. Abhängig vom Ziel des generierten Szenarios, kann dieses Objekt eine klassische randomisierte x-fache Kreuzvalidierung oder eine auf einem Quereinfluss basierende Leave-One-Group-Out Validierung beinhalten. Dadurch kann das Modell gezielt auf Robustheit gegenüber bekannten Quereinflüssen optimiert und unter möglichst realen Bedingungen auf Übertragbarkeit getestet werden.

Nachdem die entsprechende strukturelle Datenqualität sichergestellt ist, können die Daten auf Plausibilität sowie den Informationsgehalt bezüglich des definierten Lernproblems untersucht werden. Hierzu erfolgt eine explorative Datenanalyse auf Basis der in der Toolbox verfügbaren Methoden. Ziel dieser Untersuchung ist es, festzustellen, ob physikalisch begründbare Informationen über die Zielgröße in den Daten vorhanden sind. Dies kann sich beispielsweise durch charakteristische Signalverläufe oder physikalisch motivierte Frequenzen äußern. Auch hier ist die Vorgehensweise stark von der Anwendung abhängig, weshalb eine Automatisierung nur schwer zu realisieren ist. Falls bei der explorativen Datenanalyse Probleme mit der Plausibilität oder dem Informationsgehalt festgestellt werden, erfolgt auch hier ein iteratives Anpassen des Prozesses.

Wurden in ausreichend Iterationen Testdaten erzeugt und es konnte eine ausreichende Datenqualität festgestellt werden, kann eine Abnahme der Daten erfolgen. Zu diesem Zeitpunkt sollte bereits ein tieferes Daten- und Prozessverständnis vorhanden sein, um ein fundiertes DoE, basierend auf bereits identifizierten Quereinflüssen sowie einer ausreichenden Variation der Zielgröße (soweit möglich), erstellen zu können. Abhängig vom Ziel des in der entsprechenden Messkampagne generierten Szenarios können verschiedene Vorgehen zur Erstellung des DoE verwendet werden. Sollen beispielsweise systematisch Einflüsse bestimmter Quereinflüsse untersucht werden, kann ein vollfaktorielles DoE auf wenigen diskreten Stützstellen in einem kleinen Bereich des Versuchsraums verwendet werden. Wohingegen bei der Datenaufnahme für ein Modell, welches in der Anwendung betrieben werden soll, eine möglichst vollständige Abdeckung des Versuchsraums vorzuziehen wäre. Zudem können die für das DoE relevanten Quereinflüsse im Austausch mit Domänenexperten für die entsprechenden Anwendungen erarbeitet werden. Die Versuche werden dem DoE folgend durchgeführt und in einem Szenario-File abgelegt. Das im Anschluss gebildete Modell kann bei Feststellung einer ausreichenden Performanz und Stabilität in der Anwendung zur Inferenz gebracht werden. Bei negativem Ergebnis kann eine weitere Messkampagne durchgeführt werden bzw. bei Bedarf noch mal eine Änderung am Prozess erfolgen.

3.2 Adaptive Zustandsüberwachung

Wie einleitend erläutert, muss ein Zustandsüberwachungssystem in der Lage sein, sich an komplexe industrielle Umweltbedingungen und bis dato unbekannte Systemzustände anzupassen. Aus diesem Grund wurde im Rahmen dieser Arbeit eine Methodik zur adaptiven Zustandsüberwachung entwickelt, welche den zentralen Beitrag dieser Dissertation darstellt. Der Kern der Methodik besteht aus einem übergeordneten Algorithmus, welcher die Kombination generischer, teilüberwachter und überwachter Methoden des maschinellen Lernens ermöglicht. Die Entwicklung dieses Algorithmus, der verwendeten Stacks sowie der Grenzwertbestimmung für die verwendete Novelty Detection erfolgte anhand des in Anwendung 1 (siehe Abschnitt 4.1) vorgestellten Datensatzes zur Zustandsüberwachung eines Hydraulikaggregates. Dieser zeichnet sich durch eine hohe Datenqualität sowie ein umfassend, vorhandenes System- und Prozessverständnis aus, wodurch er sich in besonderem Maß für eine realitätsnahe Methodenentwicklung eignet. Die im Folgenden präsentierten Verfahren sind dabei das finale Resultat eines Entwicklungsprozesses, welcher die Erprobung und Optimierung verschiedenster Methodenkombinationen des in Abschnitt 2.3 vorgestellten Methodenpools umfasste. Die dabei identifizierten und im Folgenden vorgestellten Methoden wurden im Anschluss auf weitere industrielle Anwendungen (Anwendung 2 -4) übertragen und auf ihre generelle Leistungsfähigkeit untersucht. Im Sinne einer geschlossenen Darstellung erfolgt die Präsentation der Ergebnisse sowohl für das Entwicklungsszenario (Anwendung 1) als auch für die weiteren untersuchten industriellen Anwendungen (Anwendungen 2 – 4) in Kapitel 4.

Die adaptive Zustandsüberwachung erfolgt durch eine Kombination von Novelty Detection und überwachter Modellbildung in Form eines übergeordneten Algorithmus, dessen Ablauf in Abbildung 3-3 dargestellt ist. Zu Beginn wird ein Novelty Detection Modell auf Grundlage der verfügbaren historischen Daten trainiert. Durch das Einbeziehen gelabelter Daten bereits bekannter Fehlerzustände wird das Novelty Detection Modell für diese Fehlermechanismen desensibilisiert. Die Diagnose bzw. Quantifizierung der bekannten Zustände erfolgt durch das überwachte Modell, welches ebenfalls auf der Grundlage der verfügbaren historischen Daten trainiert wird. Neu erfasste Datenpunkte werden zunächst bezüglich ihres Novelty Scores bewertet. Wird der Datenpunkt nicht als Anomalie eingestuft, erfolgt im zweiten Schritt eine Domain Shift Detection, um die Gültigkeit des überwachten Modells sicherzustellen. Wird hierbei festgestellt, dass der neue Datenpunkt zu einem der bereits bekannten Zustände gehört, d.h. die Domain Shift Detection zeigt keine Anomalie, erfolgt die Auswertung über das überwachte Modell und der aktuelle Zustand des Systems wird dem Personal mitgeteilt. Bei negativer Prüfung der Modellgültigkeit muss die Ursache

für den neuen Zustand vom Personal festgestellt und entsprechende Maßnahmen eingeleitet werden.

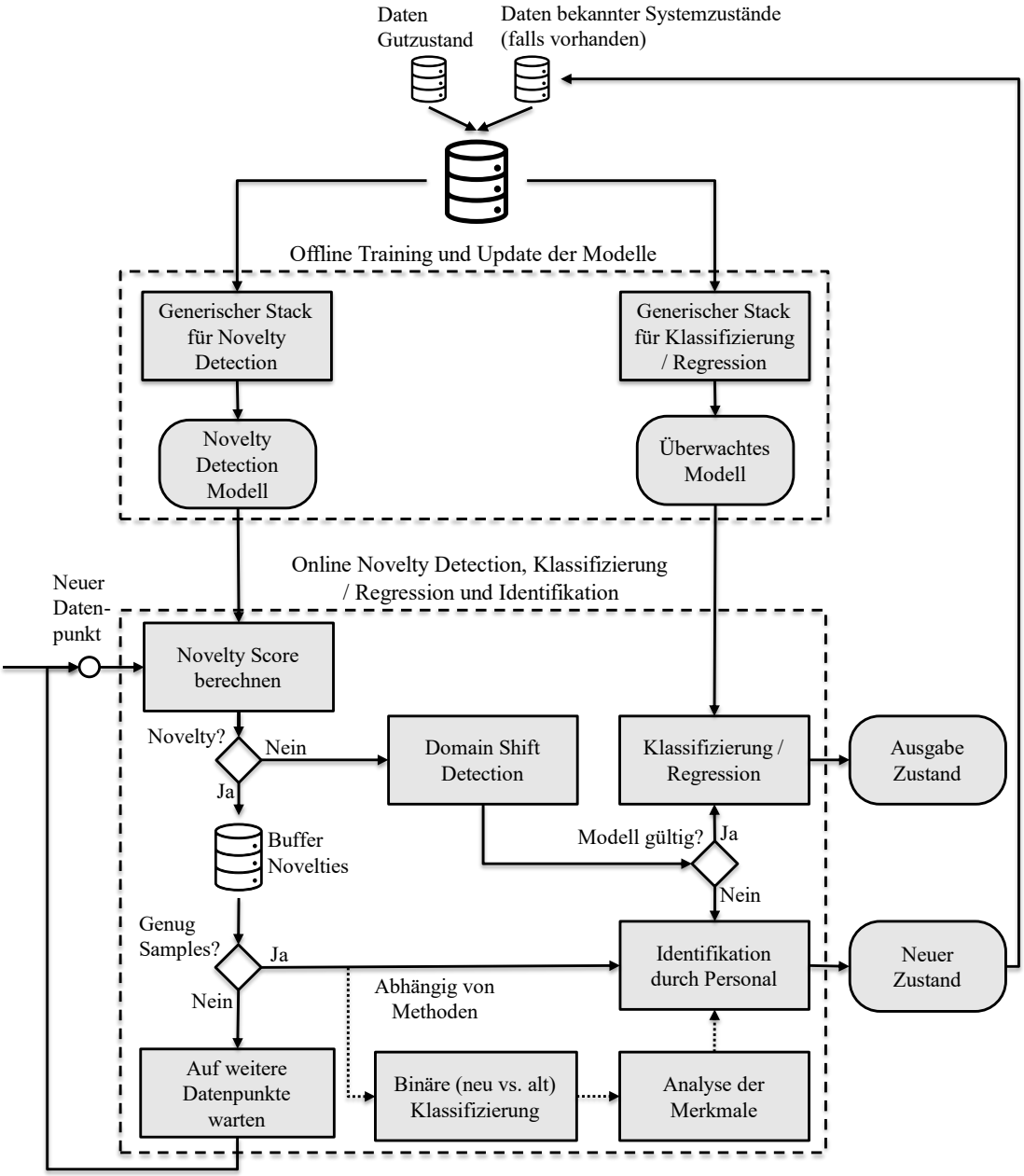


Abbildung 3-3: Ablaufdiagramm für eine adaptive Zustandsüberwachung durch Kombination von ND und überwachter Lernverfahren. Nachdem beide Modelle zunächst trainiert wurden, können neue Zustände von der ND erkannt und schrittweise in das überwachte Modell übertragen werden.

Wird der neu eintreffende Datenpunkt als Novelty identifiziert, wird er stattdessen entsprechend gelabelt und in einem Buffer abgelegt. Sind im Buffer eine definierte Anzahl von aufeinanderfolgenden Novelties aufgetreten, muss die Ursache für den neu eingetretenen Zustand von entsprechendem Personal untersucht werden. Die Identifikation der Ursache kann dabei softwareseitig unterstützt werden. Hierfür wird eine zusätzliche überwachte Klassifizierung zur Trennung der historischen und der im Buffer abgelegten neuen Daten trainiert. Eine anschließende Analyse der für diese

Klassifizierung ausgewählten Merkmale in Bezug auf Herkunft (Sensor, Position im Zyklus) und Typ (Zeit- oder Frequenzbereich, statistisch, formbeschreibend) kann zur Identifikation der Ursache des neu aufgetretenen Zustands verwendet werden. Sobald der neue Zustand identifiziert und gelabelt wurde, werden die Daten zur Trainingsdatenbasis hinzugefügt und sowohl das Novelty Detection Modell als auch das überwachte Modell werden neu trainiert. Bei erneutem Auftreten des Zustands wird er nicht mehr als Novelty erkannt und kann vom überwachten Modell ausgewertet werden.

Zusammenfassend soll ausgehend von dem vorgestellten Algorithmus die Zustandsüberwachung einer technischen Applikation, basierend auf den Daten des Gutzustandes des Systems oder aufbauend auf einer bereits vorhandenen Trainingsdatenbasis ermöglicht werden. Von der Novelty Detection erkannte, bis dato unbekannte Fehler können in das überwachte Modell übernommen werden, wodurch bei erneutem Auftreten eine Diagnose bzw. Quantifizierung durch das überwachte Modell erfolgen kann. Die Identifikation bzw. Interpretation dieser neuen, unbekannten Zustände kann durch ein weiteres überwachtes Modell zur Trennung der historischen und der neuen, unbekannten Daten unterstützt werden. Durch die Integration der Domain Shift Detection in diesen Prozess kann zudem eine Beeinflussung des überwachten Modells beispielsweise durch sich verändernde Umwelteinflüsse erkannt werden. Zudem können durch die allgemeine Formulierung des übergeordneten Algorithmus die verwendeten Stacks für Novelty Detection sowie für die überwachte Modellbildung an die jeweilige Applikation angepasst werden.

3.2.1 Time-Frequency-Extractor

Die im Methodenpool enthaltenen Merkmalsextraktionsmethoden (vgl. Abschnitt 2.3.2) basieren auf domänenspezifischen Methoden, welche entweder Merkmale im Zeit-, Frequenz- oder Zeitfrequenzbereich abbilden. Gerade für die Anwendung einer Novelty Detection ist eine generelle Empfindlichkeit gegenüber den auf das System wirkenden Einflüssen erforderlich, weshalb die Merkmalsextraktion in der Lage sein muss, breitbandig Informationen aus verschiedenen Domänen zu extrahieren. Aus diesem Grund wurde der bestehende Methodenbaukasten um eine Merkmalsextraktion erweitert, im Folgenden als Time-Frequency-Extractor (TFEx) bezeichnet. Diese folgt dem in Abbildung 3-4 dargestellten Verfahren. Ausgehend von den Daten im Zeitbereich wird das Signal eines Arbeitszyklus in N_{Zeit} äquidistante Segmente zerteilt. Auf diese Weise können die berechneten Merkmale aus charakteristischen Signalabschnitten extrahiert werden, die im Kontext der spezifischen Anwendung sowie der betrachteten Zustände von Bedeutung sind. Anschließend wird mittels FFT

das Amplitudenspektrum sowohl für jedes Teilsegment als auch für den gesamten Zyklus berechnet. Auch die berechneten Spektren werden wiederum in N_{Freq} Segmente unterteilt, wodurch in Abhängigkeit der betrachteten Anwendung Merkmale schmalbandig aus charakteristischen Frequenzen oder breitbandigen aus breiten Frequenzbändern berechnet werden können. Für den gesamten Zyklus und das gesamte Spektrum sowie für alle Segmente im Zeit- und Frequenzbereich werden die in Abschnitt 2.3.2 vorgestellten statistischen und formbeschreibenden Merkmale berechnet. Auf das Merkmal „Position der min. Werte“ wird an dieser Stelle bewusst verzichtet, da dieses bei Schwingungen im Zeitbereich vermehrt hohe Korrelationen zum Merkmal „Position der max. Werte“ aufweist und auch im Frequenzbereich kein aussagekräftiges Merkmal darstellt. Die Anzahl der Merkmale pro Sensor ergibt sich danach wie folgt:

$$N_M = (N_{Zeit} + 1) * N_{Fun} + (N_{Zeit} + 1) * (N_{Freq} + 1) * N_{Fun} \tag{3.1}$$

mit

N_{Zeit}: Anzahl der Segmente im Zeitbereich

N_{Freq}: Anzahl der Segmente im Frequenzbereich

N_{Fun}: Anzahl Merkmalsfunktionen

N_M: Anzahl Merkmale pro Sensor

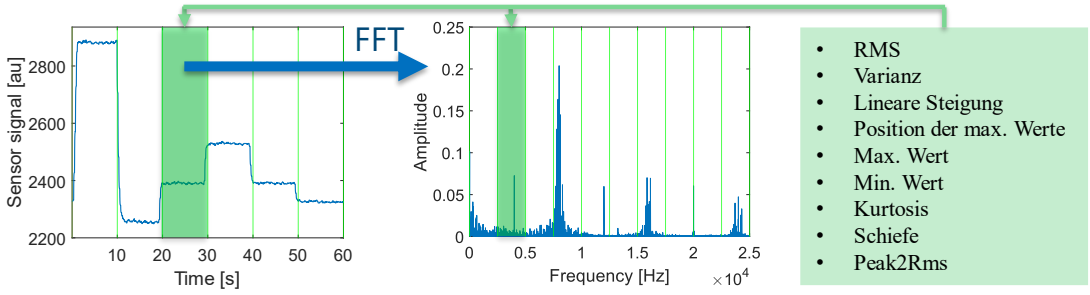


Abbildung 3-4: Schematische Darstellung des allgemeinen Ablaufs und der Funktionsweise des TFEx. Segmentierung des Zeitsignals (links) und anschließende Berechnung des Amplitudenspektrums (Mitte) für jedes Zeitsegment. Aufzählung der zu berechnenden statistischen und formbeschreibenden Merkmale (rechts).

Die Anzahl der Segmente im Zeit- und Frequenzbereich sind Hyperparameter, die vergleichbar zur STFT eine Anpassung der gewünschten Zeit- und Frequenzauflösung der Merkmale ermöglichen und stark vom jeweiligen Datensatz abhängen. Bei einem statischen Prozess, z.B. bei einer Maschine mit fester Drehzahl oder nur einem Prozessabschnitt, wäre es sinnvoll, nur ein Segment im Zeitbereich zu verwenden. Bei komplexeren Anwendungen, bei denen die gesuchte Information in zeitlich lokalisierten Abschnitten oder charakteristischen Frequenzen vermutet wird, kann eine höhere Anzahl an Segmenten sinnvoll sein. Im Fall von überwachten Lernproblemen kann zudem die in Abschnitt 2.3.8.2 vorgestellte Bayes Optimierung verwendet

werden, um eine optimale Segmentierung im Zeit- und Frequenzbereich zu ermitteln. Darüber hinaus ermöglicht die Segmentierung die Rückverfolgbarkeit der extrahierten Merkmale auf ihren Ursprung, d. h. auf bestimmte Ereignisse im Zeitbereich oder bestimmte Bänder im Frequenzbereich. Insbesondere in Kombination mit einer Merkmalsselektion kann dies genutzt werden, um die physikalische Plausibilität von Merkmalen zu überprüfen und relevante Signalabschnitte hervorzuheben oder zu visualisieren.

Bei Bedarf kann zudem eine Anpassung des TFEx an verschiedene Problemstellungen über entsprechende Transformationen der Daten erfolgen. Als Beispiele können hier die für die Vibrationsanalyse typische Berechnung der Hüllkurven und Hüllkurvenspektren sowie im Fall drehzahlvariabler Problemstellungen die in Abschnitt 2.3.2.2 erläuterten Ordnungsspektren genannt werden [183].

3.2.2 Stacks für Novelty Detection, Klassifizierung und Regression

Im Folgenden werden die für die adaptive Zustandsüberwachung verwendeten Stacks für Novelty Detection, Klassifizierung und Regression vorgestellt. Der verwendete Stack für die Novelty Detection ist in Abbildung 3-5 dargestellt. Im ersten Schritt erfolgt eine breitbandige Merkmalsextraktion mittels TFEx. Die Bestimmung der verwendeten Anzahl an Segmenten im Zeit- und Frequenzbereich erfolgt in Abhängigkeit der betrachteten Anwendung auf Basis von Prozesswissen, sodass charakteristische bzw. funktional zusammenhängende Signalabschnitte segmentiert werden. Im Anschluss erfolgt eine Standardisierung der Merkmale mittels Z-Score, wodurch verhindert werden soll, dass die nachfolgende Hauptkomponentenanalyse durch Skaleneffekte einzelner Merkmale dominiert wird. Um die Dimensionalität für die abschließende Berechnung des Novelty Scores mittels kNN zu reduzieren, werden nur die ersten 10 Hauptkomponenten für die weitere Verarbeitung verwendet. Für die Bestimmung des Grenzwertes wird eine Methode basierend auf der Verteilung der im Training vorhandenen Novelty Scores verwendet, welche in Abschnitt 3.2.3 vorgestellt wird.

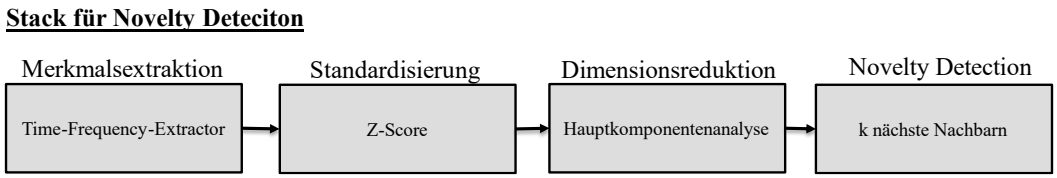


Abbildung 3-5: Stack für die im Rahmen der adaptiven Zustandsüberwachung verwendeten Novelty Detection.

In Abbildung 3-6 ist der verwendete Stack für die Klassifizierung dargestellt. In [184] wurden verschiedene Methoden der Merkmalsextraktion hinsichtlich ihrer Leistungsfähigkeit, insbesondere im Kontext der industriellen Zustandsüberwachung basierend auf überwachten Modellen, verglichen. Dabei zeigte sich, dass der TFEx eine überlegene Performanz aufwies, weshalb auch hier im ersten Schritt eine TFEx basierte Merkmalsextraktion erfolgt. Je nach Konfiguration des TFEx kann dabei eine große Zahl an teils hochkorrelierten Merkmalen entstehen. Durch die nachfolgende Merkmalselektion mittels MRMR kann die Anzahl der selektierten Merkmale unter Berücksichtigung dieser Korrelationen für die folgende Modellbildung reduziert werden. Die Bestimmung der optimalen Anzahl an Segmenten im Zeit- und Frequenzbereich sowie der Anzahl der für die Modellbildung verwendeten Merkmale erfolgt durch eine Bayes Optimierung. Die abschließende Klassifizierung erfolgt durch die Kombination von LDA und Mahalanobis Distanz. Auch die bei der Identifikation der Ursache unterstützende, überwachte Klassifizierung zur Trennung der historischen und der im Buffer abgelegten neuen Daten basiert auf diesem Stack.

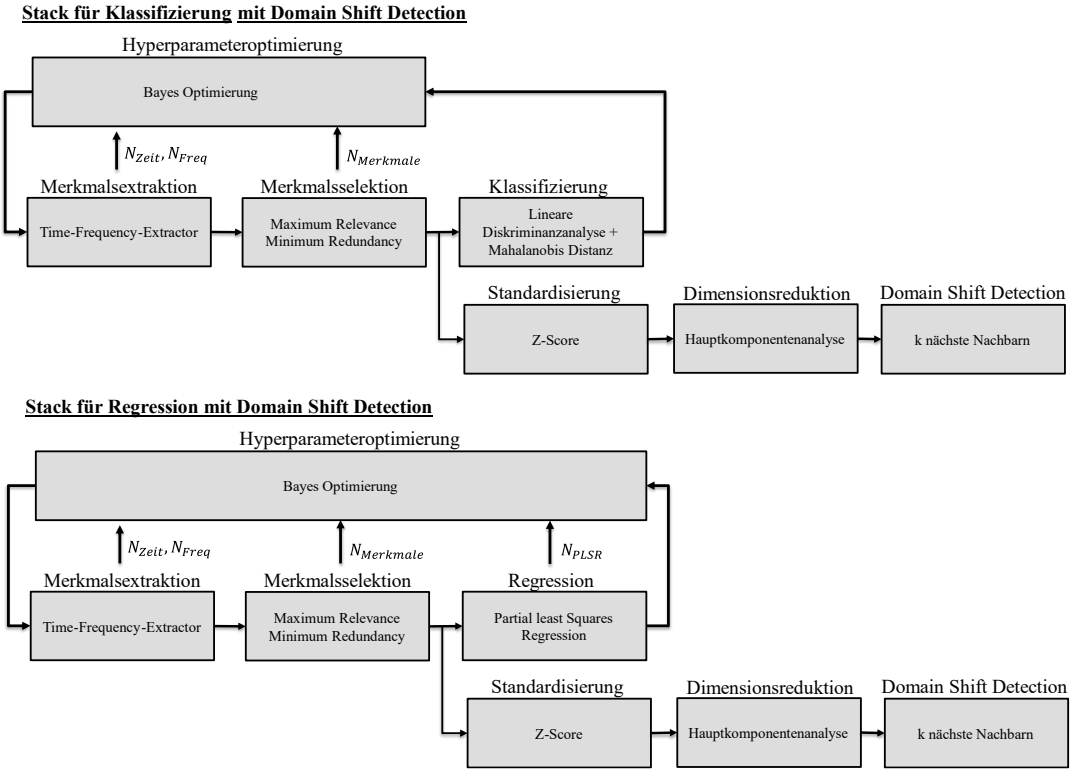


Abbildung 3-6: Stacks für die im Rahmen der adaptiven Zustandsüberwachung verwendeten Klassifizierung und Regression mit Domain Shift Detection.

Der für die Regression verwendete Stack ist ebenfalls in Abbildung 3-6 dargestellt. Dieser ist abgesehen von der verwendeten Partial Least Squares Regression identisch zum Stack für die Klassifizierung aufgebaut. Der durch die Partial Least Squares Regression hinzukommende Hyperparameter N_{PLSR} wird ebenfalls über die Bayes Optimierung bestimmt. Als Abbruchkriterium der Bayes Optimierung für

Klassifizierung und Regression werden anwendungsübergreifend 50 Iterationen definiert.

Um die Gültigkeit der überwachten Modelle sowohl im Fall einer Klassifizierung als auch der Regression festzustellen, erfolgt ausgehen von den in Abbildung 2-16 vorgestellten Eingriffsmöglichkeiten der Novelty Detection eine Domain Shift Detection. Dazu wird ein zweiter identischer Novelty Detection Stack auf den für das überwachte Modell selektierten Merkmalen trainiert. Ein erhöhter Domain Shift Score zeigt damit an, dass der betrachtete Datenpunkt keiner der im überwachten Modell enthaltenen Klassen zugeordnet werden kann. Dies kann einerseits auf einen bisher nicht im Modell berücksichtigten Fehlermechanismus hinweisen und andererseits auf veränderte Umwelteinflüsse, die die Performanz der überwachten Modelle beeinträchtigen.

3.2.3 Grenzwertbestimmung für die Novelty Detection

Eine zentrale Fragestellung bei der Anwendung teilüberwachter Verfahren stellt die Bestimmung des Grenzwertes dar, welcher definiert, ob ein unbekannter Datenpunkt als normal oder neuartig eingestuft wird. Wie in Abschnitt 2.3.6 bereits erläutert, werden hierfür typischerweise statische Grenzwerte verwendet, die auf Expertenwissen basieren. Dies führt einerseits zu einer subjektiven Komponente in der Grenzwertbestimmung und andererseits wird die Beschaffenheit der Daten nicht berücksichtigt. Zudem bestehen verschiedene überwachte Verfahren, welche eine Anpassung des Grenzwertes in Abhängigkeit bereits vorhandener Fehlerzustände ermöglichen, was im Falle der industriellen Zustandsüberwachung aufgrund der fehlenden Trainingsdaten häufig als nicht zielführend angesehen werden kann [161]. Aus diesem Grund wurde im Rahmen dieser Arbeit eine neuartige Methode zur Grenzwertbestimmung entwickelt, welche den Grenzwert in Anlehnung an die klassische Messunsicherheitsbetrachtung unter Berücksichtigung der für das Training vorhandenen Anzahl an Stichproben bestimmt. Die zugrundeliegende Annahme ist hierbei, dass bei einer höheren Anzahl an Stichproben mehr Information über die tatsächliche Verteilung der berechneten Scores vorhanden ist. Die Formulierung des Grenzwertes erfolgt durch den Mittelwert sowie ein Vielfaches der Standardabweichung wie folgt:

$$Grenzwert = \mu \pm (x \cdot \sigma) \tag{3.2}$$

Unter der Annahme, dass die Scores normalverteilt vorliegen, erfolgt die Anpassung von x dann in Abhängigkeit der Anzahl an Stichproben so, dass für eine Normalverteilung alle Novelty Scores innerhalb des Vertrauensintervalls liegen. Die daraus resultierende Entwicklung des positiven Grenzwertes in Abhängigkeit der

Anzahl an verfügbaren Stichproben ist in Abbildung 3-7 exemplarisch dargestellt. Dies führt im Falle normalverteilter Scores zu einem optimalen Grenzwert, welcher sehr empfindlich gegenüber systematischen Effekten und Ausreißern in den Daten ist.

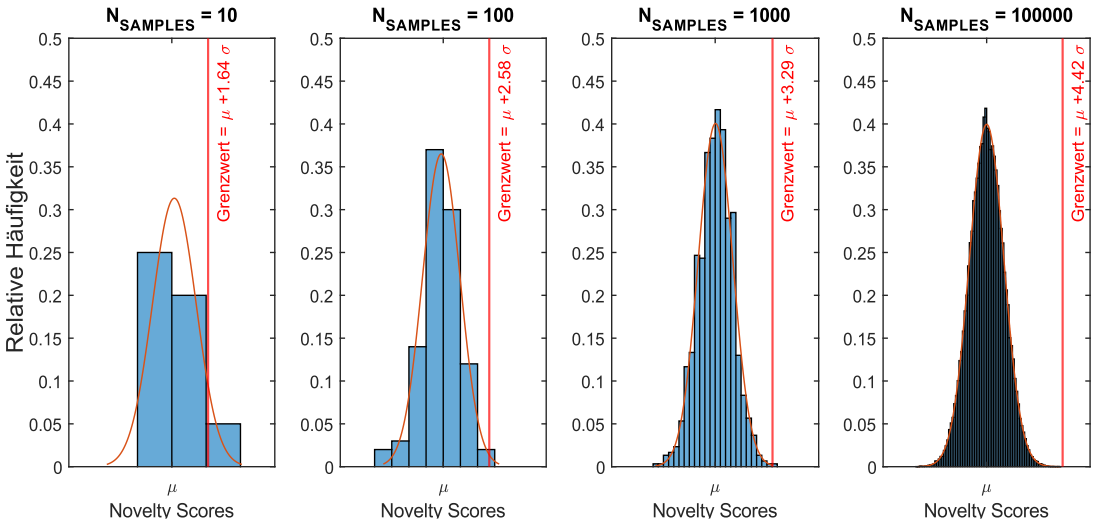


Abbildung 3-7: Entwicklung des Grenzwertes in Abhängigkeit der Anzahl der im Training verfügbaren Samples.

Die Novelty Detection der vorgestellten adaptiven Zustandsüberwachung wird im Rahmen dieser Arbeit anwendungsübergreifend mittels kNN durchgeführt. Die zugrundeliegenden Novelty Scores repräsentieren also die Summe der Distanzen zu den k nächsten Nachbarn. Da die Verteilung der Distanzen stark von den zugrundeliegenden Daten abhängt, ist eine pauschale Aussage über die Verteilung der Distanzen bei kNN nicht möglich. Jedoch kann unter der Annahme, dass keine systematischen Fehler in den Sensordaten vorhanden sind, für die berechneten Merkmale von einer Normalverteilung ausgegangen werden. In Abbildung 3-8 sind die unter Annahme normalverteilter Merkmale resultierenden Verteilungen der kNN-Distanzen mit k=5 und $N_{SAMPLES} = 1000$ dargestellt. Diese weisen im Fall niederdimensionaler Merkmalsräume ($N_{Dims} = 3$) eine deutlich unsymmetrische Verteilung (Exponential- bzw. Gammaverteilung [185]) auf, wodurch vereinzelt Datenpunkte aus dem Vertrauensintervall 3,29 σ ausreißen. Wie in Abbildung 3-8 zu sehen, nähert sich die Verteilung in Übereinstimmung mit dem zentralen Grenzwertsatz bei höheren Dimensionen ($N_{Dims} = 10$) an eine Normalverteilung an bzw. geht bei weiterem Erhöhen der Dimensionalität ($N_{Dims} = 100$) in eine Normalverteilung über [54]. Da im Rahmen der adaptiven Zustandsüberwachung anhaltende, systematische Änderungen des Systemzustandes detektiert werden sollen, wird die erhöhte Empfindlichkeit, welche bei einzelnen reversiblen Ausreißern zu False Positives führt, nicht als kritisch angesehen. Um den Einfluss der fehlerhaften Annahme der Normalverteilung abschätzen zu können, erfolgt in Kombination mit der Grenzwertbestimmung eine Prüfung der Novelty Scores mittels QQ-Diagramm.

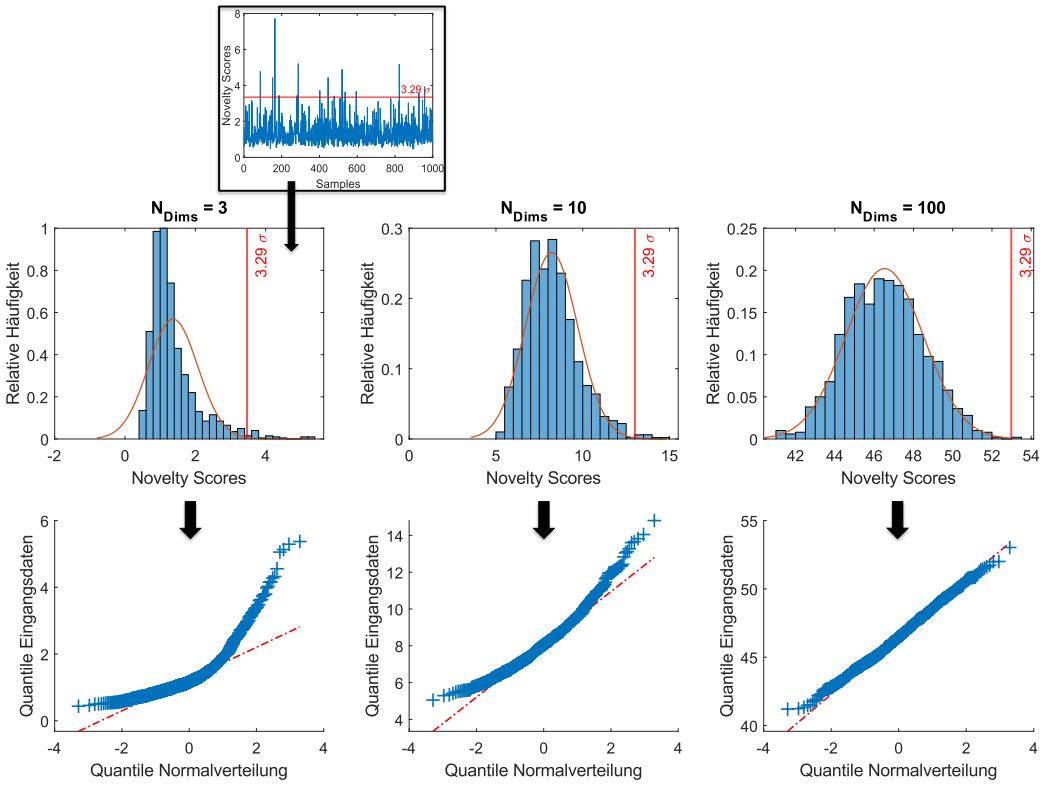


Abbildung 3-8: Entwicklung der zugrundeliegenden Verteilung der kNN basierten Novelty Scores in Abhängigkeit der Anzahl der Dimensionen des Merkmalsraums.

4 Anwendungsszenarien

Im Folgenden wird die entwickelte Methodik an verschiedenen Anwendungsszenarien validiert. Die Vorgehensweise ist dabei, wie in Kapitel 3 beschrieben, stark von der Datenlage sowie von der Art der Problemstellung abhängig. Aus diesem Grund erfolgen im ersten Schritt immer erst eine Beschreibung der Datenlandschaft und eine Definition des Ziels sowie die daraus abgeleiteten Anforderungen an ein Zustandsüberwachungssystem. Anschließend werden die Ergebnisse der einzelnen Untersuchungen vorgestellt und diskutiert. Die gezeigten Ergebnisse basieren dabei auf den für die entsprechenden Lernprobleme vorgestellten Stacks aus Kapitel 3.2.2. Ausgenommen der modellgestützten Identifikation der Fehlerursache ist die Methodik auch mit beliebigen anderen Pipelines für überwachtes Lernen bzw. Novelty Detection durchführbar.

Speziell bei Anwendung 1 „Zustandsüberwachung eines hydraulischen Systems“ handelt es sich um eine Laboranwendung, welche bereits in einer Vielzahl von Veröffentlichungen untersucht wurde [7, 11, 186, 187, 188]. Der öffentlich verfügbare Datensatz zeichnet sich vor allem durch eine hohe Datenqualität sowie ein hohes Maß an vorhandenem Prozesswissen aus. Aus diesem Grund erfolgte die Entwicklung der im Rahmen dieser Arbeit vorgestellten Methoden anhand dieses Datensatzes. Die dabei identifizierten Methoden wurden im Anschluss auf Anwendung 2 – 4 übertragen, um deren anwendungsübergreifende Leistungsfähigkeit zu validieren. Der Datensatz wird an dieser Stelle noch einmal für die Veranschaulichung der in realen industriellen Anwendungen auftauchenden Hürden für den Einsatz datengetriebener Modelle verwendet. Hierzu zählen vor allem sich ändernde Umgebungsbedingungen, welche zur Ungültigkeit der trainierten Modelle führen.

Auch bei Anwendung 2 „Zustandsüberwachung einer Schweißzange“ handelt es sich um eine Laboranwendung, deren Daten vom Projektpartner Festo SE & Co. KG aufgezeichnet und im Rahmen des Projektes MessMo zur Verfügung gestellt wurden.

Bei Anwendungen 3 „Akustische Leckagedetektion an Pneumatikzylindern“ und Anwendung 4 „Verschleißdetektion beim Wälzschälen“ wurde aktiv an der Mess- und Datenplanung sowie der Generierung der Daten mitgearbeitet. Im Falle von Anwendung 3 bezieht sich dies hauptsächlich auf iterative Auswertungen der während der Anfangsphase des Projektes von Festo zur Verfügung gestellten Daten, um die

strukturelle und informationsbezogene Datenqualität sicherzustellen. Bei Anwendung 4 wurde zudem ein Messsystem entwickelt und die Datenaufzeichnung an einer Versuchsanlage bei der Schaeffler Technologies AG & Co. KG selbst im Rahmen von insgesamt vier Messkampagnen durchgeführt.

4.1 Zustandsüberwachung eines Hydraulischen Systems

4.1.1 Einführung

Hydraulik ermöglicht eine effiziente Kraft- und Leistungsübertragung durch den statischen Druck von Fluiden. Die hohe Leistungsdichte von hydraulischen Systemen macht sie besonders für mobile Anwendungen wie Bau-, Forst- und Bergbaumaschinen interessant, wobei sie sowohl für die Arbeitsfunktionen als auch für die Fahrantriebe eingesetzt werden kann. Auch in stationären Anwendungen wie Pressen, Werkzeugmaschinen für Vorschübe und Spindelantriebe ist sie weit verbreitet. Neben der hohen Leistungsdichte bietet die Hydraulik viele weitere Vorteile wie beispielsweise eine stufenlose Regelung von Geschwindigkeit und Drehzahlen in weiten Bereichen, das Anfahren unter Last sowie eine hohe Flexibilität bei der Konstruktion, da die Komponenten räumlich verteilt angeordnet werden können, wobei die Verbindung über flexible Schläuche erfolgt. Als Nachteil hydraulischer Systeme werden häufig Leistungsverluste, hohe Anforderungen an und damit Kosten für die Genauigkeit der Bauteile sowie die temperaturabhängigen Viskositätsänderungen des Fluids und die dadurch benötigte Kühlung genannt. [189]

Durch die kontinuierliche Schmierung weisen hydraulische Systeme in der Regel geringe Verschleißraten auf. Trotzdem ist eine Überwachung spezieller, häufig auftretender Fehlermechanismen zur Steigerung der Anlagenverfügbarkeit wünschenswert. Des Weiteren entstehen in vielen Anwendungen während des Betriebs hohe Kräfte, was aus Sicherheitsgründen eine Überwachung einzelner funktionaler Komponenten erfordert.

Von Helwig wurde die Zustandsüberwachung hydraulischer Systeme auf Basis datengetriebener Modelle detailliert untersucht [186]. Dafür wurden im Labor eigene Prüfumgebungen bestehend aus dem mechanischen Aufbau, der Steuerung sowie der sensorischen Instrumentierung entwickelt, die das Generieren einer umfangreichen Trainingsdatenbasis gestatten. Untersucht wurde dabei vor allem die Detektion typischer hydraulischer Fehlermechanismen auf Basis überwachter Lernverfahren,

welche mit hoher Sensitivität auch bei gleichzeitigem Auftreten quantifiziert werden konnten. Nähere Informationen zu den erzielten Ergebnissen können [7, 8, 11, 186, 187] entnommen werden.

Ziel der Anwendung ist die Zustandsüberwachung eines hydraulischen Systems auf Basis der installierten Prozesssensoren. Hierbei sollen typische hydraulische Fehlermechanismen klassifiziert werden. Im Folgenden werden der Aufbau und die Funktion der Systeme sowie die Mess- und Datenplanung vorgestellt. Die Informationen dafür wurden aus [186] entnommen.

4.1.2 Experimenteller Aufbau

Das für die Datengenerierung verwendete System folgt dem Aufbau, wie in Abbildung 4-1 anhand des hydraulischen Schaltplans dargestellt. Das Hydraulikaggregat besteht aus einem Arbeitskreislauf (a) sowie einem sekundären Kühl- und Filtrierkreislauf (b). Beide Kreisläufe beziehen das Hydrauliköl aus einem Tank (Volumen 40 l), in den es nach Durchströmen des jeweiligen Kreislaufs wieder zurückfließt. Der Aufbau des Primärkreislaufes ist in Abbildung 4-1 dargestellt. Die Hauptkomponenten des Primärkreislaufs sind eine Innenzahnradpumpe (MP1), welche einen konstanten Volumenstrom von 7,5 l/min bereitstellt, und der zugehörige Elektromotor. Bei dem Elektromotor handelt es sich um einen 4-pol Asynchronmotor mit einer Leistung von 3 kW und einer festen Drehzahl von 1410 min^{-1} (50 Hz). Die Druckbegrenzungsventile V7 und V8 sind zu Sicherheitszwecken installiert und schützen das System vor Überdruck. Ein Rückschlagventil verhindert nach dem Abschalten den Rückfluss der Hydraulikflüssigkeit und somit ein Rückwärtsdrehen der Pumpe. Mit Hilfe des 2/2-Wegeventils V10 ist ein Zu- oder Abschalten der Last möglich. Die Last wird durch das Proportionaldruckbegrenzungsventil V11 simuliert, welches einen Gegendruck proportional zum Eingangsdruck aufbaut. Bei Überschreiten des eingestellten Drucks wird das Hydrauliköl weiter zum Tank geleitet, wodurch im Lastventil hauptsächlich eine Umwandlung hydraulischer Leistung in Wärme erfolgt. In den über die Ventile V1-V4 zuschaltbaren Membran-Hydraulikspeichern A1-A4 kann ein gewisses Maß an Druckenergie gespeichert werden, welche bei Lastsprüngen wieder an das System abgegeben werden kann.

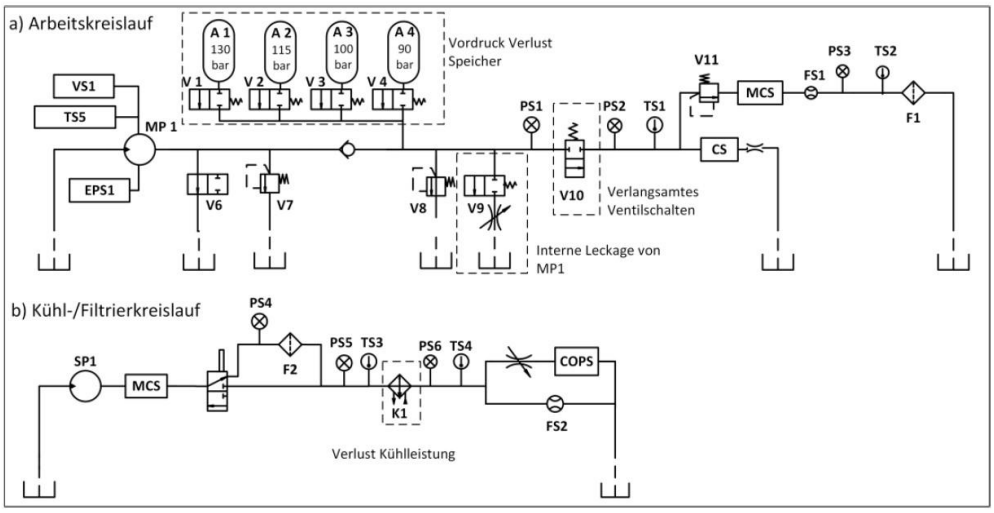


Abbildung 4-1: Hydraulischer Schaltplan der Prüfstände [186].

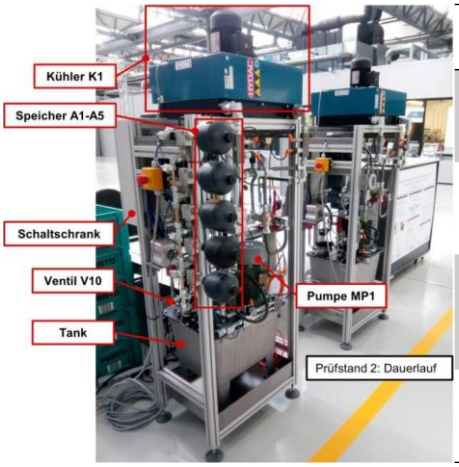
Der Sekundärkreis besteht hauptsächlich aus einer Außenzahnradpumpe SP1 mit dem zugehörigen 8-pol Asynchronmotor mit einer Leistung von 0,37 kW bei einer festen Drehzahl von 660 rpm. Die Pumpe liefert dabei einen konstanten Volumenstrom von 15 l/min, wodurch das gesamte Ölvolumen in ca. drei Minuten umgewälzt wird. Beim Umwälzen wird das Hydrauliköl durch einen Nebenstromfilter F2 sowie einen Öl-Luft-Kühler K1 geführt.

Insgesamt ist der Prüfstand zur Prozesssteuerung mit 14 Prozesssensoren ausgestattet. Eine Prozessüberwachung auf Basis dieser Sensoren wäre wünschenswert, da keine zusätzlichen Sensoren in das System gebracht werden müssten. Aus den Signalen der physikalischen Sensoren werden zudem drei virtuelle Sensoren abgeleitet. Eine vollständige Liste der Sensorbezeichnungen mit den entsprechenden Messgrößen sowie den zugehörigen Abtastraten ist in Tabelle 4-1 gegeben. Die genaue Position der Sensoren kann Abbildung 4-1 entnommen werden.

Tabelle 4-1: Liste der im hydraulischen System installierten Prozesssensoren.

Sensorbezeichnung	Messgröße	Abtastrate
TS1 – TS4	Temperatur [°C]	1 Hz
VS1	Vibration Effektivwert [mm/s]	1 Hz
FS1 und FS2	Volumenstrom [l/min]	10 Hz
EPS1	El. Motorleistung MP1 [W]	100 Hz
PS1 – PS6	Druck [bar]	100 Hz
CE	Virtuell: Kühlereffizienz [%]	1 Hz
CP	Virtuell: Kühlleistung [kW]	1 Hz
SE	Virtuell: System-Wirkungsgrad [%]	1 Hz

Um Trainingsdaten für typische hydraulische Fehlermechanismen zu generieren, wurden vier Komponenten des Systems dahingehend modifiziert, dass die Komponentenzustände gezielt über die Prüfstands-Steuerung variiert werden können. Die Auswahl der Fehlermechanismen zielt dabei primär auf die Zustandsüberwachung der Komponenten Pumpe, Wegeventil, Hydrospeicher und Kühler ab. Eine Zusammenfassung der steuerbaren Fehlermechanismen kann der Übersicht in Abbildung 4-2 entnommen werden. Eine detaillierte Beschreibung der simulierten Fehlermechanismen ist in [186] zu finden. Um die Langzeitstabilität sowie die Übertragbarkeit der Modelle zu testen, wurde ein zweites baugleiches System, allerdings ohne die Modifikationen zur Schadenssimulation, aufgebaut.



Komponente	Zielzustand	Steuerparameter	Möglicher Bereich
Kühler K1	Reduzierte Kühlleistung	Gepulste Ansteuerung des Lüfters	0…100 %
Wegeventil V10	Verzögertes Schaltverhalten	Ansteuerstromstärke V10	0…100 %
Pumpe MP1	Interne Leckage	Zuschaltbare Leckageströme (V9)	Keine, 3 x 0.2 ⌀, 3 x 0.25 ⌀
Hydro-speicher	Verlust Gasfülldruck	Umschalten der Speicher A1-A4	90, 100, 115, 130 bar

Abbildung 4-2: Hydraulische Systeme, welche in [186] für die Generierung der Trainingsdaten verwendet wurden. Prüfstand 1 wurde dabei für die umfangreiche Charakterisierungsmessung mit einzelnen und überlagerten Fehlermechanismen verwendet, wobei Prüfstand 2 für Stabilitätsuntersuchungen im Dauerbetrieb verwendet wurde [186].

Während eines Zyklus durchläuft das Aggregat ein definiertes Lastprofil, welches in Abbildung 4-3 anhand des Drucksensors P1 dargestellt ist. In der ersten Phase ist das 2/2-Wegeventil (V10) geschlossen, wodurch das 190 bar Druckbegrenzungsventil (V8) schaltet. In der zweiten Phase (nach 10 s) ist das 2/2-Wegeventil geöffnet und es werden mit Hilfe des Proportional-Druckbegrenzungsventils (V11) verschiedene Lastzustände simuliert. Der konstante Arbeitszyklus repräsentiert dabei den Ablauf einer typischen industriellen Anwendung wie beispielsweise eines zyklischen Umformvorgangs oder einer hydraulisch betriebenen Spritzgussmaschine.

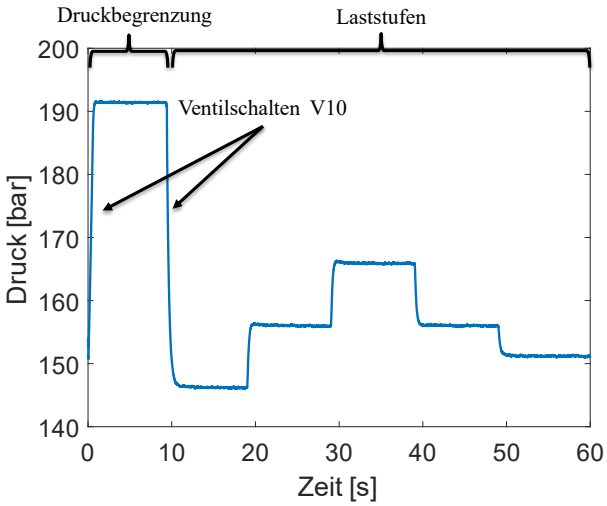


Abbildung 4-3: Arbeitszyklus des hydraulischen Aggregats dargestellt anhand des Signals des Drucksensors P1.

4.1.3 Mess- und Datenplanung

Da es sich bei diesem Anwendungsfall um einen öffentlich verfügbaren Datensatz handelt, liegen die Daten bereits in strukturierten Textdateien vor [188]. Eine systematische Mess- und Datenplanung zur Sicherstellung der Datenqualität wird dadurch nicht mehr benötigt.

Der Datensatz besteht aus insgesamt 2205 aufgezeichneten Zyklen, wobei ein Zyklus einer Dauer von 60 Sekunden entspricht. Während der Versuche werden diskrete Zustände der Komponenten Kühler, Ventil, Pumpe und Speicher nach einem vollfaktoriellen DoE variiert, wobei jede Konfiguration für 10 Zyklen konstant gehalten wird. Die diskreten Zustände der einzelnen Komponenten sowie ihre Bedeutung können Tabelle 4-2 entnommen werden. Der Ablauf des DoE ist in Abbildung 4-4 dargestellt. Die Intensitäten der simulierten Fehlermechanismen der einzelnen Komponenten sind dabei zwischen 100 und 0 skaliert, wobei 100 einer fehlerfreien Komponente und 0 einer Komponente kurz vor dem Ausfall entsprechen. Neben dem Zustand der Komponenten enthalten die Labels des Datensatzes ein Ignore Flag. Dieses gibt Bereiche der Messung an, in denen aufgrund der teilweise großen Zeitkonstanten des Systems kein klar definierter Zustand eingenommen wird. Hauptsächlich ist dies auf die langsam einlaufende Öltemperatur bei Variieren der Kühleffizienz zurückzuführen. Für die folgenden Auswertungen werden diese Zyklen entfernt, wodurch insgesamt 1449 Zyklen mit 144 verschiedenen Kombinationen zur Verfügung stehen.

Tabelle 4-2: Zusammenfassung der vorhandenen Zustände für die manipulierten Komponenten.

Komponente	Zustand	Klassen	Bedeutung
Kühler C1	Kühlereffizienz	100	Optimale Kühlereffizienz (100 %)
		20	Verringerte Kühlereffizienz (20 %)
		3	Kurz vor Ausfall (3 %)
Ventil V10	Schaltverhalten	100	Optimales Schaltverhalten (100 %)
		90	Leicht verzögert (90 %)
		80	Stark verzögert (80 %)
		73	Kurz vor Ausfall (73 %)
Pumpe MP1	Interne Pumpenleckage	100	Keine Leckage
		50	Schwache Leckage (⌀ 3 × 0.2 mm)
		0	Starke Leckage (⌀ 3 × 0.25 mm)
Speicher A1-A4	Vordruck	100	Optimaler Vordruck (130 bar)
		66	Leicht verringert (115 bar)
		28	Stark verringert (100 bar)
		0	Kurz vor Ausfall (90 bar)

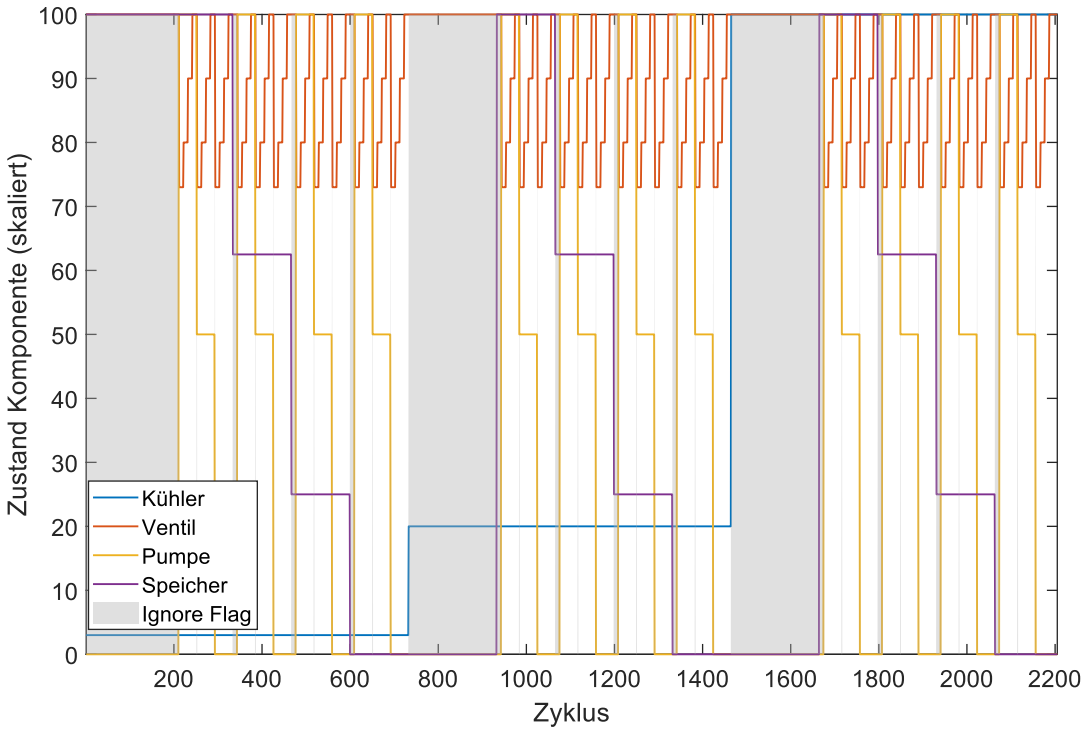


Abbildung 4-4: Vom DoE vorgegebene, gesteuerte Zustände der Komponenten Kühler, Ventil, Pumpe und Speicher während der Charakterisierungsmessung. In den grauen Bereichen ist der Zustand aufgrund größerer Zeitkonstanten des Systems z.B. bei einer sich ändernden Öltemperatur durch Variieren der Kühlereffizienz noch nicht eingeschwungen, weshalb der Systemzustand nicht klar definiert ist.

4.1.4 Modelle für überlagerte Fehler

Mittels des in Abbildung 4-5 gezeigten Stacks werden auf Basis des gesamten Datensatzes Modelle für eine überwachte Klassifizierung der einzelnen Fehlermechanismen gebildet. Für die Bayes Optimierung wird eine maximale Anzahl an 20 Iterationen definiert. Als Validierungsmethode wird hierbei eine randomisierte 10-fache Kreuzvalidierung verwendet. Wie in Abbildung 4-6 zu sehen ist, kann für jeden Fehlermechanismus ein nahezu fehlerfreies Modell hergeleitet werden. Einzig bei dem Modell für die Vordruckbestimmung des Hydraulikspeichers (d) kommt es zu einzelnen Fehlklassifizierungen.

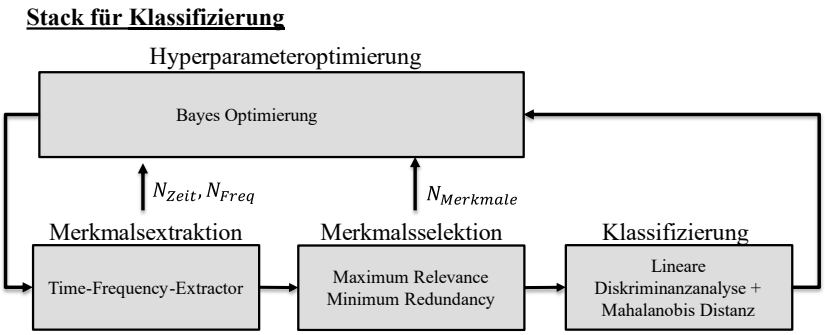


Abbildung 4-5: Verwendeter Stack für die überwachte Klassifizierung im Hydraulik Use Case.

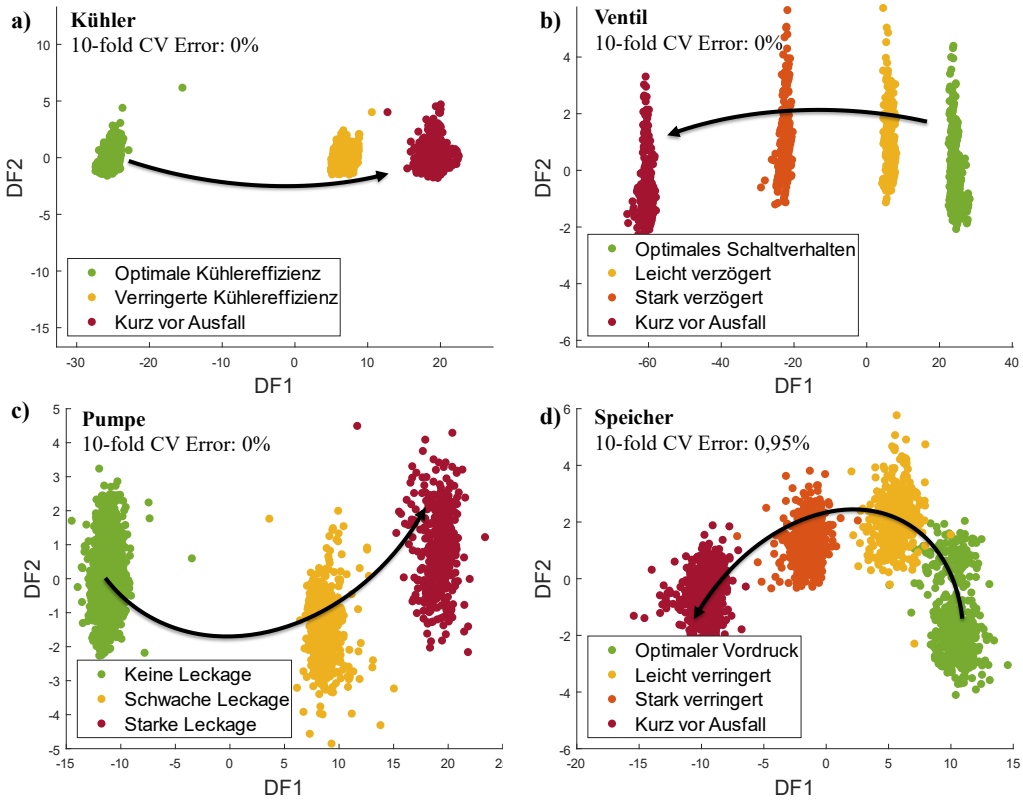


Abbildung 4-6: Ergebnisse der überwachten Modellbildung bei Nutzung des gesamten Datensatzes und einer randomisierten 10-fachen Kreuzvalidierung.

Die finalen, für die Modellbildung verwendeten Stacks mit den gefundenen Hyperparametern für die einzelnen Fehlermechanismen sind in Tabelle 4-3 zusammengefasst. Eine zusätzliche Übersicht über die Herkunft der selektierten Merkmale bezüglich der Position im Arbeitszyklus und der zugrundeliegenden Sensorsignale ist in Abbildung 4-7 dargestellt.

Tabelle 4-3: Durch Bayes Optimierung gefundene Hyperparameter für die Modellbildung im Hydraulik Use Case.

	N_{Zeit}	N_{Freq}	$N_{Merkmale}$
Kühler	4	0	3
Ventil	23	0	4
Pumpe	1	0	66
Speicher	19	0	204

Wie hier zu sehen ist, handelt es sich bei der Erkennung einer verringerten Kühlereffizienz durch die im System verbauten Temperatursensoren um ein triviales Lernproblem, welches ohne Methoden des maschinellen Lernens zu lösen ist. Aus diesem Grund werden die Kühlereffizienz und damit die Öltemperatur im Folgenden hauptsächlich als ein überlagerter Quereinfluss betrachtet, welcher verwendet werden kann, um die Robustheit der Modelle zu testen, bzw. den Einfluss solcher im Realfall häufig unbekannter Einflüsse zu veranschaulichen. Auch bei dem verschlechterten Ventilschaltverhalten handelt es sich um ein eher simples Lernproblem, welches sich im ersten Teil des Arbeitszyklus lokal stark begrenzt während der Schaltflanken zeigt. Aus diesem Grund werden auch nur 4 Merkmale für eine zuverlässige Bestimmung des Schaltverhaltens benötigt. Für die Fehlermechanismen der Pumpe und des Speichers werden die Modelle komplexer, was sich vor allem in der Anzahl der selektierten Merkmale und der Anzahl der zugrundeliegenden Sensorsignale zeigt. Wie in Abbildung 4-7 zu sehen, werden die Merkmale zur Bestimmung der internen Pumpenleckage aus dem gesamten Arbeitszyklus extrahiert. Dies ist dadurch zu erklären, dass die interne Pumpenleckage die Gesamteffizienz des Aggregates beeinflusst, weshalb die Merkmale vor allem auch aus dem Sensorsignal EPS1, also der Leistung des Antriebs der Hauptpumpe MP1 stammen. Im Fall des Speichers wird eine große Anzahl an Merkmalen verwendet, welche einerseits aus dem gesamten Arbeitszyklus extrahiert werden, andererseits vorwiegend aus kleineren Segmenten, welche Lastwechsel bzw. Drucksprünge beinhalten. Besonders hervorzuheben sind hierbei die Signale der Drucksensoren, des Flusssensors FS1 sowie des virtuellen Sensors SE, welcher den Gesamtwirkungsgrad des Systems beschreibt. Eine ausführlichere Plausibilisierung der Merkmale und Modelle erfolgt im Rahmen der adaptiven Zustandsüberwachung in den folgenden Abschnitten.

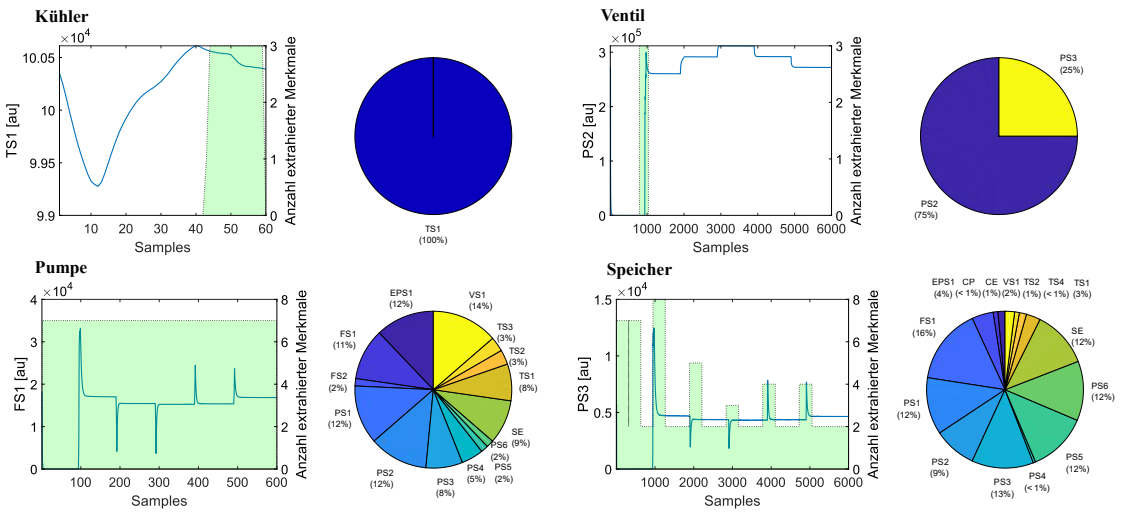


Abbildung 4-7: Herkunft der selektierten Merkmale für die Modelle der einzelnen Fehlermechanismen bezüglich der Position im Arbeitszyklus (links) und der zugrundeliegenden Sensorsignale (rechts). Die Darstellung der Segmente, aus denen die Merkmale extrahiert werden, erfolgt dabei anhand des Sensorsignals, welches den höchsten Anteil an der Gesamtkorrelation aufweist.

4.1.5 Einfluss von Domain Shifts

Im Folgenden soll untersucht werden, wie stabil sich die Modelle bei schwankenden Umgebungsbedingungen verhalten. Hierzu wird anhand der Kühlereffizienz eine steigende Öltemperatur im Aggregat simuliert. Denkbar wäre hier beispielsweise, dass die Trainingsdaten im Winter aufgezeichnet wurden und die Zustandsüberwachung im Sommer erfolgt. Um dies zu erreichen, werden die Modelle wie in Abschnitt 4.1.4 allerdings nur auf Basis der Daten mit optimaler Kühlereffizienz (100%) trainiert und anschließend erfolgt die Projektion der Daten mit verringerter Kühlereffizienz (20%) wie in Abbildung 4-8 a) dargestellt. Wie in Abbildung 4-8 b) bis d) zu erkennen, hat die Öltemperatur einen sehr unterschiedlichen Einfluss auf die Vorhersagegenauigkeit der Modelle für die einzelnen Fehlermechanismen. Im Falle des Ventilschaltverhaltens ist beispielweise kein Einfluss der Öltemperatur erkennbar. Der Einfluss bei der Erkennung der internen Pumpenleckage zeigt sich vor allem in einem Offset in der zweiten Diskriminanzfunktion, wodurch die Klassifizierungsrate allerdings noch nicht beeinflusst wird (Skalierung der Achsen beachten). Im Realfall wiederum wäre die Information über ein wegdriftendes Modell wünschenswert, um äußere Einflüsse beispielsweise mittels Offsetkorrektur bzw. Kalibrierung des Modells zu kompensieren. Den größten Einfluss zeigt die Öltemperatur auf die Detektion des Speichervordrucks. Hier werden für die gesamten Testdaten unabhängig vom Speicherdruck die Klasse mit optimalem Vordruck vorhergesagt. Im Realfall könnte ein sich ändernder Vordruck durch den überlagerten Einfluss der Öltemperatur nicht mehr erkannt werden. Hier stellt die Möglichkeit, im Betrieb weiterzulernen, die

einzige Möglichkeit dar, den Vordruck des Hydraulikspeichers weiterhin auf Basis datengetriebener Methoden zu überwachen. Unter Berücksichtigung der in Tabelle 4-3 dargestellten Komplexität der Modelle kann mit steigender Komplexität eine steigende Empfindlichkeit gegenüber Domain Shifts festgestellt werden.

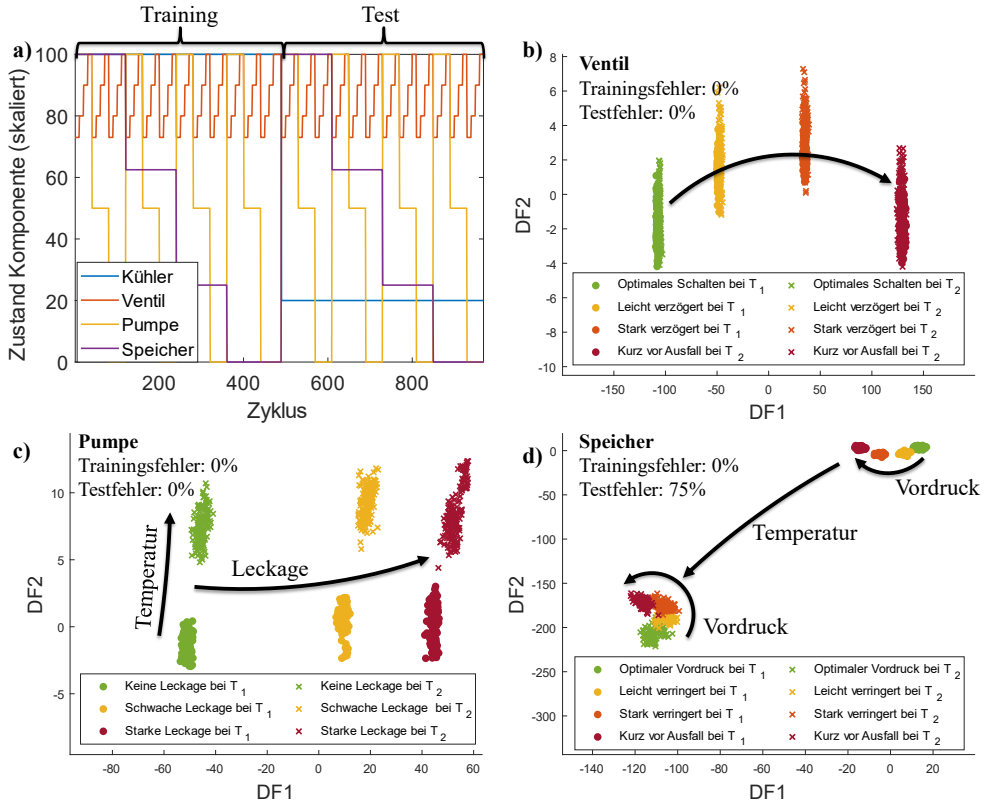


Abbildung 4-8: Einfluss einer sinkenden Kühlereffizienz bzw. der daraus resultierenden steigenden Öltemperatur auf die Vorhersagegenauigkeit der Modelle für die einzelnen Fehlermechanismen.

Im Kontext einer robusten Zustandsüberwachung müssen im ersten Schritt der Einfluss von Störgrößen und ihre Auswirkungen auf die Leistungsfähigkeit der überwachten Modelle mittels Domain Shift Detection festgestellt werden. Dafür wird mittels des in Abbildung 4-9 dargestellten Stacks eine Domain Shift Detection auf Basis der selektierten Merkmale bei voller Kühlereffizienz trainiert. Als Eingangsdaten für das kNN-Verfahren werden die 10 ersten Hauptkomponenten verwendet. Die Anzahl der nächsten Nachbarn wird auf k = 5 festgelegt.

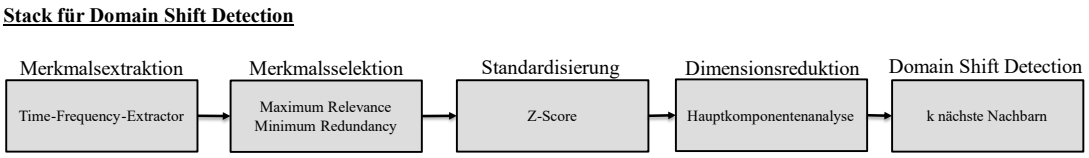


Abbildung 4-9: Stack für die Domain Shift Detection.

Die Ergebnisse sind anhand der Domain Shift Scores in Abbildung 4-10 dargestellt. Entsprechend der Resultate der überwachten Klassifizierung zeigen die Domain Shift Scores der Modelle für Ventil, Pumpe und Speicher eine sehr unterschiedliche

Empfindlichkeit gegenüber der verringerten Kühlereffizienz sowie der anderen überlagerten Fehlermechanismen (Skalierung der y-Achse beachten). Im Fall des Ventils zeigt sich insgesamt ein sehr geringer Einfluss, wobei sich eine systematische Änderung durch die überlagerten Fehlermechanismen ergibt. Im Gegensatz dazu zeigt sich bei den Modellen für die interne Pumpenleckage und den verringerten Vordruck der Speicher ein deutliches Abheben der Domain Shift Scores in Abhängigkeit der Kühlereffizienz. Das unterschiedliche Grundniveau der Domain Shift Scores lässt sich durch die unterschiedliche Komplexität der Modelle erklären. Da für das Modell des Ventils lediglich 4 Merkmale verwendet werden, stellen die Domain Shift Scores die Summe zu den 5 nächsten Nachbarn in einem 4-dimensionalen Merkmalsraum dar. Bei beiden anderen Modellen ist die Dimensionalität des Merkmalsraums deutlich höher, weshalb diese durch die Dimensionsreduktion mittels PCA auf die ersten 10 Hauptkomponenten begrenzt werden.

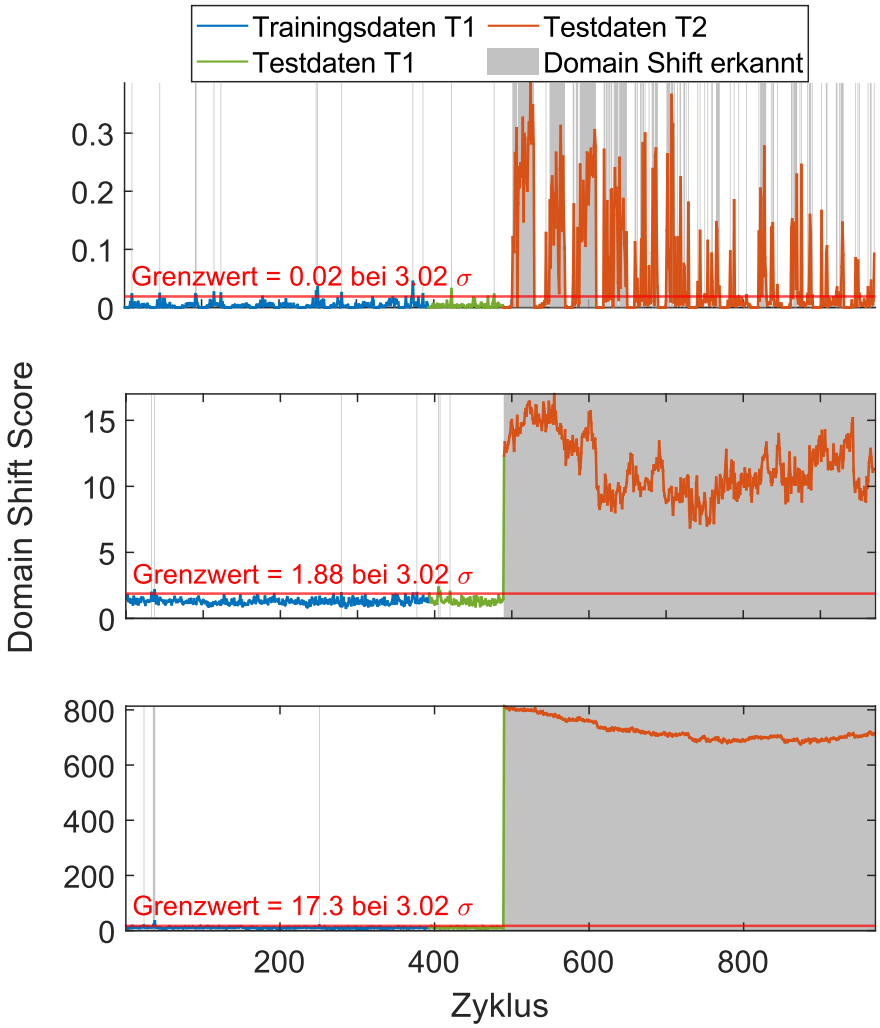


Abbildung 4-10: Domain Shift Scores für die überwachten Modelle der kombinierten Fehlermechanismen.

Als Problem zeigt sich an dieser Stelle das Festlegen eines Grenzwertes, bei dessen Überschreitung der Domain Shift Scores das Modell als ungültig angenommen werden kann. Die in Abschnitt 3.2.3 vorgestellte Methodik zeigt hierbei eine deutlich zu hohe Empfindlichkeit, was vor allem bei der Betrachtung der Domain Shift Scores für das fehlerhafte Ventilschaltverhalten deutlich wird. Obwohl hier ein offensichtlicher Anstieg der Domain Shift Scores über den definierten Grenzwert zu beobachten ist, zeigt das Modell in Abbildung 4-8 eindeutig, dass eine zuverlässige Bewertung des Ventilzustands auch bei verringerter Kühlereffizienz erhalten bleibt. Ein ähnliches Bild ergibt sich bei der Betrachtung der Domain Shift Scores für die Überwachung der Pumpe. Auch hier zeigt die Domain Shift Detection eine deutliche Beeinflussung des selektierten Merkmalspools an, obwohl das Klassifizierungsergebnis noch nicht beeinflusst wird. Lediglich bei den Domain Shift Scores der Leckage des Hydrospeichers zeigt sich eine Übereinstimmung mit dem Ergebnis der überwachten Klassifizierung.

Um den Einfluss eines Domain Shifts auf eine Novelty Detection zu demonstrieren, wird eine Novelty Detection für die Ventilüberwachung auf Basis des in Abbildung 4-12 dargestellten Stacks trainiert, wobei angelehnt an die Domain Shift Detection 23 Segmente im Zeitbereich und kein Segment im Frequenzbereich für den TFEx verwendet werden. Folglich werden keine Merkmale aus dem Frequenzbereich verwendet. Auch hier werden die ersten 10 Hauptkomponenten beibehalten und die Anzahl der nächsten Nachbarn wird ebenfalls auf k = 5 festgelegt. Die Darstellung erfolgt hier nur für einen Fehlermechanismus (Ventilschaltverhalten), da die Auswirkungen auf alle Modelle identisch sind.

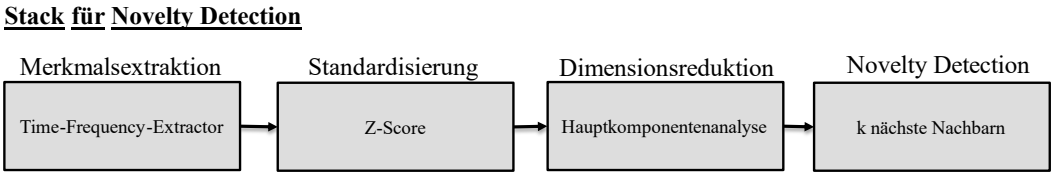


Abbildung 4-11: Stack für die Novelty Detection für das Ventilschaltverhalten.

Die resultierenden Novelty Scores sind in Abbildung 4-12 dargestellt. Im Vergleich zu dem Ergebnis der Domain Shift Detection ist hier deutlich zu sehen, dass die Novelty Scores massiv von der Kühlereffizienz beeinflusst werden. Der Grund liegt hierbei in der Herkunft der Merkmale. Im Gegensatz zur Domain Shift Detection, welche auf lokal beschränkten Merkmalen basiert (Schaltflanke Ventil), werden die Merkmale in diesem Fall breitbandig aus dem gesamten Arbeitszyklus extrahiert. Dies begründet auch die Ähnlichkeit der hier gezeigten Novelty Detection mit der Domain Shift Detection für das Modell des Speichers. Durch die hohe Komplexität des überwachten Modells des Speichers und der verteilt über den gesamten Arbeitszyklus

extrahierten Merkmale, nähert sich die Domain Shift Detection der Novelty Detection an.

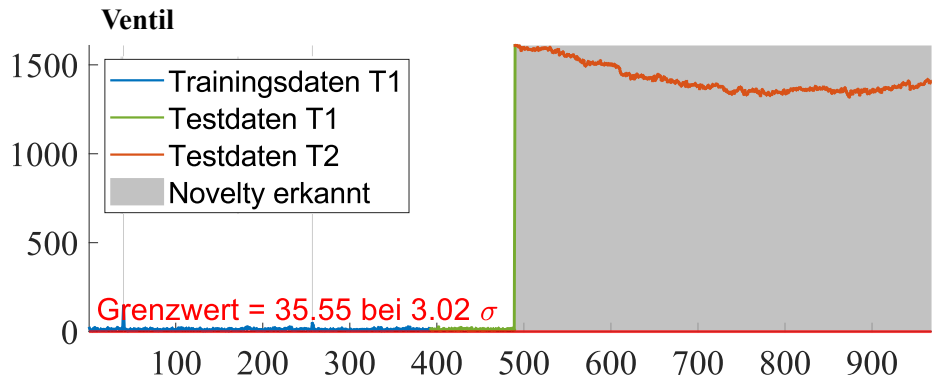


Abbildung 4-12: Novelty Detection Modell für das Ventilschaltverhalten und der Einfluss auf die Novelty Scores bei Eintreten einer verringerten Kühlereffizienz.

4.1.6 Adaptive Zustandsüberwachung

Auf Basis des in Abschnitt 3.2 vorgestellten Konzepts wird nun eine realitätsnahe Zustandsüberwachung des Hydraulikaggregates simuliert. Hierzu wird der Datensatz zunächst anhand der Schadensfälle sowie deren Intensitäten sortiert. Dadurch soll ein sequenzielles Auftreten einzelner Fehlermechanismen mit fortschreitender Intensität simuliert werden. Zusätzlich werden die Daten jedes Schadenfalls in 80 % Trainings- und 20 % Testdaten aufgeteilt, um eine Validierung der Modelle beim fortschreitenden Lernen der einzelnen Fehlermechanismen durchzuführen. Da nur das Auftreten isolierter Fehlermechanismen betrachtet wird, steht im Vergleich zur in Abbildung 4-4 gezeigten Charakterisierungsmessung eine erheblich reduzierte Anzahl an Stichproben zur Verfügung. Die Kühlereffizienz wird in diesem Fall nicht mehr als gesonderter Fehlermechanismus, sondern lediglich als überlagerter Quereinfluss betrachtet. Bei Start der Zustandsüberwachung wird dabei eine optimale Kühlereffizienz angenommen. In Abbildung 4-13 ist die Sortierung des Datensatzes für die Simulation isoliert auftretender Fehlermechanismen in steigender Intensität dargestellt. Für die adaptive Zustandsüberwachung wird davon ausgegangen, dass noch kein überwachtes Modell, also noch keine Trainingsdaten für bekannte Fehlermechanismen vorhanden sind. Die grafische Darstellung der adaptiven Zustandsüberwachung erfolgt dabei nur bis zum Eintreten der verringerten Kühlereffizienz. Der weitere Lernprozess wird in tabellarischer Form zusammengefasst dargestellt.

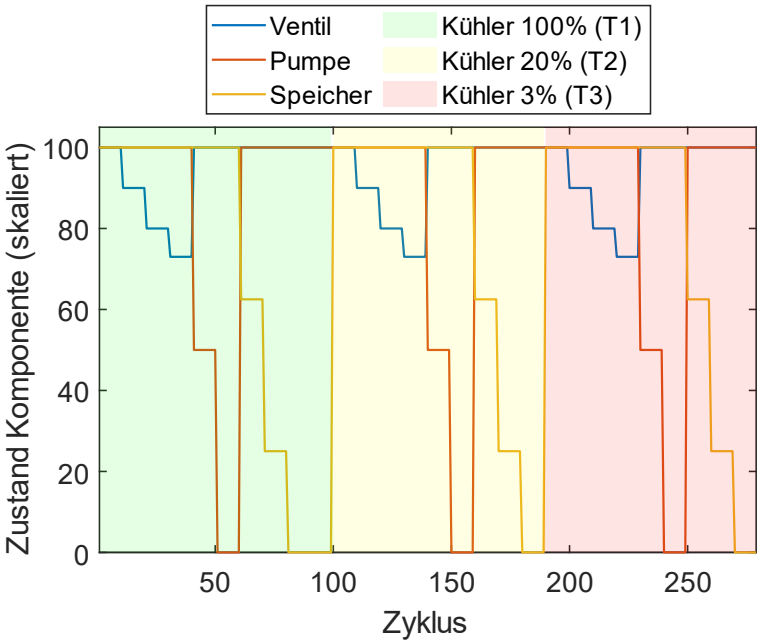


Abbildung 4-13: Reduzierter Datensatz für die Simulation isoliert auftretender Fehlermechanismen in steigender Intensität.

Im ersten Schritt erfolgt die Konfiguration des in Abschnitt 3.2.2 vorgestellten Stacks für die Novelty Detection. Für den TFEx wird hierbei eine Segmentierung in 13 Intervalle im Zeitbereich verwendet. Dadurch wird eine ausreichende zeitliche Auflösung erzeugt, um gezielt Merkmale aus einzelnen Abschnitten (Flanken und Plateaus vgl. Abbildung 4-3) des Arbeitszyklus zu extrahieren. Da keiner der Sensoren Signale periodischer Natur liefert, werden 0 Segmente im Frequenzbereich verwendet, was bedeutet, dass die Merkmalsextraktion rein im Zeitbereich stattfindet. An dieser Stelle sei nochmals angemerkt, dass es sich bei der Anzahl an Intervallen in Zeit- und Frequenzbereich um Hyperparameter handelt, welche vor dem Training der Novelty Detection auf Basis von Prozesswissen definiert werden müssen. Im Fall einer überwachten Modellbildung kann ebenfalls die in Abschnitt 2.3.8.2 vorgestellte Bayes Optimierung verwendet werden. Aufgrund der begrenzten Anzahl an Stichproben (ca. 10 pro Variation, gesamt 279) wird in diesem Fall für kNN ein k von 3 verwendet.

4.1.6.1 Initialisierung des Normalzustands

Die Initialisierung des Modells im Normalzustand der Anlage auf Basis der Novelty Detektion ist in Abbildung 4-14 a) anhand des Histogramms dargestellt. Das Training erfolgt hier mit 80 % der verfügbaren Normaldaten (8 Samples) und der Grenzwert wird auf 122,23 bei 1,53 σ festgelegt. Lediglich der erste Zyklus überschreitet den Grenzwert und würde als Anomalie detektiert. Da es sich hierbei aber nicht um eine anhaltende Änderung des Systemzustandes handelt, kann dieser Ausreißer

vernachlässigt werden. Wie in Abbildung 4-14 b) zu sehen ist, liegen die Daten annähernd auf einer Geraden im QQ-Diagramm. Auch der SW-Test verwirft die Hypothese auf Normalverteilung nicht, weshalb von einer Normalverteilung der Daten ausgegangen werden kann.

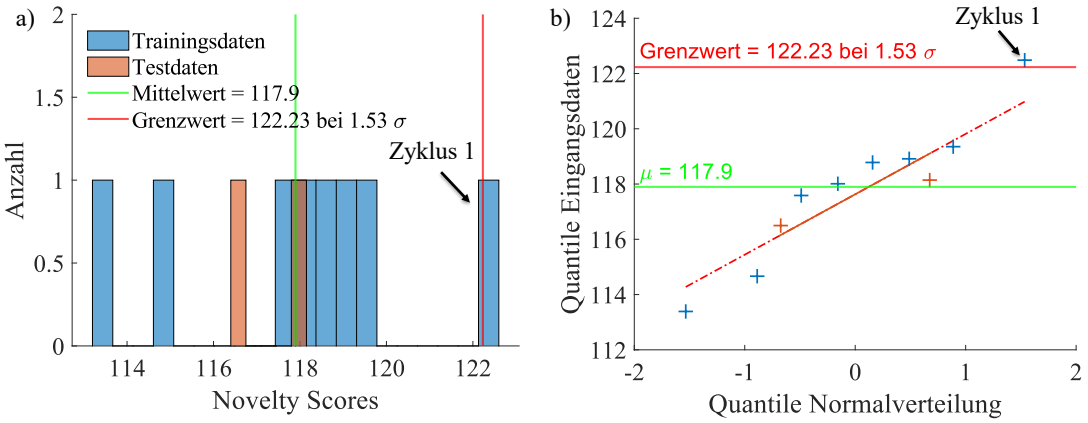


Abbildung 4-14: a) Histogramm der Novelty Scores der Trainingsdaten im Normalzustand, b) QQ-Diagramm der Novelty Scores der Trainingsdaten im Normalzustand.

4.1.6.2 Verschlechtertes Ventilschaltverhalten

Nach der Initialisierung des Modells wird das Auftreten des ersten Fehlerzustandes, in diesem Fall das verzögerte Schaltverhalten des Ventils V10, simuliert. Die zugehörigen Novelty Scores mit bekannter Zielgröße und entsprechender Ausgabe der Novelty Detection sind in Abbildung 4-15 a) dargestellt. Wie hier zu sehen ist, wird bereits das leicht verzögerte Schaltverhalten als Anomalie erkannt. Des Weiteren ist eine deutliche Korrelation zwischen dem Novelty Score und dem Schaltverhalten des Ventils zu beobachten.

Im nächsten Schritt wird das überwachte Modell zur Unterscheidung des Normalzustands und des Fehlers auf Basis der Ausgabe der Novelty Detektion trainiert. Das Ergebnis der Modellbildung ist in Abbildung 4-15 b) anhand der ersten Diskriminanzfunktion sowie der vorhergesagten Zustände des Modells dargestellt. Das Ergebnis wird unter Verwendung von 13 Zeitsegmenten des TFEx und einem selektierten Merkmal erreicht. Die Klassifizierungsrate auf Trainings- und Testdaten beträgt 100 %.

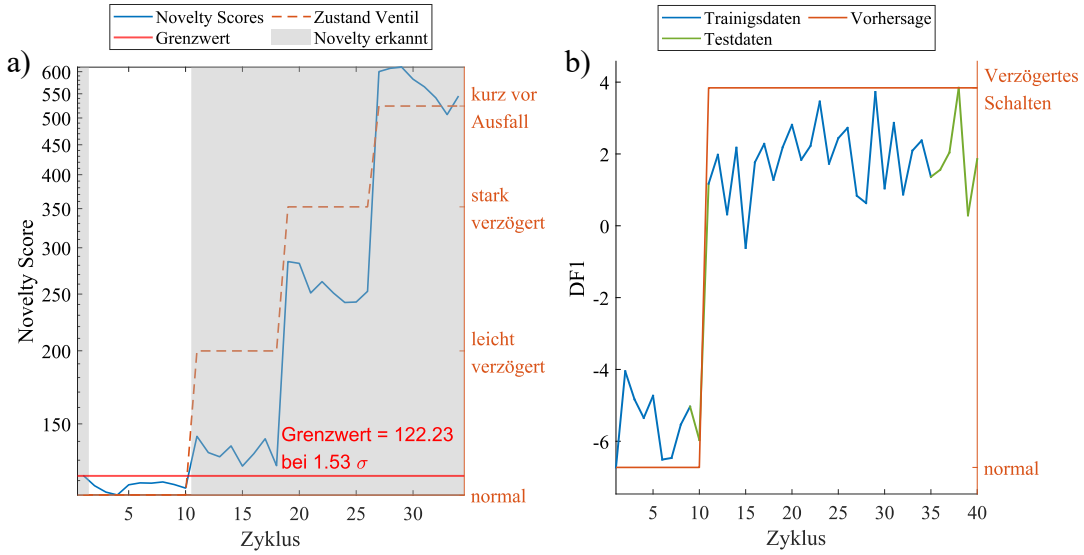


Abbildung 4-15: a) Reaktion der Novelty Detection bei Auftreten des verzögerten Schaltens von V10, b) Ergebnis der überwachten Klassifizierung anhand der ersten Diskriminanzfunktion (DF1) für Training (blau) und Test (grün) auf Basis der Ausgabe der Novelty Detection sowie der Vorhersage des Modells (orange).

Eine Analyse des für die Modellbildung verwendeten Merkmals zeigt, dass dieses aus dem in Abbildung 4-16 dargestellten Zeitsegment stammt. Berechnet wird das Merkmal mittels Kurtosis aus dem Signal des Drucksensors PS2 und kann damit als Indikator für die Steilheit der Schaltflanke gesehen werden. In der Anwendung wäre dies ein deutlicher Hinweis, dass der von der Novelty Detection erkannte Fehlerzustand vom Ventil V10 verursacht wird.

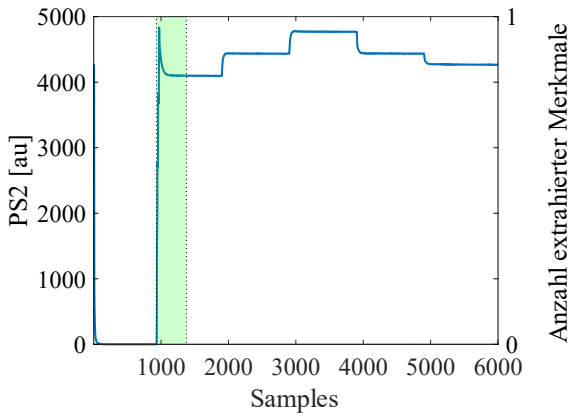


Abbildung 4-16: Position des höchstkorrelierten, identifizierten Merkmals für die Erkennung eines fehlerhaften Schaltverhaltens von V10, berechnet mittels Kurtosis aus dem Signal des Sensors PS2, dem Drucksensor, der sich unmittelbar hinter dem Ventil befindet.

Nach der Identifikation der Ursache des Fehlerzustandes und dem Training des überwachten Modells kann die Novelty Detection unter Einbezug der Daten des nun bekannten Fehlers erneut trainiert werden. Dadurch wird die Novelty Detection für diesen Fehlerzustand desensibilisiert, weshalb bei erneutem Auftreten des Fehlers eine Klassifizierung durch das überwachte Modell erfolgen würde. Für das Training der

Novelty Detection bzw. die Bestimmung des Grenzwertes, muss wiederum überprüft werden, ob die Novelty Scores einer Normalverteilung folgen. In Abbildung 4-17 a) ist das entsprechende QQ-Diagramm mit resultierendem Grenzwert dargestellt. Auch in diesem Fall kann anhand des QQ-Diagramms sowie der nicht zurückgewiesenen Nullhypothese des SW-Tests davon ausgegangen werden, dass die Novelty Scores normalverteilt sind.

Neben der Novelty Detection zur Identifikation neuer Systemzustände erfolgt an dieser Stelle das erste Training der Domain Shift Detektion. Die entsprechenden Domain Shift Scores sind in Abbildung 4-17 b) in Form eines weiteren QQ-Diagramms dargestellt. Abgesehen von wenigen Ausreißern liegen die Domain Shift Scores innerhalb eines bestimmten Grenzwertes. Zu beachten ist die Skalierung der x-Achse, welche zeigt, dass die Absolutwerte der Domain Shift Scores auf einer Skala kleiner 1 liegen. Dies ist durch die geringe Dimensionalität (ein selektiertes Merkmal) des Merkmalsraums zu erklären. Zusätzlich ist eine deutliche Abweichung der Domain Shift Scores von der Normalverteilung zu beobachten, weshalb auch der SW-Test in diesem Fall die Nullhypothese verwirft.

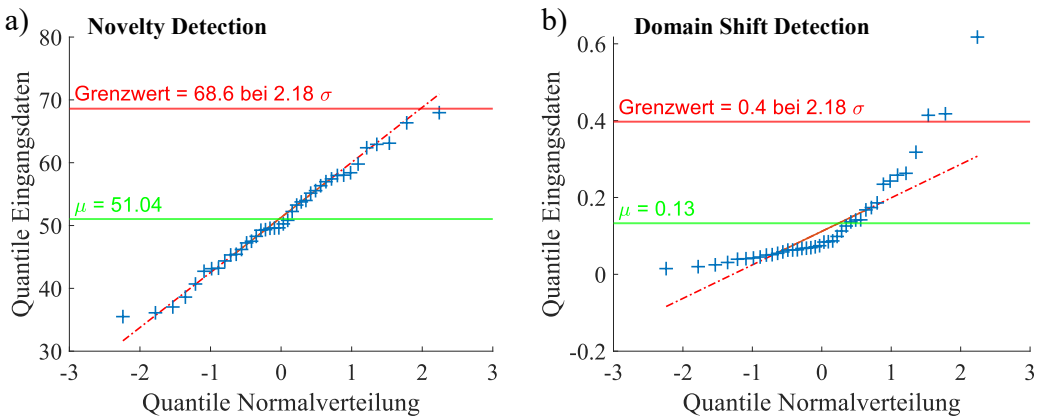


Abbildung 4-17: a) Initialisierung der Novelty Detection a) und der Domain Shift Detection b) nach der Detektion eines verschlechterten Ventilschaltverhaltens von V10.

4.1.6.3 Interne Pumpenleckage

Nach dem Update der Modelle wird das Auftreten einer internen Pumpenleckage simuliert. Hierzu erfolgt im ersten Schritt die Auswertung mittels Novelty Detection. Das Ergebnis ist wiederum anhand der sich ergebenden Novelty Scores in Abbildung 4-18 dargestellt. Auch in diesem Fall wird der fehlerhafte Systemzustand zuverlässig von der Novelty Detection erkannt und die Novelty Scores weisen eine klare Korrelation mit der Intensität der internen Pumpenleckage auf.

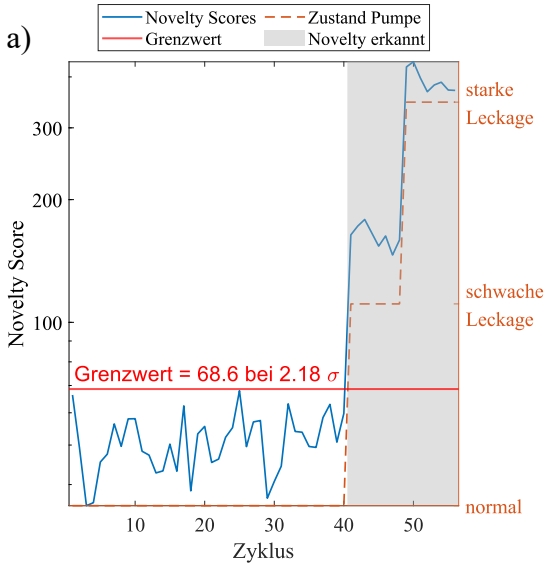


Abbildung 4-18: Reaktion der Novelty Detection auf eine interne Pumpenleckage.

Um die Reaktion einer statischen Ventilüberwachung, basierend auf einem rein überwachten Modell, zu demonstrieren, ist in Abbildung 4-19 a) die Reaktion des überwachten Modells dargestellt. Hier würde das System fälschlicherweise den Normalzustand des Hydraulikaggregates anzeigen. Vergleichbar zur Reaktion der Novelty Detektion zeigt die Domain Shift Detection eine Beeinflussung des überwachten Modells, weshalb dieses auch für eine belastbare Überwachung des Ventilschaltverhaltens nicht mehr zu verwenden wäre. Die Ausgabe der Domain Shift Detection ist in Abbildung 4-19 b) dargestellt.

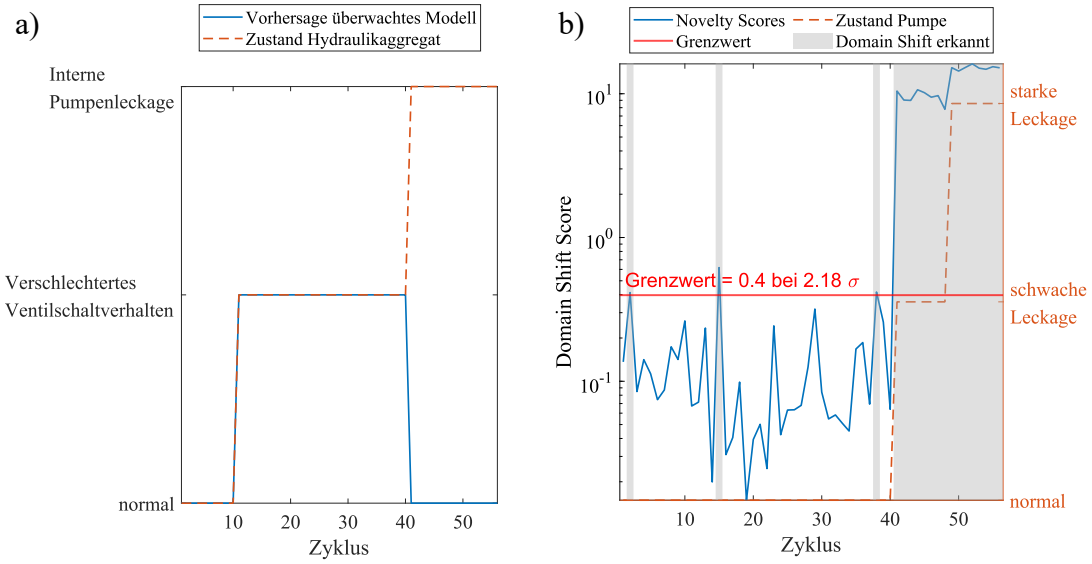


Abbildung 4-19: a) Reaktion der überwachten Klassifizierung und b) der Domain Shift Detection auf eine interne Pumpenleckage.

Im nächsten Schritt erfolgt wiederum die Rückführung des von der Novelty Detection erkannten neuen Zustands in das überwachte Modell. Zur Unterstützung der Identifikation der Ursache des neuen Zustands wird wiederum ein Modell zur Separierung der bekannten Daten und der Daten des neuen, unbekannten Zustands trainiert. Die Bayes Optimierung ergibt hier die Verwendung von 20 Zeitsegmenten und 5 selektierten Merkmalen. Die selektierten Merkmale der unterstützenden, überwachten Klassifizierung stammen dabei aus den Signalen der Fluss- und Drucksensoren im Bereich der Laststufen sowie der Motorleistung aus dem Gesamtzyklus. Die Intervalle, aus denen die höchstkorrelierten Merkmale (RMS und maximal Wert) aus dem Signal des Flusssensors FS1 extrahiert werden, sind in Abbildung 4-20 a) dargestellt. Die Kombination der Merkmale aus Fluss, Druck und der Motorleistung kann in diesem Fall einen Hinweis auf eine Problematik mit der Pumpe deuten. Dies setzt allerdings ein tieferes Systemverständnis bzw. Prozesswissen voraus. Nach der erfolgreichen Identifikation der Pumpenleckage, kann der Fehler in das überwachte Modell übernommen werden. Das Ergebnis der Modellbildung ist in Abbildung 4-20 b) anhand eines Streudiagramms dargestellt. Das Ergebnis wird in diesem Fall unter Verwendung von 11 Zeitsegmenten und 14 Merkmalen erreicht.

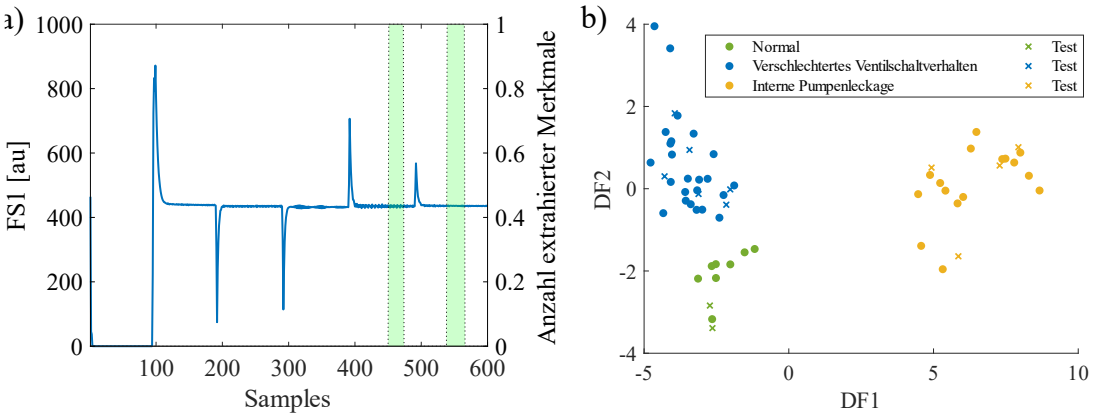


Abbildung 4-20: a) Herkunft des höchstkorrelierten Merkmals des überwachten Modells zur Trennung der bekannten Zustände vom neuen unbekannten Zustand. b) Ergebnis des überwachten Modells zur Trennung der bekannten Zustände des Hydraulikaggregats.

Nach der Rückkopplung des nun bekannten Fehlers in das überwachte Modell erfolgt wiederum das erneute Training der Novelty Detection und der Domain Shift Detection. Die Resultate für beide Modelle sind in Abbildung 4-21 in Form eines QQ-Diagramms dargestellt. Wie hier zu sehen, weichen sowohl die Novelty Scores als auch die Domain Shift Scores von der Normalverteilung ab und auch der SW-Test verwirft in beiden Fällen die Nullhypothese. Ursache hierfür stellt die in Abschnitt 3.2.3 vorgestellte systematische Abweichung der kNN-Distanzen von der Normalverteilung

dar, weshalb der Grenzwert zu sensitiv konfiguriert wird. Hierdurch werden bereits vereinzelt Stichproben als Novelty erkannt.

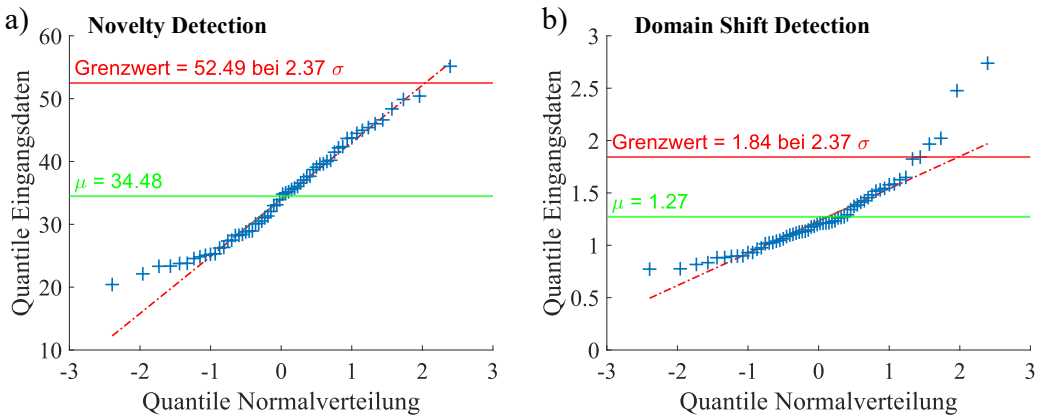


Abbildung 4-21: Initialisierung der Novelty Detection a) und der Domain Shift Detection b) nach der Detektion einer internen Pumpenleckage.

4.1.6.4 Verringerter Vordruck des Hydrospeichers

Im folgenden Schritt wird das Auftreten des nächsten Fehlers, in diesem Fall eines verringerten Vordrucks des Hydrospeichers, simuliert. Die resultierenden Novelty Scores sind in Abbildung 4-22 dargestellt. Im Vergleich zu den vorherigen Schadenfällen wird in diesem Fall der verringerte Vordruck des Hydrospeichers erst kurz vor dem Ausfall sicher detektiert. Im Bereich des leicht verringerten Vordrucks lassen sich deutlich drei Ausreißer erkennen. Die Analyse der entsprechenden Zyklen zeigt, dass diese auf Unregelmäßigkeiten im Ablauf des Arbeitszyklus zurückzuführen sind. Da diese offensichtlich keinen anhaltenden Wechsel des Systemzustandes einleiten, werden diese nicht weiter berücksichtigt.

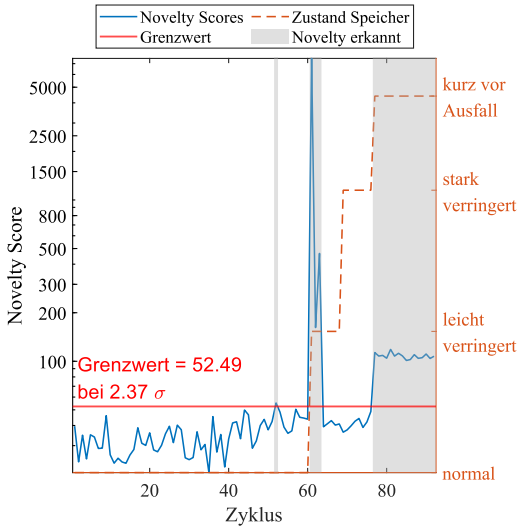


Abbildung 4-22: Reaktion der Novelty Detektion auf einen verringerten Vordruck des Hydrospeichers.

In Abbildung 4-23 a) ist die entsprechende Ausgabe des überwachten Modells dargestellt. In diesem Fall würde das überwachte Modell ein fehlerhaftes Ventilschaltverhalten anzeigen. Die Reaktion der Domain Shift Detection auf das Auftreten des verringerten Speichervordrucks ist in Abbildung 4-23 b) dargestellt. Diese zeigt bereits bei leicht verringertem Vordruck die Beeinflussung des überwachten Modells an. Die Scores zeigen dabei aber lediglich einen Zusammenhang zwischen den Zuständen „leicht verringerter“ und „kurz vor dem Ausfall“. Zusätzlich wird die Beeinflussung des überwachten Modells bereits mehrere Samples vor dem Eintreten des verringerten Vordrucks angezeigt. Auch an dieser Stelle wäre eine weitere Nutzung des bis dahin trainierten, überwachten Modells fragwürdig und die Funktion müsste erneut validiert werden.

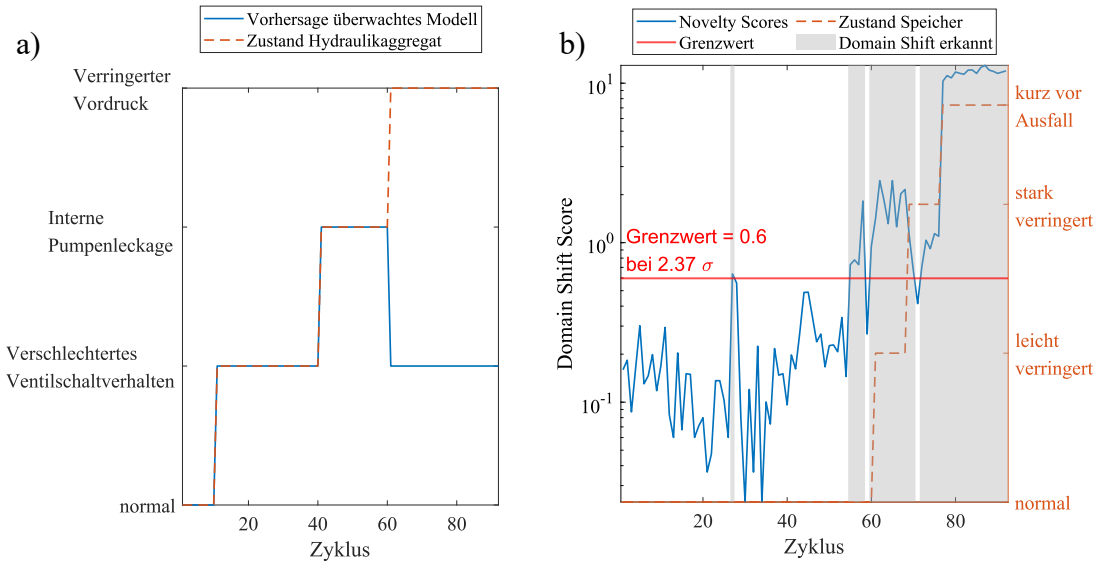


Abbildung 4-23: a) Reaktion der überwachten Klassifizierung und b) der Domain Shift Detektion auf einen verringerten Vordruck des Hydrospeichers.

Auch hier erfolgt wieder die Rückkopplung des unbekannten Fehlermechanismus in das überwachte Modell. Auf Basis der Ausgabe der Novelty Detection wird zunächst ein Modell zur Identifikation des neuen Zustands trainiert. Da die Novelty Detection den verringerten Vordruck erst kurz vor dem Ausfall detektiert, werden auch nur diese Daten für das Training des Plausibilisierungsmodells verwendet. In der Anwendung kann das Modell dann ebenfalls zur Analyse der Daten vor Eintritt des neuen Zustands verwendet werden. Das Ergebnis ist in Abbildung 4-24 a) anhand der ersten Diskriminanzfunktion dargestellt. Hier ist deutlich zu sehen, dass die erste Diskriminanzfunktion im Gegensatz zur Novelty Detection bereits eine Korrelation mit dem leicht und stark verringerten Vordruck aufweist, obwohl die Daten dieser Zustände nicht für das Training verwendet wurden. Dies deutet darauf hin, dass das Modell einen physikalischen Zusammenhang gelernt hat, welcher sich ebenfalls in den selektierten Merkmalen widerspiegelt. Insgesamt wurden durch die Bayes

Optimierung 8 Merkmale selektiert, wobei eine Segmentierung in ebenfalls 8 Intervalle verwendet wurde. Der Ursprung der beiden höchstkorrelierten Merkmale ist in Abbildung 4-24 b) dargestellt, welche beide die Varianz des Drucksensors PS1 während Drucksprüngen repräsentieren. Da sich gerade in diesen Bereichen das Lade- und Entladeverhalten der Hydraulikspeicher zeigt, kann dies als Hinweis auf eine Problematik mit den Speichern verstanden werden.

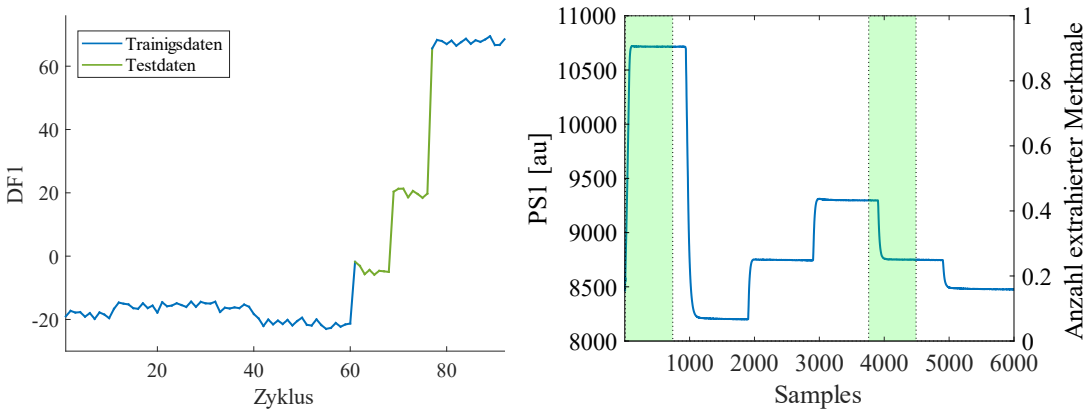


Abbildung 4-24: Modell zur Trennung des unbekannten Zustands von den bereits bekannten.

Im Anschluss an die Identifikation der Fehlerursache erfolgt das Training der überwachten Klassifizierung mit allen bekannten Fehlermechanismen. Das Ergebnis der Modellbildung ist in Abbildung 4-25 anhand der ersten beiden Diskriminanzfunktionen in Form eines Streudiagramms dargestellt und wird unter Verwendung von 15 Merkmalen und einer Segmentierung in 15 Intervalle erzielt. Die Klassifizierungsrate auf Trainings- und Testdaten beträgt 100 %.

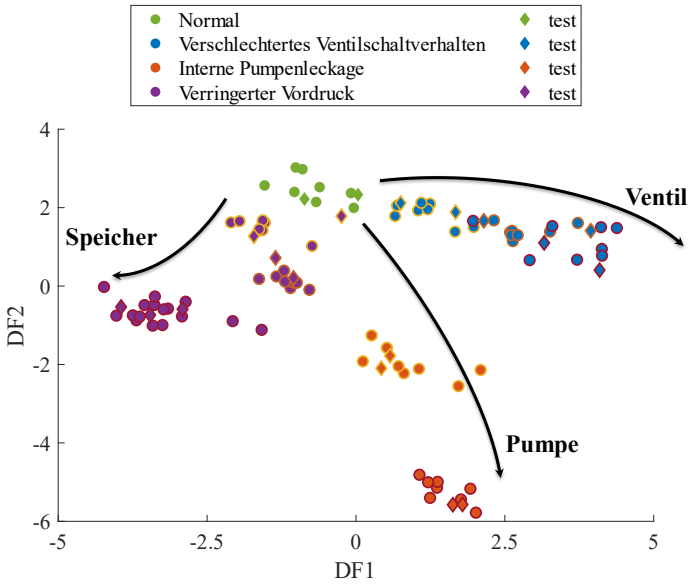


Abbildung 4-25: Überwachtes Modell mit allen Fehlermechanismen des Hydraulikaggregats nach stufenweiser Erweiterung ausgehend vom Gutzustand. Die Umrandung der Marker (von Gelb zu Rot) beschreibt die Intensität der einzelnen Fehlermechanismen.

Aufgrund der steigenden Zahl der im Training berücksichtigten Fehlermechanismen nimmt auch die Komplexität des Modells bezüglich der Anzahl der selektierten Merkmale über den gesamten Lernprozess zu. Neben der fehlerfreien Trennung der verschiedenen Zustände des Hydraulikaggregates lassen sich ebenfalls Informationen über die Intensität der einzelnen Fehlermechanismen ableiten, obwohl diese Information nicht im Training vorhanden war. Auch hier kann daher vermutet werden, dass für die Modellbildung physikalisch begründbare Merkmale selektiert wurden.

In Abbildung 4-26 ist die Initialisierung der Novelty Detection und der Domain Shift Detection nach Anlernen aller Fehlermechanismen dargestellt. Im Fall der Novelty Detection sind deutlich die drei Ausreißer zu erkennen, welche bereits in Abbildung 4-22 a) zu sehen sind. In den Domain Shift Scores tauchen diese Ausreißer nicht auf, weshalb angenommen werden kann, dass sie die überwachte Modellbildung nicht beeinflussen.

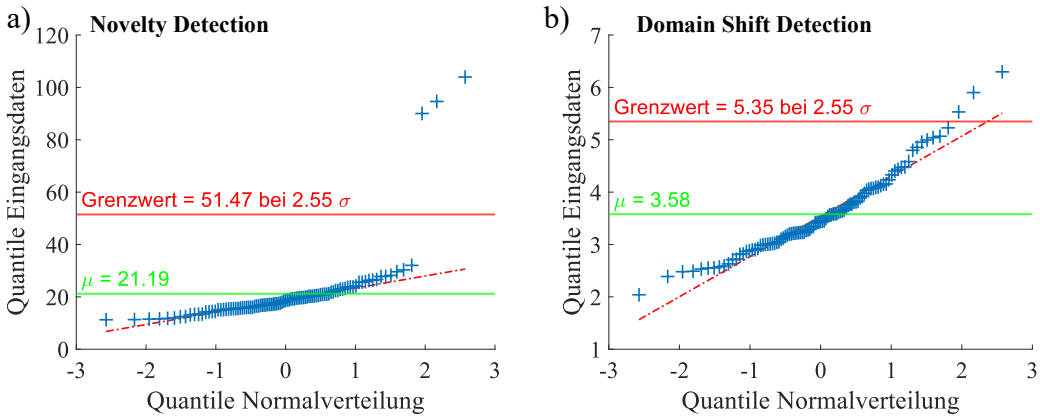


Abbildung 4-26: Initialisierung der Novelty Detection a) und der Domain Shift Detection b) nach der Detektion eines verringerten Speichervordrucks.

Im Fall der Domain Shift Detection liegen lediglich drei Samples geringfügig oberhalb des Grenzwertes (Nummern der Samples: 93, 95, 98). In beiden Fällen lassen sich durch die QQ-Diagramme Abweichungen von der Normalverteilung nachweisen und auch der SW-Test verwirft die Nullhypothese für die Novelty Detection und die Domain Shift Detection. Auch in diesem Fall kann deshalb von einem zu sensitiven Grenzwert ausgegangen werden, wodurch bereits einige Samples als Novelty erkannt werden.

4.1.6.5 Kühlereffizienz

Im nächsten Schritt wird das Auftreten eines Domain Shifts in Form einer verringerten Kühlereffizienz und dadurch steigenden Öltemperatur simuliert. Die entsprechenden Novelty Scores sind in Abbildung 4-27 a) dargestellt. Die erhöhte Öltemperatur wird von der Novelty Detection deutlich als unbekannter Zustand angezeigt. Im Falle einer

Klassifizierung durch das überwachte Modell würde die Zustandsüberwachung, wie in Abbildung 4-27 b) zu sehen, einen verringerten Speichervordruck anzeigen. Auch die Domain Shift Detection zeigt deutlich den Eintritt eines unbekannten Zustands bzw. die Beeinflussung des überwachten Modells an.

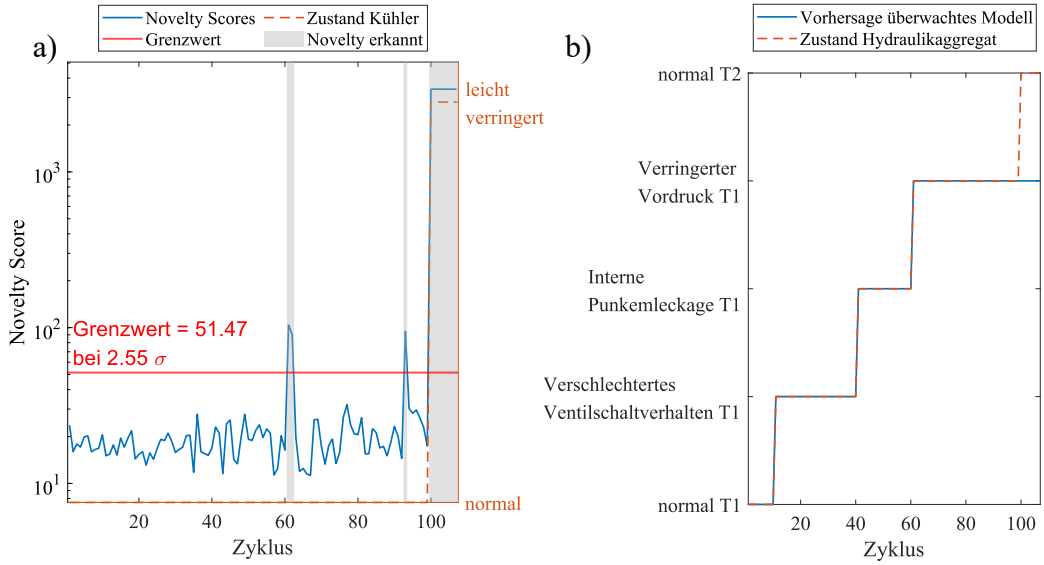


Abbildung 4-27: Reaktion der Novelty Detektion a) und der überwachten Klassifizierung b) auf eine erhöhte Öltemperatur.

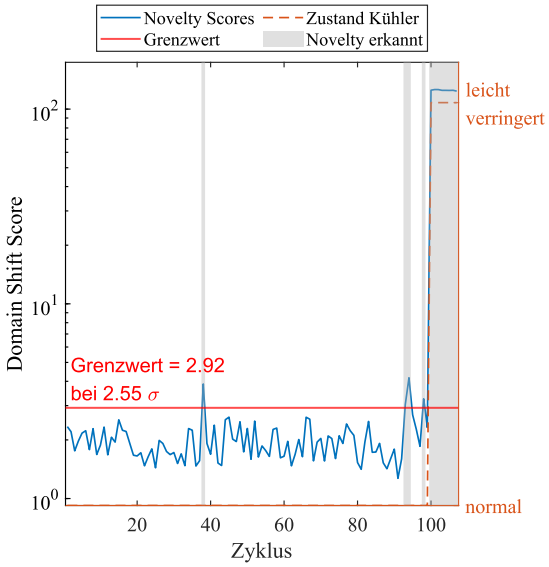


Abbildung 4-28: Reaktion der Domain Shift Detection auf eine erhöhte Öltemperatur.

4.1.6.6 Zusammenfassung der Ergebnisse der adaptiven Zustandsüberwachung

In Tabelle 4-4 sind die Ergebnisse für die adaptive Zustandsüberwachung für alle Öltemperaturstufen zusammengefasst. Als Kennzahlen für die Performanz der

Novelty Detection erfolgt hier die Angabe der TP, TN, FP und FN. Wie im vorherigen Kapitel erfolgt die Unterscheidung zwischen einer Initialisierungs- bzw. Trainingsphase, in der neue Fehlermechanismen in das Novelty Detection Modell übernommen werden, und einer Detektionsphase, in der neue Fehlermechanismen erkannt werden. Auch für die überwachte Klassifizierung erfolgt die Angabe der erreichten Klassifizierungsraten auf Trainings- und Testdaten für die verschiedenen Fehlermechanismen sowie die für die Modellbildung verwendeten Hyperparameter.

Wie der Tabelle entnommen werden kann, zeigt sowohl die Novelty Detection als auch die überwachte Klassifizierung, ausgenommen einzelner Ausreißer, eine hervorragende Performanz über den gesamten Lernprozess. Eine Ausnahme bezüglich der Novelty Detection bildet die Erkennung des verringerten Speichervordrucks bei T1 (FN: 13), der, wie bereits in Abschnitt 4.1.6.4 gezeigt, erst kurz vor dem Ausfall detektiert werden konnte.

Tabelle 4-4: Zusammenfassung der Ergebnisse der adaptiven Zustandsüberwachung eines hydraulischen Systems.

Kühler	**Komponente**	**Novelty Detection**						**Überwachte Klassifizierung**			
		Detektion				**Training**		**Stack**		**Klassifizierungsrate**	
		TP	**TN**	**FP**	**FN**	**TN**	**FP**	N_{Zeit}	N_{Merkma}	**Training**	**Test**
T1	**Normalzustand**	-	-	-	-	9	1	-	-	-	-
	Ventil	24	9	1	0	40	0	18	1	100	100
	Pumpe	16	40	0	0	59	1	23	5	96.6	100
	Speicher	19	59	1	13	96	3	23	12	100	100
T2	**Normalzustand**	8	96	3	0	106	3	11	14	100	100
	Ventil	21	106	3	3	135	4	24	14	99.2	100
	Pumpe	16	135	4	0	157	2	11	49	94.3	100
	Speicher	24	157	2	0	183	6	30	38	100	100
T3	**Normalzustand**	8	183	6	0	194	5	11	54	96.4	100
	Ventil	24	194	5	0	223	6	27	53	97.8	100
	Pumpe	16	223	6	0	246	3	19	62	99.2	100
	Speicher	23	246	3	1	274	5	28	64	81.0	100

Bezüglich der Leistungsfähigkeit des überwachten Modells bildet der verringerte Speichervordruck bei T3 mit einer Klassifizierungsrate von lediglich 81 % eine Ausnahme. Eine Analyse der Daten zeigt, dass in diesem Bereich aufgrund der erhöhten Öltemperatur vermehrt Fehlfunktionen des Ventils V11 zu beobachten sind. In den entsprechenden Arbeitszyklen öffnet das Ventil V11 nicht, wodurch das Öl direkt über ein Druckbegrenzungsventil zurück in den Tank fließt. Dadurch ergibt sich ein völlig anderes Lastprofil, wie in Abbildung 4-29 a) anhand des Sensorsignals PS1 dargestellt. Dieses Lastprofil führt, wie in Abbildung 4-29 b) zu sehen, zu stark erhöhten Novelty Scores (gleichermaßen in der Domain Shift Detection). Diese Ausreißer bleiben auch nach dem Training der Novelty Detection erhalten, was zu einem deutlich erhöhten Grenzwert (95.90) nach Detektion des verringerten Speichervordrucks bei T3 führt. Diese tritt bereits in deutlich geringerem Maße in den Daten der internen Pumpenleckage bei T3 auf, was hier bereits zu einem leicht erhöhten Grenzwert führt (43.75). Auch dieser nicht in den Daten annotierte Fehler wird sicher von der verwendeten Novelty Detection erkannt.

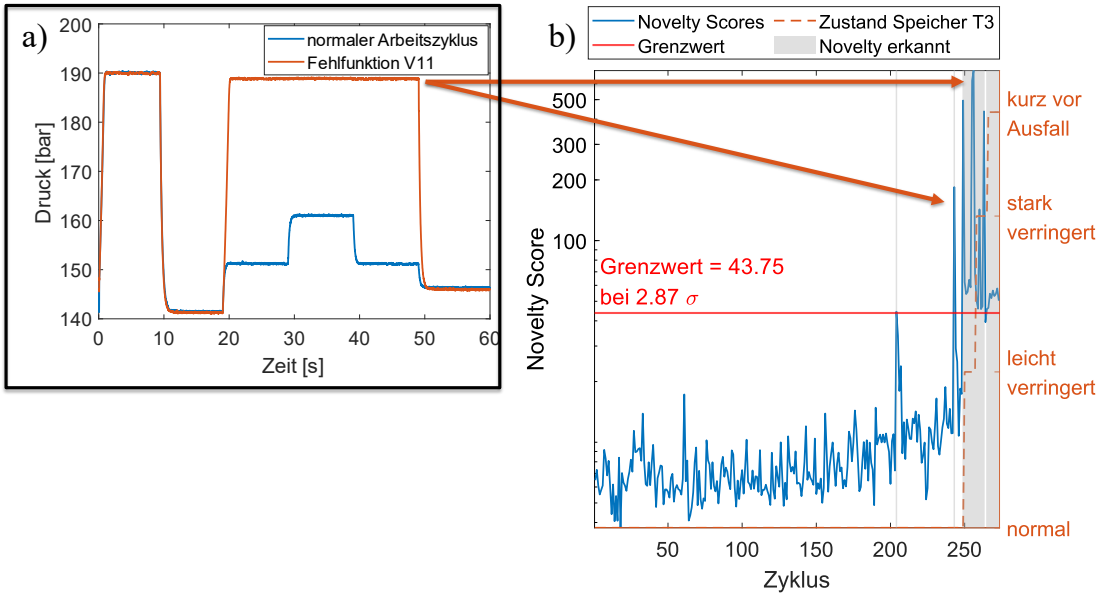


Abbildung 4-29: Reaktion der Novelty Detektion auf einen verringerten Speichervordruck bei T3 und die Abweichung der Arbeitszyklen bei Fehlfunktion von Ventil V11 aufgrund einer zu hohen Öltemperatur.

In Abbildung 4-30 ist das vollständige Modell nach Anlernen aller Fehlermechanismen anhand der ersten drei Diskriminanzfunktionen dargestellt. Aufgrund der häufig auftretenden Fehlfunktionen im Bereich des verringerten Speichervordrucks bei T3 wurden diese Daten für das Training ausgeschlossen. Die erzielten Klassifizierungsraten auf Trainings- und Testdaten betragen dabei 99,2 und 100 Prozent. Es ist deutlich zu erkennen, dass die erste Diskriminanzfunktion den dominierenden Effekt der Öltemperatur abbildet, wohingegen die Information der

einzelnen Fehlermechanismen erst in höheren (für Speicher größer 3) Dimensionen enthalten ist.

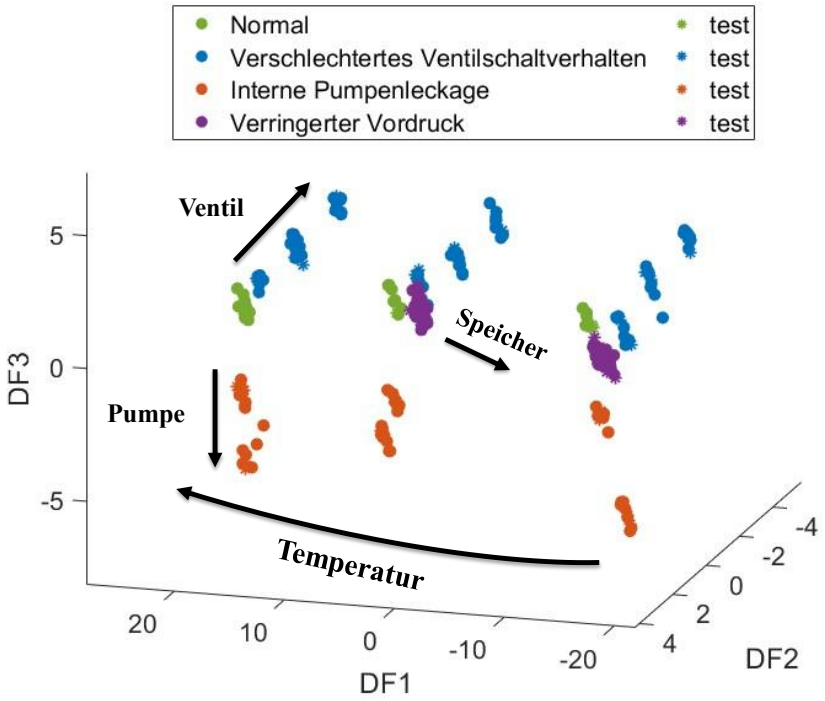


Abbildung 4-30: Vollständig angelerntes Modell zur Überwachung des Hydraulikaggregats (ausgenommen verringerter Vordruck des Speichers bei T3).

4.1.7 Zusammenfassung und Diskussion der Anwendung

Bei dem betrachteten Datensatz handelt es sich um eine bereits vielfach untersuchte und gut verstandene Anwendung. Die umfangreiche, nach DoE aufgezeichnete Datenbasis liegt bereits in strukturierter Form vor und eignet sich hervorragend für die realitätsnahe Methodenentwicklung und Validierung. Auf Basis dieses Datensatzes konnte die Leistungsfähigkeit der vorgestellten automatisierten überwachten Modellbildung demonstriert werden. Die Methoden erlauben dabei eine nahezu fehlerfreie Beurteilung des Zustands des Hydraulikaggregats hinsichtlich der in den Trainingsdaten enthaltenen, überlagerten Fehlermechanismen. Durch die Transparenz und gute Visualisierbarkeit der gebildeten Modelle kann eine nachgelagerte physikalische Plausibilisierung der Modelle und Merkmale durchgeführt werden, was die Akzeptanz und breite Anwendung datengetriebener Verfahren im industriellen Umfeld fördert.

Als großes Hindernis für die breite Anwendung datengetriebener Verfahren ist die Anfälligkeit der Modelle gegenüber dem Einfluss unbekannter, nicht kontrollierbarer Umweltbedingungen und deren Auswirkungen auf die Vorhersagequalität der Modelle

zu nennen, was in diesem Fall anhand einer variierenden Öltemperatur gezeigt werden konnte. Ausgehend von den dabei erzielten Ergebnissen kann bei höherer Komplexität der Modelle von einer gesteigerten Empfindlichkeit gegen Quereinflüsse ausgegangen werden. Im einfachsten Fall kann dies bei Modellen mit erhöhter Komplexität durch die breitbandige Herkunft der Merkmale aus dem gesamten Arbeitszyklus erklärt werden.

Im Falle des hydraulischen Systems zeigen sich die Fehler als abrupte Zustandsänderung, wohingegen in der Realität eher ein langsames Wegdriften des Systemzustands zu beobachten wäre. Aus diesem Grund wird ein Nachtrainieren des Gutzustands als kritisch angesehen, da dadurch bereits beginnende Symptome in das Modell des Gutzustandes mit einfließen würden, was einer Detektion entgegenwirken würde.

Die Domain Shift Detection zeigt zuverlässig an, wenn neu eintreffende Datenpunkte zu keiner der im überwachten Modell bis dahin enthaltenen Klassen gehören. Inwiefern die Domain Shifts das überwachte Modell bei der Diagnose bereits bekannter Fehler beeinflussen, müsste allerdings noch genauer untersucht werden.

Neben der Problematik der Domain Shifts besteht bei der Mehrheit der industriellen Anwendungen nicht die Möglichkeit, Trainingsdaten in dem hier vorliegenden Umfang zu generieren. Hier konnte anhand eines realitätsnahen Ablaufs das Potential der vorgeschlagenen adaptiven Zustandsüberwachung demonstriert werden. Die verwendete Novelty Detection Pipeline detektiert dabei sämtliche Fehlermechanismen der Komponenten Ventil, Pumpe und Speicher. Die vorgestellte automatisierte Methodik zur Grenzwertbestimmung zeigt dabei trotz der in den meisten Fällen nichtzutreffenden Annahme auf Normalverteilung der Novelty Scores eine zuverlässige Detektion der anhaltenden Zustandsänderungen. Die unregelmäßig auftauchenden Ausreißer werden aufgrund ihrer geringen Häufigkeit und Reversibilität als nicht problematisch eingestuft. Die aktuelle Implementierung, basierend auf der Standardabweichung der im Training vorhandenen Novelty Scores zeigt dabei aber eine starke Empfindlichkeit gegenüber den Ausreißern. Um den Grenzwert über längere Zeit empfindlich zu halten, könnten diese Ausreißer aus den Trainingsdaten entfernt werden.

Nach der Identifikation der jeweiligen Fehlermechanismen konnten diese alle erfolgreich in das überwachte Modell übernommen werden. Trotz der sehr geringen Anzahl an Stichproben gelingt es der Kombination aus Bayes optimiertem TFEx mit MRMR-Merkmalsselektion, relevante Sensoren und Signalbereiche zu identifizieren, wodurch die resultierenden Merkmale in vielen Fällen eine physikalische Interpretierbarkeit aufweisen. Aus diesem Grund sind die Modelle in der Lage,

Informationen über die Intensität der fehlerhaften Maschinenzustände abzuleiten, ohne dass diese Information bei der Modellbildung mit eingegeben wurde, was auf eine umfangreiche Generalisierung hindeutet. Darüber hinaus kann die Kenntnis der extrahierten Merkmale dazu verwendet werden, Hinweise auf die Ursache des fehlerhaften Maschinenzustands abzuleiten. Die Qualität der Hinweise ist dabei stark vom betrachteten Fehlermechanismus abhängig. Im Fall des Ventils kann die Herkunft der für die Modellbildung verwendeten Merkmale aus der Schaltflanke des Ventils als eindeutiger Hinweis für die Fehlerursache angesehen werden. Für die interne Pumpenleckage und den verringerten Vordruck des Hydrospeichers ist für eine zielführende Interpretation der Merkmale ein tieferes Systemverständnis notwendig, das bei einem entsprechenden Prozessingenieur vorausgesetzt werden darf.

4.2 Zustandsüberwachung einer servopneumatischen Schweißzange

4.2.1 Einführung

Die schnell voranschreitende Elektrifizierung stellt die Automobilbranche vor beträchtliche Herausforderungen. Eine Möglichkeit, diesen Herausforderungen zu begegnen, liegt in der Steigerung der Automatisierung der Fertigungsprozesse mit einer einhergehend gesteigerten Effizienz und Zuverlässigkeit [190]. Gerade im Bereich der Automobilindustrie, die branchenübergreifend schon einen der höchsten Automatisierungsgrade aufweist, steigt damit der Bedarf an Methoden des maschinellen Lernens weiter an, um automatisierte Prozesse zu überwachen und zu optimieren [191]. Ein Beispiel für einen hochautomatisierten Prozess stellt der Rohkarosseriebau dar. Hierbei werden durch eine Vielzahl an aufeinanderfolgenden Prozessen Fahrzeugkarossen gefertigt. Das Fügen der einzelnen Karosserieteile wird dabei typischerweise mittels Schweißzangen umgesetzt. Um ein präzises und kontrolliertes Arbeiten zu gewährleisten, werden diese Systeme durch ein servopneumatisches Antriebssystem angetrieben [192]. Im Gegensatz zu einfachen pneumatischen Komponenten verfügen servopneumatische Systeme über eine Reihe von Sensoren, die umfangreiche Messungen der Systembedingungen sowie eine Regelung von Kräften und Positionen im geschlossenen Regelkreis ermöglichen. Ein schematischer Aufbau eines servopneumatischen Antriebssystems, verbaut in einer Euro-X-Schweißzange, ist in Abbildung 4-31 dargestellt. Das Antriebssystem der Schweißzange gliedert sich in das eigentliche Hauptantriebssystem und ein Kompensationssystem. Das Antriebssystem besteht aus dem doppeltwirkenden Zylinder (Hauptzylinder, HZ), der durch ein servopneumatisches Proportionalventil (MPYE von Festo) gesteuert wird. Der HZ verfügt über einen Drucksensor für jede Zylinderkammer sowie einen Positionssensor zur Bestimmung der absoluten Position der Kolbenstange. Angetrieben durch das servopneumatische System führen die Arme der Schweißzange eine Drehbewegung um eine zentrale Achse aus. Das Ausgleichssystem besteht ebenfalls aus einem doppeltwirkenden Zylinder, der von einem servopneumatischen Differenzdruckregelventil (MPYD von Festo) gesteuert wird. Durch die Belüftung des Ausgleichszylinders (AZ) wird eine weiche Luftfeder aufgebaut, die bei asynchronem Kontakt und ungleichmäßigem Kraftaufbau eine Verformung der Bleche verhindert. Die Regelkreise des Hauptantriebssystems sowie des Kompensationssystems werden von einem gemeinsamen Controller angesteuert.

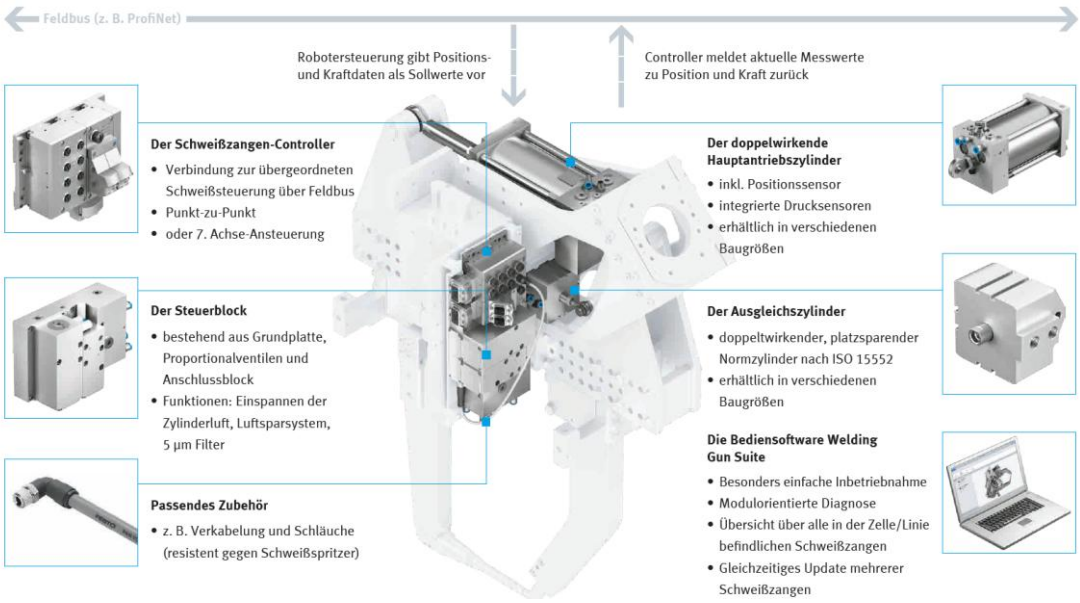


Abbildung 4-31: Schematische Abbildung einer Euro-X-Schweißzange mit integriertem servopneumatischem Antriebssystem, wie es beispielsweise im Karosserierohbau verwendet wird [13].

Ziel des Anwendungsfalls ist die Überwachung des servopneumatischen Antriebssystems der Schweißzange, um einen effizienten und, bezogen auf die gefertigten Karosserien, fehlerfreien Betrieb zu ermöglichen. Fehler sollen dabei nicht nur erkannt, sondern das entsprechende Wartungspersonal zusätzlich durch eine Klassifizierung und Diagnose der fehlerhaften Komponente unterstützt werden.

4.2.2 Experimenteller Aufbau

In Abbildung 4-32 ist der Aufbau einer servopneumatischen Euro-X-Schweißzange und ihrer Komponenten dargestellt, wie er für die Datengenerierung im Labor verwendet wird [193]. Für die Auswertungen stehen die Signale der beiden Drucksensoren des HZ, die Position der Kolbenstange, der Differenzdruck des MPYD-Ventils und dessen Sollwert sowie die Schieberstellung des MPYE-Ventils und dessen Sollwert zur Verfügung. Alle Signale werden mit einer Abtastrate von 100 Hz aufgezeichnet. In Tabelle 4-5 sind die für die Zustandsüberwachung der Schweißzange zur Verfügung stehenden Signale und die zugehörigen Komponenten zusammengefasst. Neben der synchronen Erfassung der Daten ermöglicht der Aufbau gezielt das Einbringen typischer Fehler wie Leckagen an den Druckanschlüssen der einzelnen Zylinderkammern von HZ und AZ sowie eine erhöhte Reibung des zentralen Gelenks der Schweißzange.

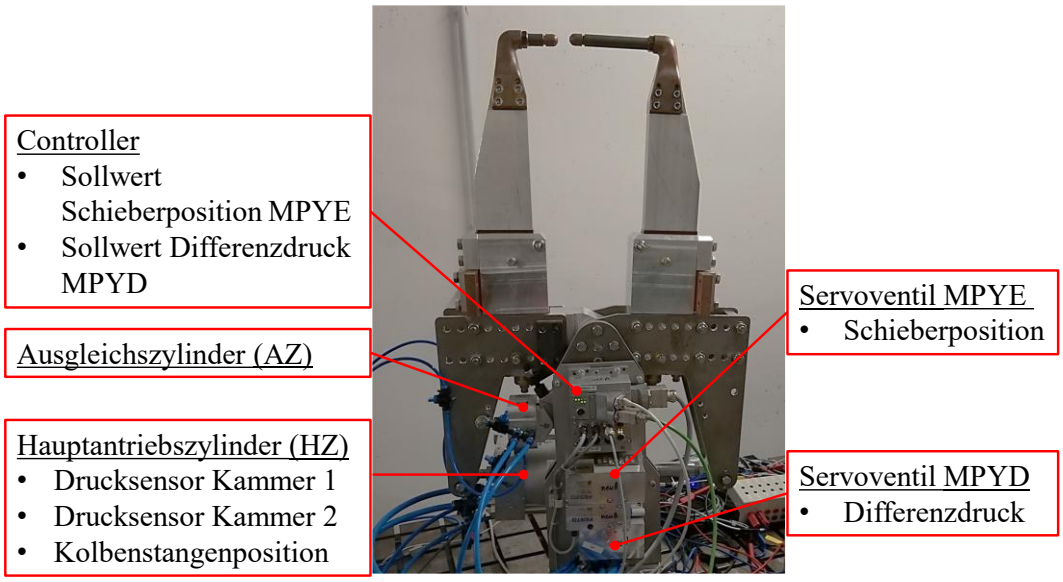


Abbildung 4-32: Versuchsaufbau bei Festo zum Aufzeichnen der Daten im Labor [193].

Tabelle 4-5: Verfügbare Sensoren bzw. Messsysteme für die Überwachung der Schweißzange.

Sensor / Messgröße	Bezeichnung	Bezug	Abtastrate
HZ Kammer 1 Drucksensor	DHZ1	Hauptzylinder / MPYE	100 Hz
HZ Kammer 2 Drucksensor	DHZ2	Hauptzylinder / MPYE	100 Hz
Position Kolbenstange	PKS	Hauptzylinder / MPYE	100 Hz
Sollwert Schieberposition	PSsoll	Hauptzylinder / MPYD	100 Hz
Istwert Schieberposition	PSist	Ausgleichszylinder / MPYD	100 Hz
Sollwert Differenzdruck	Pdiffsoll	Ausgleichszylinder / MPYD	100 Hz
Istwert Differenzdruck	Pdiffist	Ausgleichszylinder / MPYD	100 Hz

4.2.3 Mess- und Datenplanung

Die Datenaufzeichnung erfolgte bei dem Industriepartner Festo und wurde im Rahmen des Projektes MessMo für eine Auswertung zur Verfügung gestellt. Die Daten des Anwendungsfalls liegen daher bereits in einer strukturierten .h5 Datei vor. Eine systematische Mess- und Datenplanung zur Sicherstellung der Datenqualität wird dadurch nicht mehr benötigt.

Um einen effizienten und störungsfreien Betrieb der Schweißzange zu gewährleisten, wird ein aktiver Fehlerdiagnoseansatz verfolgt, d. h. die Schweißzange wird in einem individuell gestalteten Prüfzyklus betrieben. Dies ermöglicht das Erfassen relevanter und reproduzierbarer Informationen über den Systemzustand anhand der aufgezeichneten Sensorsignale. Dadurch besteht abgesehen von der Einbausituation kein Unterschied zwischen im Labor generierten Daten und den Daten einer in der Serienfertigung verwendeten Schweißzange. Der Aufbau des Prüfzyklus sieht dabei mehrere funktionale Blöcke vor, in denen spezifische Komponenten bzw. Kombinationen von Komponenten angesteuert werden.

Der zur Verfügung gestellte Datensatz besteht insgesamt aus 9506 unter Laborbedingungen aufgezeichneten Prüfzyklen. Während der Experimente werden Leckagen an den Druckanschlüssen der Kammern 1 und 2 von HZ und AZ künstlich in das Hauptantriebssystem der Schweißzange eingebracht. Die eingebrachten Leckagen können in einem Bereich von 0 bis 20 l/h variiert werden. Zudem kann eine erhöhte Reibung des zentralen Drehgelenks der Schweißzange simuliert werden, indem das Lagerdrehmoment von den spezifizierten 30 Nm (Normal) auf 60 Nm (Reibung erhöht) gesteigert wird. Neben den Variationen der eigentlichen Zielgröße wurde im Rahmen der Versuche das den HZ ansteuernde Ventil MPYE getauscht, was für die nachfolgende adaptive Zustandsüberwachung als Simulation eines Domain Shifts verwendet wird. In Abbildung 4-33 ist das dem Datensatz zugrundeliegende DoE dargestellt.

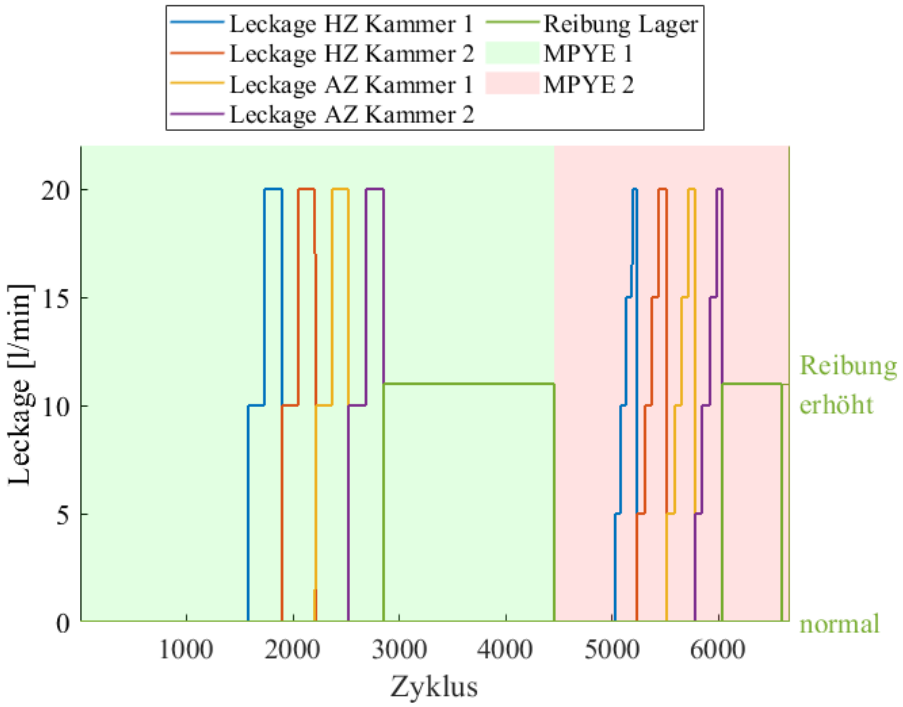


Abbildung 4-33: DoE für die aufgezeichneten Daten der Schweißzange.

4.2.4 Adaptive Zustandsüberwachung

Auch für die Schweißzange wird eine realitätsnahe Zustandsüberwachung auf Basis der vorgeschlagenen adaptiven Zustandsüberwachung durchgeführt. Das Auftreten der Fehler wird dabei in der Reihenfolge, wie in Abbildung 4-33 dargestellt, simuliert. Auch hier werden die Daten jedes Schadenfalls in 80 % Trainings- und 20 % Testdaten aufgeteilt, um eine Validierung der Modelle beim fortschreitenden Lernen der einzelnen Fehlermechanismen durchzuführen. Die grafische Darstellung der adaptiven Zustandsüberwachung erfolgt dabei nur bis zum Wechsel des Ventils MPYE. Der weitere Lernprozess wird in tabellarischer Form zusammengefasst. Da es sich bei dem komplexen Prüfzyklus um Know-how des Maschinenbetreibers handelt, wird an dieser Stelle auf eine detaillierte Darstellung der Sensorsignale beispielsweise bei der Plausibilisierung der Merkmale und deren Herkunft verzichtet. Die Beschreibung der selektierten Merkmale und ihres Ursprungs erfolgt aus diesem Grund nur textuell.

Die verwendetet Novelty Detection basiert wiederum auf dem in Abschnitt 3.2.2 vorgestellten Stack. Für den TFEx wird hierbei eine Segmentierung von 20 Intervallen im Zeitbereich verwendet, wodurch eine ausreichende zeitliche Auflösung erreicht wird, um die extrahierten Merkmale spezifischen Komponenten bzw. Funktionalitäten der Schweißzange zuzuordnen. Auch hier handelt es sich ausschließlich um Prozesssensorik, deren Informationsgehalt vollständig im Zeitbereich angenommen wird, weshalb wieder 0 Segmente im Frequenzbereich, also keine Merkmale aus dem Frequenzbereich, verwendet werden. Aufgrund der erhöhten Anzahl verfügbarer Stichproben und um ein Gleichgewicht zwischen der Anpassung an die Trainingsdaten und geglätteten Entscheidungsgrenzen zu erzielen, wird die Anzahl der nächsten Nachbarn für das kNN-Verfahren in diesem Fall auf $k = 5$ festgelegt.

4.2.4.1 Initialisierung des Normalzustands

In Abbildung 4-34 ist die Initialisierung der Novelty Detection anhand des Histogramms und des QQ-Diagramms für Trainings- und Testdaten dargestellt. Für diese Darstellungen sowie für die folgende Auswertung wurde aus den Normaldaten ein Ausreißer (letzter Zyklus) entfernt, bei welchem es sich um eine Fehlmessung handelte. Bereits an dieser Stelle ist eine deutliche Abweichung der Trainingsdaten von der Normalverteilung zu beobachten. Auch der SW-Test weist die Nullhypothese an dieser Stelle zurück. Durch die Abweichung von der Normalverteilung liegen bereits einige Samples über dem Grenzwert von 3,36 σ. Da es sich hier ausschließlich um vereinzelte reversible Novelties (siehe Abbildung 4-35) handelt, werden diese als nicht problematisch angesehen.

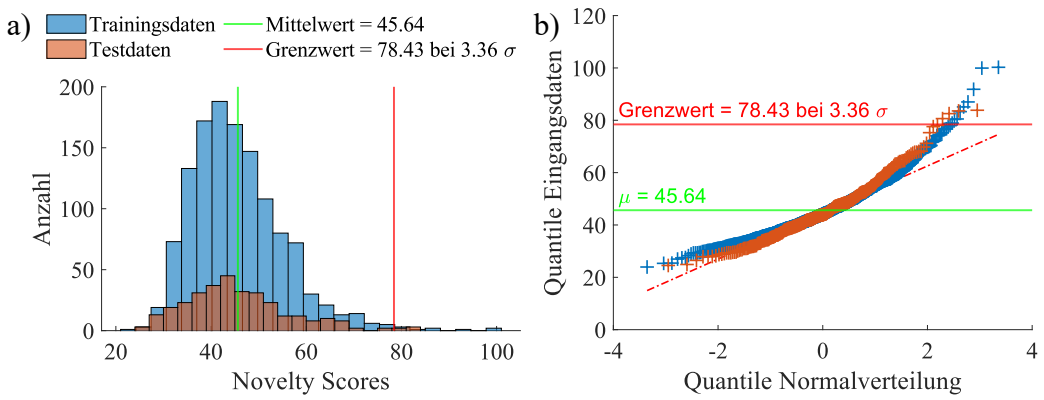


Abbildung 4-34: a) Histogramm der Novelty Scores der Trainingsdaten im Normalzustand, b) QQ-Diagramm der Novelty Scores der Trainingsdaten im Normalzustand.

4.2.4.2 Leckage Hauptzylinder

In Abbildung 4-35 a) ist die Reaktion der Novelty Detection für das Auftreten einer Leckage am HZ Kammer 1 dargestellt. Alle Samples des fehlerhaften Zustands werden detektiert und es ist deutlich eine Korrelation zwischen den Novelty Scores und der Leckagerate zu erkennen. Das Ergebnis der überwachten Klassifizierung ist anhand der ersten Diskriminanzfunktion in Abbildung 4-35 b) zu sehen. Die Bayes optimierte überwachte Klassifizierung erzielt 100 % Klassifizierungsrate auf Trainings- und Testdaten unter Verwendung von 29 Segmenten im Zeitbereich und einem selektierten Merkmal.

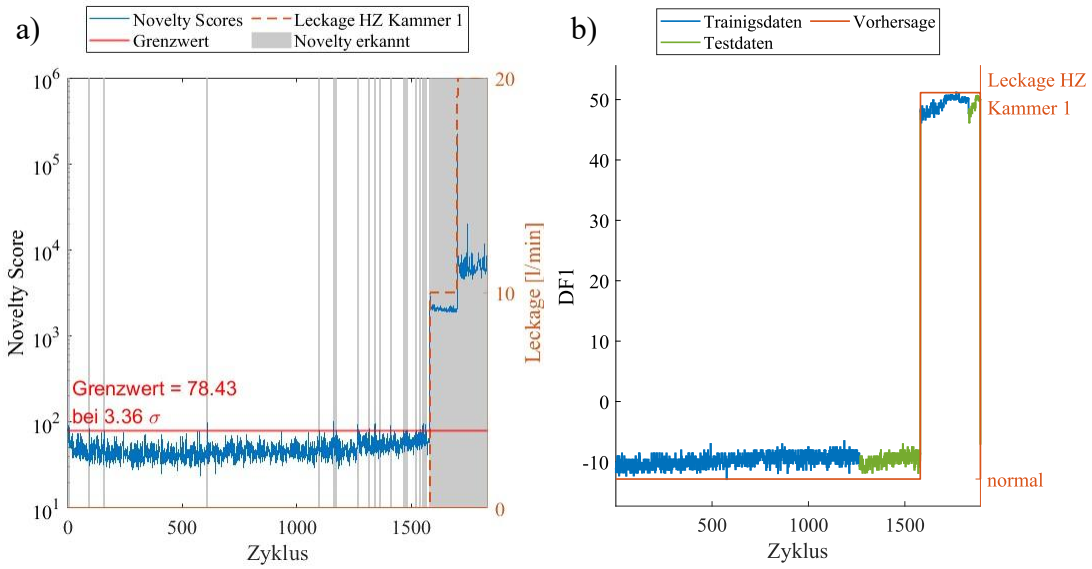


Abbildung 4-35: a) Reaktion der Novelty Detection bei Auftreten einer Leckage am HZ Kammer 1, b) Ergebnis der überwachten Klassifizierung anhand der ersten Diskriminanzfunktion (DF1) für Training (blau) und Test (grün) auf Basis der Ausgabe der Novelty Detection sowie der Vorhersage des Modells (orange).

Da es sich um den ersten in das überwachte Modell übernommenen Fehlermechanismus handelt, ist dieses Modell auch gleichzeitig das unterstützende Modell zur Identifikation der Ursache. Das selektierte Merkmal stammt aus dem Signal des Sensors DHZ1 und stammt aus dem Kraftaufbaubereich des HZ im Testzyklus. Das Merkmal deutet dadurch auf ein Problem mit dem zur Verfügung stehenden Arbeitsdruck zum Ausfahren des HZ hin und kann damit die Identifikation der Fehlerursache beschleunigen.

Nach der Identifikation erfolgt die Übernahme des Zustands in die Novelty Detection und die überwachte Klassifizierung. Zu diesem Zeitpunkt erfolgt das erste Training der Domain Shift Detection. Wie in Abbildung 4-36 zu sehen, liegen die Absolutwerte der Domain Shift Scores deutlich unter denen der Novelty Scores, was auf die geringe Dimensionalität (ein selektiertes Merkmal) sowie die sehr kompakte Darstellung der Daten im Merkmalsraum zurückzuführen ist. Neben diesem kompakten Cluster gibt es bereits einige Ausreißer, welche den definierten Grenzwert überschreiten. Des Weiteren ergibt sich auch an dieser Stelle bereits eine deutliche Abweichung von der Normalverteilung, welche ebenfalls durch das Zurückweisen der Nullhypothese des SW-Tests bestätigt wird.

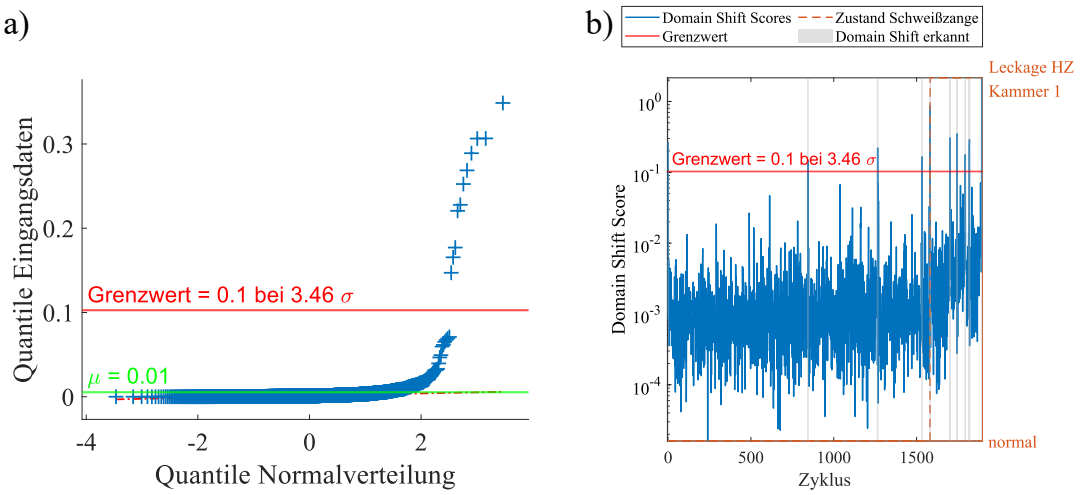


Abbildung 4-36: a) QQ-Diagramms der Initialisierung der Domain Shift Detection nach Detektion der Leckage HZ Kammer 1, b) zugehörige Domain Shift Scores.

Im Anschluss erfolgt wiederum die Initialisierung der Novelty Detection unter Einbezug der Daten des ersten Fehlermechanismus. Auch hier ist das QQ-Diagramm wie auch die Novelty Scores in Abbildung 4-37 dargestellt. Aus Gründen der besseren Visualisierbarkeit wurden im Fall des QQ-Diagramms zwei Ausreißer mit sehr hoher Amplitude entfernt. Grundsätzlich zeigt sich allerdings ein ähnliches Bild wie bei der Initialisierung des Normalzustandes. Abgesehen von einigen reversiblen Ausreißern liegen die Novelty Scores innerhalb des Grenzwertes, wobei ein leichtes Abheben der Scores vom Zustand Leckage HZ Kammer 1 beobachtet werden kann. Ebenfalls kann

wieder eine Abweichung der Novelty Scores von der Normalverteilung beobachtet werden, was ebenfalls durch den SW-Test bestätigt wird. Da sich ein ähnliches Bild für alle folgenden Initialisierungen nach Detektion eines neuen Fehlermechanismus ergibt, wird im Folgenden auf die Darstellung der QQ-Diagramme bzw. der Novelty Scores in den Initialisierungsphasen verzichtet.

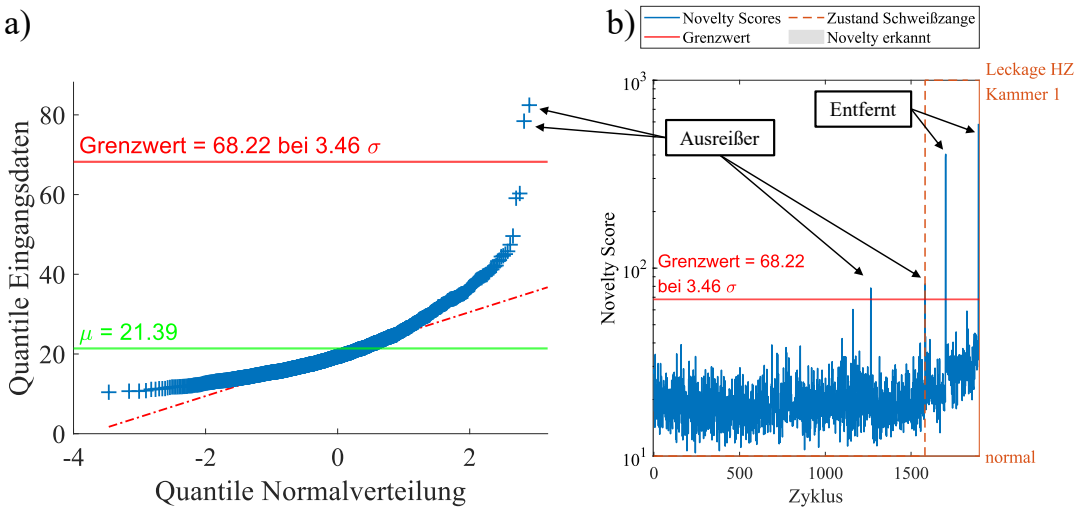


Abbildung 4-37: a) QQ-Diagramm der Initialisierung der Novelty Detection nach Detektion der Leckage HZ Kammer 1, b) die entsprechenden Novelty Scores.

Im nächsten Schritt wird das Auftreten einer Leckage am HZ Kammer 2 simuliert. Die entsprechende Reaktion der Novelty Detection, der überwachten Klassifizierung sowie der Domain Shift Detection sind in Abbildung 4-38 dargestellt. Die Novelty Detection zeigt dabei deutlich das Eintreten eines neuen Zustands an, wobei die Novelty Scores eine klare Korrelation mit der Leckagerate aufweisen. Da der neue Zustand in dem überwachten Modell noch nicht enthalten ist, würde dieses im Wechsel den Normalzustand bzw. eine Leckage am HZ Kammer 1 der Schweißzange anzeigen. Auch die Domain Shift Detection zeigt einen neu eingetretenen Zustand und damit die Beeinflussung der überwachten Klassifizierung. Auch hier ist eine Korrelation der Domain Shift Scores mit der Leckagerate zu beobachten, was vermuten lässt, dass die Information über die Leckage der Kammer 2 bereits in dem für die Erkennung der Leckage an Kammer 1 selektiertem Merkmal enthalten ist.

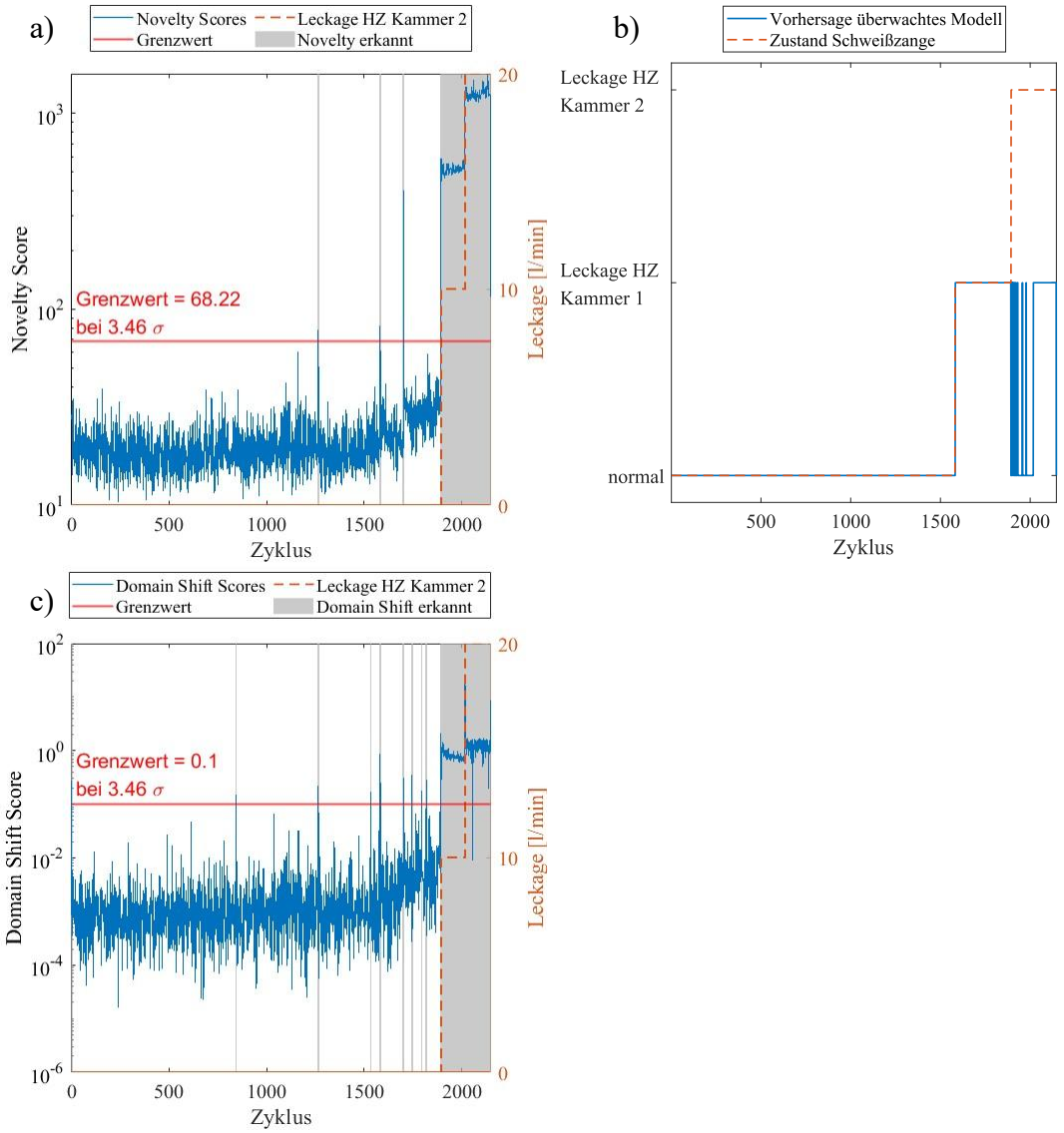


Abbildung 4-38: a) Reaktion der Novelty Detektion, b) der überwachten Klassifizierung sowie c) der Domain Shift Detection auf die Leckage des HZ Kammer 2.

Das Modell zur Unterstützung der Identifikation des neuen Zustands erreicht eine Klassifizierungsrate von 100 % bei Verwendung von 10 Segmenten im Zeitbereich. Es wird wiederum lediglich ein Merkmal selektiert, welches aus dem Signal des Sensors DHZ2 stammt. Eine genauere Analyse der Herkunft des Merkmals zeigt, dass es aus der Kraftabbauflanke des HZ extrahiert wird. Nach der Identifikation der Fehlerursache werden die Daten entsprechend gelabelt und in das überwachte Modell übernommen. In Abbildung 4-39 ist das zu diesem Zeitpunkt vollständige überwachte Modell zur Erkennung der Leckagen am HZ dargestellt. Erreicht wird das Ergebnis mit 10 Zeitsegmenten und vier Merkmalen, die alle aus den Signalen der Drucksensoren von Kammer 1 und Kammer 2 extrahiert werden. In den Clustern der beiden Leckagen ist dabei deutlich der Einfluss der Leckagerate zu erkennen, obwohl die Information bei der Modellbildung nicht verwendet wurde. In den Daten der Klasse

Leckage HZ Kammer 1 ist zusätzlich ein kleines Cluster erkennbar, in dem sich sowohl Trainings- als auch Testdaten befinden, deren Ursache allerdings nicht aufgeklärt werden konnte.

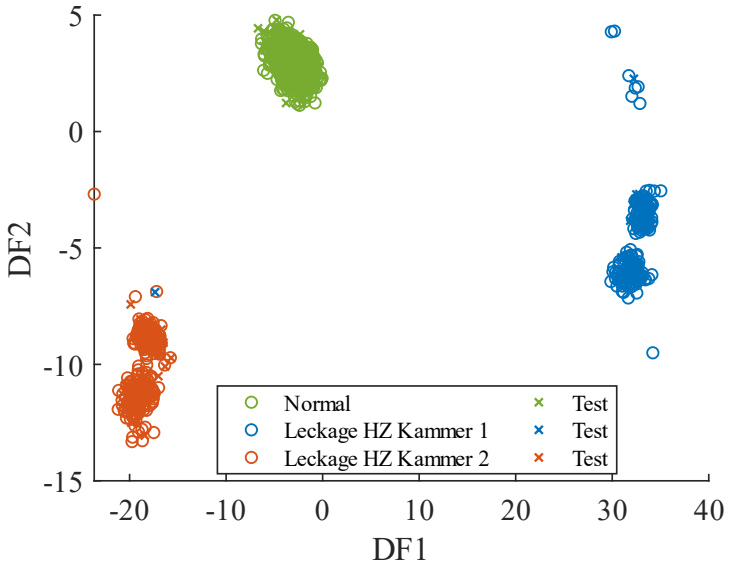


Abbildung 4-39: Überwachtes Modell für die Leckagedetektion beider Kammern des HZ.

Im Anschluss an das Training des überwachten Modells erfolgt die Initialisierung der Novelty Detection und Domain Shift Detection unter Einbezug der bis dahin bekannten Fehlermechanismen.

4.2.4.3 Leckage Ausgleichszylinder

Abbildung 4-40 zeigt die Resultate der Novelty Detection, Domain Shift Detection und überwachten Klassifizierung bei Eintreten einer Leckage am AZ Kammer 1. Vergleichbar zu den Leckagen am Hauptzylinder, zeigt die Novelty Detection deutlich das Eintreten eines neuen Zustands an, wobei auch an dieser Stelle wieder ein klarer Zusammenhang zwischen der Höhe der Scores und der Leckagerate erkennbar ist. Das überwachte Modell würde entsprechend der bis zu diesem Zeitpunkt bekannten Klassen einen Normalzustand der Schweißzange anzeigen. Die Domain Shift Detection zeigt in diesem Fall ein von der Novelty Detection abweichendes Verhalten und zeigt für die ersten Leckagerate sporadisch das Eintreten eines neuen Zustands an. Dadurch, dass bis zu diesem Zeitpunkt nur Merkmale aus dem Sensor des HZ im überwachten Modell vorhanden sind, in denen die Funktionalität des HZ überprüft wird, ist die Domain Shift Detection nicht sensibel für Effekte, welche sich auf andere Teile des Testzyklus auswirken. Die kombinierte Aussage aus Novelty Detection und Domain Shift Detection würde im Realfall bedeuten, dass zwar ein neuer unbekannter Fehler im pneumatischen Antriebssystem vorhanden ist, dieser sich aber nicht auf das

bisherige überwachte Modell auswirkt, weshalb es weiter für eine Überwachung des HZ verwendet werden könnte.

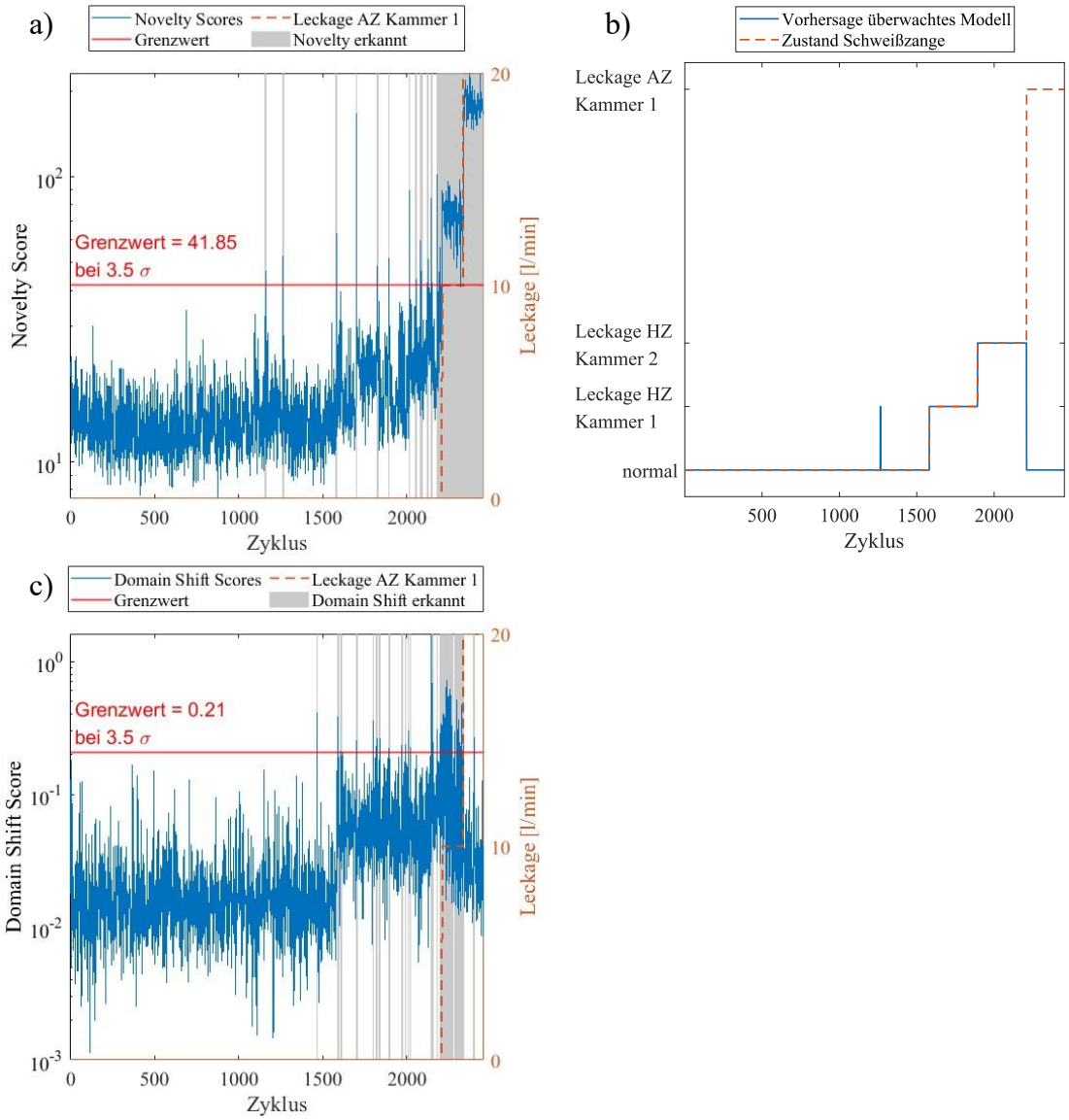


Abbildung 4-40: a) Reaktion der Novelty Detektion, b) der überwachten Klassifizierung sowie c) der Domain Shift Detection auf ein erhöhtes Lagermoment.

Auch die unterstützende Klassifizierung zur Identifikation der Ursache des neuen Zustands verdeutlicht diese Annahme. Das Modell erreicht hier eine Klassifizierungsrate von 100 % bei Verwendung von 12 Segmenten im Zeitbereich und 10 selektierten Merkmalen. 8 dieser Merkmale stammen aus dem Signal des Differenzdrucksensors, welcher zur Regelung des Federverhaltens des AZ verwendet wird. Zusätzlich stammen diese alle aus Teilen des Testzyklus, in dem die Funktionalität des AZ geprüft wird, was auf ein Problem mit dem Versorgungsdruck des AZ hindeutet. Nach Identifikation wird der Fehler in das überwachte Modell übernommen.

Im nächsten Schritt wird das Auftreten einer Leckage am AZ Kammer 2 simuliert. In Abbildung 4-41 kann die Reaktion der Novelty Detection, überwachten Klassifizierung und Domain Shift Detection entnommen werden. Der unbekannte Zustand wird deutlich von Novelty Detection angezeigt, wobei das überwachte Modell eine Leckage am AZ Kammer 1 anzeigen würde. Da nun auch Merkmale mit Bezug zum AZ im überwachten Modell vorhanden sind, zeigt die Domain Shift Detection ebenfalls deutlich einen neuen Zustand an. Die Auswertung der unterstützenden Klassifizierung zeigt hierbei ähnliche Ergebnisse wie bereits bei der Leckage an AZ Kammer 1.

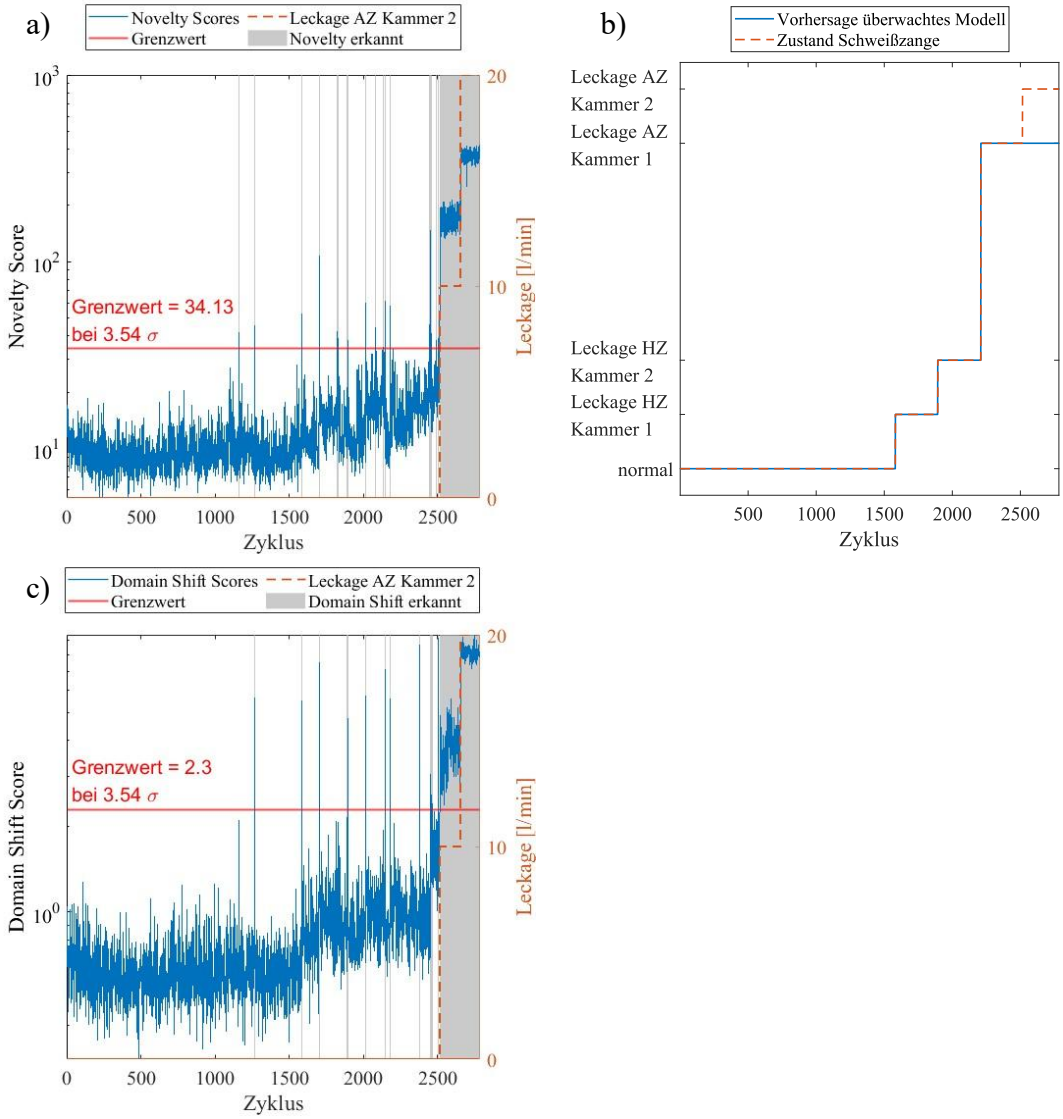


Abbildung 4-41: a) Reaktion der Novelty Detektion, b) der überwachten Klassifizierung sowie c) der Domain Shift Detection auf die Leckage des AZ Kammer 1.

Nach der Identifikation der Leckage am AZ Kammer 2 kann ein überwachtes Modell mit den Leckagen des HZ und AZ trainiert werden. Das Ergebnis ist in Abbildung 4-42 anhand eines dreidimensionalen Streudiagramms dargestellt. Grundsätzlich kann hier

eine nahezu optimale Trennung der einzelnen Fehler beobachtet werden. Die Klassifizierungsrate auf Trainings- und Testdaten beträgt hierbei jeweils 100 %. Es kommt lediglich zu vereinzelten Fehlklassifizierungen, welche aufgrund ihrer deutlichen Lage in einem anderen Cluster aber auch auf falsch gelabelte Datenpunkte zurückzuführen sein könnten.

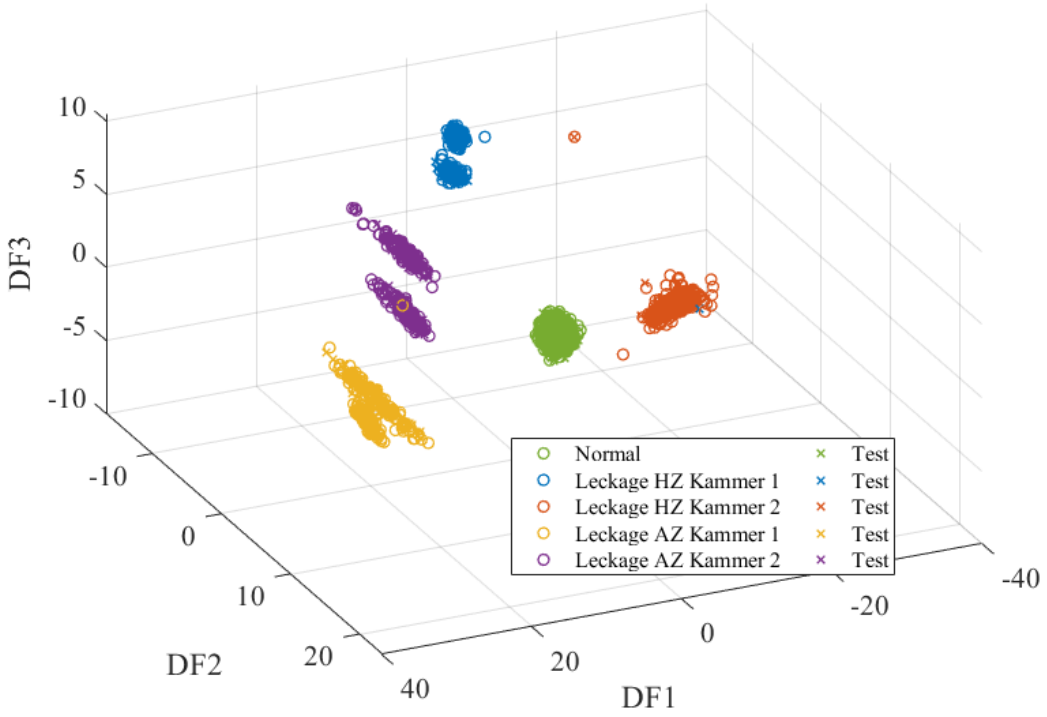


Abbildung 4-42: Überwachtes Modell für die Leckagen an HZ und AZ, jeweils für Kammer 1 und 2.

4.2.4.4 Erhöhte Lagerreibung

Im nächsten Schritt wird das Auftreten einer erhöhten Lagerreibung im zentralen Gelenk der Schweißzange simuliert. In Abbildung 4-43 sind die entsprechenden Reaktionen der Novelty Detection, Domain Shift Detection und überwachten Klassifizierung dargestellt. Im Gegensatz zu den Leckagen zeigt sich in diesem Fall keine klare Detektion des neuen Zustands durch die Novelty Detection. Lediglich im Bereich um Sample 4000 liegen insgesamt ca. 150 Samples in Folge über dem Grenzwert, was je nach Konfiguration der Puffergröße des übergeordneten Algorithmus für eine Detektion ausreichen würde. Ein ähnliches Verhalten zeigt sich für die Domain Shift Detection. Das überwachte Modell würde hier abgesehen vom letzten Zyklus einen Normalzustand der Schweißzange anzeigen. Bereits die zur Unterstützung der Identifikation trainierte binäre Klassifizierung zeigt, dass es sich hier um ein komplexeres Lernproblem handelt. Das Modell erreicht zwar eine optimale Trennung der Daten, verwendet dazu aber 70 Merkmale, welche aus allen 7

verfügbaren Sensorsignalen extrahiert werden, weshalb an dieser Stelle kein Hinweis auf eine Problematik mit dem Gelenk geliefert wird.

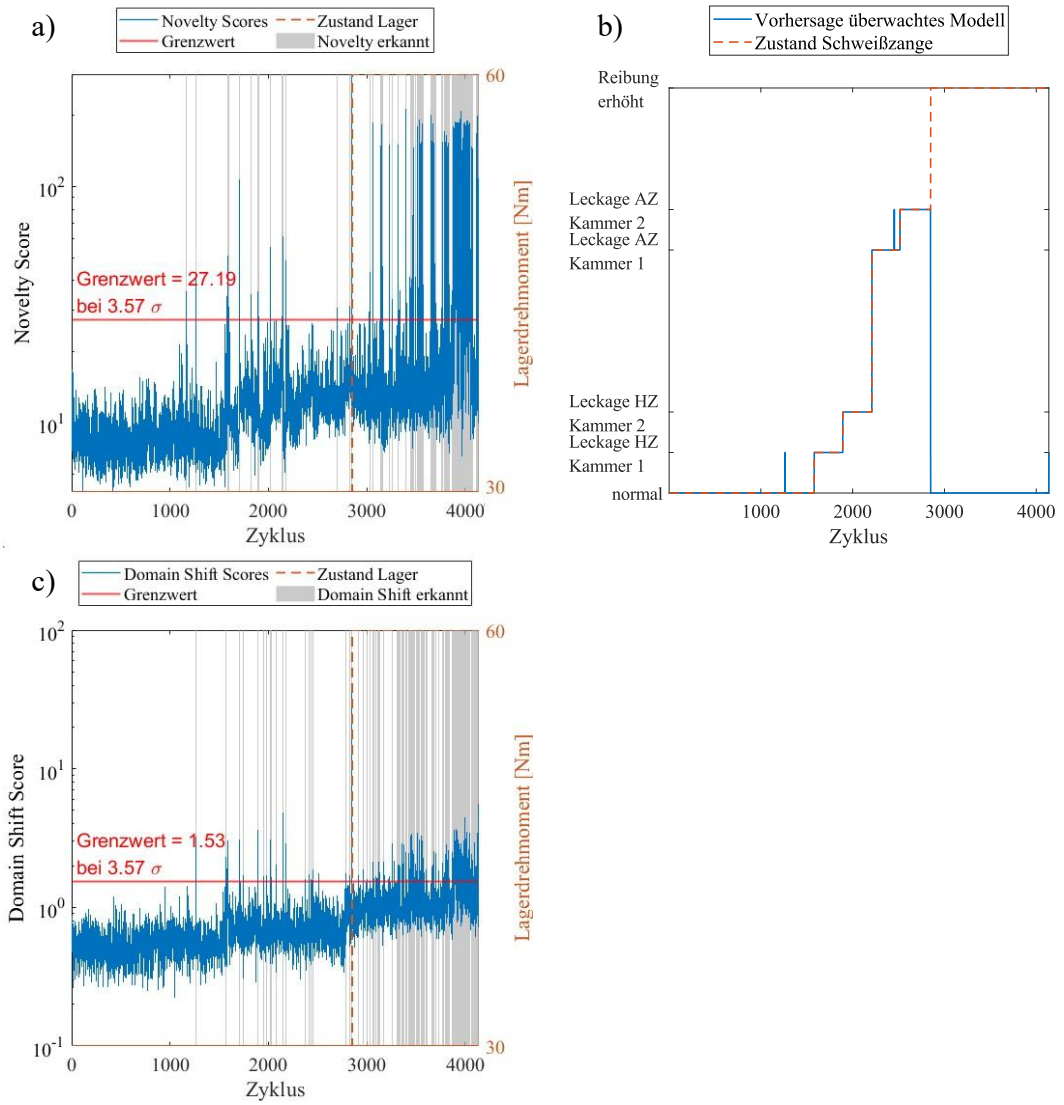


Abbildung 4-43: a) Reaktion der Novelty Detektion, b) der überwachten Klassifizierung sowie c) der Domain Shift Detection auf die erhöhte Lagerreibung.

Auch beim Training des überwachten Modells zeigt sich die erhöhte Komplexität des Lernproblems. Die Anzahl der verwendeten Merkmale steigt von 9 auf 132 Merkmale an, wobei eine Klassifizierungsrate von 99.16 % im Training und 98.48 % beim Test erreicht wird. Abbildung 4-44 zeigt das Ergebnis der Modellbildung unter Einbezug aller in den Daten verfügbaren Fehler in Form eines dreidimensionalen Streudiagramms. Hier ist zu sehen, dass sich die erhöhte Lagerreibung erst in höheren Diskriminanzfunktionen (DF5) von den Normaldaten abhebt. Aufgrund der hohen Ähnlichkeit der Daten im Normalzustand gelingt weder der Novelty Detection noch der Domain Shift Detection eine zuverlässige Detektion des Zustands, und das überwachte Modell ordnet die Daten dem Gutzustand zu (vgl. Abbildung 4-43 b).

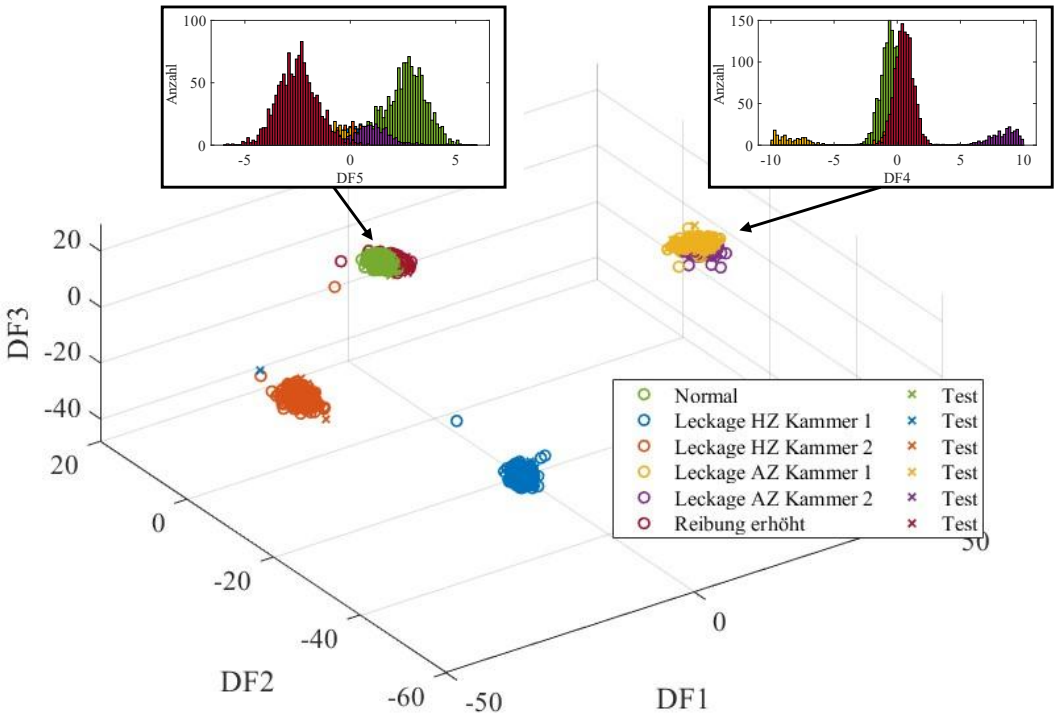


Abbildung 4-44: Überwachtes Modell der Schweißzange inklusive des Fehlers erhöhte Lagerreibung. Die erhöhte Lagerreibung hebt sich erst in höheren Diskriminanzfunktionen (DF5) von den Normaldaten ab.

4.2.4.5 Ventilwechsel MPYE

Nachdem ein Modell mit allen im Datensatz verfügbaren Fehlern trainiert wurde, wird das Eintreten eines Domain Shifts in Form eines Ventilwechsels an der Schweißzange simuliert. Die entsprechenden Reaktionen von Novelty Detection, überwachten Klassifizierung Domain Shift Detection sind in Abbildung 4-45 dargestellt. Hier ist deutlich zu erkennen, dass der Normalzustand der Schweißzange nach Wechsel des Ventils sowohl von Novelty Detection als auch von Domain Shift Detection als neuer Zustand erkannt wird. Die Ursache für das Abfallen der Scores gegen Ende der Messung konnte nicht identifiziert werden. Im Fall der Domain Shift Detection kann zudem bereits vor dem Ventilwechsel (ca. Zyklus 4100 – 4300) eine systematische Erhöhung der Scores festgestellt werden. Zu erklären ist dies durch die jetzt enthaltenen Merkmale, welche die erhöhte Lagerreibung abbilden.

Auch die Vorhersage des überwachten Modells wird deutlich beeinflusst und die Daten werden zum großen Teil der Klasse Leckage HZ Kammer 1 zugeordnet. Dadurch ist das Modell in der trainierten Version nicht mehr für eine Zustandsüberwachung der Schweißzange geeignet. Eine Möglichkeit, die Funktionsfähigkeit des Modells wiederherzustellen, könnte eine Nullpunktkalibrierung des Modells mit den Daten des Gutzustandes nach dem Ventilwechsel darstellen. Gelingt dies nicht, kann die weitere Zustandsüberwachung

nur durch das Weiterlernen im Betrieb aufrechterhalten werden, wodurch allerdings die Möglichkeit der Diagnose der bekannten Fehler zunächst verloren geht.

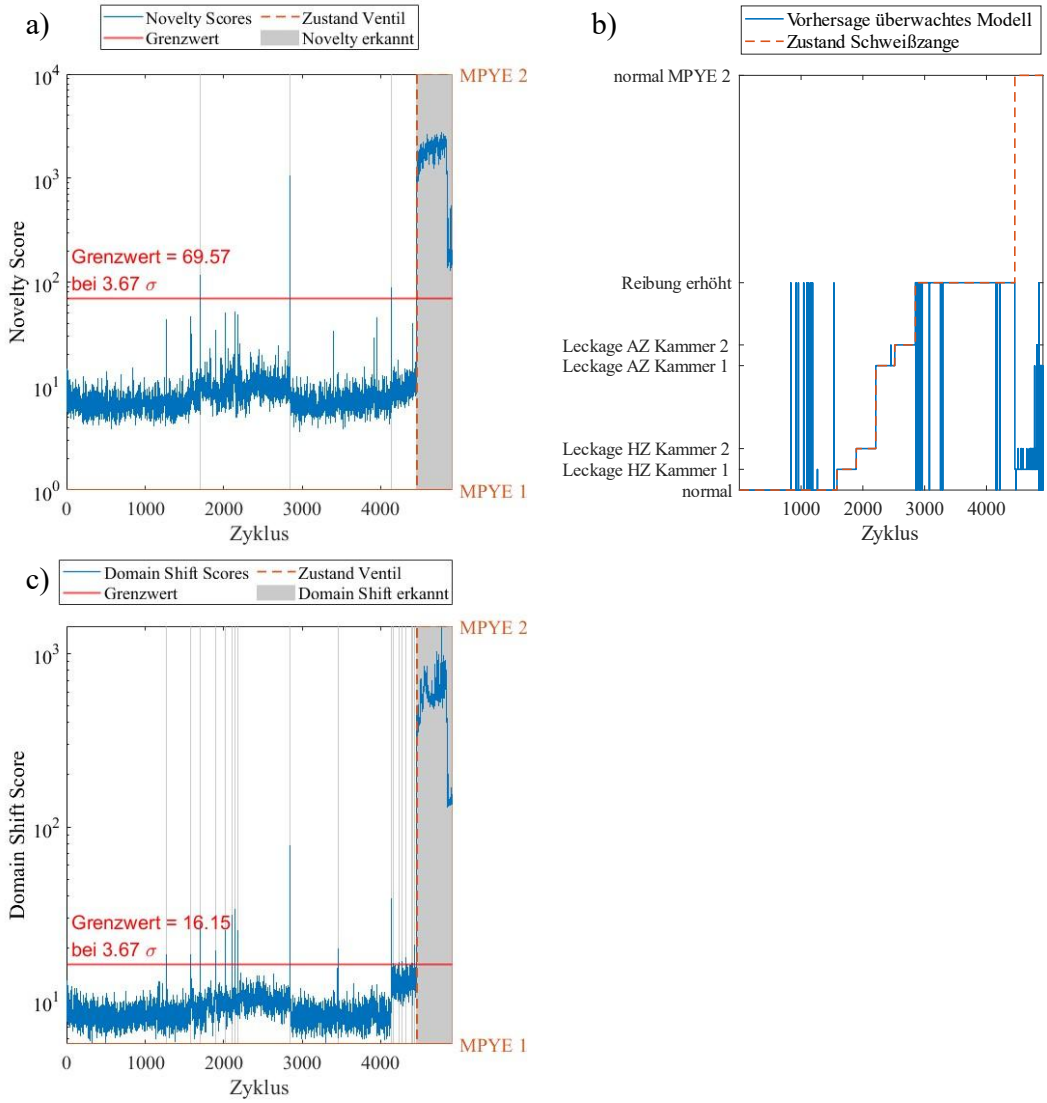


Abbildung 4-45: a) Reaktion der Novelty Detektion, b) der überwachten Klassifizierung sowie c) der Domain Shift Detection auf den Wechsel des Ventils MPYE.

4.2.4.6 Zusammenfassung der Ergebnisse der adaptiven Zustandsüberwachung

In Tabelle 4-6 sind die Ergebnisse des gesamten Lernprozesses, einschließlich des erneuten Lernens aller Fehler nach dem Wechsel des Ventils, in Form einer Tabelle zusammengefasst. Auch hier werden als Kennzahlen für die Performanz der Novelty Detection die TP, TN, FP und FN für jeden Fehler angegeben. Unterschieden wird hier wiederum zwischen einer Initialisierungs- bzw. Trainingsphase, in der neue Fehlermechanismen in das Novelty Detection Modell übernommen werden, und einer Detektionsphase, in der unbekannte Fehlermechanismen erkannt werden. Auch für die

überwachte Klassifizierung erfolgt die Angabe der erreichten Klassifizierungsraten auf Trainings- und Testdaten für die verschiedenen Fehlermechanismen sowie die für die Modellbildung verwendeten Hyperparameter.

Tabelle 4-6: Zusammenfassung der Ergebnisse der adaptiven Zustandsüberwachung einer servopneumatischen Schweißzange.

Ventil MPYE	Komponente	Novelty Detection						Überwachte Klassifizierung			
		Detektion				Training		Stack		Klassifizierungsrate	
		TP	TN	FP	FN	TN	FP	N_{Zeit}	$N_{Merkmale}$	Training	Test
1	normal	-	-	-	-	1548	31	-	-	-	-
	HZ 1	250	1548	31	0	1887	4	29	1	100	100
	HZ 2	252	1887	4	1	2169	38	10	2	100	98.3
	AZ 1	244	2169	38	1	2494	19	9	7	99.9	98.4
	AZ 2	268	2494	19	0	2830	17	28	9	100	98.4
	Reibung	193	2830	17	1095	4448	9	20	132	99.1	98.4
2	normal	455	4448	9	0	4989	36	30	76	99.1	99.0
	HZ 1	165	4989	36	0	5190	41	29	76	98.1	99,0
	HZ 2	220	5190	41	0	5454	51	30	77	98.0	99,1
	AZ 1	204	5454	51	9	5719	52	30	160	98.2	99,1
	AZ 2	198	5719	52	15	5991	46	30	181	92.9	99,8
	Reibung	443	5991	46	5	6557	54	30	172	99.3	97,6

Auch in diesem Fall kann grundsätzlich eine gute Performanz der Novelty Detection über den gesamten Lernprozess festgestellt werden. Eine Ausnahme stellt, wie bereits in Abschnitt 4.2.4.4 vorgestellt, die Detektion einer erhöhten Lagerreibung vor Wechsel des Ventils MPYE dar. Zur Veranschaulichung ist in Abbildung 4-46 der gesamte Datensatz anhand des Novelty Detection Stacks, also Merkmalsextraktion, Standardisierung und Dimensionsreduktion mittels PCA visualisiert. Hier ist deutlich der Domain Shift durch den Ventilwechsel erkennbar, welcher die Daten als dominanter Einfluss entlang der ersten Hauptkomponente teilt. Erst in der zweiten Hauptkomponente werden die Leckagen des HZ Kammer 1 und 2 sichtbar, wobei ein deutlicher Einfluss der Leckagerate (MPYE 1: 10 und 20 l/h, MPYE 2: 5, 10, 15 und 20 l/h) zu erkennen ist. Die Leckagen am AZ sowie die erhöhte Reibung werden dabei

erst in höheren Hauptkomponenten vom Gutzustand der Schweißzange trennbar. Die Klasse „erhöhte Reibung“ teilt sich dabei in ein deutlich vom Gutzustand abgesetztes Cluster (vgl. Novelty Scores in Abbildung 4-43 a) sowie ein zumindest bis Hauptkomponente 5 nicht vom Gutzustand trennbares Cluster. Dies könnte auf einen zusätzlich überlagerten, nicht annotierten Quereinfluss schließen lassen. Ähnliches zeigt sich beispielsweise im Fall des Gutzustandes nach dem Wechsel des Ventils MPYE. Auch hier teilen sich die Klassen „normal“ und „Reibung erhöht“ in mehrere kleinere Cluster, was auf einen nicht annotierten, überlagerten Quereinfluss schließen lässt. Wie Tabelle 4-6 zu entnehmen ist, stellt im Gegensatz zur erhöhten Reibung bei MPYE 1 die Erkennung der erhöhten Reibung bei MPYE 2 kein Problem dar.

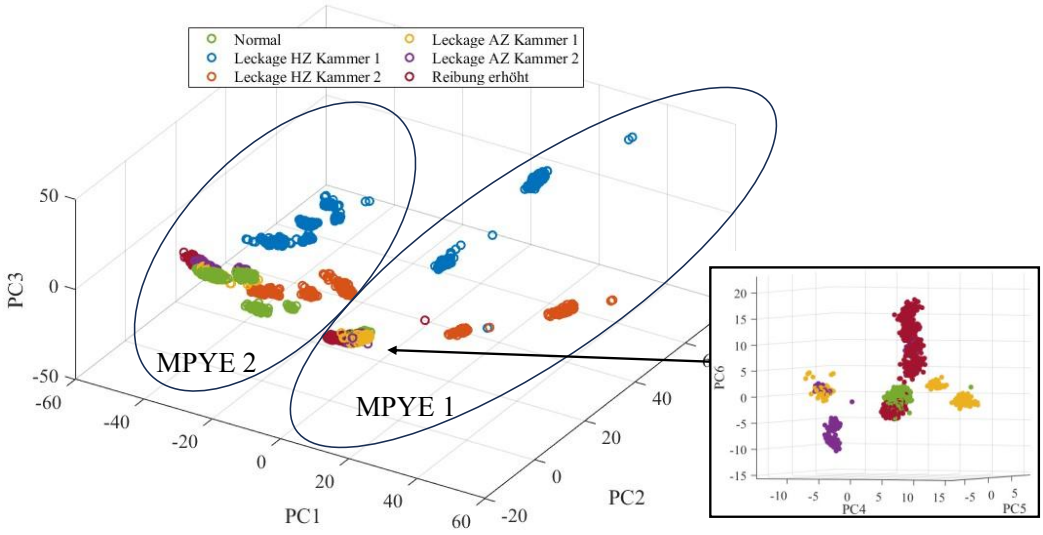


Abbildung 4-46: Visualisierung aller verfügbaren Daten der Schweißzange mittels PCA. Der Ventilwechsel ist deutlich als Domain Shift in den Daten erkennbar. Die Daten mit erhöhter Reibung heben sich erst in höheren Komponenten ab und trennen sich in zwei Cluster, wobei ein Großteil der Daten sehr nah am Cluster der Normaldaten liegt.

Auch für das überwachte Modell kann insgesamt eine gute Performanz über den gesamten Lernprozess festgestellt werden. Es zeigt sich auch hier eine erhöhte Komplexität des Lernproblems bei der Diagnose einer erhöhten Reibung, was sich neben der stark ansteigenden Anzahl an selektierten Merkmalen auch durch vereinzelte Fehlklassifizierungen zwischen den Zuständen „normal“ und „Reibung erhöht“ in Training und Test äußert. In Abbildung 4-47 ist das vollständig angelernte überwachte Modell für die Zustandsüberwachung der servopneumatischen Schweißzange dargestellt. Vergleichbar zur PCA erfolgt die Trennung der Ventile MPYE 1 und MPYE 2 über die erste Diskriminanzfunktion, die Leckagen HZ Kammer 1 und 2 über die zweite Diskriminanzfunktion und die Leckagen AZ und Reibung erhöht werden erst über höhere Diskriminanzfunktionen getrennt. Die Fehler vor (MPYE 1) und nach (MPYE 2) Wechsel des Ventils werden hier als separate Klassen

gehandhabt. In der realen Anwendung wäre ein Zusammenführen der Klassen, die den gleichen Fehler bei unterschiedlichen Umwelteinflüssen beinhalten, sinnvoll. Die dadurch stetig wachsende Varianz in den Trainingsdaten kann zu einer gesteigerten Robustheit und Stabilität der Zustandsüberwachung beitragen.

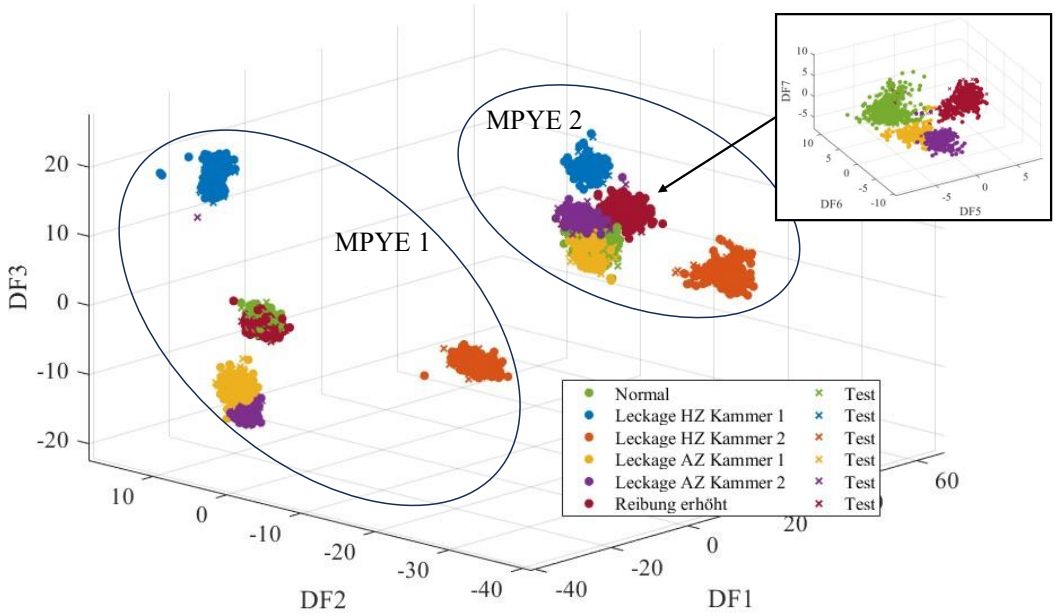


Abbildung 4-47: Vollständig angelerntes Modell mit allen verfügbaren Fehlern. Die Fehler vor (MPYE 1) und nach (MPYE 2) Wechsel des Ventils werden als separate Klassen behandelt.

4.2.5 Zusammenfassung und Diskussion der Anwendung

Die untersuchte Anwendung der servopneumatischen Schweißzange zeigt außerordentliches Potential bezüglich einer Zustandsüberwachung. Durch den Ansatz der aktiven Fehlerdiagnose, basierend auf einem einheitlichen Prüfzyklus, können qualitativ hochwertige Daten erzeugt werden, welche sich hervorragend im Kontext einer Zustandsüberwachung eignen. Die dafür genutzte Prozesssensorik scheint größtenteils ausreichend, um typische Fehler detektieren und diagnostizieren zu können. Eine Ausnahme stellt hier die Erkennung der erhöhten Lagerreibung vor dem Wechsel des Ventils da, welche auch bei der Übernahme in das überwachte Modell noch vereinzelt zu Fehlklassifizierungen führt. Dies verdeutlicht die Grenzen des teilüberwachten Lernens, welches auf die Detektion dominanter Muster in den Daten beschränkt ist.

Eine rein auf überwachten Methoden basierende Zustandsüberwachung ist allerdings als kritisch anzusehen. Durch die große Anzahl an in der Realität denkbaren Fehlern sowie den im industriellen Umfeld stark schwankenden extrinsischen und intrinsischen Einflüssen ist es nahezu unmöglich, sämtliche Events in den

Trainingsdaten abzudecken. Treten für das überwachte Modell unbekannte Fehler auf, können falsche Zustände angezeigt werden, was im schlimmsten Fall zusätzliche Kosten durch unnötige oder unpräzise Wartungsmaßnahmen nach sich zieht. Hier kann eine das überwachte Modell unterstützende Domain Shift Detection die Klassenzugehörigkeiten neu eintreffender Datenpunkte absichern.

Wie im Fall der Schweißzange gezeigt werden konnte, können Umbauten an den Betriebsmitteln einen gravierenden Domain Shift verursachen, welcher ein zuvor angelerntes Modell größtenteils unbrauchbar zurücklässt. Unabhängig von den verwendeten Methoden und der betrachteten Anwendung ist die Generalisierbarkeit eines überwachten Modells maßgeblich durch die in den Trainingsdaten enthaltene Varianz bestimmt. Das Erzeugen einer solchen Varianz im Falle der Schweißzange würde neben der Variation der zu detektierenden fehlerhaften Systemzustände eine umfassende Variation von Komponenten bzw. Umbauten an der Schweißzange erfordern. Dadurch steigen die Komplexität und damit die Kosten des DoE zur Erzeugung einer nachhaltigen Trainingsdatenbasis erheblich an. Auch in diesem Fall bietet die vorgeschlagene adaptive Zustandsüberwachung die Möglichkeit, das Modell der Schweißzange während des Betriebs zu erweitern und sukzessive neue Zustände anzulernen bzw. die Varianz der Trainingsdaten bereits bekannter Zustände stetig zu erweitern.

Speziell bei der Unterstützung der Identifikation der Leckagen konnte die unterstützende Klassifizierung eindeutige Hinweise auf die Ursache, d.h. der Druckluftversorgung der Zylinder, liefern. Im Gegensatz dazu zeigt sich die Identifikation der erhöhten Reibung als deutlich schwieriger, da das unterstützende Modell hier deutlich an Komplexität zulegt. Ob die Qualität der Hinweise beispielsweise von der Reihenfolge der angelernten Fehler abhängt, müsste in Zukunft noch genauer untersucht werden.

Auch im Falle der servopneumatischen Schweißzange zeigen sich die Fehler als abrupte Zustandsänderung, wohingegen in der Realität eher ein langsames Wegdriften des Systemzustands zu beobachten wäre. Auch hier sollte aus diesem Grund auf ein Nachtrainieren des Gutzustands verzichtet werden, um den Kontrast zwischen Gutzustand und Fehler zu maximieren.

4.3 Akustische Leckagedetektion an Pneumatikzylindern

4.3.1 Einführung

Im Hinblick auf die immer weiter steigenden Anforderungen an die Nachhaltigkeit und Effizienz beim Betrieb industrieller Anlagen nimmt die Drucklufterzeugung eine wichtige Rolle ein. Insgesamt fallen etwa 10 % des industriellen Stromverbrauchs in Industrieländern, in Deutschland sogar 14 %, bei der Drucklufterzeugung an [194]. Dabei treten bei Druckluftsystemen typischerweise Verluste von 20 % bis 40 % auf, was sie zu einem Hauptfaktor für den effektiven Betrieb von industriellen Anlagen macht [195]. In Tabelle 4-7 sind dazu exemplarisch die auftretenden Energieverluste und -kosten in Abhängigkeit der Größe der Leckage und des Versorgungsdrucks zusammengefasst. Hier wird deutlich, dass nur durch eine frühzeitige Detektion von Leckagen ein effizienter und nachhaltiger Betrieb pneumatischer Anlagen möglich ist.

Tabelle 4-7: Jährliche Energieverluste und -kosten durch Leckage an Druckluftanlagen. Basierend auf einem Preis von 0,06 € pro kWh und 8000 Betriebsstunden pro Jahr [195].

Loch-durch-messer [mm]	**Luftverlust**		**Energieverlust**		**Kosten**	
	Bei 6 bar [l/s]	Bei 12 bar [l/s]	Bei 6 bar [kWh]	Bei 12 bar [kWh]	Bei 6 bar [€]	Bei 12 bar [€]
1	1,2	1,8	0,3	1,0	144	480
3	11,1	20,8	3,1	12,7	1488	6096
5	30,9	58,5	8,3	33,7	3984	16176
10	123,8	235,2	33,0	132,0	15840	63360

In diesem Kontext spielen die Überwachung und Diagnose pneumatischer Antriebe als Teil komplexer, hochautomatisierter Montagelinien eine wichtige Rolle. Ziel des Projektpartners Festo im Rahmen des Forschungsprojektes KI-MUSIK4.0 ist aus diesem Grund der Aufbau eines akustischen Sensorsystems für die Leckagedetektion pneumatischer Antriebe. Das System sollte dabei einerseits als Option für neu angeschaffte Antriebe wie auch als Retrofit Lösung für Bestandsanlagen geeignet sein. In Abbildung 4-48 ist das Zielbild der Entwicklung in Form eines adaptiven KI-Sensorsystems dargestellt, welches an einem pneumatischen Antrieb des Typs Festo DSBC-32-100 befestigt ist.

Das System soll dabei ein parametrierbares Interface (z.B. OPC-UA) besitzen, welches eine Anpassung der Auswertealgorithmik an verschiedene Antriebstypen ermöglichen

soll. Durch die hardwarenahe Signalauswertung bzw. Modellinferenz soll das System dem Nutzer anzeigen, ob und im Idealfall in welcher Intensität eine Leckage vorhanden ist. Generell können im Kontext akustischer Überwachungssysteme die unkontrollierbaren, industriellen Hintergrundgeräusche, abhängig von der Umgebung, ein Problem für die Zuverlässigkeit des Systems darstellen.

Zum Zeitpunkt der Anfertigung dieser Arbeit befand sich das zu entwickelnde akustische Sensorsystem zur Leckagedetektion noch in einem sehr frühen Stadium. Ziel der Anwendung ist es also im ersten Schritt, grundlegende Fragen zur Machbarkeit einer akustischen Leckagedetektion zu untersuchen. Dazu müssen zunächst gemeinsam mit dem Projektpartner Festo eine Datenstruktur entwickelt und erste Versuchsreihen auf ihren Informationsgehalt bezüglich der zu bestimmenden Leckagen untersucht werden. Das übergeordnete Ziel der Anwendung ist eine Quantifizierung der auftretenden Leckagen und eine darauf basierende Aussage über den Handlungsbedarf aufgrund eines möglicherweise ineffizient laufenden Prozesses.

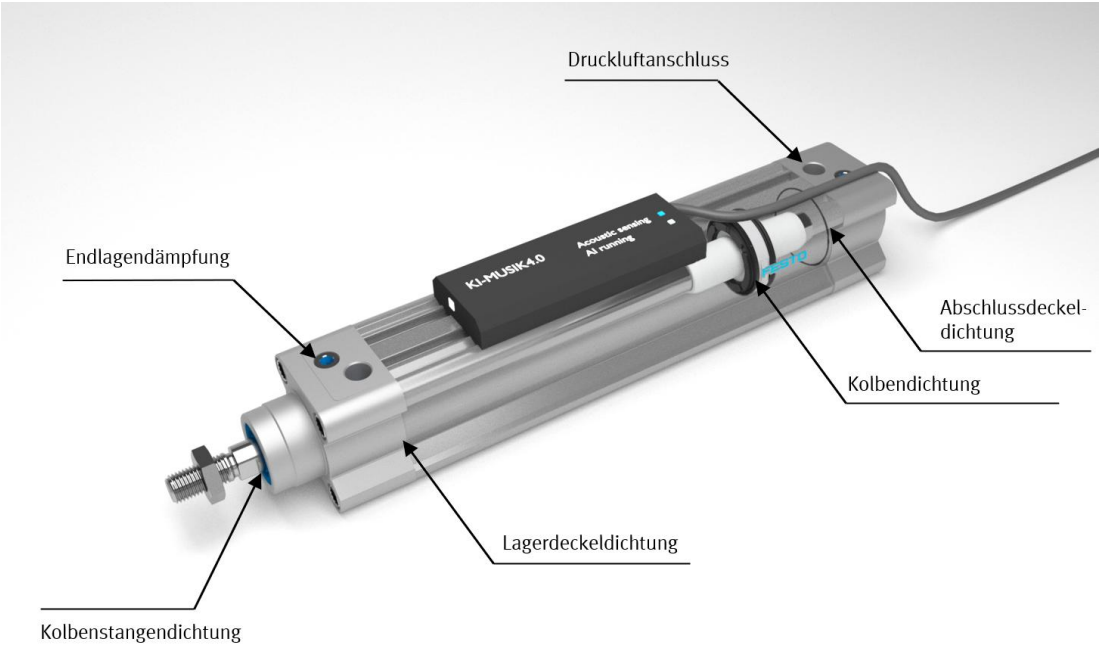


Abbildung 4-48: Zielvorstellung einer akustischen Leckagedetektion für pneumatische Antriebe. KI-Sensorsystem mit hardwarenaher Signalverarbeitung befestigt an einem pneumatischen Antrieb [196].

4.3.2 Experimenteller Aufbau

Typischerweise können Leckagen an einem Pneumatikzylinder an allen in Abbildung 4-48 dargestellten Dichtungen, also Kolbenstangendichtung, Abschlussdeckeldichtung und Lagerdeckeldichtung auftreten. Da die Dichtung am Abschlussdeckel und Lagerdeckel während des Betriebs keiner Reibung ausgesetzt sind, sind Leckagen hierbei beispielsweise auf eine fehlerhafte Montage, Verschmutzungen oder Materialfehler zurückzuführen. Diese Leckagen können

bereits bei der End of Line Prüfung detektiert und durch Nacharbeit beseitigt werden. Im Gegensatz dazu ist die Dichtung an der Kolbenstange während des Betriebs durchgängig belastet, was zu einer Degradation der Dichtung und einer damit verbundenen Leckage führen kann. Im Rahmen des Projektes wurden die zu detektierenden Leckagen von Festo bis etwa 150 l/h festgelegt. Um eine verschlissene Dichtung und die daraus resultierende Leckage nachzubilden, wurden Dichtungen mechanisch oder mittels Laser bearbeitet und anschließend in einen Zylinder montiert. Abbildung 4-49 zeigt exemplarisch eine sowohl mechanisch als auch mittels Laser bearbeitete Dichtung. Da die Leckage nur grob durch die Größe der Beschädigung gesteuert werden kann, wird nach der Montage über ein Leckagemessgerät die Leckage in Abhängigkeit des Versorgungsdrucks bestimmt. Da die Zylinder nach der Montage nicht mehr geöffnet werden können, bilden Zylinder und Dichtung für die Versuche eine feste Kombination. Bei den untersuchten Zylindern kann also bei konstantem Versorgungsdruck von einer konstanten Leckage ausgegangen werden.

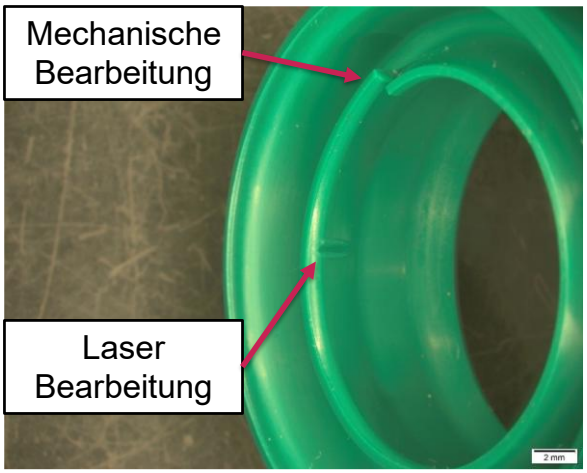


Abbildung 4-49: Manipulierte Dichtung an der Kolbenstange. Die Dichtung wird manipuliert und anschließend in den Zylinder eingesetzt. Nach der Montage wird die Leckage bestimmt [196].

Für die akustische Detektion der Leckagen werden zwei MEMS- Mikrofone der Firma STMicroelectronics durch Festo erprobt. Hierbei handelt es sich um das low-cost Mikrofon MP23ABS1 mit einer schmalen Bandbreite bis etwa 10 kHz und das Ultraschall Mikrofon IMP23ABSU mit einer Bandbreite bis 80 kHz [197, 198]. Neben den Signalen dieser dedizierten Sensoren sind die digitalen Schaltsignale der für die Kammerbelüftung verantwortlichen Ventile sowie die digitalen Schaltsignale der beiden Endlagenschalter verfügbar.

Um Daten für den Entwicklungsprozess des Sensorsystems zu generieren, wurde vom Projektpartner Festo ein Prüfstand entwickelt. Der Prüfstand besteht, wie in Abbildung 4-50 abgebildet, aus drei Ebenen. Die oberen beiden Ebenen stellen dabei Prüfebenen dar, die mit Prüflingen verschiedener Baugrößen bestückt werden können. Darunter befindet sich eine Versorgungsebene, welche die Steuerung in Form einer

speicherprogrammierbaren Steuerung sowie die pneumatische Versorgung beinhaltet. Hier können beispielsweise der Versorgungsdruck oder der Durchfluss konfiguriert werden. Zylinder verschiedener Baugrößen können auf zwei Ebenen parallel betrieben werden, wodurch in den Messkampagnen komplexe industrielle Hintergrundgeräusche simuliert werden können. Während der Versuche können die Prüflinge automatisiert verfahren werden, wobei die Signale der verwendeten Sensoren sowie die Steuersignale über einen Datenlogger erfasst werden.

Tabelle 4-8: Verfügbare Sensoren für die Leckagedetektion.

Sensor / Messgröße	Bezeichnung	Bandbreite	Abtastrate
Mikrofon 1: ST MP23ABS1	MP23ABS1	10 kHz	200 kHz
Mikrofon 2: ST IMP23ABSU	IMP23ABSU	80 kHz	200 kHz
Ventil 1	V1	-	5 kHz
Ventil 2	V2	-	5 kHz
Endlagenschalter 1	ES1	-	5 kHz
Endlagenschalter 1	ES2	-	5 kHz

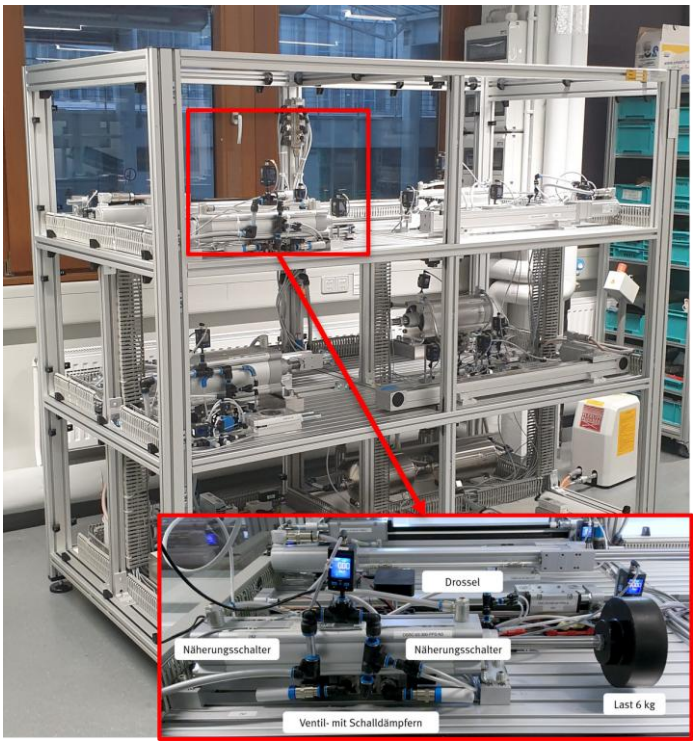


Abbildung 4-50: Prüfstand zum Generieren von Daten für die Leckagedetektion an Pneumatikzylindern. Ein oder mehrere Prüflinge können zum Testen montiert und über eine SPS automatisiert betrieben werden. Parallel werden die Daten der verwendeten Sensoren und Steuersignale über einen Datenlogger erfasst (angepasst nach [196]).

4.3.3 Mess- und Datenplanung

4.3.3.1 Strukturelle Datenqualität

Die am Prüfstand aufgezeichneten Daten liegen in binären .mdf Dateien vor und beinhalten jeweils 5 Arbeitszyklen eines pneumatischen Zylinders. In Abbildung 4-51 sind exemplarisch die Daten eines Arbeitszyklus, bestehend aus Ein- und Ausfahrt des Pneumatikzylinders, dargestellt. Im ersten Teil des Arbeitszyklus belüftet Ventil 1 die Kammer des Zylinders, wodurch die Kolbenstange einfährt und in Endlage 2 stehen bleibt. Im Anschluss belüftet Ventil 2 die Zylinderkammer, wodurch die Kolbenstange ausfährt und in Endlage 2 stehen bleibt. In einer industriellen Anwendung ist der Ablauf der Bewegungsphasen und damit auch der emittierte Luftschall sehr stark von der Konfiguration des Arbeitszyklus abhängig. Mittels des zur Verfügung gestellten Durchflusses und Drucks können beispielsweise die Zylindergeschwindigkeit und die Zylinderkraft konfiguriert werden. Im Gegensatz dazu sind konfigurationsunabhängig bei Richtungswechseln der Kolbenstange kurze Ruhephasen in den Endlagen vorhanden. Dadurch hat man gleiche Ausgangsbedingungen für eine Leckagedetektion, unabhängig von der verrichteten Aufgabe des Zylinders. Zudem könnten die im Vergleich zur Leckage hohen Lautstärken beim Losreißen und Einfahren in die Endlage (bzw. Endlagendämpfung) sowie beim Be- und Entlüften der Kammern eine Detektion der Leckagen deutlich erschweren. Aus diesem Grund wird nach Vorgabe des Projektpartners Festo für die Leckagedetektion die Endlage des Zylinders, also die Ruhephase, verwendet.

Wie in Abbildung 4-51 sind für eine Erkennung bzw. Quantifizierung einer Leckage an der Kolbenstangendichtung nur die Daten der Endlage 2 relevant, da nur in diesem Zeitraum der richtige Teil der Zylinderkammer mit Druck beaufschlagt ist. Um die strukturelle Datenqualität in Form segmentierter und synchronisierter Daten sicherzustellen, müssen also die Ruhephasen in Endlage 2 aus den Daten extrahiert und in ein Szenario-Format gebracht werden.

Neben den eigentlichen Sensordaten der Mikrofone sind in den Dateien die Zeitstempel zu den einzelnen Messwerten aller Sensorsignale verfügbar. Um die Daten in das für das maschinelle Lernen nutzbare Szenario-Format zu bringen, wird die in Abbildung 4-52 dargestellte Datenvorverarbeitung durchlaufen. Im ersten Schritt können hier durch Ableiten des Signals von Endschalter 2, gefolgt von einer Peakdetektion, die Indizes bzw. Zeitstempel der steigenden und fallenden Flanken detektiert werden. Das Übertragen der dabei identifizierten Zeitstempel auf die Daten der Mikrofone erlaubt das Segmentieren der Gesamtdaten in die einzelnen Ruhephasen.

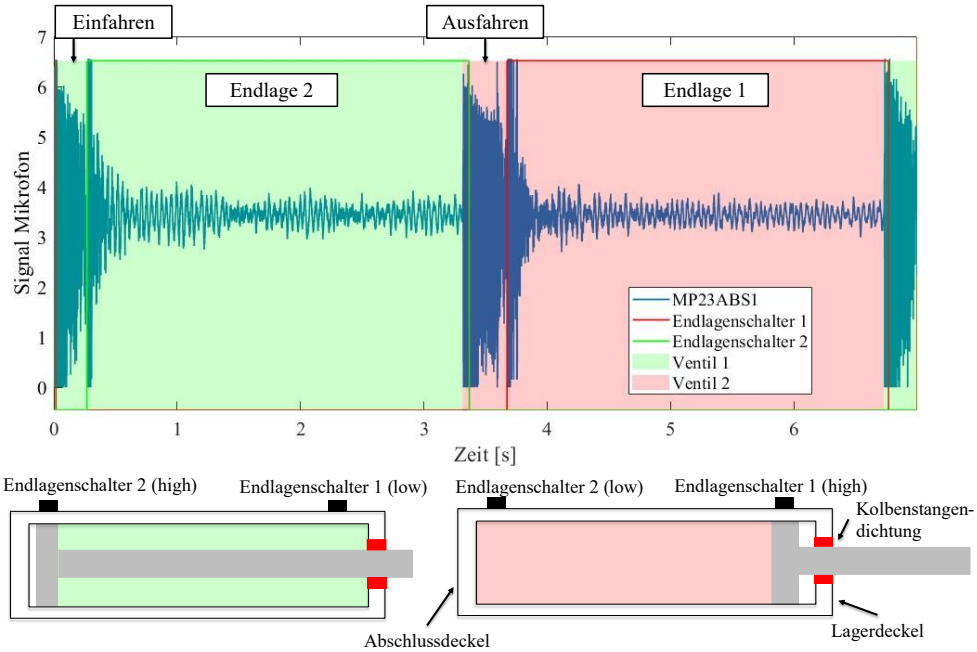


Abbildung 4-51: Arbeitszyklus des Pneumatikzylinders anhand des Mikrofonsignals von MP23ABS1 (10 kHz) und der zusätzlich verfügbaren Steuersignale der beiden Endschalter und der Ventile zum Belüften der Kammern des doppelwirkenden Zylinders.

Zu den Sensordaten werden die von Festo in Form von Excel Tabellen zur Verfügung gestellten Ergebnisse der Referenzmessung mittels Leckagemessgerät als Zielgröße in das Szenario-Format abgelegt. Als einziger bekannter Quereinfluss wird die Zylinder bzw. Dichtungsnummer aufgenommen, welche für die Auswertung eine realitätsnahe Leave-One-Cylinder-Out (LOCO) Validierung gestattet. Durch die Segmentierung ergeben sich ca. 2,5 Sekunden lange Zeitreihen aus den Ruhephasen. Diese werden wiederum in 0,5 Sekunden lange Segmente unterteilt, welche dann zum Training der Modelle verwendet werden. Dadurch kann einerseits die Anzahl der für das Training zur Verfügung stehenden Stichproben erhöht werden. Andererseits sind abhängig von der jeweiligen Anwendung keine 2,5 Sekunden lange Ruhephasen vorhanden, weshalb die Auswertung der verkürzten Zeitreihen eine breitere Palette an industriellen Anwendungen abdeckt.

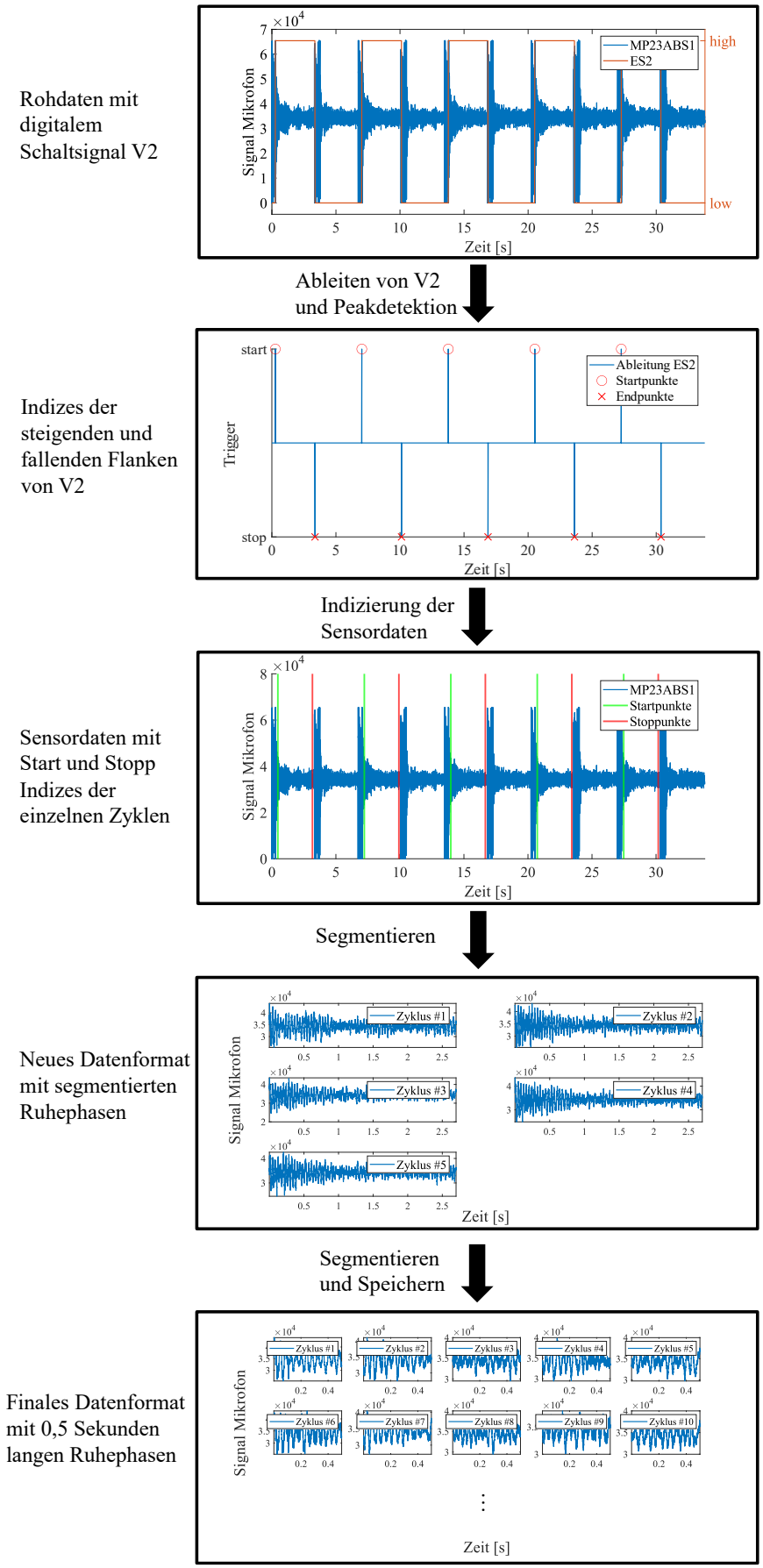


Abbildung 4-52: Datenvorverarbeitung, um die Mikrofondaten in den einzelnen Ruhephasen der Endlage 2 in ein Szenario-Format zu bringen.

4.3.3.2 Explorative Datenanalyse

Nachdem die Daten in strukturierter Form vorliegen, kann eine Prüfung auf Plausibilität bzw. Information in Form einer explorativen Datenanalyse durchgeführt werden. Im Rahmen des Projektes wurden von Festo mehrere Iterationen an Testdaten aufgezeichnet und zur Verfügung gestellt, bevor mit der eigentlichen, umfangreichen Datengenerierung begonnen wurde. Unter anderem wurde dadurch die Verstärkung der Mikrofone an die Lautstärkepegel der Leckagen in der Ruhephase angepasst sowie ein Datenverlust bei der Übertragung der Daten vom Datenlogger identifiziert. Da es sich dabei um einen lang andauernden, iterativen Prozess handelt, wird im Folgenden exemplarisch die Prüfung der Datenqualität anhand der Daten eines Zylinders durchgeführt.

Im ersten Schritt wird die Auswirkung der Leckage auf das Signal der Mikrofone bzw. deren Spektren untersucht. Um relevante Frequenzbereiche zu identifizieren, werden Modelle für die einzelnen Zylinder auf Basis des in Abbildung 4-53 dargestellten Stacks trainiert und die für die Modellbildung selektierten Merkmale analysiert. Die Darstellung erfolgt dabei exemplarisch anhand der Daten des Zylindertyps DSBC-32-100.

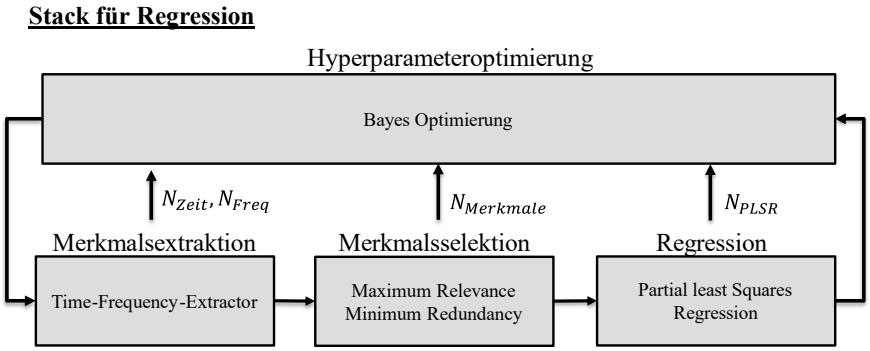


Abbildung 4-53: Stack für die Auswertung der Szenarien der einzelnen Zylinder.

In Abbildung 4-54 a) ist das Ergebnis einer 10-fachen Kreuzvalidierung dargestellt. Der RMSE beträgt hierbei 1 l/h und wird unter Verwendung von 0 Intervallen im Zeitbereich und 40 Intervallen im Frequenzbereich erzielt. Für die PLSR werden 28 Merkmale und 20 Komponenten verwendet. In Abbildung 4-54 b) sind die für die Merkmale zugrunde liegenden Frequenzbereiche dargestellt. Wie hier zu sehen ist, ergeben sich bereits für das Modell eines Zylinders komplexe Merkmalszusammensetzungen aus verschiedensten Frequenzintervallen. Neben dem Bereich der hörbaren Frequenzen bis 20 kHz, sind hier auch Merkmale aus dem Ultraschall-Bereich bis über 60 kHz vertreten.

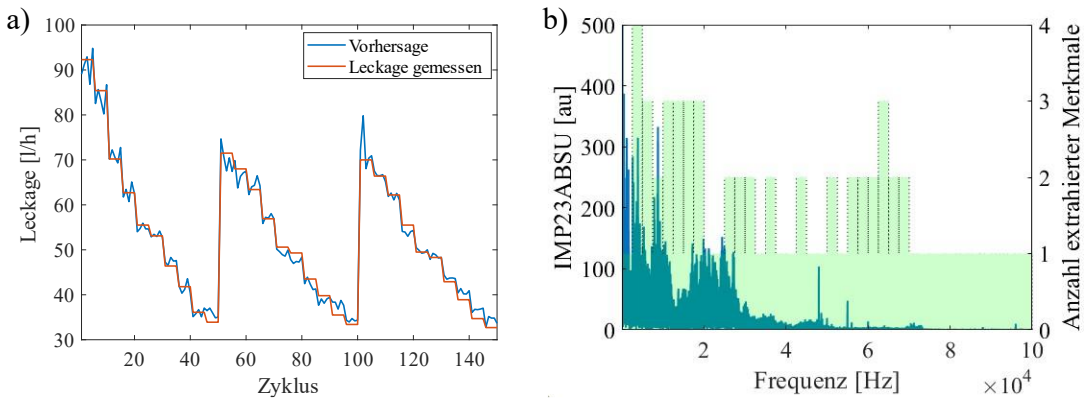


Abbildung 4-54: a) Modell für Zylinder 4 (DSBC-32-100) auf Basis von IMP23ABSU, b) Frequenzintervalle der für die Modellbildung verwendeten Merkmale.

Werden die Daten nach einer steigenden Leckage sortiert zeigen die Frequenzbereiche, aus denen die höchstkorrelierten Merkmale extrahiert werden, ein mit der Leckage zunehmendes breitbandiges Rauschen, wie in Abbildung 4-55 a) zu sehen. Das Merkmal wird mittels RMS berechnet und zeigt, wie in Abbildung 4-55 b) dargestellt, eine nahezu optimale Korrelation zur Leckage von 0,99 (Pearson). Ein vergleichbares Ergebnis ergibt sich für das Modell, basierend auf den Daten des Mikrofons MP23ABS1, sowie für die Modelle der Zylinder des Typs 63-300, weshalb auf eine Darstellung hier verzichtet wird.

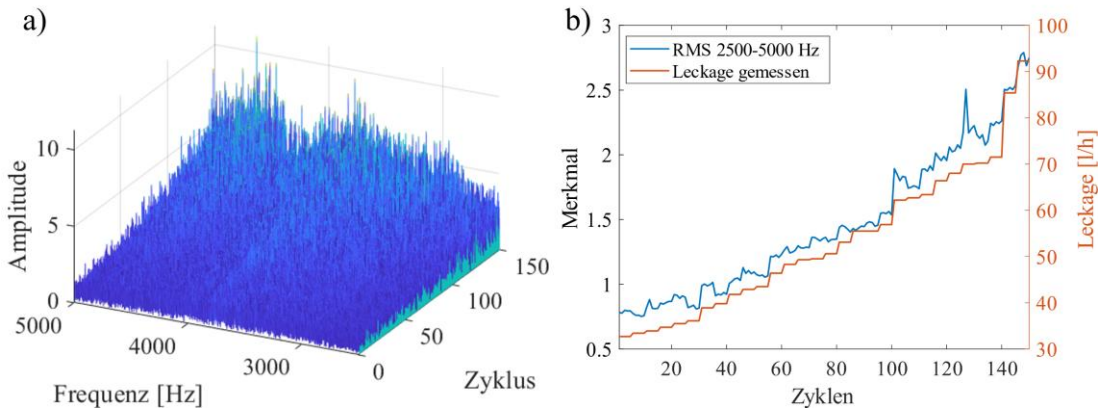


Abbildung 4-55: a) Frequenzbereich, aus dem das höchstkorrelierte Merkmal extrahiert wird, b) Verlauf des höchstkorrelierten Merkmals, berechnet mittels RMS im Bereich von 2500 – 5000 Hz.

4.3.3.3 Design of Experiment

Nachdem eine grundlegende Datenqualität sichergestellt wurde, kann mit einer umfangreicheren Datengenerierung gestartet werden. Im Rahmen dieser Arbeit wurde die akustische Leckagedetektion anhand der Datensätze von zwei unterschiedlichen Baugrößen des DSBC-Zylinders untersucht. Insgesamt wurden 8 Zylinder-Dichtungskombinationen für den Typ DSBC-32-100 (Kolbenstangendurchmesser – Hub) und 10 für den Typ 63-300 vermessen. In Abbildung 4-56 sind die DoEs für

beide Zylindertypen mit den durch das Leckagemessgerät bestimmten Leckagen dargestellt. Um ein vom Versorgungsdruck unabhängiges Modell bilden zu können, wurde für die Datenaufzeichnung der Versorgungsdruck zwischen 4 und 8 bar variiert. Statt der erwarteten positiven Korrelation kann ausgenommen von Zylinder 3 eine negative Korrelation zwischen Versorgungsdruck und gemessener Leckage beobachtet werden. Nach Rücksprache mit dem Projektpartner Festo ist dies durch einen selbstabdichtenden Effekt der Dichtung bei steigendem Druck zu erklären, wodurch weniger Luft aus dem Zylinder entweichen kann. Bei Zylinder 1 des Typs DSBC-32-100 handelt es sich um einen unbeschädigten Zylinder sowie bei Zylinder 1 und 11 des Typs DSBC-63-300.

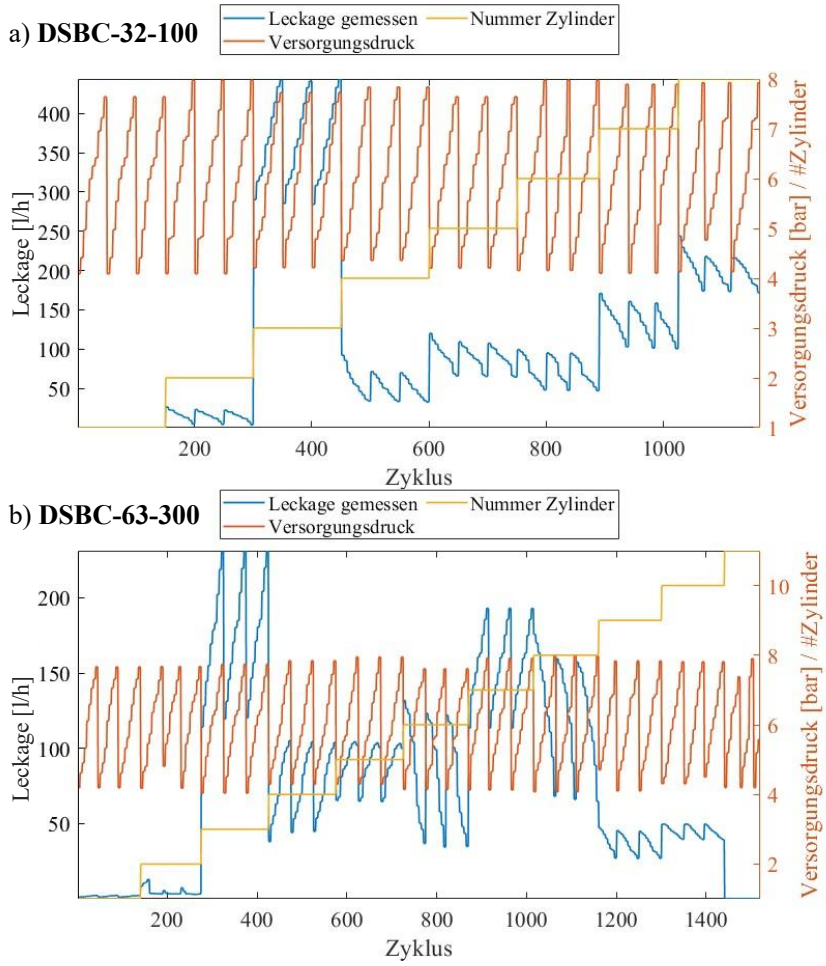


Abbildung 4-56: DoE für die Leckagedetektion für die untersuchten Baugrößen DSBC-32-100 und DSBC-63-100.

4.3.4 Auswertung und Modellbildung

4.3.4.1 Modelle einzelner Zylinder des Typ DSBC-32-100

Auf Basis des in Abbildung 4-53 gezeigten Stacks werden, basierend auf den Daten des Ultraschall Mikrofons IMP23ABSU, Modelle für die einzelnen Zylinder zur Quantifizierung der Leckage trainiert und mittels einer randomisierten 10-fachen Kreuzvalidierung validiert. Um zusätzlich die Übertragbarkeit der Modelle einzelner Zylinder auf die Daten der restlichen Zylinder zu überprüfen, wurden die trainierten Modelle der einzelnen Zylinder für die Vorhersage der Leckagen der restlichen Zylinder verwendet. Die resultierenden Fehler sind in Form des RMSE in Abbildung 4-57 zusammengefasst. Die Diagonale enthält dabei die Validierungsfehler der einzelnen Zylindermodelle. Auch hier ergeben sich vergleichbare Ergebnisse für den Zylindertyp DSBC-62-300, weshalb auf eine explizite Darstellung verzichtet wird.

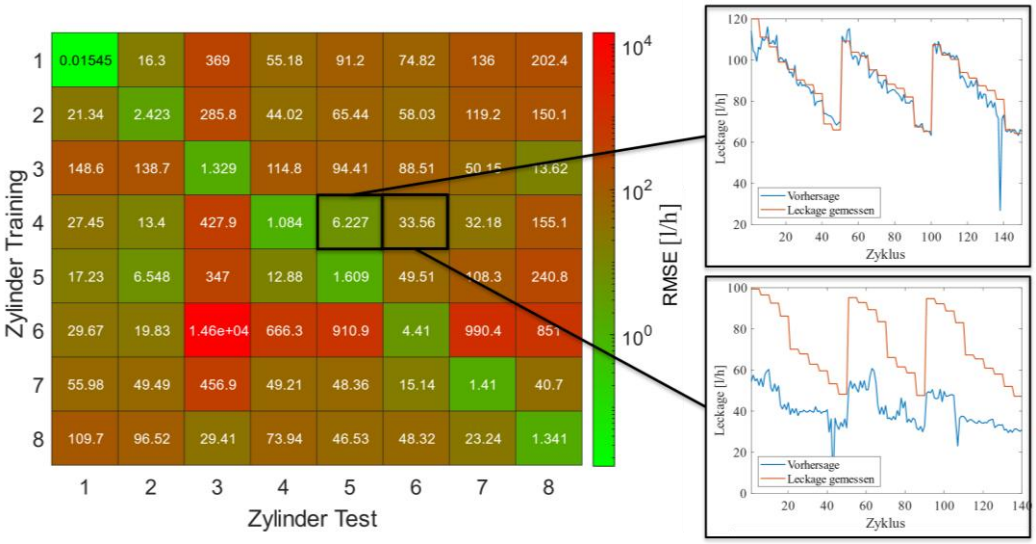


Abbildung 4-57: Übertragbarkeit der Modelle zwischen Zylindern anhand des Test RMSE. Das Modell von Zylinder 4 lässt sich vergleichsweise gut auf Zylinder 5 übertragen, wohingegen eine schlechte Übertragbarkeit zu Zylinder 6 gegeben ist.

Für die einzelnen Zylinder können jeweils sehr präzise Vorhersagen der Leckagen getroffen werden. Der maximale Fehler tritt bei Zylinder 6 auf und beträgt 4,41 l/h. Bei der Übertragbarkeit der Modelle zwischen den Zylindern sind allerdings erhebliche Unterschiede festzustellen. Während Zylinder 4 und 5 eine gute Übertragbarkeit der Modelle aufweisen, versagen in den meisten anderen Fällen die Vorhersagen vollständig. Dies kann einerseits durch die sich nicht überschneidenden Leckagewerte der einzelnen Zylinder erklärt werden, wodurch die Modelle zur Extrapolation gezwungen werden. Andererseits hat die Analyse der für die einzelnen Zylinder selektierten Merkmale gezeigt, dass sich die Leckagen in abweichenden Frequenzbereichen äußern. Dies könnte durch einen dichtungsabhängigen Klang der

Leckage erklärt werden. Durch die fehlende Information während der Trainingsphase sind die für die Quantifizierung der Leckage benötigten Klangmuster im Modell noch nicht vorhanden.

4.3.4.2 Zylinderübergreifende Modelle

In Abbildung 4-58 a) ist das Ergebnis einer randomisierten 8-fachen Kreuzvalidierung eines zylinderübergreifenden Modells für den Typ 32-100 auf Basis der Daten des Ultraschall Mikrofons IMP23ABSU dargestellt. Das beste Ergebnis wird unter Verwendung von einem Zeitintervall, 24 Frequenzintervallen, 208 Merkmalen und 100 PLSR-Komponenten erzielt und der verbleibende RMSE beträgt 5,08 l/h. Der Fehler ist hier bei allen Zylindern im Bereich unter 7 l/h, was als ausreichende Genauigkeit für eine Quantifizierung der Leckage angesehen werden kann. In Abbildung 4-58 b) ist das Ergebnis der LOCO Kreuzvalidierung dargestellt. Die Bayes Optimierung erzielt das beste Ergebnis in diesem Fall unter Verwendung von einem Zeitintervall, 40 Frequenzintervallen, 250 Merkmalen und 36 PLSR-Komponenten. Der verbleibende RMSE beträgt dabei 18,98 und liegt damit deutlich über dem der randomisierten Kreuzvalidierung. Auch hier kann für die meisten Zylinder eine Quantifizierung der Leckagen mit ausreichender Genauigkeit erfolgen. Eine Ausnahme bildet hier Zylinder 6, dessen Leckage offensichtlich mittels der für die restlichen Zylinder selektierten Merkmale nicht quantifiziert werden kann.

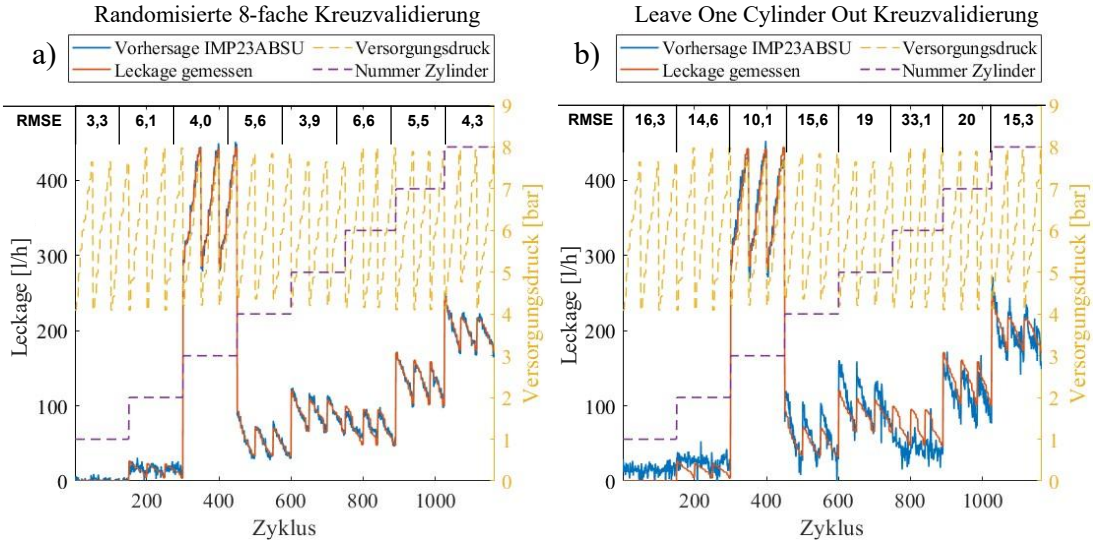


Abbildung 4-58: Ergebnisse der überwachten Modellbildung für a) eine randomisierte Kreuzvalidierung und b) eine LOCO Kreuzvalidierung mit Angabe der Fehler (RMSE) für die einzelnen Zylinder.

Bemerkenswert an dieser Stelle ist vor allem, dass für Zylinder 4, der die höchsten Leckagen aufweist, die beste Quantifizierung erreicht wird, obwohl das Modell an dieser Stelle weit im extrapolierenden Bereich ist. Da lediglich ein intakter Zylinder

(Zylinder 1) im Datensatz vorhanden ist, kann das Modell bei der LOCO Validierung diesen Zustand nicht erkennen und zeigt auch hier eine Leckage an.

In Tabelle 4-9 sind die Ergebnisse für die Quantifizierung der Leckagen für beide Mikrofone, Zylindertypen und Validierungsszenarien zusammengefasst. Das beste Ergebnis wird für alle Modelle unter Verwendung von keinem oder einem Segment im Zeitbereich identifiziert. Kein Segment bedeutet dabei, dass lediglich ein Spektrum aus dem gesamten verfügbaren Zeitbereich berechnet wird und ausschließlich Merkmale im Frequenzbereich für die Modellbildung verwendet werden. Auch bei einem Segment wird lediglich ein Spektrum aus dem gesamten Zeitberiech berechnet, es werden aber zusätzlich Merkmale aus dem Zeitbereich bestimmt. Grundsätzlich zeigen hier die Modelle für den Zylindertyp DSBC-32-100 geringere Fehler. Auch beim Vergleich der Mikrofone kann festgestellt werden, dass die Modelle, basierend auf dem Ultraschall Mikrofon IMP23ABSU in allen Fällen einen etwas geringeren Fehler aufweisen. Dies lässt vermuten, dass ein Teil der Informationen über die Leckage in einem Frequenzbereich außerhalb der Bandbreite des MP23ABS1 liegt.

Tabelle 4-9: Zusammenfassung der Ergebnisse für die Quantifizierung der Leckage für beide Zylindertypen und Validierungsszenarien.

Zylinder Typ	**Mikrofon**	**Validierung**	**Stack Parameter**				**RMSE**
			N_{Zeit}	$N_{Frequenz}$	$N_{Merkmal}$	N_{PLSRK}	
32-100	MP23ABS1	RKV8	1	32	194	100	5,56
32-100	MP23ABS1	LOCO	0	40	220	95	22,81
32-100	IMP23ABSU	RKV8	1	24	208	100	5,08
32-100	IMP23ABSU	LOCO	1	40	250	36	18,98
63-300	MP23ABS1	RKV8	1	39	247	98	7,14
63-300	MP23ABS1	LOCO	1	42	200	89	29,04
63-300	IMP23ABSU	RKV8	0	40	225	100	6,95
63-300	IMP23ABSU	LOCO	1	29	233	31	19,32

Eine Erklärung für die Verschlechterung der Quantifizierung bei einer LOCO-Validierung lässt sich durch die Analyse der Merkmale finden. In Abbildung 4-59 ist der Frequenzbereich dargestellt aus dem ein hochkorreliertes Merkmal der Modelle des Zylindertyps 32-100, basierend auf den Daten des Ultraschall Mikrofons extrahiert wird. Es ist deutlich zu erkennen, dass die hier durch die Leckage emittierte Frequenz nicht für alle Zylinder vorhanden ist. Dies deutet wiederum auf ein zylindertypisches Klangbild hin, welches nur bekannt ist, wenn ein Teil der Daten des entsprechenden Zylinders im Training enthalten ist.

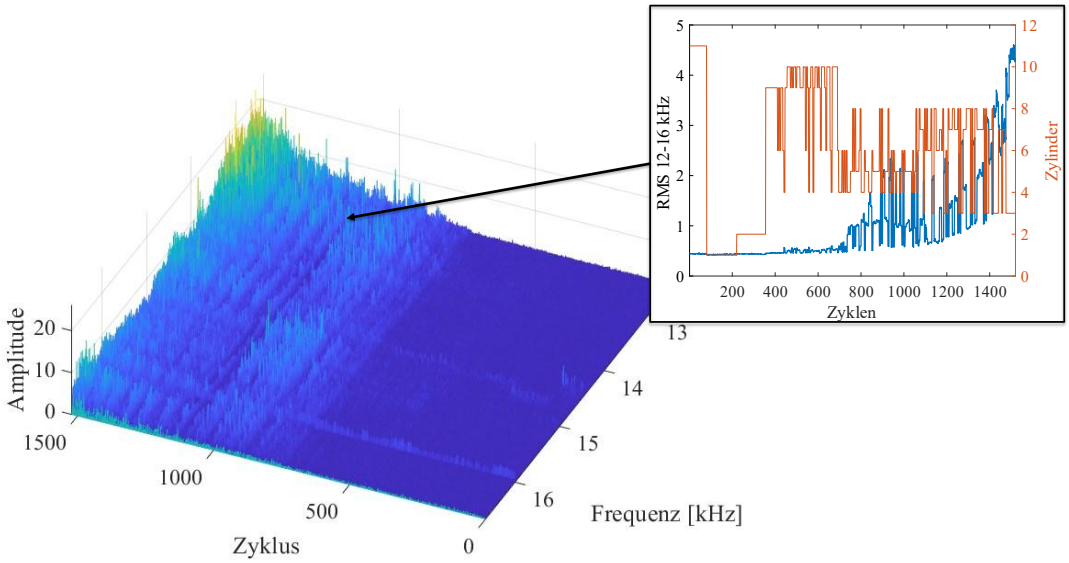


Abbildung 4-59: Frequenzbereich aus dem eines der höchskorrelierten Merkmale im Bereich zwischen 12 und 16 kHz extrahiert wird.

4.3.5 Adaptive Zustandsüberwachung

Im Folgenden wird eine adaptive Zustandsüberwachung für die Leckagedetektion an Pneumatikzylindern simuliert. Verwendet werden hierfür wiederum die Daten des Ultraschallmikrofons IMP23ABSU zur Überwachung des Zylindertyps DSBC-32-100. Als Ausgangssituation wird angenommen, dass noch kein Modell für eine Leckage vorhanden ist (Zylinder 1). Dann erfolgt die Simulation einer fortschreitenden Leckage auf Basis der Daten von Zylinder 4, 5 und 3. Mittels Zylinder 4 wird das Auftreten einer ersten Leckage simuliert, welche mittels Novelty Detection erkannt und folgend in das überwachte Modell übernommen wird. Tritt im Anschluss eine vergleichbare Leckage auf (Zylinder 5) kann diese mittels des überwachten Modells bestimmt und hinsichtlich des Handlungsbedarfs bewertet werden. Bei Auftreten einer Leckage, deren Klang oder Intensität dem Modell bis dahin nicht bekannt ist (Zylinder 3), kann diese von der Novelty Detection erkannt und wiederum in das überwachte Modell übernommen werden. Für die Novelty Detection und Regression werden die in Abbildung 3-6 dargestellten Stacks verwendet. Im Sinne einer kompakten, anschaulichen Darstellung wird auf die Darstellung der Domain Shift Detection in dieser Anwendung verzichtet. Die Daten jedes Zylinders werden in 80 % Trainings- und 20 % Testdaten aufgeteilt, um eine Validierung der Modelle beim fortschreitenden Lernen der Leckagen durchzuführen. Da es sich, bezogen auf die Dynamik innerhalb eines Zyklus, um einen stationären Prozess handelt, wird für die Merkmalsextraktion der Novelty Detektion mittels TFEx ein Segment im Zeitbereich definiert. Die Segmentierung im Frequenzbereich wird aufgrund der hohen Abtastrate der Mikrofonsignale von 200 kHz auf 100 Segmente festgelegt, wodurch sich 2,5 kHz

breite Frequenzsegmente ergeben. Auch hier ist das Ziel ein Gleichgewicht zwischen der Anpassung an die Trainingsdaten und geglätteten Entscheidungsgrenzen, weshalb die Anzahl der nächsten Nachbarn für das kNN-Verfahren ebenfalls auf k = 5 festgelegt wird.

4.3.5.1 Initialisierung des Normalzustands

In Abbildung 4-60 ist die Initialisierung der Novelty Detection anhand des Histogramms und des QQ-Diagramms für Trainings- und Testdaten dargestellt. Aus Gründen der Visualisierung wurde ein Ausreißer aus den Daten entfernt. Abgesehen von einzelnen Ausreißern liegen sowohl Trainings- als auch Testsamples unter dem festgelegten Grenzwert von 2,63 σ. Bei Betrachtung des QQ-Diagramms kann festgestellt werden, dass die Novelty Scores bereits im Gutzustand von der Normalverteilung abweichen, was ebenfalls durch den SW-Test bestätigt wird. Auch hier ergibt sich bezüglich der Verteilung der Daten bei Initialisierung der Modelle immer ein ähnliches Bild mit geringfügigen Abweichungen von der Normalverteilung, weshalb auf eine explizite Darstellung der QQ-Diagramme im weiteren Lernprozess verzichtet wird.

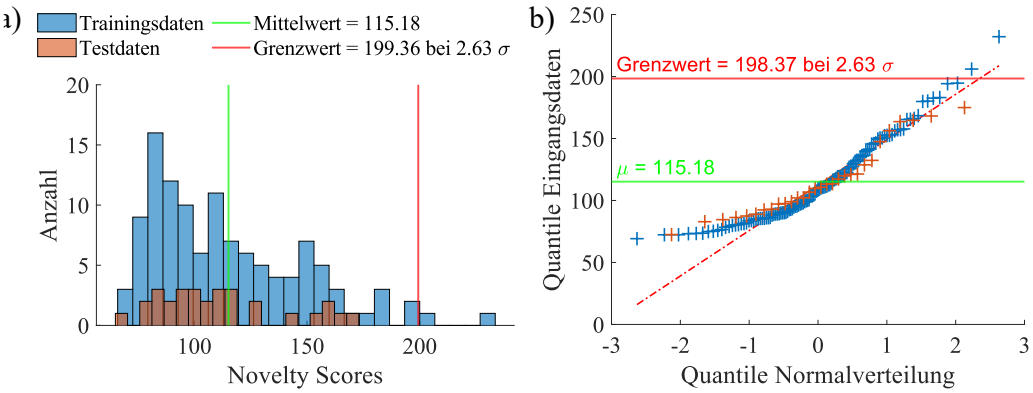


Abbildung 4-60: a) Histogramm der Novelty Scores der Trainingsdaten im Normalzustand, b) QQ-Diagramm der Novelty Scores der Trainingsdaten im Normalzustand.

4.3.5.2 Detektion einer unbekannten Leckage

Im nächsten Schritt wird das Auftreten einer kleinen Leckage durch die Daten von Zylinder 4 simuliert. Die entsprechenden Novelty Scores sind in Abbildung 4-61 dargestellt. Die Leckage wird in diesem Fall deutlich von der Novelty Detection angezeigt, wobei ebenfalls eine klare Korrelation zwischen Leckage und Novelty Scores beobachtet werden kann. In Anschluss wird die erkannte Leckage in das überwachte Modell übernommen. Das Ergebnis ist in Abbildung 4-61 b) dargestellt. Der RMSE auf Trainings- und Testdaten beträgt dabei 0,96 l/h und 1,92 l/h. Da die Leckage sich hier noch in einem wenig kritischen Bereich befindet, bestände für den

Anlagenbetreiber hier die Möglichkeit, den Zylinder weiter zu betreiben und für zukünftige Wartungsarbeiten vorzumerken.

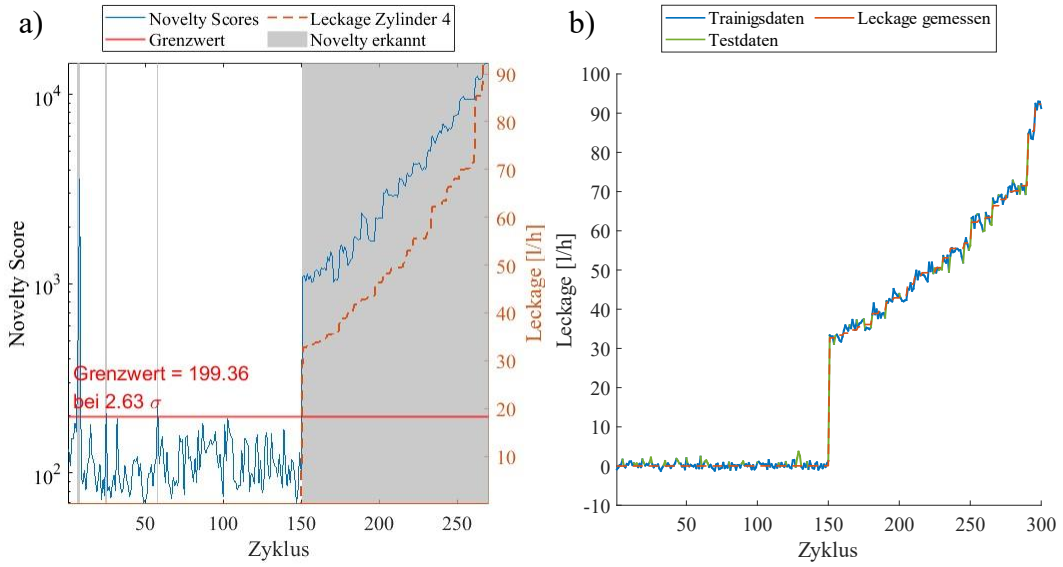


Abbildung 4-61: a) Reaktion der Novelty Detection auf das Auftreten einer Leckage, b) Training des überwachten Modells mit der erkannten Leckage von Zylinder 4.

4.3.5.3 Quantifizierung einer Leckage

Mittels der Daten von Zylinder 5 wird eine sich durch Degradation der Kolbenstangendichtung weiterentwickelnde Leckage simuliert. Wie in Abbildung 4-62 a) zu sehen ist, würde die Novelty Detection die Leckage erst nach dem Anstieg auf 120 l/h anzeigen, weshalb im vorherigen Bereich eine Quantifizierung der Leckage durch das überwachte Modell erfolgen würde. Durch den Ausreißer (Sample 5) sowie die systematischen Erhöhungen des Novelty Scores (Samples 260 – 270) in den Trainingsdaten der Novelty Detection ist hier bereits eine deutliche Erhöhung des Grenzwertes festzustellen. Abbildung 4-62 b) zeigt die entsprechende Quantifizierung durch das überwachte Modell. Der RMSE beträgt hierbei 10,7 l/h. Der Fehler entsteht dabei vor allem in einem Bereich um Zyklus 300 sowie ab Zyklus 370, in dem sich das Modell im Bereich einer Extrapolation befindet.

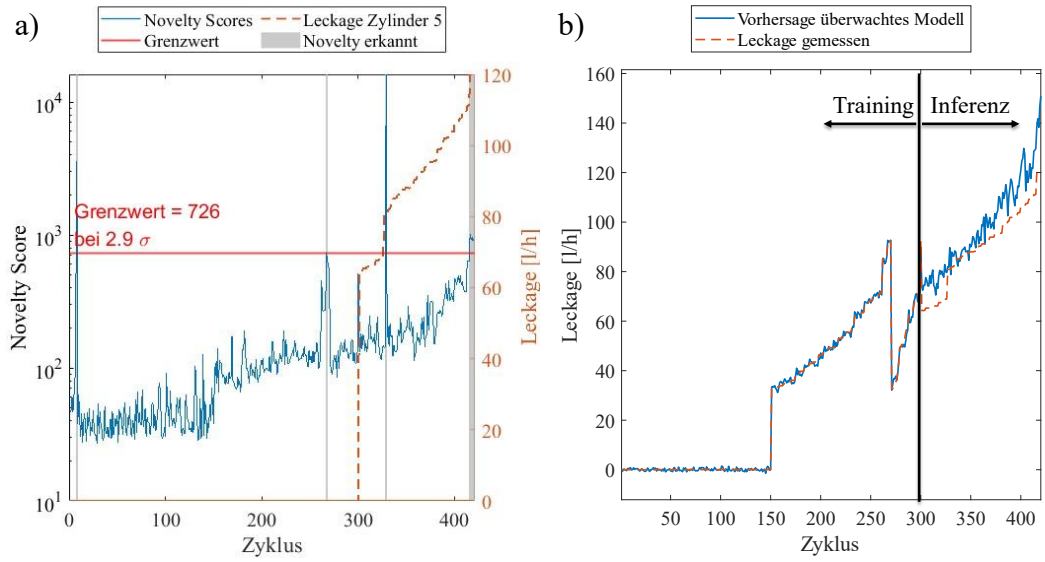


Abbildung 4-62: a) Reaktion der Novelty Detection auf das Auftreten der Leckage von Zylinder 5, b) Vorhersage der Leckage auf Basis des überwachten Modells für Zylinder 5.

4.3.5.4 Erweiterung des Leckage Modells

Ein weiteres Fortschreiten der Leckage wird mittels der Daten von Zylinder 3 simuliert. Diese überschreitet die im Modell bis dahin enthaltenen Leckagen deutlich, was sich ebenfalls in den in Abbildung 4-63 a) abgebildeten Novelty Scores zeigt. Die neue Leckage wird hier deutlich von der Novelty Detection angezeigt. Ebenfalls wird dies anhand der Qualität der vorhergesagten Leckage des überwachten Modells deutlich. Der RMSE steigt hier auf 103 l/h an, was keine ausreichende Genauigkeit darstellt.

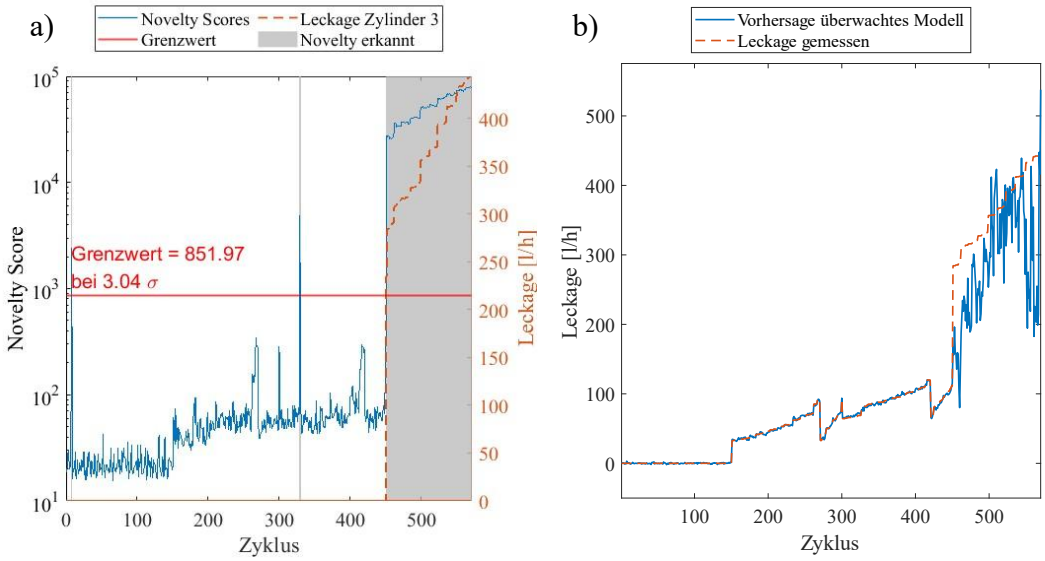


Abbildung 4-63: a) Reaktion der Novelty Detection auf das Auftreten einer Leckage, b) Vorhersage der Leckage auf Basis des überwachten Modells für Zylinder 3.

In diesem Fall müsste der Zylinder ausgebaut und eine Bestimmung der Leckage mittels Leckagemessgerät durchgeführt werden. Im Anschluss kann die bestimmte Leckage in das überwachte Modell übernommen werden

4.3.6 Zusammenfassung und Diskussion der Anwendung

Gemeinsam mit dem Projektpartner Festo konnte ein grundlegendes Konzept für eine dezentrale Leckagedetektion an pneumatischen Systemen erarbeitet und im Rahmen erster Versuche validiert werden. Im ersten Schritt erfolgte dazu eine iterative Mess- und Datenplanung, welche die für die Algorithmen benötigte Struktur auf Basis der Signale der Endlagenschalter generierte. Im Rahmen einer explorativen Datenanalyse konnten bereits relevante Signalbereiche identifiziert und somit der für die Leckagedetektion benötigte Informationsgehalt nachgewiesen werden. Erste Analysen von umfangreicheren im Labor generierten Daten zeigen hierbei vielversprechende Resultate. Für einzelne Zylinder können dabei hochpräzise Vorhersagen der Leckage (mittlerer RMSE: 1,70 l/h) getroffen werden, wobei sich auch hier bereits das Problem der Übertragbarkeit der Modelle zeigt. Als Ursache hierfür können unterschiedliche Klangbilder der Zylinder bzw. Dichtungen identifiziert werden, welche sich in unterschiedlichen Frequenzbändern niederschlagen und dadurch dem Modell noch nicht bekannt sind. Aber auch für zylinderübergreifende Modelle können größtenteils ausreichende Genauigkeiten der Vorhersagen erreicht werden. Besonders hervorzuheben sind hier die positiven Ergebnisse der realitätsnahen LOCO Validierung, welche im Mittel eine Quantifizierung unbekannter Leckagen von unter 20 l/h verspricht. Durch die große Streuung der Fehler für die individuellen Zylinder (von 10,1 l/h bis 33,0 l/h) zeigen sich auch hier teilweise Probleme bei der Übertragbarkeit der Modelle, basierend auf den unterschiedlichen Klangbildern der Leckagen. Ausgehend von den erzielten Ergebnissen ist für die Umsetzung des Systems die Verwendung eines Ultraschallmikrofons zu empfehlen, welches durchweg bessere Vorhersagen ermöglicht.

Resultierend aus diesen Erkenntnissen können für die adaptive Zustandsüberwachung zwei mögliche Ansatzpunkte identifiziert werden. Einerseits können neue, bis dahin dem Modell unbekannte Klangbilder erkannt und in das überwachte Modell übertragen werden. Andererseits kann die Erweiterung des Messbereichs des Modells bezüglich der Leckagen erfolgen. Obwohl die Modelle teilweise im Rahmen der Untersuchungen eine gute Fähigkeit zur Extrapolation aufweisen, ist dies von theoretischer Seite nicht begründbar. Die Erweiterung des Modells um unbekannte Leckagen kann dann, wie in Abschnitt 4.3.5 anhand der Daten von Zylinder 3 dargestellt, erfolgen. Dies setzt allerdings nach der Detektion durch die Novelty Detection den Ausbau des Zylinders

und die Bestimmung der vorhandenen Leckage mittels gängiger Messverfahren voraus.

Für den Einsatz des Systems unter realen industriellen Bedingungen besteht noch ein erheblicher Entwicklungsaufwand. Als kritischste Komponente werden hierbei die unkontrollierbaren Hintergrundgeräusche in industriellen Umgebungen gesehen. Der Einfluss auf die überwachte Modellbildung muss im Rahmen weiterer Versuche oder durch die Überlagerung synthetischer Geräuschprofile detailliert untersucht werden. Ebenfalls sind die für die Absicherung der strukturellen Datenqualität verwendeten Steuersignale nur bei der Integration des Systems in die prozessführende Steuerung vorhanden. Ist dies nicht der Fall, beispielsweise bei der Nachrüstung pneumatischer Zylinder, müsste eine andere Triggerung der Leckagedetektion in den Ruhephasen des Zylinders erfolgen. Auch die zur Verfügung stehende Dauer der Ruhephase ist von der Anwendung abhängig. Der Einfluss dieser Dauer kann beispielsweise durch eine künstliche Verkürzung bzw. Verlängerung der Ruhephasen der bestehenden Datensätze untersucht werden.

Die im Rahmen dieser Arbeit verwendeten Methoden, speziell die Novelty Detection (kNN), sind nicht für eine Ausführung auf dezentralen Systemen mit stark begrenztem Speicher und geringer Rechenkapazität ausgelegt. Durch die Flexibilität des übergeordneten Algorithmus in Bezug auf die verwendeten Stacks könnte aber eine Umsetzung der adaptiven Zustandsüberwachung auf der Edge mittels anderer in der Toolbox implementierter Methoden wie beispielsweise Autoencodern umgesetzt werden.

4.4 Verschleißdetektion beim Wälzschälen

4.4.1 Einführung

Beim Wälzschälen handelt es sich um ein leistungsfähiges, spanabhebendes Verfahren zur Fertigung von Innenverzahnungen, wie sie für leistungsstarke elektrische Antriebsstränge benötigt werden. Aufgrund seiner hohen Präzision und Geschwindigkeit hat das Verfahren in den letzten Jahren stark an Aufmerksamkeit gewonnen [199]. Die komplexe Prozesskinematik verursacht jedoch starke Schwankungen der effektiven Schnittparameter während des Werkzeugeingriffs. Daher muss insbesondere das Werkzeug hohen Anforderungen genügen, um lange Standzeiten bei der geforderten Werkstückqualität zu erreichen. Diese Anforderungen werden bei der Bearbeitung von massiven Rohlingen aus vergüteten Werkstoffen noch verstärkt. Eine Steigerung der Wertschöpfung des Prozesses durch eine effiziente Nutzung der Werkzeuge hinsichtlich des verbleibenden Abnutzungsvorrats sowie das frühzeitige Erkennen von Schädigungen, idealerweise bevor diese zur Fertigung mangelhafter Produkte führen, sind von hohem Interesse [200, 201].

Der grundlegende Aufbau der Werkzeugmaschine mit der Ausrichtung von Werkzeug und Werkstück sowie der relevanten Dreh- und Bewegungsachsen ist in Abbildung 4-64 dargestellt. Während der Bearbeitung werden das Werkzeug sowie das Werkstück über die Spindeln C1 und C4 in eine synchronisierte Drehung versetzt, wobei die Achsen der beiden Spindeln um einen Achskreuzwinkel $\sum$ zueinander gekippt sind. Das Werkzeug erhält dabei zusätzlich eine Vorschubbewegung in x-Richtung des Werkzeugkoordinatensystems. Die Drehzahlen von Werkzeug und Werkstück stehen dabei im umgekehrten Verhältnis wie das Verhältnis der Zähnezahlen [202]. Um die Werkzeugbelastung gering zu halten, erfolgt die Bearbeitung dabei typischerweise in mehreren Schritten, während derer abhängig von den Prozessparametern Material abgetragen wird. Diese materialabtragenden Schnitte werden als Schruppen oder Schruppschnitte bezeichnet. Im letzten Schritt erfolgt eine Feinbearbeitung mittels eines im Vergleich zu den Schruppschnitten längeren Schlichtschnitts, wodurch die benötigte Oberflächenqualität sichergestellt wird. Eine detaillierte Beschreibung des Prozesses und der dazugehörigen Kinematik kann [202] entnommen werden.

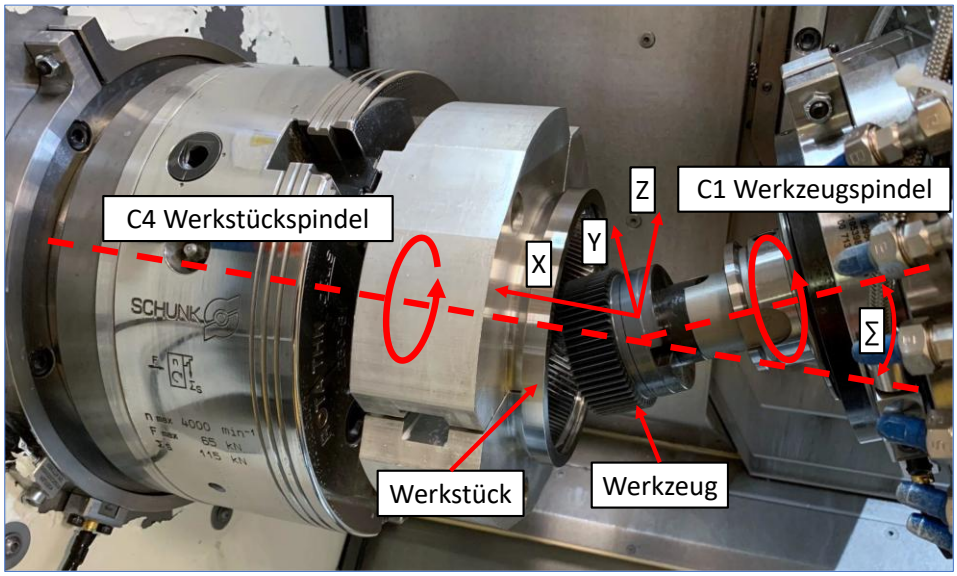


Abbildung 4-64: Grundlegender Aufbau des Wälzschälprozesses mit relevanten Achsen für translatorische und rotatorische Bewegungen (angepasst nach [203]).

Obwohl das Verfahren schon seit Jahrzehnten bekannt ist, gelten bestimmte Bereiche des Schneidprozesses sowie die Einflüsse bestimmter Prozessparameter in Bezug auf die Oberflächenqualität und die Werkzeuglebensdauer als noch nicht vollständig verstanden. Als für den Werkzeugverschleiß relevante Parameter gelten unter anderem die Anzahl der Schruppschnitte, die Kopfspandicke, die Schnittgeschwindigkeit sowie die Kühlschmierstrategie [200].

Ebenfalls steht zu dem Prozess bis dato nur eine geringe Anzahl öffentlich verfügbarer Datensätze zur Verfügung, die eine Untersuchung des Einflusses der verschiedenen Parameter erlauben. Ziel dieser Untersuchungen ist es also im ersten Schritt, eine strukturierte Datenbasis zur Analyse verschiedener Prozessvariationen zu erstellen und daraus ein Modell zur Prognose des Werkzeugverschleißes zu erzeugen. Andererseits ist das Ziel, ein Prozessvariation-unabhängiges Modell für den Werkzeugverschleiß herzuleiten, was eine möglichst unkorrelierte Variation der Prozessparameter in der Datenbasis erfordert. Die Versuche wurden gemeinsam mit dem Projektpartner Schaeffler an einer Versuchsanlage in der Fertigung bei Schaeffler durchgeführt.

4.4.2 Experimenteller Aufbau

Die Versuche wurden an einer Dreh-Fräs-Maschine für die Komplettbearbeitung des Typs CTX Beta 1250 TC von DMG MORI durchgeführt. Diese verfügt über die Möglichkeit, die Daten der internen Steuer- und Regelgrößen darzustellen und in Form von CSV-Dateien zu exportieren. Neben diesen internen Signalen wurden aber vor allem dedizierte Beschleunigungs- und Acoustic Emission Sensoren auf ihren

Informationsgehalt bezüglich des Werkzeugverschleißes untersucht. Zu diesem Zweck wurden die Spindeln von Werkstück (WST) sowie Werkzeug (WZK) mit zusätzlichen Sensoren ausgestattet. Die Übersicht über die verfügbaren Sensoren kann Tabelle 4-10 entnommen werden.

In Absprache mit den Prozessexperten der Projektpartner sowie unter Berücksichtigung der in Kapitel 3.1 beschriebenen Überinstrumentierung wurde die Abtastrate für die Beschleunigungssensoren auf 100 kHz und die der Acoustic Emission Sensoren auf 1 MHz festgelegt.

Tabelle 4-10: Verfügbare Sensoren zur Verschleißdetektion beim Wälzschälen.

Sensor / Messgröße	Be-zeichnung	Ausgang	Position	Bandbreite / Resonanzfrequenz
Momentbildender Strom- Istwert Spindel C1	IC1	Ampere [A]	Intern	-
Momentbildender Strom- Istwert Spindel C4	IC4	Ampere [A]	Intern	-
Positionierachse Position-Istwert C1	PosC1	Position [mm]	Intern	-
Triaxialer IEPE Beschleunigungssensor **PCB 356A45** [204]	ACCXC1 ACCYC1 ACCZC1	Spannung [V]	Spindel C1	±5%: 0,7…7000 Hz ±10%: 0,4… 10000 Hz > 30 kHz
Triaxialer IEPE Beschleunigungssensor **PCB 356A15** [205]	ACCXC4 ACCYC4 ACCZC4	Spannung [V]	Spindel C4	±5%: 2…5000 Hz ±10%: 1,4… 6500 Hz > 25 kHz
Acoustic Emission **Kistler 8152C0050500** [206]	AEC1	Ladung [V]	Spindel C1	50 …400 kHz
Acoustic Emission **Kistler 8152C0050500**	AEC4	Ladung [V]	Spindel C4	50 …400 kHz

Die Befestigung der zusätzlichen Sensoren an Werkstückspindel (Abbildung 4-65, a) und Werkzeugspindel (Abbildung 4-65, c) wurde dabei jeweils über zwei verschraubbare Klemmringe am nicht rotierenden Spindelgehäuse (Abbildung 4-65, b) realisiert. Die Positionierung erfolgt auf beiden Spindeln im Bereich der Schneideneingriffszone unter Berücksichtigung des Positionswinkels K von ~ 25° (Abbildung 4-65, d). Die Sensoren auf der Werkstückspindel sind zusätzlich über den Achskreuzwinkel Σ von ~ 20° zur Werkstückspindel geschwenkt.

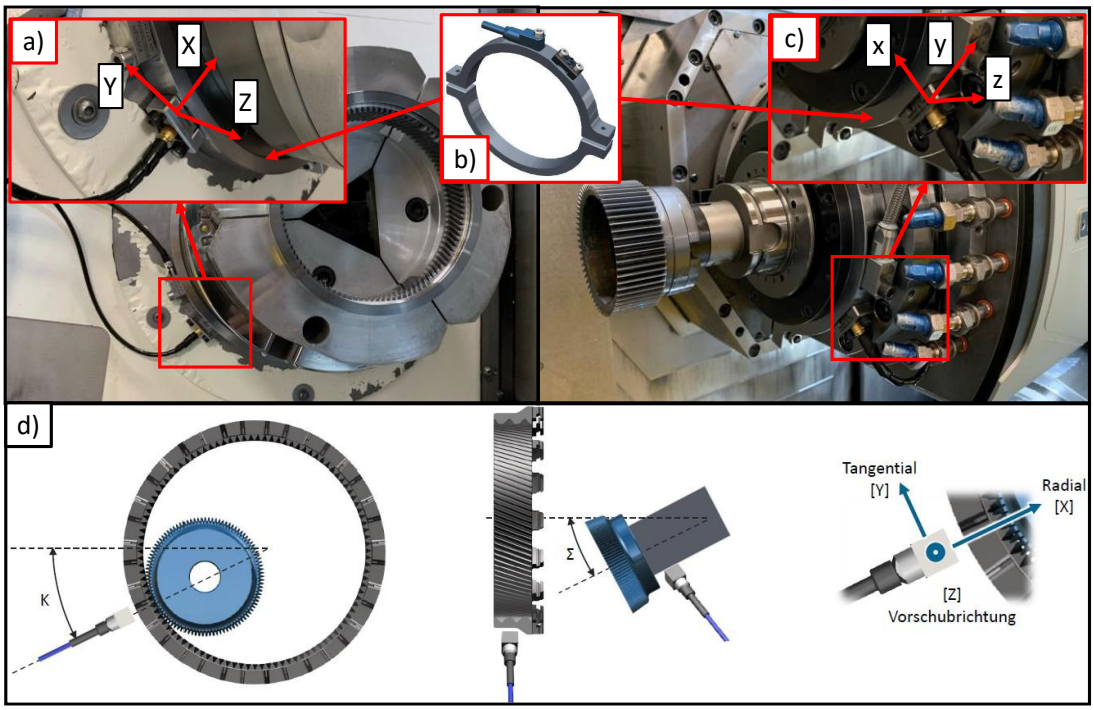


Abbildung 4-65: Ausstattung der Wälzschälmaschine mit zusätzlichen Beschleunigungs- und Acoustic Emission Sensoren mittels Klemmringen an den Spindelgehäusen (angepasst nach [203]).

Bei dem gefertigten Produkt handelt es sich um ein Hohlrad mit Innenverzahnung. Das hierfür verwendete Werkzeug, siehe Abbildung 4-64, verfügt über 60 Zähne, wodurch sich in Abhängigkeit von der Werkzeugdrehzahl (f_{WKZ}) die Zahneingriffsfrequenz nach [207] wie folgt berechnet:

$$f_{Zahneingriff} = f_{WKZ} * n_{Zähne} \tag{4.1}$$

Im betrachteten Anwendungsfall ist die Werkzeugdrehzahl variable und abhängig von der Konfiguration des Prozesses. Typische Drehzahlen liegen im Bereich von 15 Hz, wodurch sich eine Zahneingriffsfrequenz von 900 Hz ergeben würde.

4.4.3 Mess- und Datenplanung

4.4.3.1 Strukturelle Datenqualität der dedizierten Sensoren

Zur strukturierten Erfassung der Daten wurde ein Messsystem aufgebaut, welches eine den festgelegten Abtastraten entsprechende, hochfrequente Abtastung der Sensoren ermöglicht. Verwendet wurde hierzu der in Abbildung 4-66 dargestellte Messkoffer. Dieser basiert auf einem CompactRIO-9040 (cRIO) Chassis der Fa. National Instruments, welches einen Embedded Controller, bestehend aus einem Real-Time-Prozessor (RT) und einer FPGA, bereitstellt [208]. Das Chassis des Controllers bietet Platz für insgesamt vier Messmodule der C-Serie und besitzt zusätzlich einen dedizierten Triggereingang zum Steuern der Messung über ein externes Triggersignal.

Für die Beschleunigungssensoren wurden zwei Module des Typs NI-9232 verwendet, welche drei simultane Kanäle für IEPE-Sensoren mit einer maximalen Abtastrate von 102,4 kHz bei einer Auflösung von 24 Bit zur Verfügung stellen [209]. Zur Aufzeichnung der Acoustic Emission Sensoren wurde das Spannungseingangsmodul NI-9223 genutzt, welches vier differenzielle Kanäle mit einer maximalen Abtastrate von 1 MHz und einer Auflösung von 16 Bit bereitstellt [210]. Um den Ladungsausgang der Acoustic Emission Sensoren in ein Spannungssignal zu wandeln, wurden zusätzlich zwei AE-Piezotron-Kuppler des Typs 5125C der Fa. Kistler verwendet [211].

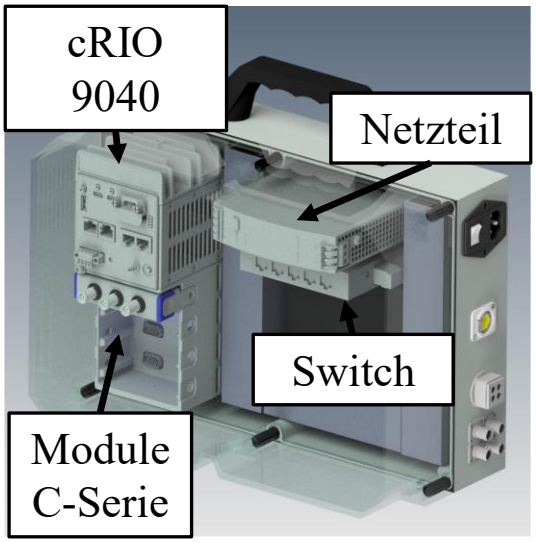


Abbildung 4-66: Verwendete Messkoffer zum Aufzeichnen der Daten der Beschleunigungs- und Acoustic Emission Sensoren.

Erste Versuche, bei denen alle Module in einem Chassis betrieben wurden, zeigten, dass die hohen Datenmengen zu Laufzeitproblemen beim Schreiben der Dateien führen. Aus diesem Grund musste die Erfassung der Signale von Beschleunigungssensoren und Acoustic Emission Sensoren auf zwei Messkoffer aufgeteilt werden. Dadurch ergeben sich zusammen mit den internen Sensordaten der Anlage drei verschiedene Datenquellen, deren Daten für eine Auswertung strukturiert zusammengeführt werden müssen. Da die Länge des Gesamtprozesses stark von dessen Konfiguration abhängt (Anzahl Schruppschnitte, Vorschub, Kopfspandicke), ist für die Datenverarbeitung mittels der in Kapitel 2.3 vorgestellten Methoden eine Segmentierung der Daten in die einzelnen Schnitte notwendig.

Im Fall der Messkoffer wurden hierfür zwei unterschiedliche Hardware-Triggersignale der Anlage verwendet. Das erste Triggersignal zeigt dabei den Prozessstart über eine steigende Flanke und das Prozessende über eine fallende Flanke an und wurde direkt mit dem Triggereingang beider cRIOs zum Starten und Stoppen der Messung verbunden. Ein zweiter Trigger, der mit einer steigenden bzw. fallenden

Flanke den Start bzw. das Ende jedes einzelnen Schrupp- und Schlichtschnitts anzeigt, wird über einen zusätzlichen Spannungseingang an beiden Chassis erfasst. Dieser Trigger dient als Orientierungssignal, welches im Nachhinein eine softwareseitige Segmentierung der einzelnen Schnitte ermöglicht (siehe Abbildung 4-69). Das Aufzeichnen dieses Triggers erfolgt im Falle des AE-Messkoffers mittels des ohnehin vorhandenen Spannungseingangsmoduls NI-9223. Für den Messkoffer ACC wurde ein zusätzliches Spannungseingangsmodul NI-9775 verwendet [212]. Ein Überblick über das Gesamtsystem ist in Abbildung 4-67 gegeben.

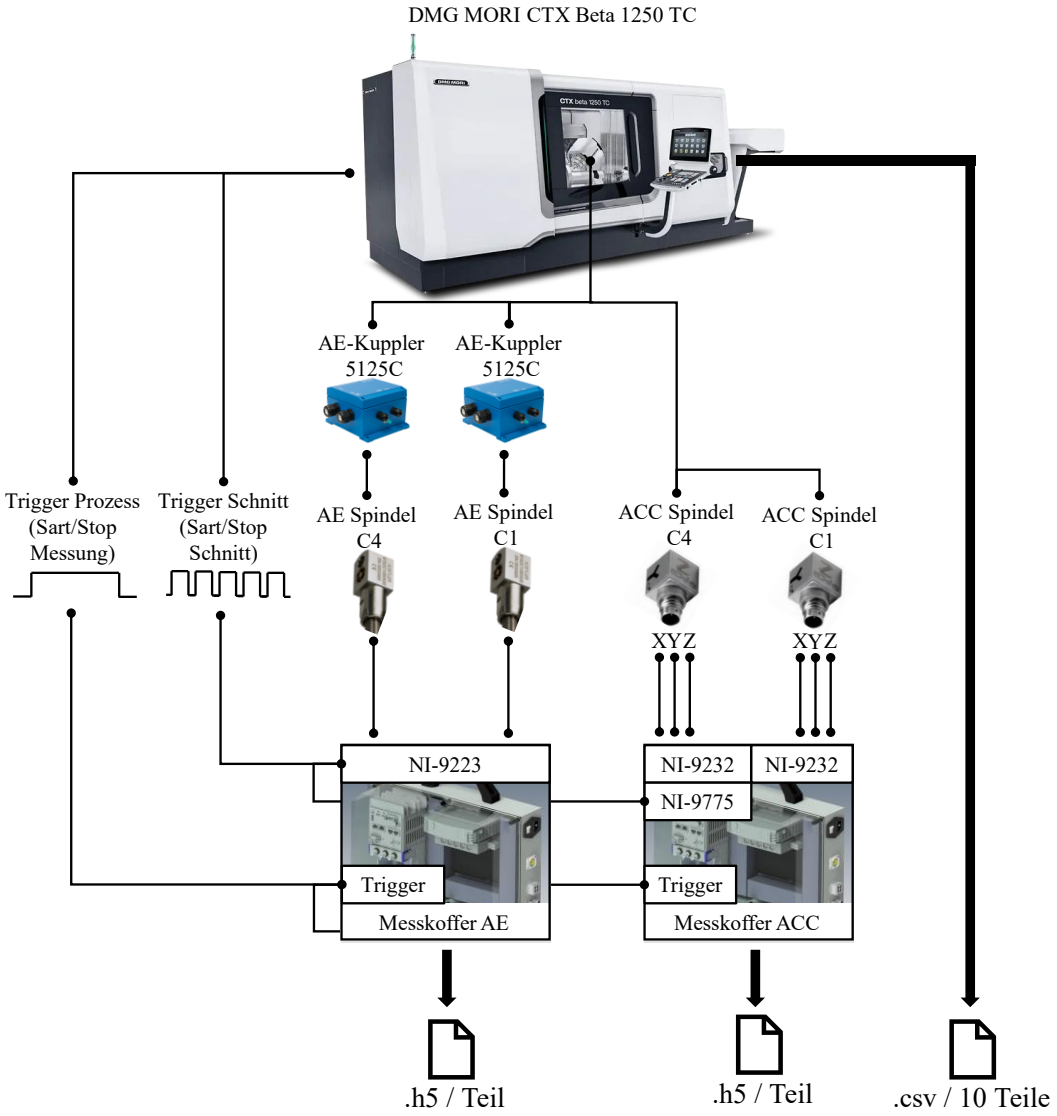


Abbildung 4-67: Übersicht über das Messsystem zum Aufzeichnen der Daten von Beschleunigungs- und Acoustic Emission Sensoren sowie der Prozessdaten, die direkt von der Anlage erfasst werden. Die Daten der verschiedenen Datenquellen werden in separaten Dateien abgelegt und im Nachhinein softwareseitig mit Hilfe der Triggersignale synchronisiert und segmentiert.

Die LabVIEW Software zum Betreiben der Messkoffer arbeitet in einer typischen Producer-Consumer-Architektur. Die Daten von Beschleunigungs- und Acoustic Emission Sensoren werden in einer ersten Schleife vom FPGA des cRIO erfasst und

in einer parallel laufenden zweiten Schleife in separaten hdf5-Dateien auf einer externen Festplatte, welche am Messkoffer angeschlossen ist, gespeichert.

Zur Überprüfung der strukturellen Datenqualität wurden mit dem vorgestellten Messsystem Testdaten aufgezeichnet. In Abbildung 4-68 ist exemplarisch das Beschleunigungssignal der X-Achse des dreiachsigen Beschleunigungssensors an der Werkzeugspindel für einen gesamten Bearbeitungsprozess dargestellt. Bei der gefertigten Variante sind deutlich die sieben Schruppschnitte und der länger dauernde Schlichtschnitt zu erkennen. Ebenfalls ist hier das Triggersignal zum Anzeigen der einzelnen Schnitte sichtbar. Hier zeigen sich allerdings leichte Abweichungen zu dem erwarteten nahezu idealen Rechtecksignal. Die Ursache hierfür konnte nicht aufgeklärt werden.

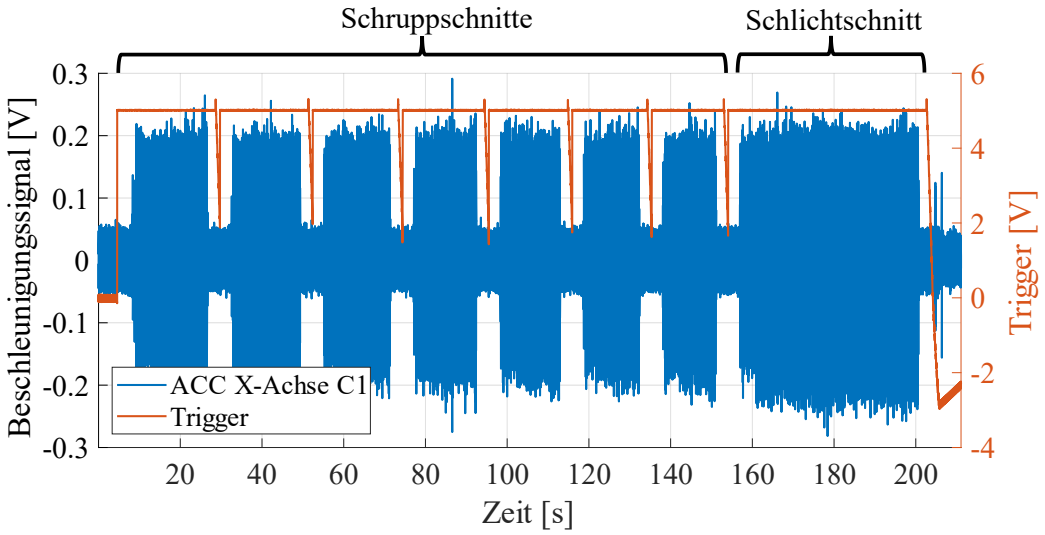


Abbildung 4-68: Exemplarisches Signal der X-Achse des Beschleunigungssensors an der Werkzeugspindel und der Trigger zum Anzeigen der Start- und Endzeitpunkte der Einzelschnitte.

Um den Trigger für die Segmentierung der Daten zu verwenden, sind einige Vorverarbeitungsschritte notwendig. Ausgehend von dem Triggersignal, welches in Abbildung 4-68 zu sehen ist, erfolgt im ersten Schritt durch einen Schwellwertvergleich (Trigger > 4V ≙ high) eine Überführung des Signals in ein ideales Rechtecksignal. Im Anschluss können durch Ableiten des Triggersignals, gefolgt von einer Peakdetektion, die Indizes der steigenden und fallenden Flanken detektiert werden. Das Übertragen der dabei identifizierten Indizes auf die Daten der Beschleunigungs- und Acoustic Emission Sensoren erlaubt das Segmentieren des Gesamtprozesses in die einzelnen Schnitte. Die schrittweise Verarbeitung des Triggersignals sowie das Segmentieren der Daten des Gesamtprozesses in einzelne Schnitte sind in Abbildung 4-69 dargestellt.

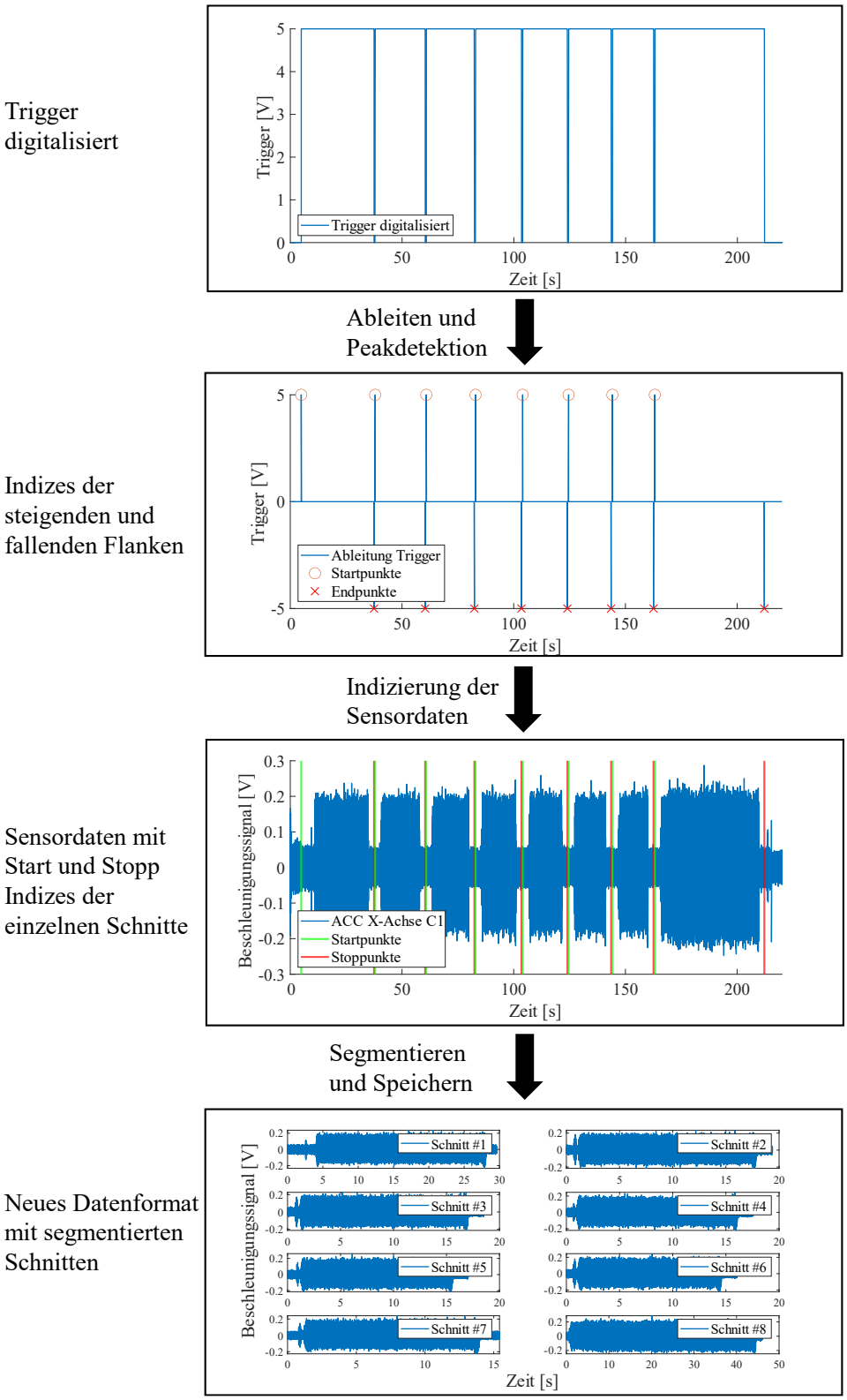


Abbildung 4-69: Vorverarbeitung des Triggersignals und Segmentierung der Sensordaten in einzelne Schnitte.

4.4.3.2 Strukturelle Datenqualität der Prozessdaten

Die Aufzeichnung der internen Prozessgrößen erfolgt jeweils für den gesamten Prozess. Da hier keine Triggersignale verfügbar sind, erfolgt die Segmentierung in Einzelschnitte anhand eines künstlichen Triggersignals, welches mittels der Position des Werkzeugs in Vorschubrichtung generiert wird. Auch hier erfolgt die Erzeugung eines idealisiert rechteckigen Triggersignals durch einen Schwellwertvergleich. Als Schwellwert wurde hier eine Position von 0 mm gewählt, welcher den Eingriffspunkt vom Werkzeug in das Werkstück beschreibt. Bei negativen Positionen befindet sich das Werkzeug dann im Eingriff und bei positiven Positionen im Bewegungsraum der Anlage. Hier wird deutlich, dass der Werkzeugeingriff erst bei ca. der Hälfte der Schnittdauer erfolgt. Zusätzlich ist zu erkennen, dass der Schlichtschnitt aus zwei Teilen besteht. Im ersten Teil erfolgt dabei die eigentliche Oberflächenbearbeitung des Werkstücks. Bei dem zweiten Teil handelt es sich um einen Leerschnitt. Hier ist das Werkzeug nicht mehr im Eingriff und das Werkstück wird lediglich durch ein Durchfahren der Verzahnung von Spänen befreit, wodurch das Risiko für Oberflächenbeschädigungen bei nachfolgenden Bearbeitungsschritten verringert wird. Für die Auswertungen und Modellbildung werden die Daten des Leerschnitts nicht verwendet.

Das Vorgehen zum Segmentieren der Daten erfolgt dann in denselben Schritten, wie bereits in Kapitel 4.4.3.1 für die Daten der dedizierten Sensoren beschrieben. Abbildung 4-70 zeigt das Positionssignal sowie die daraus abgeleiteten Start- und Stopppunkte der einzelnen Schnitte. Aufgrund der Prozesskonfiguration erfolgt zwischen den Schnitten eine Neupositionierung des Werkzeugs, wodurch die ruckartigen Bewegungen zwischen den Schnitten im Positionierungssignal zu sehen sind.

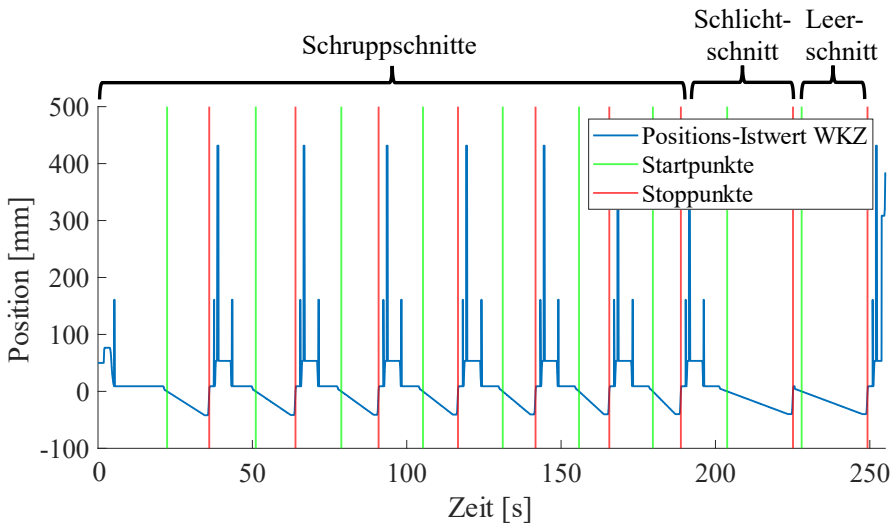


Abbildung 4-70: Erzeugen eines künstlichen Triggers anhand der Werkzeugposition und die daraus abgeleiteten Start- und Stopppunkte der einzelnen Schnitte.

4.4.3.3 Explorative Datenanalyse

Die Plausibilitätsprüfung erfolgt sowohl durch die Betrachtung des Gesamtprozesses als auch durch die Analyse einzelner Schnitte. Eine Darstellung der Analyse aller Sensorsignale ist aufgrund des Umfangs nicht sinnvoll, weshalb die Darstellung hier exemplarisch anhand einzelner Sensoren bzw. Sensorachsen erfolgt.

In Abbildung 4-71 sind das Zeitsignal der Y-Achse des Beschleunigungssensors an der Werkzeugspindel C1, das Signal mit Downsampling um Faktor 10 (10 kHz) sowie das zugehörige Spektrogramm der Originaldaten bis 5 kHz dargestellt. Hier wird deutlich, dass sich zu Beginn der einzelnen Schnitte scharf definierte Frequenzen ausbilden, obwohl das Werkzeug noch nicht im Eingriff mit dem Werkstück ist. Als Ursache für diese Frequenzen werden leichte Unwuchten sowie Einstreuungen durch die benachbarten Elektromotoren vermutet. Bei Eingriff des Werkzeugs entstehen zusätzliche Frequenzanteile, welche dann auf Schwingungen zurückzuführen sind, die durch den Bearbeitungsprozess entstehen. Dies ist auch deutlich im Zeitsignal mit verringerter Abtastrate erkennbar, was darauf schließen lässt, dass ein Großteil der durch die Bearbeitung eingebrachten Schwingungsenergie im Frequenzbereich bis 10 kHz liegt.

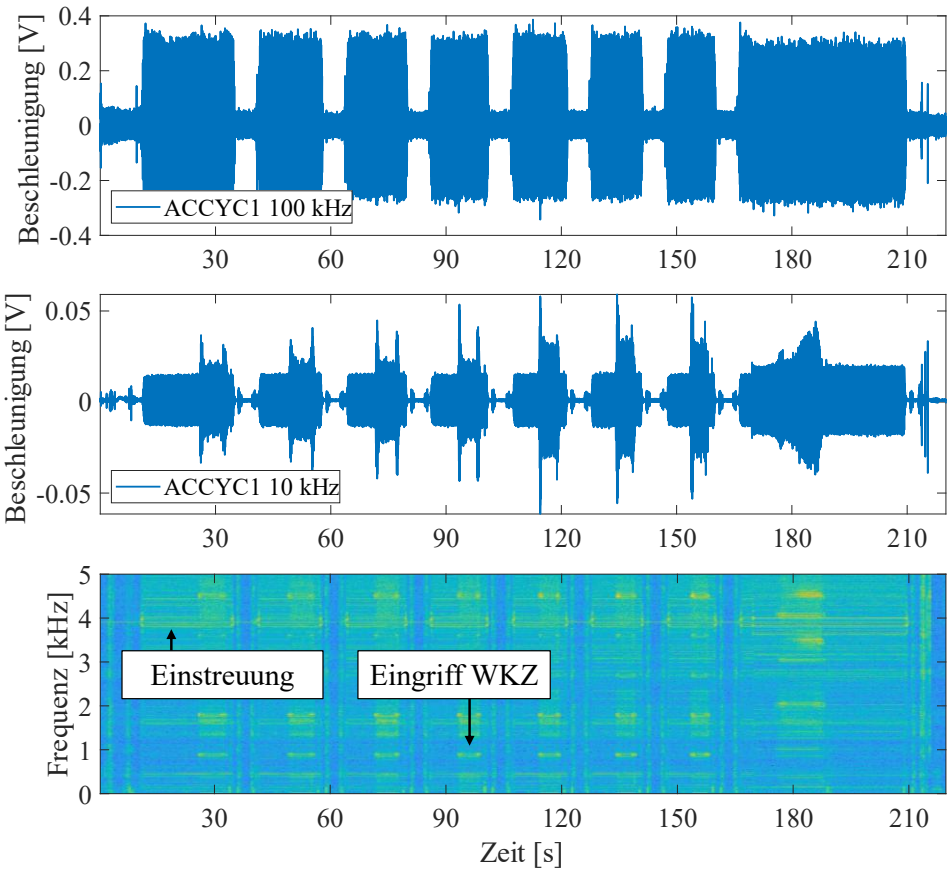


Abbildung 4-71: Zeitsignal der Y-Achse des Beschleunigungssensors an der Werkzeugspindel C1 mit original Abtastrate von 100 kHz (oben), Signal mit Downsampling um Faktor 10 (Mitte), und das Spektrogramm der Originaldaten im Bereich 1 - 5 kHz (unten).

Eine Aussage über den Informationsgehalt spezifischer Frequenzbereiche im Hinblick auf die Verschleißdetektion lässt sich aus diesen Erkenntnissen jedoch nicht ableiten. Auch geringere Amplituden in höheren Frequenzbereichen können relevante Informationen über den Werkzeugverschleiß beinhalten.

Noch deutlicher wird die Entstehung der zusätzlichen Frequenzinhalte bei Betrachtung des Spektrogramms beispielsweise des ersten Schnitts anhand der Daten der Y-Achse des Beschleunigungssensors an der Werkzeugspindel C1, wie in Abbildung 4-72 a) dargestellt. Neben drehsynchronen Frequenzen im Frequenzbereich bis 400 Hz zeigt sich vor allem die Zahneingriffsfrequenz bei ca. 900 Hz als dominierende Frequenz im Bereich bis 1000 Hz. In Analogie zu Schwingungen an Getrieben bilden sich um diese Mittenfrequenz die typischen Seitenbänder im Abstand der Drehfrequenz (Δf) der jeweiligen Spindel für Werkstück (Abbildung 4-72 b) und Werkzeug (Abbildung 4-72 c). Die Zahneingriffsfrequenz und die entsprechenden Seitenbänder werden im Bereich der Getriebeüberwachung genutzten und können auch im Kontext einer Werkzeugverschleißerkennung als potenzieller Informationsursprung angesehen werden [213].

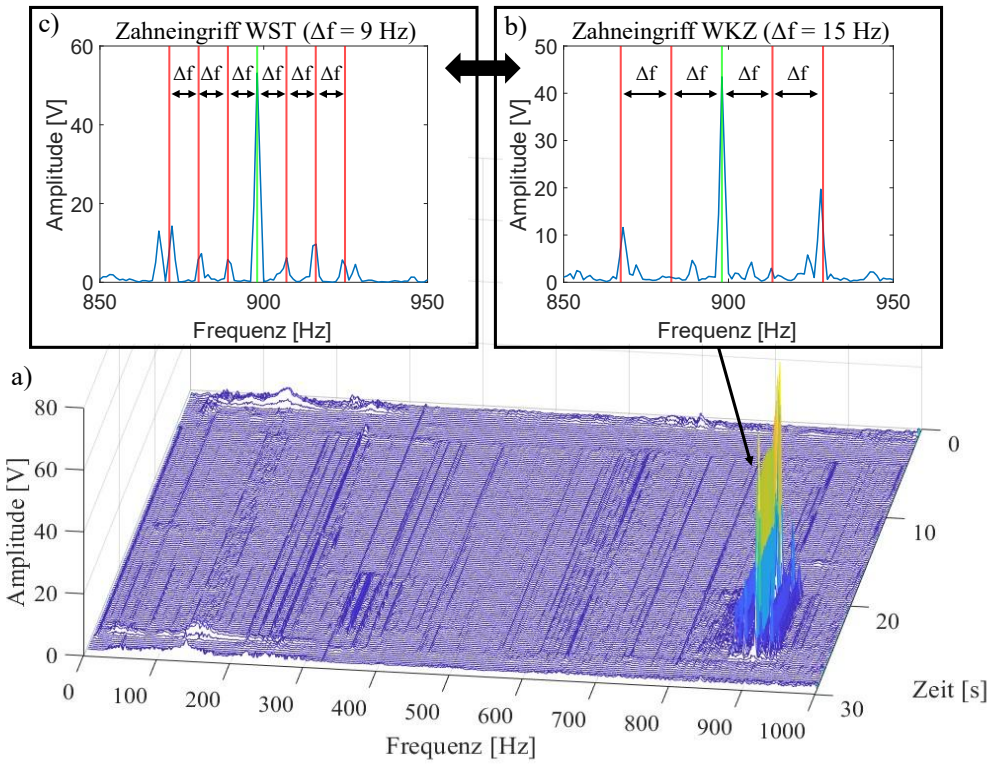


Abbildung 4-72: a) Spektrogramm des ersten Schruppschnitts der Y-Achse des Beschleunigungssensors an der Werkstückspindel. Bei Werkzeugeingriff ist deutlich die Entstehung der Zahneingriffsfrequenz zu erkennen, b) Zahneingriffsfrequenz mit Seitenbändern des Sensors ACCYC4, c) Zahneingriffsfrequenz mit Seitenbändern des Sensors ACCYC1.

Um abzusichern, dass es sich bei den detektierten Frequenzen um die Zahneingriffsfrequenz handelt, kann die Lage der Peaks in Abhängigkeit der Drehzahl

untersucht werden. Abbildung 4-73 a) zeigt dazu die Position der Zahneingriffsfrequenz für sechs unterschiedliche Prozessvariationen, die sich unter anderem in der Drehzahl unterscheiden. Die Berechnung der Soll-Zahneingriffsfrequenz erfolgt hierbei durch die im Versuchsprotokoll notierten Drehzahlen des Werkzeugs sowie der Anzahl der Zähne des Werkzeugs (60) nach Formel (4.1) aus Abschnitt 4.4.1. Die systematische Abweichung der berechneten und gemessenen Zahneingriffsfrequenzen wird in der Regelung der Anlage vermutet, konnte aber nicht vollständig aufgeklärt werden.

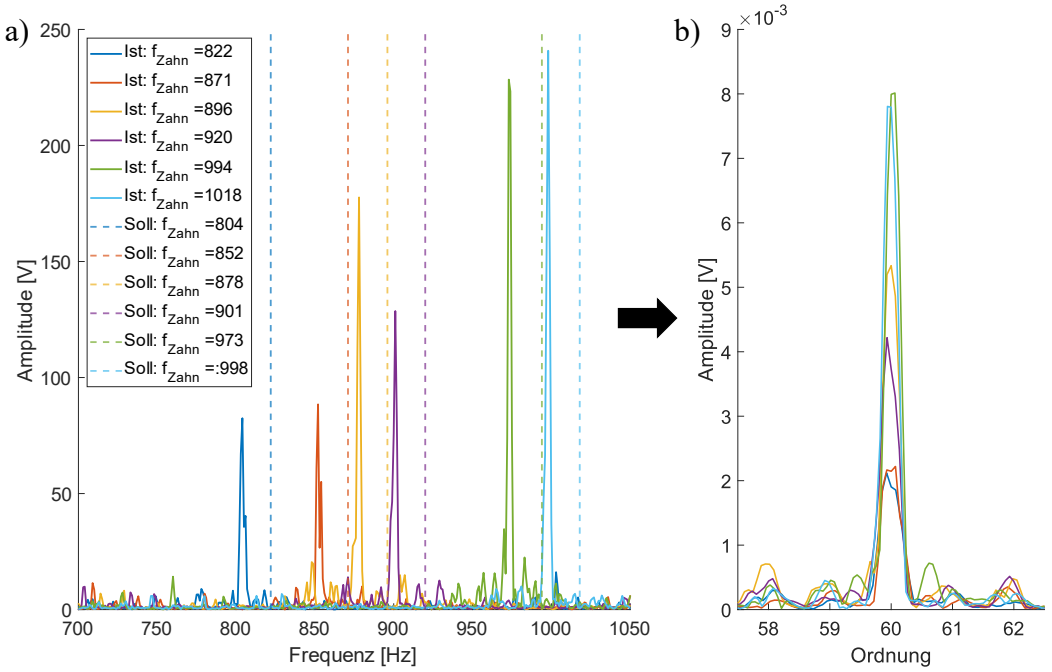


Abbildung 4-73: Von der Werkzeugdrehzahl abhängige Zahneingriffsfrequenz. Die Abweichung zwischen berechneter (Soll) und gemessener (Ist) Frequenz wird in der Regelung vermutet, die abhängig von den Regelparametern die Schnittgeschwindigkeit anpasst.

Aus Abbildung 4-73 lassen sich zwei wichtige Erkenntnisse hinsichtlich eines Prozessvariation-unabhängigen Werkzeugverschleißmodells ableiten. Durch die veränderliche Schnittgeschwindigkeit muss das Modell in der Lage sein, die drehsynchronen, physikalisch motivierten Frequenzen zu verfolgen. Hierzu kann, wie in Abbildung 4-73 b) dargestellt, die Umwandlung der Frequenzspektren in Ordnungsspektren (siehe Abschnitt 2.3.2.2) erfolgen. Ebenfalls ist eine starke Abhängigkeit der aufgezeichneten Amplituden der einzelnen Frequenzen festzustellen, weshalb eine einfache Amplitudenbetrachtung für eine Verschleißdetektion als nicht ausreichend erachtet wird.

In Abbildung 4-74 sind die Ordnungsspektren über den gesamten Ordnungsbereich dargestellt. Im unteren Bereich sind der Zahneingriff (60. Ordnung) sowie deren erste (120. Ordnung) und zweite (180. Ordnung) Harmonische deutlich anhand scharf

definierter Peaks erkennbar. Beginnend von der 200. Ordnung bis zur maximalen Ordnung beginnen die Peaks für die verschiedenen Drehzahlen auseinanderzulaufen, was auf Strukturresonanzen hindeutet. Da auch die Spektren der restlichen Beschleunigungssignale in diesem Bereich massiv von den Amplituden der Strukturresonanzen dominiert werden, erscheint eine Auswertung der Ordnungsspektren nur im Bereich der niedrigeren Ordnungen bis 200 sinnvoll.

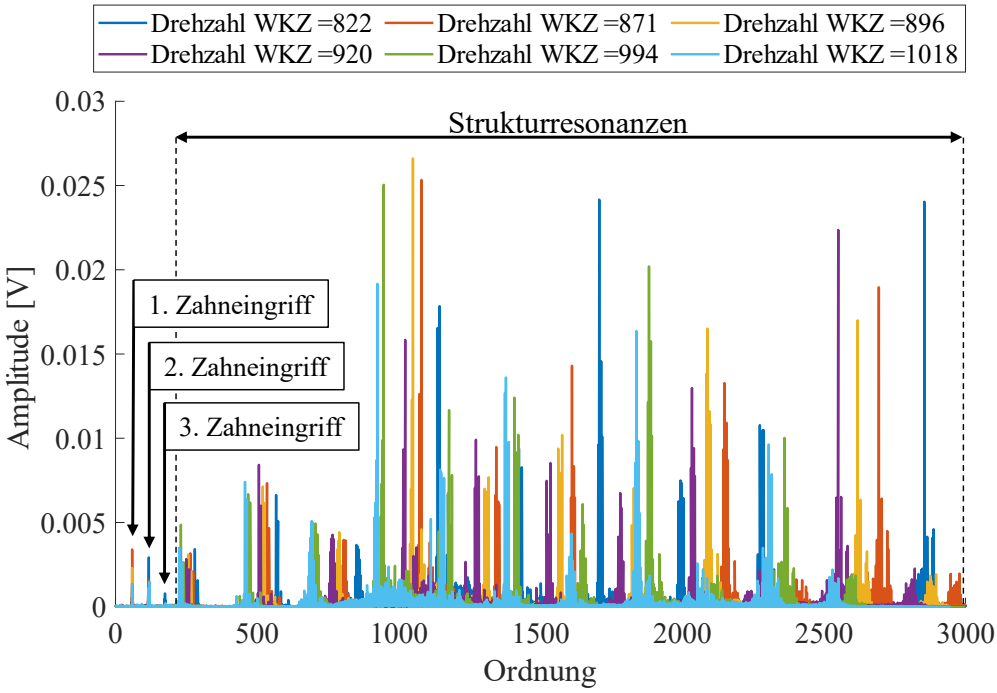


Abbildung 4-74: Ordnungsspektren des Sensors ACCYC1 über den gesamten Ordnungsbereich. Im unteren Ordnungsbereich zeigen sich deutlich die Zahneingriffsfrequenz sowie deren erste und zweite Harmonische. Im höheren Ordnungsbereich dominieren Strukturresonanzen das Spektrum.

Im Falle der Acoustic Emission Daten zeigt sich ein ähnliches Verhalten, wie anhand der Beschleunigungssensoren gezeigt, wobei die Ausprägung charakteristischer Frequenzen während des Werkzeugeingriffs deutlich geringer ausfällt. In Abbildung 4-71 sind exemplarisch das Signal des Acoustic Emission Sensors an der Werkzeugspindel sowie das zugehörige Spektrogramm im Bereich von 100 bis 250 kHz für einen gesamten Bearbeitungsprozess dargestellt.

Insgesamt erscheint eine Einordnung des Potentials der Acoustic Emission Sensoren deutlich schwerer, da aufgrund der hohen unteren Grenze der Bandbreite der Sensoren (bei ca. 50 kHz) kein direkter Bezug zu drehsynchronen Frequenzen hergestellt werden kann.

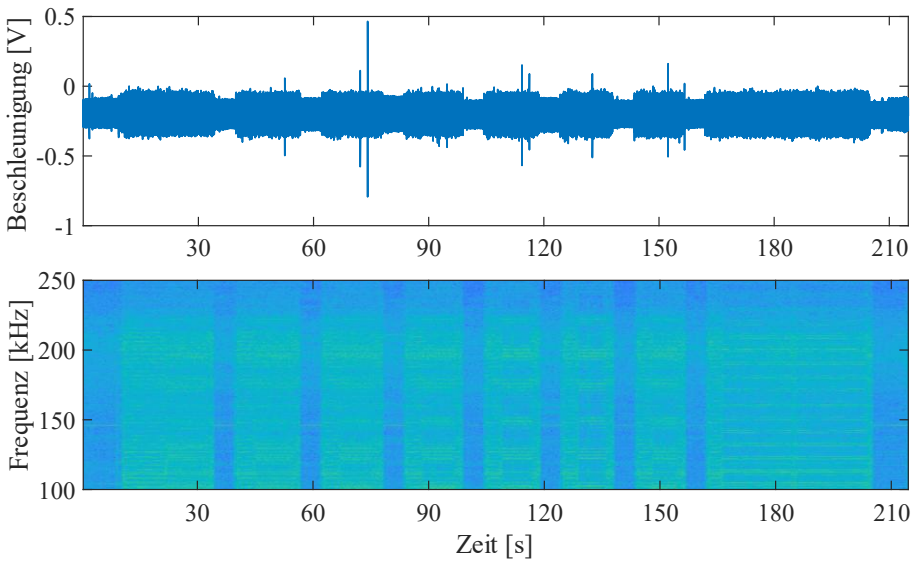


Abbildung 4-75: Zeitsignal des Acoustic Emission Sensors an der Werkzeugspindel C1 und das zugehörige Spektrogramm im Bereich von 100 bis 250 kHz.

In Abbildung 4-76 sind die verfügbaren Prozessdaten der Anlage dargestellt. Neben dem Positionssignal, welches hauptsächlich für die Segmentierung der Prozessdaten verwendet wird, stehen die momentbildenden Ströme der Spindeln von Werkstück und Werkzeug zur Verfügung. Vergleichbar zu den in Abbildung 4-72 (Mitte) dargestellten Beschleunigungssignalen steigt der momentbildende Strom-Istwert der Werkzeugspindel über die Schruppschnitte an. Dies ist dadurch zu erklären, dass das Werkzeug mit jedem Schruppschnitt tiefer in das Werkstück eintaucht.

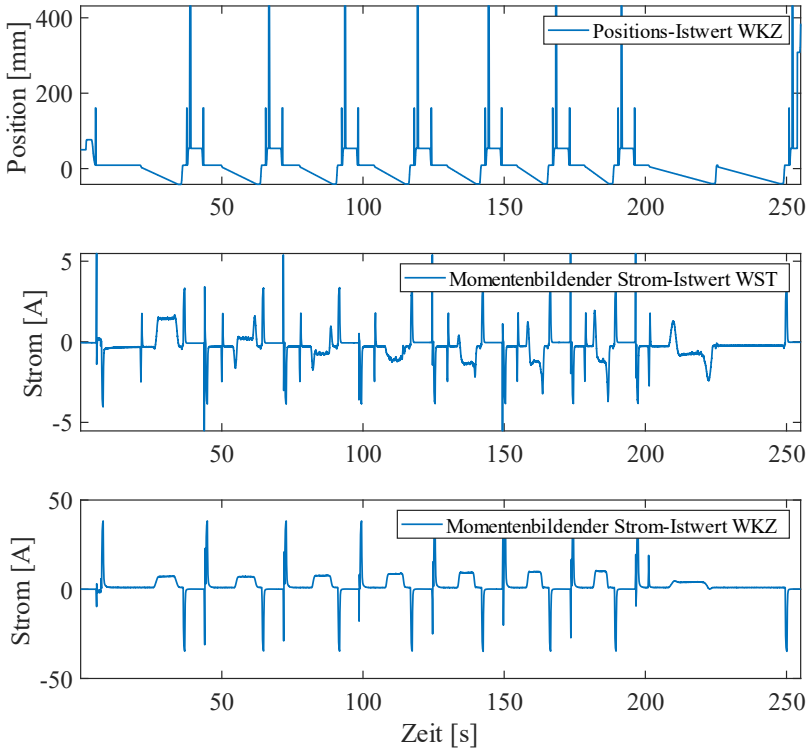


Abbildung 4-76: Positions- und Stromsignale für einen gesamten Bearbeitungsprozess.

Durch die plausiblen Verläufe der Prozessdaten sowie das Auftreten physikalisch erklärbarer Frequenzen in den Daten der Vibrationssensoren wird die Datenqualität als ausreichend angenommen.

4.4.3.4 Design of Experiment

Die Generierung der Daten erfolgt, wie in 4.4.2 beschrieben, anhand einer Dreh-Fräs-Maschine für die Komplettbearbeitung des Typs CTX Beta 1250 TC von DMG MORI, welche bei dem Projektpartner Schaeffler im Versuch sowie in der Serienfertigung genutzt wird. Das Ziel der Datengenerierung ist einerseits eine physikalisch motivierte Analyse der Daten. Dadurch soll Prozesswissen aufgebaut und Werkzeugverschleiß beschreibende Symptome identifiziert werden. Andererseits soll versucht werden, eine Datenbasis für ein Prozessvariation-unabhängiges Verschleißmodell aufzubauen. Unter Prozessvariation ist eine Kombination der für die Bearbeitungsgeschwindigkeit, Produktqualität sowie Verschleißentwicklung relevanten Prozessparameter zu verstehen. Eine Übersicht über die Prozessparameter sowie die mit den Prozessexperten von Schaeffler festgelegten Prozessfenster sind in Tabelle 4-11 zusammengefasst.

Tabelle 4-11: Für die Bearbeitungsgeschwindigkeit, die Produktqualität, die Verschleißentwicklung relevanten Prozessparameter sowie die mit den Prozessexperten von Schaeffler festgelegten Prozessfenster.

Prozessparameter	Spalte im DoE	Eingriffsgrenzen
Kopfspandicke	A	0,2 – 0,3 mm
Anzahl Schruppschnitte	B	7 oder 8
Schnittgeschwindigkeit Schruppen	C	65 – 85 m/min
Vorschubrate Schlichten	D	0,2 – 0,3 mm/U
Schnittgeschwindigkeit Schlichten	E	75 – 95 m/min

Während der Versuche werden Werkzeuge im Neuzustand in die Anlage eingesetzt und Hohlrad-Innenverzahnungen gefertigt, bis das Werkzeug verschlissen ist. Für die Versuche stehen zwei Werkzeuge zur Verfügung, wobei jedes der Werkzeuge zweimal nachgeschliffen werden kann und damit wieder als neu gilt. In den Versuchen 1, 3 und 5 wurde das Werkzeug 1 und in den Versuchen 2, 4 und 6 das Werkzeug 2 verwendet. Im Folgenden werden die Werkzeuge angelehnt an die Nummerierung der Versuche als Werkzeug 1 – 6 bezeichnet.

Für die Versuche wurde mittels LHC (siehe Abschnitt 2.2.1) ein DoE generiert, wobei als Optimierungsziel eine möglichst geringe Korrelation zwischen den Parametern verwendet wurde. Die Parameter von Schrupp- (A, B und C) und Schlichtzyklus (D und E) wurden getrennt betrachtet, da sich die Konfiguration des Schruppzyklus maßgeblich auf die Entwicklung des Werkzeugverschleiß auswirkt, wohingegen die Konfiguration des Schlichtens die Oberflächenqualität der gefertigten Innenverzahnungen beeinflusst. Die resultierenden DoEs für die Versuche 1 – 6 sind in Tabelle 4-12 dargestellt. Da die Parametrierung der gefertigten Variante nur manuell an der Steuerung der Anlage geändert werden kann, erfolgt der Wechsel der Variante jeweils nach 10 gefertigten Teilen. Die gelb markierte Variante befindet sich mittig im Parameterraum und wurde in jedem Versuch gefertigt. Mit Werkzeug 6 wurde ausschließlich diese Prozessvariante gefertigt.

Tabelle 4-12: DoE für die Datengenerierung von Versuch 1 - 6. Die gelb markierte Variante befindet sich mittig im Parameterraum und ist bei allen Versuchen vorhanden.

Versuchsmatrix **WKZ 1**					
Zyklus	Schruppzyklus			Schlichtzyklus	
Variante	Werkzeugverschleiß			Qualität	
Nr.	A	B	C	D	E
1	0,26	8	67	0,26	77
2	0,25	7	75	0,25	85
3	0,23	7	83	0,23	93
4	0,27	7	81	0,27	91
5	0,22	7	71	0,22	81
6	0,30	8	79	0,30	89
7	0,26	7	67	0,26	77
8	0,28	7	73	0,28	83
9	0,20	8	76	0,20	86

Pearson-Korrelations-koeffizient	
A-B	0,03
A-C	0,01
B-C	-0,09
D-E	0,01

Versuchsmatrix **WKZ 2**					
Zyklus	Schruppzyklus			Schlichtzyklus	
Variante	Werkzeugverschleiß			Qualität	
Nr.	A	B	C	D	E
1	0,25	7	82	0,25	92
2	0,29	8	66	0,29	76
3	0,24	8	80	0,24	90
4	0,25	7	75	0,25	85
5	0,27	8	77	0,27	87
6	0,23	7	73	0,23	83
7	0,21	8	69	0,21	79
8	0,26	8	84	0,26	94
9	0,29	7	74	0,29	84

Pearson-Korrelations-koeffizient	
A-B	-0,02
A-C	-0,05
B-C	-0,07
D-E	-0,05

Versuchsmatrix **WKZ 3**					
Zyklus	Schruppzyklus			Schlichtzyklus	
Nr.	A	B	C	D	E
1	0,30	8	80	0,30	90
2	0,26	7	76	0,26	86
3	0,20	8	69	0,20	79
4	0,22	7	78	0,22	88
5	0,25	7	70	0,25	80
6	0,25	7	75	0,25	85
7	0,29	8	67	0,29	77
8	0,28	7	72	0,28	82
9	0,24	8	82	0,24	92

Pearson-Korrelations-koeffizient	
A-B	0,09
A-C	0,01
B-C	0,03
D-E	0,01

Versuchsmatrix **WKZ 4**					
Zyklus	Schruppzyklus			Schlichtzyklus	
Nr.	A	B	C	D	E
1	0,25	7	74	0,25	84
2	0,22	8	77	0,22	87
3	0,27	8	66	0,27	76
4	0,23	7	71	0,23	81
5	0,29	8	85	0,29	95
6	0,20	8	82	0,20	92
7	0,26	7	80	0,26	90
8	0,25	7	75	0,25	85
9	0,24	8	68	0,24	78

Pearson-Korrelations-koeffizient	
A-B	-0,07
A-C	0,02
B-C	0,05
D-E	0,02

Versuchsmatrix **WKZ 5**					
Variante	Schruppzyklus			Schlichtzyklus	
Nr.	A	B	C	D	E
1	0,27	8	80	0,27	90
2	0,21	7	68	0,21	78
3	0,26	8	84	0,26	94
4	0,29	7	70	0,29	80
5	0,24	7	77	0,24	87
6	0,25	7	75	0,25	85
7	0,23	8	82	0,23	92
8	0,28	8	66	0,28	76
9	0,22	8	73	0,22	83

Pearson-Korrelations-koeffizient	
A-B	-0,10
A-C	0,09
B-C	0,06
D-E	0,19

Versuchsmatrix **WKZ 6**					
Zyklus	Schruppzyklus			Schlichtzyklus	
Nr.	A	B	C	D	E
1	0,25	7	75	0,25	85

Pearson-Korrelations-koeffizient

4.4.3.5 Bestimmung des Werkzeugverschleiß

Die Beurteilung des Werkzeugzustands in der Serienfertigung basiert häufig auf der subjektiven Wahrnehmung des Werkers, der das Werkzeug beispielsweise auf Basis einer steigenden Geräuschentwicklung wechselt. Andere Maschinenführer wechseln das Werkzeug nach einer festen Anzahl an Teilen, im Fall der Hohlradinnenverzahnungen beispielsweise nach maximal 90 gefertigten Teilen. Um eine objektive Erfassung des Werkzeugverschleißes zu ermöglichen, werden während der Versuche Messungen mittels des in Abbildung 4-77 dargestellten Mikroskops durchgeführt. Hierzu wird das Werkzeug mit Hilfe eines Anschlags unter dem Mikroskop positioniert. Dadurch ist es möglich, reproduzierbare Bilder von sechs definierten Messpunkten zu machen. Bei den Messpunkten handelt es sich um jeweils drei um 180° versetzte Zähne, bei denen jeweils vom linken Zahn die rechte Flanke, vom mittleren Zahn die Spanfläche und vom rechten Zahn die linke Flanke fotografiert werden. Neben der Möglichkeit, hoch aufgelöste Bilder der Zähne aufzuzeichnen, erlaubt die Software des Mikroskops die Modellierung der Zähne auf Basis der Aufnahmen. Hierzu werden manuell charakteristische Kanten der Zähne markiert und anschließend in ein 3d Modell überführt. Durch eine Differenzbildung des modellierten Zahnvolumens nach jeder Variante zu dem Zahnvolumen des Werkzeugs im Anfangszustand soll der Volumenverlust und somit der Verschleiß des Werkzeugs gemessen und maschinenlesbar gemacht werden.

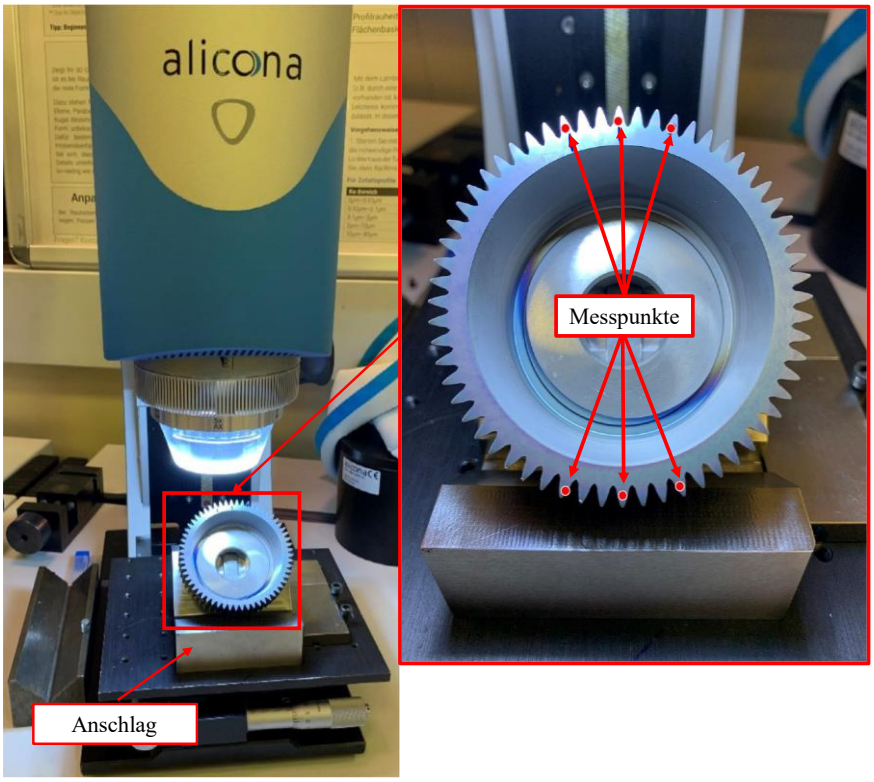


Abbildung 4-77: Messaufbau zum Bestimmen des Werkzeugverschleiß nach einer Variante, also 10 gefertigten Innenverzahnungen [203].

4.4.4 Auswertung und Modellbildung

4.4.4.1 Übersicht über die Versuchsreihen

Während der Versuchsreihen traten technische Probleme sowohl mit dem Messsystem als auch mit der Anlage auf, weshalb die Daten einiger Teile nicht für eine Modellbildung zur Verfügung stehen. Aufgrund eines Hardwarefehlers des USB-Controllers des für die Aufzeichnung der Acoustic Emission Daten verwendeten cRIO, mussten die Daten im internen Speicher des cRIO gebuffert und über eine TCP/IP Verbindung zum Host PC übertragen werden. Durch die hohen Datenraten in Kombination mit einer nicht ausreichenden Kommunikationsgeschwindigkeit führte dies in einzelnen Fällen zum Überlaufen des internen Speichers des cRIO, wodurch die Messung während des Prozesses abgebrochen ist. Dieses Problem trat vermehrt bei der Aufzeichnung von Varianten mit erhöhter Prozessdauer auf.

Auch seitens der Anlage traten während der Versuche Probleme auf, die zum Verlust von Daten führten. Die Ursache für diese Abbrüche konnte nicht abschließend identifiziert werden. Als mögliche Ursachen wurden verkantete Späne im Werkzeug oder Werkstück genannt, die zu einem Über- oder Unterschreiten der in der Anlage berechneten Grenzen der Stellgrößen geführt haben. Ebenfalls gab es teilweise zu spät einsetzende Trigger der einzelnen Schnitte, weshalb die Daten einzelner Schnitte nur teilweise vorhandenen waren.

Um eine einheitliche Modellbildung zu ermöglichen, werden zur Generierung der Szenarien nur die Teile verwendet, für die vollständige Daten, also die Signale aller Sensoren über den gesamten Prozess, verfügbar sind. Tabelle 4-13 gibt eine Übersicht über die für die Auswertungen zur Verfügung stehenden Daten. Dargestellt sind hierbei die verfügbaren Teile je Variante, die gesamten Teile für die einzelnen Werkzeuge sowie das Volumen der Rohdaten. Insgesamt stehen dadurch die Daten von 524 gefertigten Teilen zur Verfügung. Ausgenommen für die Variante 9, sollten für alle Varianten bei allen Werkzeugen 10 Teile vorhanden sein. Weicht die Anzahl der Teile aus der Tabelle ab, ist dies auf Probleme bei der Datenerfassung oder der Anlage zurückzuführen.

Der Abbruch der Versuche erfolgte in allen Fällen aufgrund einer vom Werker festgestellten steigenden Geräuschentwicklung. Im Falle der Werkzeuge 1, 2 und 4 wurden aus diesem Grund in der letzten Variante nicht mehr alle 10 Teile gefertigt. Die Versuchsreihe von Werkzeug 3 musste sogar schon nach Variante 8 beendet werden. Nur im Fall von Werkzeug 5 erreichte das Werkzeug die maximale Versuchsdauer von 90 Teilen. Bei Werkzeug 6 konnte der Versuch deutlich über die übliche Standzeit des Werkzeugs von 90 Teilen hinaus durchgeführt werden.

Aufgrund der starken Auslastung der Anlage in der Serienfertigung sowie eines Ausfalls der Anlage aufgrund eines Defekts am Werkzeugwechsler wurden die Versuche über einen längeren Zeitraum in insgesamt vier Versuchsreihen durchgeführt. In Vorversuchen (Versuch 0) wurden dabei nur die Testdaten vereinzelter Teile aufgezeichnet, um eine Grundlage für die Mess- und Datenplanung zu generieren. In den Versuchen 1, 2 und 3 wurden dann jeweils die Daten von zwei Werkzeugen aufgezeichnet.

Tabelle 4-13: Ergebnis der Messkampagnen in Form der verfügbaren Daten für jede Variante sowie die dabei entstandenen Datenmengen.

Mess-kampagne	**Werk-zeuge**	**Anzahl Teile pro Variante**									**Teile gesamt**	**Vol. Rohdaten**
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		
1	1	10	9	9	10	10	10	10	9	7	84	545 GB
	2	8	10	9	10	10	10	8	8	6	79	530 GB
2	3	9	10	9	9	10	10	9	10	-	76	480 GB
	4	8	10	9	10	10	10	10	9	9	85	537 GB
3	5	8	10	9	10	9	10	10	10	10	86	526 GB
	6	114	-	-	-	-	-	-	-	-	114	619 GB

4.4.4.2 Auswertung der optischen Verschleißmessung

In Abbildung 4-78 sind exemplarisch die mittels des in Abschnitt 4.4.3.5 vorgestellten Mikroskops gemachten Aufnahmen der oberen rechten und linken Zahnflanken von Werkzeug 1 über die gesamte Messkampagne dargestellt. Bild 0 beschreibt hierbei den Werkzeugzustand vor den Versuchen, wohingegen die restlichen Bilder (1 – 9) den Zustand nach Fertigung der jeweiligen Variante zeigen. Auf eine Darstellung des mittleren Zahns wird aufgrund der Ähnlichkeit zu den bereits gezeigten Zähnen an dieser Stelle verzichtet. Beide Zähne weisen eine langsame Entwicklung der Verschleißmarkenbreite an der Spanfläche sowie an der Schneide auf. Insgesamt ist allerdings nur eine sehr geringe verschleißbedingte Veränderung des Werkzeugzustands festzustellen. Eine plötzliche Veränderung des Werkzeugzustandes, die einen Ausfall des Werkzeugs oder einen Anstieg der Bearbeitungsgeräusche erklären würde, konnte nicht gefunden werden.

In Abbildung 4-79 sind die Aufnahmen des oberen linken und rechten Zahns für Werkzeug 2 dargestellt. Der linke Zahn zeigt dabei ein ähnliches Bild wie bereits bei Werkzeug 1 festgestellt, mit einem sich langsam entwickelnden Freiflächenverschleiß an der Spanfläche sowie der Schneide. Im Gegensatz zum ersten Werkzeug ist allerdings am rechten Zahn ein deutlicher Ausbruch nach der Fertigung der dritten Variante festzustellen. Eine genauere Untersuchung des Werkzeugs zeigt, dass der

Ausbruch nicht nur an diesem Zahn, sondern mit einer Periodizität von 12 auftritt. Dies deutet auf einen Materialeinschluss im Rohling des Hohlrads hin, welcher von jedem 12. Zahn überrollt wurde. Wie anhand von Tabelle 4-13 erkennbar, konnte trotz des Ausbruchs weiter gefertigt werden. Auch eine akute Veränderung des Fertigungsgeräusches aufgrund des Ausbruchs konnte während der Versuche nicht festgestellt werden.

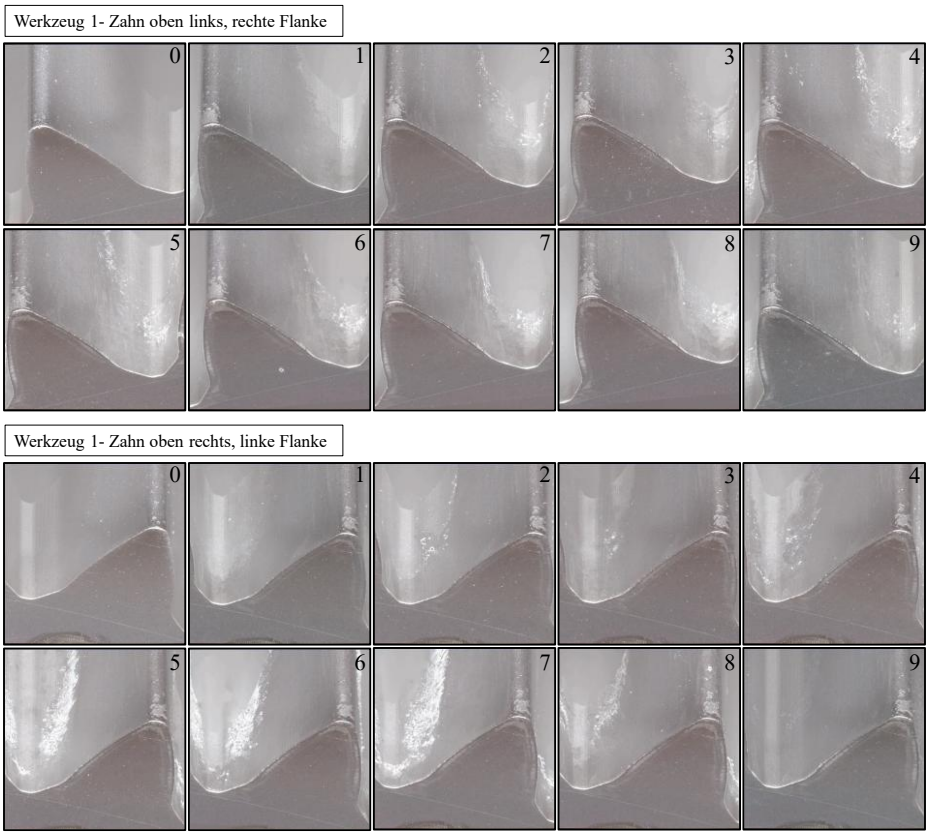


Abbildung 4-78: Mikroskop Aufnahmen Zahn oben rechts und Zahn oben links des ersten Werkzeugs, jeweils zu Beginn der Messkampagne und jeweils nach 10 gefertigten Teilen [203].

Die dreidimensionale Modellierung der Zähne auf Basis der mikroskopischen Aufnahmen und die dadurch geplante Verschleißbestimmung anhand der Bestimmung des am Werkzeug verlorenen Materials zeigte aufgrund der hohen Ungenauigkeit der manuellen Modellierung eine zu hohe Streuung, wodurch für die ersten beiden Werkzeuge trotz des Ausbruchs an Werkzeug 2 keine signifikante Änderung des Zahnvolumens gemessen werden konnte. Des Weiteren erfolgt die Verschleißbestimmung der Werkzeuge, nach jeder gefertigten Variante, weshalb das Werkzeug nach jeder Variante aus der Maschine ausgebaut, im Labor vermessen und danach wieder in die Maschine eingesetzt und eingerichtet werden muss. Aufgrund des hohen Zeitdrucks der in der Fertigung eingesetzten Anlage sowie der ungenügenden Ergebnisse der Verschleißbestimmung für Werkzeug 1 und 2 wurde die optische Vermessung ab Werkzeug 3 eingestellt. Insgesamt stehen für die überwachte

Modellbildung keine gemessenen Zielgrößen des Verschleißfortschritts zur Verfügung.

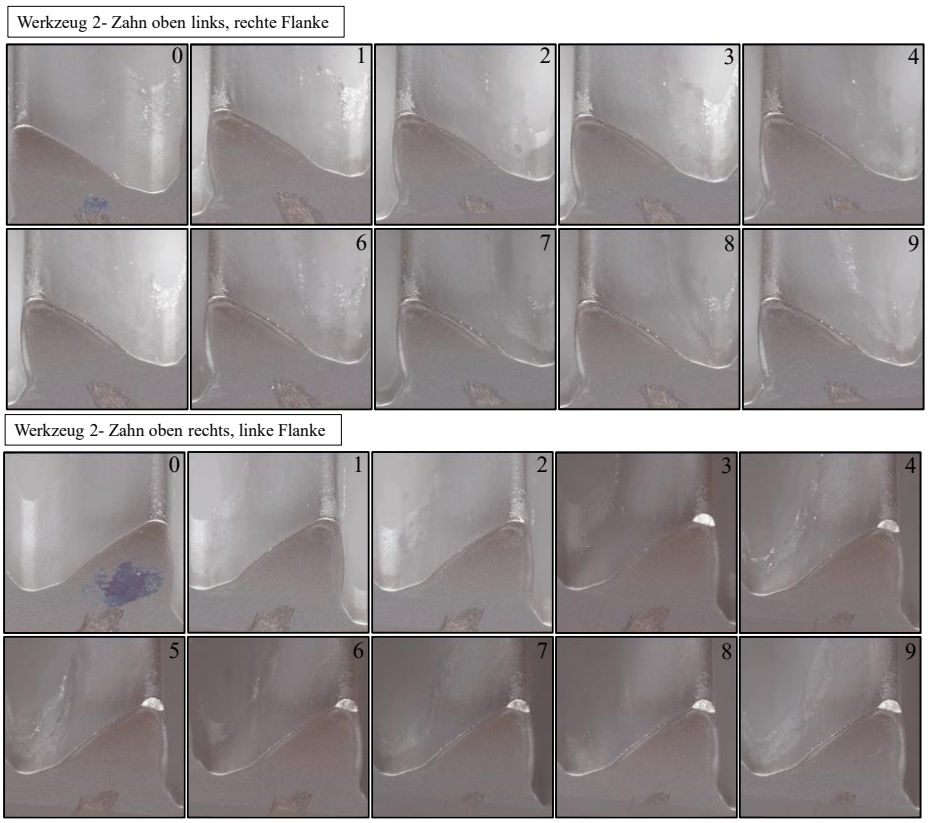


Abbildung 4-79: Mikroskop Aufnahmen Zahn oben rechts und Zahn oben links des zweiten Werkzeugs, jeweils zu Beginn der Messkampagne und jeweils nach 10 gefertigten Teilen. Am Zahn oben rechts ist nach Variante 3 ein deutlicher Ausbruch am Werkzeug erkennbar [203].

4.4.4.3 Modell Werkzeug 6

Im Gegensatz zu den Werkzeugen 1 – 5 erfolgte die Versuchsdurchführung von Werkzeug 6 mit einer konstanten Prozessvariante. Dadurch soll eine physikalische Betrachtung der Daten ohne die Einflüsse der Prozessparameter ermöglicht werden, weshalb die Ergebnisse von Werkzeug 6 an dieser Stelle noch einmal separat präsentiert werden. Zudem soll untersucht werden, ob sich bestimmte Schnitte oder Sensoren in besonderem Maße für eine Verschleißdetektion eignen. Zu diesem Zweck wurde eine erste überwachte Auswertung mit dem in Abbildung 4-80 abgebildeten Stack auf Basis der Acoustic Emission, Beschleunigungs- und Prozessdaten des Werkzeugs 6 durchgeführt. Als Zielgröße wurde ein linearer Verschleißfortschritt über die Anzahl der gefertigten Teile angenommen. Aufgrund der großen Datenmengen erfolgt die Auswertung für jeden Schnitt separat. Die Beschränkung des Suchraums für die Bayes Optimierung wurde hier auf 12 Intervalle im Zeitbereich, 500 Intervalle im Frequenzbereich, 300 Merkmale und 150 PLSR-Komponenten festgelegt. Durch die hohe Anzahl an Segmenten ergeben sich nach Formel (3.1) maximal ca. 600000

Merkmale, wodurch die Berechnungszeit einer multivariaten Merkmalsselektion nicht mehr zielführend wäre, weshalb in diesem Fall eine Begrenzung des Merkmalspools mittels Pearson Selektion auf 5000 Merkmale durchgeführt wird.

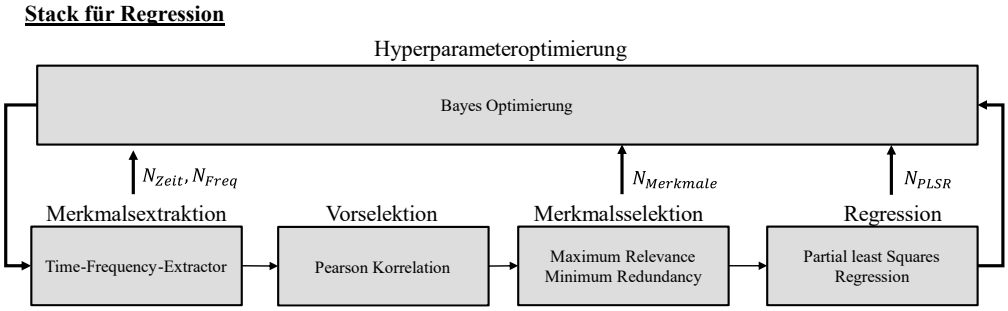


Abbildung 4-80: Stack für die überwachte Modellbildung für Werkzeug 1.

Das Ergebnis einer 10-fach randomisierten Kreuzvalidierung für den zweiten Schnitt (bestes Ergebnis) und den siebten Schnitt (schlechtestes Ergebnis) sowie der Ursprung der Merkmale bezüglich der Sensoren für Schnitt 2 sind in Abbildung 4-81 dargestellt. Insgesamt wurden 72 Merkmale selektiert und 43 PLSR-Komponenten verwendet. Der RMSE beträgt dabei 0,023, also 2,3% der Lebensdauer bzw. 2,7 gefertigte Teile. Die gesammelten Ergebnisse aller Schnitte einschließlich der Hyperparameter der optimierten Stacks sind in Tabelle 4-14 zusammengefasst.

Obwohl viele Merkmale aus dem Signal der Acoustic Emission Daten selektiert werden, stammen die höchstkorrelierten Merkmale alle aus den Signalen der Beschleunigungssensoren, weshalb diese im nächsten Schritt noch einmal genauer betrachtet werden.

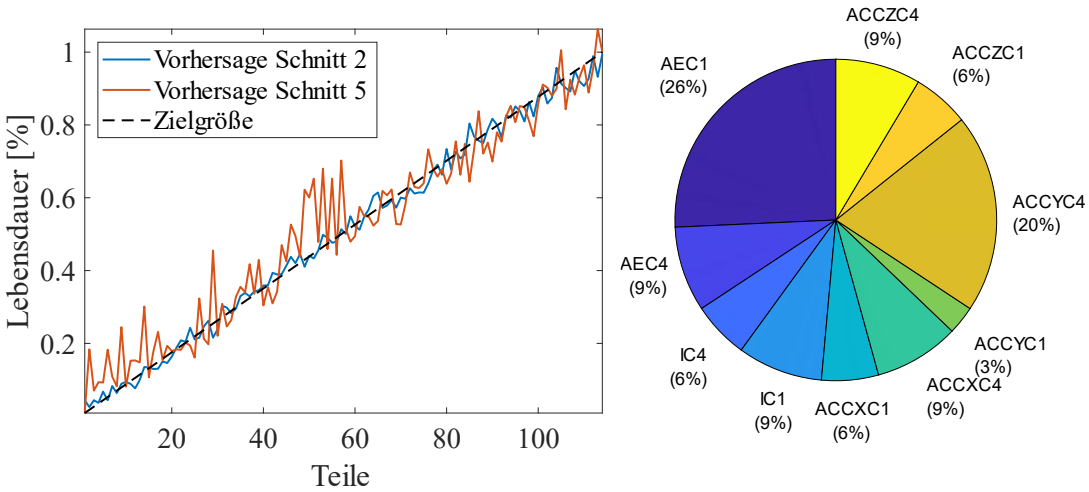


Abbildung 4-81: Das Ergebnis einer 10-fachen randomisierten Kreuzvalidierung für Schnitt 2 und Schnitt 5 sowie die Herkunft der für die Modellbildung verwendeten Merkmale für Schnitt 2.

Tabelle 4-14: Parameter der optimierten Stacks sowie die Ergebnisse der 10-fachen randomisierten Kreuzvalidierung für die einzelnen Schnitte von Werkzeug 6 bei Vorgabe eines linearen Verschleißfortschritts.

Schnitt	Stack Parameter				RMSE
	N_{Zeit}	$N_{Frequenz}$	$N_{Merkmale}$	N_{PLSR}	
Schruppen 1	9	437	35	72	0,038
Schruppen 2	8	486	25	74	0,023
Schruppen 3	10	147	23	25	0, 025
Schruppen 4	7	448	117	45	0, 023
Schruppen 5	10	271	124	27	0,072
Schruppen 6	10	421	123	18	0,027
Schruppen 7	6	497	59	27	0,042
Schlichten	4	267	38	52	0,032

Durch die Analyse der selektierten Merkmale wird abgesichert, dass die Merkmale aus plausiblen Zeit- und Frequenzbereichen der Daten extrahiert werden. Dies ist insbesondere relevant, da keine objektive verschleißbeschreibende Zielgröße vorhanden ist, sondern ein einfacher linear fortschreitender Verschleißfortschritt für die Modellbildung angenommen wurde.

In Abbildung 4-82 sind die Zeitbereiche dargestellt, aus denen die selektierten Merkmale der Beschleunigungssensoren stammen. Obwohl 7 Merkmale aus dem gesamten Zeitbereich extrahiert werden, stammt ein Großteil der Merkmale aus dem Teil des Schnitts, in dem sich das Werkzeug im Eingriff befindet. Besonders hervorgehoben ist hierbei der Bereich, in dem das Werkzeug in Kontakt mit dem Werkstück tritt.

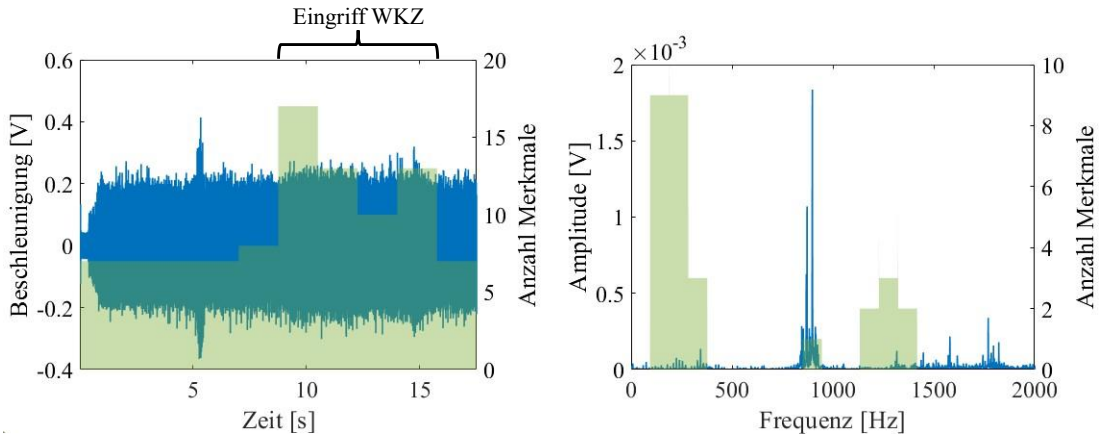


Abbildung 4-82: Zeit- und Frequenzbereiche, aus denen die selektierten Merkmale der Beschleunigungssensoren stammen. Die Plots sind exemplarisch anhand des Zeitsignals und Spektrums des Sensors ACCYC1 erstellt.

Das Frequenzintervall, aus dem das höchstkorrelierte Merkmal mittels RMS aus dem Signal des Sensors ACCY1 extrahiert wird, ist in Abbildung 4-83 dargestellt. In dem Intervall befinden sich mehrere scharf definierte Frequenzen, die einen Anstieg der Amplitude über die Anzahl der gefertigten Teile aufweisen. Besonders hervorzuheben ist hierbei die Frequenz von 360 Hz, welche gerade gegen Ende der Messreihe einen starken Amplitudenanstieg verzeichnet.

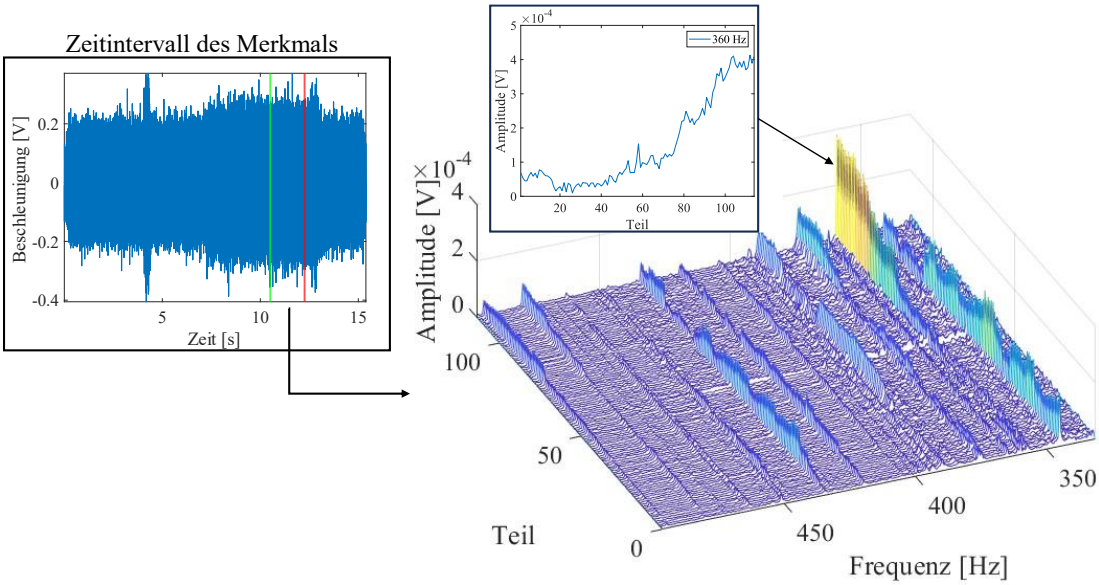


Abbildung 4-83: Zeit- und Frequenzbereich, aus dem das höchstkorrelierte Merkmal mittels RMS extrahiert wird. Es zeigt sich deutlich eine Frequenz bei 360 Hz, deren Amplitude gerade gegen Ende der Messreihe einen deutlichen Anstieg verzeichnet.

4.4.4.4 Werkzeugübergreifende Auswertung

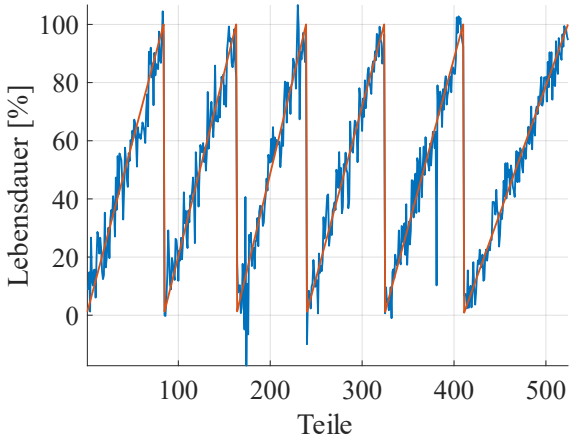
Die werkzeugübergreifende Auswertung wird anhand von zwei unterschiedlichen Validierungsszenarien durchgeführt. Im ersten Schritt wird anhand einer 6-fachen randomisierten Kreuzvalidierung (RKV6) versucht, ein für alle Werkzeuge gültiges Modell herzuleiten. Im Anschluss erfolgt die Darstellung der Ergebnisse einer realitätsnahen Leave-One-Tool-Out Kreuzvalidierung (LOTO). Bei beiden Validierungsszenarien wurde der Stack aus Abbildung 4-80 verwendet. Auch hier erfolgt die Auswertung aufgrund der großen Datenmengen für jeden Schnitt separat.

Die Ergebnisse der randomisierten Kreuzvalidierung sind in Tabelle 4-15 zusammengefasst. Abbildung 4-84 zeigt dazu exemplarisch das Ergebnis der Vorhersage basierend auf den Daten von Schruppschnitt 2. Im Vergleich zum Modell basierend auf den Daten von Werkzeug 6, verschlechtert sich der RMSE hierbei auf 0,068. Ebenso wird anhand der Anzahl der verwendeten Merkmale klar, dass das Modell zunehmend komplexer werden muss, um die Verschleißentwicklungen aller Werkzeuge abzubilden. Zusammenfassend kann jedoch anhand Tabelle 4-15

festgestellt werden, dass ein für alle Werkzeuge gültiges Modell hergeleitet werden kann.

Tabelle 4-15: Parameter der optimierten Stacks sowie die Ergebnisse der 6-fachen randomisierten Kreuzvalidierung für die einzelnen Schnitte aller bei Vorgabe eines linearen Verschleißfortschritts.

Schnitt	**Stack Parameter**				**RMSE / RKV6**
	N_{Zeit}	$N_{Frequenz}$	$N_{Merkmale}$	N_{PLSRK}	
Schruppen 1	9	153	247	88	0,097
Schruppen 2	8	497	238	94	0,068
Schruppen 3	10	188	240	100	0,074
Schruppen 4	10	494	242	76	0,088
Schruppen 5	9	306	250	87	0,091
Schruppen 6	7	354	250	98	0,092
Schruppen 7	9	398	100	68	0,10
Schlichten	6	356	239	72	0,11



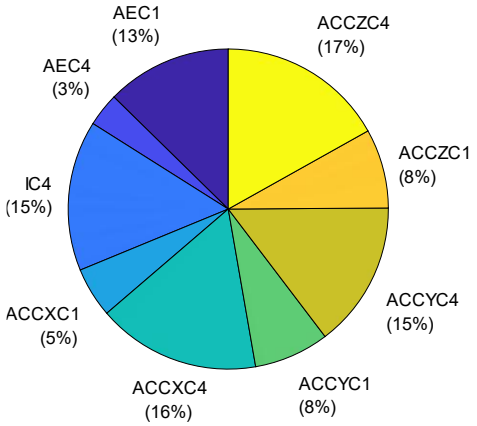


Abbildung 4-84: Ergebnis einer 6-fachen randomisierten Kreuzvalidierung auf Basis der Daten von Schruppschnitt 2 sowie die Herkunft der für die Modellbildung verwendeten Merkmale bezüglich der Sensoren.

Die Ergebnisse der deutlich realitätsnahen LOTO-Validierung sind in Abbildung 4-85 dargestellt. Bei der Vorhersage ist ein massiver Einfluss der gefertigten Variante festzustellen, welcher zu dominierenden Schwankungen in der Lebensdauervorhersage führt. Die Ergebnisse der Modellbildung für die einzelnen Schnitte sind in Tabelle 4-16 zusammengefasst. Anhand des RMSE wird deutlich, dass kein Modell für eine werkzeugübergreifende Vorhersage des Werkzeugverschleiß gefunden werden kann. Einerseits kann dies durch die in Abschnitt 4.4.3.3 vorgestellten Abhängigkeiten der Position und Amplitude drehsynchroner Frequenzen begründet werden. Andererseits deutet die Auswertung der in Abschnitt 4.4.4.2 vorgestellten optischen Verschleißbestimmung von Werkzeug 1 und 2 auf

unterschiedliche Verschleißmechanismen hin, welche dann zu spezifischen Symptomen in den Daten führen.

Tabelle 4-16: Parameter der optimierten Stacks sowie die Ergebnisse der Leave-One-Tool-Out Kreuzvalidierung für die einzelnen Schnitte aller bei Vorgabe eines linearen Verschleißfortschritts.

Schnitt	Stack Parameter				RMSE / LOTO
	N_{Zeit}	$N_{Frequenz}$	$N_{Merkmale}$	N_{PLSRK}	
Schruppen 1	7	373	214	53	0,36
Schruppen 2	8	470	257	71	0,29
Schruppen 3	6	92	165	86	0,38
Schruppen 4	5	261	197	67	0,42
Schruppen 5	9	213	220	75	0,39
Schruppen 6	10	314	223	93	0,31
Schruppen 7	9	289	190	74	0,30
Schlichten	7	315	165	69	0,38

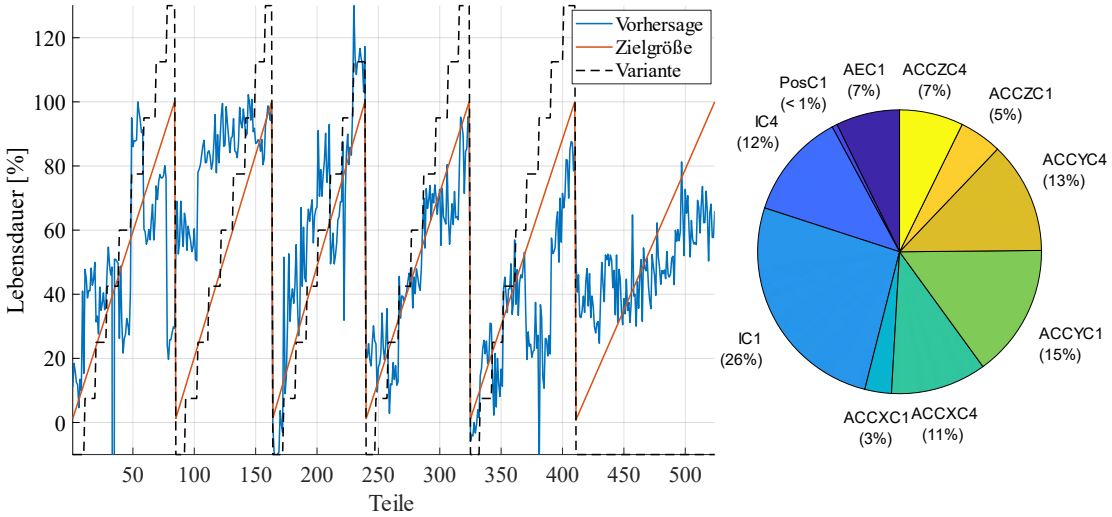


Abbildung 4-85: Ergebnis einer Leave-One-Tool-Out Kreuzvalidierung auf Basis der Daten von Schruppschnitt 2 sowie die Herkunft der für die Modellbildung verwendeten Merkmale bezüglich der Sensoren.

Da es sich um ein drehzahlvariables Verfahren handelt, bietet sich eine in den TFEx integrierte Normierung der Spektren auf die im DoE hinterlegten Drehfrequenzen von Werkzeug bzw. Werkstück an. Dadurch bleibt die Position der Ordnungen im Frequenzbereich konstant, wodurch der Einfluss der Prozessvariation verringert werden sollte. In Abbildung 4-86 ist exemplarisch das Ergebnis der LOTO-Kreuzvalidierung auf Basis der Daten von Schruppschnitt 2 dargestellt, wobei für die Merkmalsextraktion der Beschleunigungssensoren die Ordnungsspektren verwendet wurden. Die Normierung erfolgte dabei sensorspezifisch auf Basis der Spindeldrehfrequenz der Spindel, an der sich der Sensor befindet. Der verbleibende

RMSE beläuft sich auf 0,34 und zeigt dadurch sogar noch eine Verschlechterung im Vergleich zur Merkmalsextraktion basierend auf den herkömmlichen Spektren.

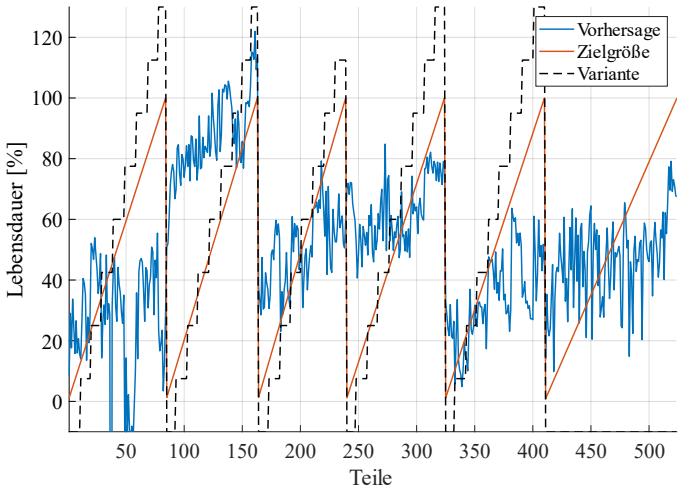


Abbildung 4-86: Ergebnis einer Leave-One-Tool-Out Kreuzvalidierung auf Basis der Daten von Schruppschnitt 2, wobei für die Merkmalsextraktion der Beschleunigungssensoren die Ordnungsspektren verwendet wurden. Für die restlichen Sensoren bleibt die Merkmalsextraktion unverändert.

4.4.5 Zusammenfassung und Diskussion der Anwendung

Im Rahmen der vorgestellten Untersuchungen zur Verschleißdetektion beim Wälzschälen wurde ein strukturierter Datensatz für das maschinelle Lernen geschaffen. Ausgehend von einer umfangreichen Mess- und Datenplanung wurde ein Konzept für eine Datenerfassung breitbandiger Sensoren entworfen und aufgebaut. Mittels Hardwaretriggerung wurde die benötigte Synchronität der externen Datenerfassungssysteme sichergestellt und deren Daten wiederum auf Basis eines Softwaretriggers mit internen Prozessdaten kombiniert. Im Anschluss an die Sicherstellung der strukturellen Datenqualität wurden erste Testdaten auf ihren Informationsgehalt untersucht, wobei speziell in den Spektren der Beschleunigungssensoren physikalisch und auf den Fertigungsprozess rückführbare Frequenzen identifiziert werden konnten. Im Anschluss an die Mess- und Datenplanung erfolgte eine umfassende Datengenerierung auf Basis eines komplexen Design of Experiments, welches gemeinsam mit den Prozessexperten von Schaeffler erarbeitet wurde. Um die nötige Varianz in den Trainingsdaten für das Training eines Prozessparameter-unabhängigen Verschleißmodells zu erzeugen, wurden die verschleißbestimmenden Parameter während der Versuche regelmäßig angepasst. Bei den insgesamt vier Wochen dauernden Versuchen wurden die Daten von 6 Werkzeugen aufgezeichnet, mit denen vom Neuzustand bis zum Ausfall gefertigt wurde. Durch die hohen Abtastraten der Beschleunigungs- und Acoustic Emission

Sensoren von bis zu 1MS/s fielen bei den Versuchen große Datenmengen von insgesamt mehr als 3 TB an. Um eine physikalische Untersuchung der Auswirkungen des Verschleißfortschritts zu ermöglichen, wurde Werkzeug 6 mit einer festen Prozesskonfiguration betrieben. Da die ursprünglich geplante Methode zur optischen Bestimmung des Werkzeugverschleißes keine genügende Auflösung gezeigt hat, stehen für eine Modellbildung nur geschätzte Verschleißverläufe auf Basis der Standzeit der Werkzeuge zur Verfügung.

Erste Analysen einzelner Werkzeuge auf Basis eines linearen Verschleißfortschritts zeigen eine gute Auflösung der Lebensdauer (Werkzeug 6, Schruppschnitt 2: 2,7 %). Auch mittels randomisierter Kreuzvalidierung gebildete werkzeugübergreifende Modelle zeigen eine moderate Auflösung (Schruppschnitt 2: 6,8 %). Bei der realitätsnahen Leave-One-Tool-Out Validierung zeigen sich jedoch die für industrielle Anwendungen typischen Probleme bei der Übertragbarkeit der Modelle. Ursachen hierfür können beispielsweise unterschiedliche Verschleißmechanismen oder wechselnde Umgebungsbedingungen darstellen. Die vermutete Drehzahlvariabilität kann auf Basis der Analyse mittels Ordnungsspektren ausgeschlossen werden. Hier werden neben den sich verschiebenden Frequenzen weitere, Prozessparameter bedingte Einflüsse vermutet.

Gerade im Hinblick auf die fehlende verschleißbeschreibende Zielgröße erscheint eine Auswertung mittels un- bzw. teilüberwachter Verfahren als zielführend. Durch die hohe Komplexität des Lernproblems und den dominierenden Einfluss der Prozesskonfiguration zeigte sich eine Analyse der Daten auf Basis un- bzw. teilüberwachter Verfahren jedoch als nicht geeignet, weshalb an dieser Stelle auf eine Darstellung der adaptiven Zustandsüberwachung verzichtet wurde. Für eine Erkennung spontaner Verschleißerscheinungen wie Materialausbrüche aus oder für das Erlernen bis dato unbekannter Verschleißmechanismen im Betrieb könnten aber beispielweise teilüberwachte Verfahren, basierend auf wenigen, physikalisch motivierten Merkmalen, untersucht werden. Hier könnte eine Modellbildung auf Basis einzelner Sensorgruppen und eine umfassende Analyse der selektierten Merkmale einen Ansatz zur Identifikation liefern.

5 Diskussion

Der zentrale Beitrag dieser Dissertation stellt die Entwicklung und Erprobung einer Methodik zur adaptiven Zustandsüberwachung industrieller Applikationen dar. Die auf Grundlage von Anwendung 1 entwickelte Methodik basiert auf einer Kombination von teilüberwachten und überwachten Verfahren des maschinellen Lernens, welche im Rahmen der durchgeführten Untersuchungen erfolgreich auf drei weitere industrielle Anwendungen übertragen werden konnte. Dadurch konnte die grundlegende, anwendungsübergreifende Funktionsweise der Methodik nachgewiesen werden. Für drei der vier untersuchten industriellen Applikationen konnte dabei eine Zustandsüberwachung ausgehend vom Gutzustand des Systems erzielt werden, in dem neue Zustände mittels Novelty Detection erkannt und sukzessive in das überwachte Modell übernommen wurden. Im Gegensatz zu einer statischen Zustandsüberwachung auf Basis überwachter Modelle entfällt dadurch die Notwendigkeit, im Vorfeld umfangreiche Trainingsdaten durch kostspielige Versuche zu erheben. Folglich rücken bis dato ökonomisch unattraktive Anwendungen in das potenzielle Spektrum einer wertschöpfungssteigernden Zustandsüberwachung.

Im Sinne einer generellen Anwendbarkeit der adaptiven Zustandsüberwachung äußern sich Einschränkungen hauptsächlich in den folgenden Aspekten. Die untersuchten Fehlerzustände zeigen sich teilweise in verschleierten Effekten in den Daten, wodurch sie nur schwer mittels teilüberwachter Verfahren zu detektieren sind. Als Beispiel kann hier die in Anwendung 2 auftretende erhöhte Lagerreibung genannt werden. Hier zeigen die verwendeten teilüberwachten Verfahren bestehend aus TFEx, Z-Score, PCA und kNN eine zu geringe Sensitivität, wodurch der Zustand nur sporadisch detektiert werden kann. Sind Fehlerzustände dieser Art bekannt, können dafür im Labor Trainingsdaten erzeugt und die adaptive Zustandsüberwachung aufbauend auf diesen gestartet werden. In Anwendung 4 zeigt sich dieser Sachverhalt in einer deutlich stärkeren Ausprägung. Aufgrund der hohen Komplexität und Varianz in der Prozessführung stellt die Konfiguration der Prozessparameter den dominanten Einfluss in den Daten dar. In diesem Fall ist sogar das überwachte Modell nicht in der Lage, diese den Verschleißsymptomen überlagerten Effekte zu unterdrücken und ein Prozessvariation-unabhängiges Verschleißmodell herzuleiten.

Bei der untersuchten Domain Shift Detection handelt es sich um ein effektives Instrument zur Absicherung der überwachten Modellbildung. In den vorliegenden Untersuchungen konnte die Beeinflussung der überwachten Modelle zuverlässig erkannt werden. Aufgrund des Eingriffspunkts nach der Merkmalsselektion und der dadurch bereits fortgeschrittenen Dimensionsreduktion kann der zusätzliche Berechnungsaufwand als gering eingestuft werden. Speziell im industriellen Umfeld, in dem aufgrund unvorhergesehener Ereignisse immer neue Systemzustände eintreten können, sind datengetriebene Modelle niemals als vollständig anzusehen, weshalb eine Domain Shift Detection in jedem Fall als Ergänzung einer auf überwachten Modellen basierenden Zustandsüberwachung empfohlen wird. Um eine Kompensation von beispielsweise sich ändernden Umwelteinflüssen zu erreichen, müsste jedoch zusätzlich eine Modellkalibrierung in den übergeordneten Algorithmus integriert werden, wie beispielsweise in [40] gezeigt. Dies erfordert jedoch an erster Stelle Mechanismen, welche sicherstellen, dass die Beeinflussung des überwachten Modells durch einen Umgebungseinfluss und nicht durch einen noch unbekannten Fehlerzustand entsteht.

Ausgenommen der Identifikation der Fehlerursache zeigt sich die entwickelte, adaptive Zustandsüberwachung als vollständig automatisierbar, wobei eine erhebliche Abhängigkeit von den für die Novelty Detection und überwachte Modellbildung verwendeten Methoden vorliegt. Im Fall des hier verwendeten Novelty Detection Stacks muss initial bei Inbetriebnahme die zu verwendende Segmentierung des TFEx sowie die Anzahl der nächsten Nachbarn (kNN) konfiguriert werden. Im Betrieb handelt es sich einschließlich der neu entwickelten Grenzwertbestimmung sowie der verwendeten überwachten Methoden um einen automatisierten Ablauf. Eine vollständige Automatisierung wird allerdings aufgrund des Identifizierungsprozessen aktuell als kritisch angesehen. Zwar bietet die in dieser Arbeit verwendete Pipeline für das überwachte maschinelle Lernen die Möglichkeit, den Identifikations- und Interpretationsprozess zu unterstützen und somit zu beschleunigen. Eine vollständig softwarebasierte Identifikation würde jedoch eine detaillierte Semantik erfordern, was umfangreiche und kostspielige Vorarbeiten voraussetzt. Aus diesem Grund ist die vorgestellte Methodik als ein Assistenzsystem für Instandhaltungspersonal und zuständige Prozessingenieure zu verstehen und nicht als ein autonom agierendes Zustandsüberwachungssystem.

Neben der adaptiven Zustandsüberwaschung wurde im Rahmen dieser Arbeit ein grundlegendes Konzept zur Mess- und Datenplanung vorgestellt, welches die Grundlage für weiterführende Arbeiten in diesem Bereich darstellt. Besteht die Möglichkeit, im Rahmen von Versuchen Trainingsdaten spezifischer Fehlerzustände zu erheben, kann dadurch die benötigte strukturelle und informationsbezogene

Datenqualität sichergestellt werden. Wie in Anwendung 3 und 4 ersichtlich handelt es sich allerdings bei der Sicherstellung der strukturellen Datenqualität speziell bei Bestandsanlagen um einen individuellen Prozess, welcher beispielsweise von den verfügbaren Sensorsignalen bzw. weiteren Steuer-, Positionierungs- oder Triggersignalen abhängt. Eine generelle Methodik ist aus diesem Grund nur begrenzt abzuleiten.

Die im Rahmen dieser Arbeit vorgestellten Methoden zur überwachten Modellbildung zeichnen sich durch eine hohe Transparenz aus und eignen sich somit hervorragend zur Analyse des Informationsgehalts der aufgezeichneten Daten sowie zur Identifikation von Problemen, beispielsweise im Zusammenhang mit der Datenaufzeichnung.

6 Zusammenfassung und Ausblick

Trotz des bereits lang anhaltenden Hypes im Bereich der künstlichen Intelligenz und im speziellen des maschinellen Lernens sind im realen industriellen Umfeld aktuell wenige datengetriebene Modelle tatsächlich in der Anwendung, um die Wertschöpfung industrieller Prozesse zu steigern. Die Ursachen hierfür stellen hauptsächlich die Stabilität der Modelle sowie die mangelnde Qualität und Verfügbarkeit von Trainingsdaten dar. Aus diesem Grund wurde in der vorliegenden Arbeit eine Methode zur adaptiven Zustandsüberwachung in Form eines übergeordneten Algorithmus, basierend auf der Kombination von teilüberwachten und überwachten Methoden des maschinellen Lernens vorgestellt. Dieser erlaubt die Zustandsüberwachung industrieller Anwendungen ausgehend vom Gutzustand der Anlage, wobei wiederholt auftretende Anomalien erkannt, identifiziert bzw. interpretiert und sukzessive in das überwachte Modell übertragen werden können. Treten diese Fehler das nächste Mal auf, können sie vom überwachten Modell diagnostiziert und entsprechend ihrer Dringlichkeit in die nächsten Wartungs- bzw. Instandhaltungsmaßnahmen eingeplant werden.

Der für die adaptive Zustandsüberwachung verwendete, übergeordnete Algorithmus erlaubt dabei die Kombination beliebiger Methoden der Novelty Detection und des überwachten Lernens, was eine schnelle Anpassung an die jeweilige Anwendung ermöglicht. Die Implementierung beruht auf einer MATLAB-basierten open-source Toolbox zum maschinellen Lernen für zyklische Sensordaten, welche eine breite Palette an Methoden für alle Ebenen des klassischen maschinellen Lernens zur Verfügung stellt. Im Rahmen der Arbeit wurde diese Toolbox um eine breitbandige Merkmalsextraktion ergänzt, welche eine Extraktion von Merkmalen aus Zeit- und Frequenzbereich ermöglicht und somit gerade im Hinblick auf eine Novelty Detection eine generelle Empfindlichkeit gegenüber unbekannten Systemzuständen bietet. Speziell in Kombination mit einer Merkmalsselektion ermöglicht die Segmentierung der Signale im Zeit- und Frequenzbereich die Rückverfolgbarkeit der Merkmale auf ihren Ursprung, d. h. auf bestimmte Ereignisse im Zeitbereich oder bestimmte Bänder im Frequenzbereich. Im Falle der adaptiven Zustandsüberwachung kann dies nicht nur für eine Plausibilisierung der Merkmale, sondern auch zur Identifikation der Ursache des neu eingetretenen Systemzustands eingesetzt werden. Hierzu wird ein zusätzliches, unterstützendes Modell zur Trennung des neu eingetretenen Zustands

von den bereits bekannten Zuständen trainiert. Im Rahmen der vorgestellten Untersuchungen konnten durch die Analyse der dafür selektierten Merkmale konkrete Hinweise auf die Fehlerursache bereitgestellt werden, was eine effektive Interpretation in der jeweiligen Anwendung ermöglichen würde.

Je nach Konfiguration der Merkmalsextraktion entsteht eine sehr große Anzahl an Merkmalen ($N_{Zeit} = 10, N_{Freq} = 10$ entsprechen 1188 Merkmalen pro Sensor), welche abhängig von der Dynamik des betrachteten Prozesses eine hohe Redundanz aufweisen können. Die neu implementierte, verwendete Merkmalsselektion mittels Maximum Relevance Minimum Redundancy schafft es, diese großen Merkmalspools bei vergleichsweise geringen Berechnungszeiten nach relevanten Merkmalen zu durchsuchen und die Redundanz im selektierten Merkmalssubset zu minimieren. Sowohl die Anzahl der Segmente für die Merkmalsextraktion als auch die Anzahl der für die Modellbildung verwendeten Merkmale stellen Hyperparameter da, welche mittels Bayes Optimierung bestimmt werden können.

Neben der Novelty Detection zur Erkennung neuer Zustände wurde in die vorgestellte adaptive Zustandsüberwachung eine Domain Shift Detection integriert und erprobt. Im Vergleich zur Novelty Detection erfolgt der Eingriff dieser Methode im Modellbildungsprozess erst nach der Merkmalsselektion und ist daher hoch sensitiv gegenüber Effekten, die das überwachte Modell beeinflussen. Dadurch erlaubt die Domain Shift Detection beispielsweise eine zuverlässige Überprüfung der Zugehörigkeit unbekannter Datenpunkte zu bereits im überwachten Modell enthaltener Zustände. Wird ein Modell in Folge eines Domain Shifts beispielsweise durch sich ändernde Umgebungsbedingungen unbrauchbar, erlaubt die adaptive Zustandsüberwachung den Neuaufbau des Modells. Im Zwischenschritt sollte jedoch unbedingt versucht werden, das bestehende Modell an die sich ändernden Umgebungsbedingungen anzupassen, weshalb eine Kombination der aktuellen Methodik mit Domain Adaptation Methoden und deren Erprobung empfohlen wird.

Den zentralen Hyperparameter der Novelty Detection und Domain Shift Detection stellt der zu bestimmende Grenzwert dar, welcher definiert, ob ein eintreffender Datenpunkt als normal oder neu eingestuft wird. Zur Automatisierung dieses Prozesses wurde eine Methode auf Basis der Verteilung der Daten in Anlehnung an die klassische Messunsicherheitsbetrachtung vorgeschlagen. Trotz der von der Normalverteilung abweichenden kNN basierten Novelty Scores zeigt die Methode eine hohe Empfindlichkeit bei einer geringen Anzahl an Fehlalarmen in allen untersuchten Anwendungen.

Die Entwicklung der vorgestellten Methoden erfolgte anhand des qualitativ hochwertigen Datensatzes stammend aus Anwendung 1 Zustandsüberwachung eines

Hydraulischen Systems. Zur Demonstration der Leistungsfähigkeit wurden die Methoden im Anschluss auf die Daten weiterer Anwendungen übertragen. Anwendung Zustandsüberwachung einer servopneumatischen Schweißzange ist dabei thematisch im Bereich der klassischen Zustandsüberwachung angesiedelt. Das Ziel ist dabei die Klassifizierung bzw. Diagnose von anwendungsspezifischen Fehlern, welche synthetisch auf Basis komplexer Prüfaufbauten simuliert werden können. Hier konnte mit Hilfe der präsentierten Methodik eine Zustandsüberwachung ausgehend vom Gutzustand der Systeme begonnen und für alle verfügbaren spezifischen Fehler weiterentwickelt werden. Eine Ausnahme stellte hierbei das erste Auftreten einer erhöhten Lagerreibung der Schweißzange dar, was die Grenzen der teilüberwachten Methoden verdeutlicht.

Zudem konnte gezeigt werden, welchen gravierenden Einfluss Domain Shifts auf die Funktionalität bestehender überwachter Modelle haben können und wie eine adaptive Zustandsüberwachung ein erneutes Anlernen der Zustände ermöglicht. Ein Lernen der Zustände im laufenden Betrieb würde die Einstiegshürde für den Einsatz datengetriebener Modelle in vielen Anwendungsbereichen senken. Dadurch könnten diese Methoden auch in Bereichen Anwendung finden, die bisher aufgrund der hohen Kosten für die Generierung von Trainingsdaten als nicht rentabel gelten.

Eine weitere Hürde für den Einsatz datengetriebener Methoden stellt häufig die benötigte hohe Qualität der Daten in Bezug auf Struktur und Information dar. Zur Absicherung der Datenqualität wird im Rahmen dieser Arbeit eine der Datengenerierung bzw. Modellbildung vorgelagerte bzw. integrierte Mess- und Datenplanung vorgeschlagen. Diese stellt einen iterativen Prozess zur Überprüfung, Analyse und Validierung der experimentell generierten Daten dar, bevor mit einer umfangreicheren, nach DoE geplanten Datengenerierung begonnen werden kann.

In Anwendung 3 und 4 steht folgerichtig neben der Erprobung der adaptiven Zustandsüberwachung zudem die Generierung von qualitativ hochwertigen Datensätzen im Fokus. In Anwendung 3 wurden hierzu im ständigen Austausch mit dem Projektpartner Festo Analysen zur Anpassung von Datenstruktur und Datenerfassung durchgeführt und im Anschluss eine erste umfangreiche Datengenerierung gestartet. Im Rahmen der Analyse dieser Datensätze konnte ebenfalls das Potential der adaptiven Zustandsüberwachung zur Erweiterung des Messbereichs eines Regressionsmodells demonstriert werden.

Bei Anwendung 4 wurde zudem eine Datenerfassung für die dedizierten Sensoren der Versuchsanlage entwickelt, welches eine synchrone, hochfrequente Abtastung der verwendeten Beschleunigungs- und Acoustic Emission Sensoren ermöglicht. Die

Analyse des resultierenden Datensatzes offenbarte die hohe Komplexität der Daten sowie des zugrundeliegenden Lernproblems.

Im Rahmen dieser Arbeit erfolgte die Erprobung der Methoden bereits an einer breiten Palette industrieller Anwendungen, wodurch unter anderem auch Potential für die Weiterentwicklung der Methoden aufgedeckt wurde. Die eingesetzten Stacks für Klassifizierung und Regression zeigen bei den durchgeführten Untersuchungen bei Anwendung 1 – 3 grundsätzlich eine ausreichende Performanz. Eine Ausnahme stellt hier die werkzeugübergreifende Verschleißdetektion bei Anwendung 4 dar. Hier sollte einerseits versucht werden, ein tieferes Verständnis für das zugrundeliegende Problem aufzubauen, um dadurch mögliche Lösungsansätze für die Übertragbarkeit zu identifizieren wie beispielsweise eine weiterführende Datenvorverarbeitung oder eine synthetische Erweiterung der Trainingsdatenbasis. Grundsätzlich ist allerdings die Frage zu stellen, ob das Problem rein auf algorithmischer Seite zu lösen ist, oder ob im ersten Schritt eine noch größere Datenbasis geschaffen werden muss.

Bei den vorgestellten Untersuchungen zeigte sich das große Potential einer adaptiven Zustandsüberwachung, speziell beim Anlernen diskreter Fehlermechanismen an komplexen Systemen. Die Stärken und Schwächen sind dabei jedoch stark von den verwendeten Stacks abhängig. Um ein vollständigeres Bild der Leistungsfähigkeit zu bekommen, müssen also weitere Untersuchungen auf Basis unterschiedlicher teilüberwachter und überwachter Stacks durchgeführt werden. Zudem könnte beispielsweise die Erweiterung der Methodik um eine umfassende Methodensuche an Stelle des überwachten Stacks erfolgen.

Literaturverzeichnis

[1] C. P. Keferstein, M. Marxer, and C. Bach, “Einführung,” in *Fertigungsmesstechnik: Alles zu Messunsicherheit, konventioneller Messtechnik und Multisensorik*, C. P. Keferstein, M. Marxer, and C. Bach, Eds., Wiesbaden: Springer Fachmedien Wiesbaden, 2018, pp. 1–15. doi: 10.1007/978-3-658-17756-0_1.

[2] T. Żabiński, T. Mączka, J. Kluska, M. Madera, and J. Sęp, “Condition monitoring in Industry 4.0 production systems - the idea of computational intelligence methods application,” *Procedia CIRP*, vol. 79, pp. 63–67, 2019, doi: 10.1016/j.procir.2019.02.012.

[3] R. Randall, “Introduction and Background,” in *Vibration–based Condition Monitoring*, 2021, pp. 1–24. doi: 10.1002/9781119477631.ch1.

[4] R. Randall, “Fault Trending and Prognostics,” in *Vibration–based Condition Monitoring*, 2021, pp. 355–392. doi: 10.1002/9781119477631.ch9.

[5] R. Randall, “Diagnostic Techniques for Particular Applications,” in *Vibration–based Condition Monitoring*, 2021, pp. 231–307. doi: 10.1002/9781119477631.ch7.

[6] N. Gupta *et al.*, “Data Quality for Machine Learning Tasks,” in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, in KDD ’21. New York, USA, 2021, pp. 4040–4041. doi: 10.1145/3447548.3470817.

[7] N. Helwig, E. Pignanelli, and A. Schütze, “Condition monitoring of a complex hydraulic system using multivariate statistics,” in *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 2015, pp. 210–215. doi: 10.1109/I2MTC.2015.7151267.

[8] T. Schneider, N. Helwig, and A. Schütze, “Industrial condition monitoring with smart sensors using automated feature extraction and

selection," *Meas Sci Technol*, vol. 29, no. 9, p. 094002, Sep. 2018, doi: 10.1088/1361-6501/aad1d4.

[9] T. Schneider, N. Helwig, and A. Schütze, "Automatic feature extraction and selection for condition monitoring and related datasets," in *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2018, pp. 1–6. doi: 10.1109/I2MTC.2018.8409763.

[10] Ö. Gözüaçık and F. Can, "Concept learning using one-class classifiers for implicit drift detection in evolving data streams," *Artif Intell Rev*, vol. 54, no. 5, pp. 3725–3747, 2021, doi: 10.1007/s10462-020-09939-x.

[11] N. Helwig, S. Klein, and A. Schütze, "Identification and Quantification of Hydraulic System Faults Based on Multivariate Statistics Using Spectral Vibration Features," *Procedia Eng*, vol. 120, pp. 1225–1228, 2015, doi: 10.1016/j.proeng.2015.08.835.

[12] T. Schneider, S. Klein, A. Blum, L. Schirmer, R. Müller, and A. Schütze, "Combination of Human and Machine Intelligence to Optimize Assembly," in *2019 First International Conference on Societal Automation (SA)*, 2019, pp. 1–4. doi: 10.1109/SA47457.2019.8938082.

[13] A. Blum *et al.*, "Ganzheitlicher Ablaufplan für wissensgetriebene Projekte des maschinellen Lernens in der Produktion," *tm - Technisches Messen*, vol. 89, no. 5, pp. 363–383, 2022, doi: 10.1515/teme-2022-0027.

[14] Bundesministerium für Bildung und Forschung, "KI-MUSIK 4.0, Mikroelektronik-basierte universelle Sensor-Schnittstelle mit Künstlicher Intelligenz für Industrie 4.0." Accessed: Sep. 08, 2024. [Online]. Available: https://www.elektronikforschung.de/projekte/ki-musik4.0

[15] I. Döbel *et al.*, "Maschinelles Lernen–Kompetenzen, Anwendungen und Forschungsbedarf," 2018. Accessed: Sep. 16, 2024. [Online]. Available: file:///C:/Users/27ste/Downloads/Fraunhofer_Studie_ML_201809.pdf

[16] S. Kamm, S. S. Veekati, T. Müller, N. Jazdi, and M. Weyrich, "A survey on machine learning based analysis of heterogeneous data in industrial automation," *Comput Ind*, vol. 149, p. 103930, 2023, doi: 10.1016/j.compind.2023.103930.

[17] M. Bertolini, D. Mezzogori, M. Neroni, and F. Zammori, "Machine Learning for industrial applications: A comprehensive literature review,"

Expert Syst Appl, vol. 175, p. 114820, 2021, doi: 10.1016/j.eswa.2021.114820.

[18] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Comput Sci*, vol. 2, no. 3, p. 160, 2021, doi: 10.1007/s42979-021-00592-x.

[19] J. M. Ramírez-Sanz, J.-A. Maestro-Prieto, Á. Arnaiz-González, and A. Bustillo, “Semi-supervised learning for industrial fault detection and diagnosis: A systemic review,” *ISA Trans*, vol. 143, pp. 255–270, 2023, doi: 10.1016/j.isatra.2023.09.027.

[20] O. Surucu, S. A. Gadsden, and J. Yawney, “Condition Monitoring using Machine Learning: A Review of Theory, Applications, and Recent Advances,” *Expert Syst Appl*, vol. 221, p. 119738, 2023, doi: 10.1016/j.eswa.2023.119738.

[21] M. Iliyas Ahmad, Y. Yusof, M. E. Daud, K. Latiff, A. Z. Abdul Kadir, and Y. Saif, “Machine monitoring system: a decade in review,” *The International Journal of Advanced Manufacturing Technology*, vol. 108, no. 11, pp. 3645–3659, 2020, doi: 10.1007/s00170-020-05620-3.

[22] International Organization for Standardization (ISO), “3372, Condition monitoring and diagnostics - Vocabulary,” 2012, *Geneva, Switzerland*.

[23] M. S. H. Bhuiyan, I. A. Choudhury, M. Dahari, Y. Nukman, and S. Z. Dawal, “Application of acoustic emission sensor to investigate the frequency of tool wear and plastic deformation in tool condition monitoring,” *Measurement*, vol. 92, pp. 208–217, 2016, doi: 10.1016/j.measurement.2016.06.006.

[24] J. KOLERUS, “Teil 2. Verfahren und Geräte / Acoustic emission analysis,” *tm - Technisches Messen*, vol. 47, no. JG, pp. 427–434, 1980, doi: 10.1524/teme.1980.47.jg.427.

[25] R. Randall, “Vibration Signals from Rotating and Reciprocating Machines,” in *Vibration–based Condition Monitoring*, 2021, pp. 25–61. doi: 10.1002/9781119477631.ch2.

[26] T. Tinga and R. Loendersloot, “Physical Model-Based Prognostics and Health Monitoring to Enable Predictive Maintenance,” in *Predictive Maintenance in Dynamic Systems: Advanced Methods, Decision Support Tools and Real-World Applications*, Cham: Springer International Publishing, 2019, pp. 313–353. doi: 10.1007/978-3-030-05645-2_11.

[27] N. Vojdani and B. Erichsen, “Anwendungspotenziale von maschinellem Lernen in der Produktion und Logistik,” *Logistics Journal : Proceedings*, vol. 2020, no. 12, 2020, doi: 10.2195/lj_Proc_erichsen_de_202012_01.

[28] M. Goldstein and S. Uchida, “A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data,” *PLoS One*, vol. 11, no. 4, pp. e0152173-, Apr. 2016, doi: 10.1371/journal.pone.0152173.

[29] A. K. Shakya, G. Pillai, and S. Chakrabarty, “Reinforcement learning algorithms: A brief survey,” *Expert Syst Appl*, vol. 231, p. 120495, 2023, doi: 10.1016/j.eswa.2023.120495.

[30] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. da P. Francisco, J. P. Basto, and S. G. S. Alcalá, “A systematic literature review of machine learning methods applied to predictive maintenance,” *Comput Ind Eng*, vol. 137, Nov. 2019, doi: 10.1016/j.cie.2019.106024.

[31] S. Weyer *et al.*, “KI in der laufenden Produktion – eine Gratwanderung zwischen tiefer Datenwissenschaft und intuitivem Plug & Play trainierter Modelle,” *VDI-Berichte 2351, Automation 2019*, pp. 731–738, Jan. 2019, doi: 10.51202/9783181023518-731.

[32] L. Tarassenko, D. A. Clifton, P. R. Bannister, S. King, and D. King, “Novelty Detection,” in *Encyclopedia of Structural Health Monitoring*, John Wiley & Sons, Ltd, 2009. doi: 10.1002/9780470061626.shm183.

[33] S. Huang, I. Arpaci, M. Al-Emran, S. Kılıçarslan, and M. A. Al-Sharafi, “A comparative analysis of classical machine learning and deep learning techniques for predicting lung cancer survivability,” *Multimed Tools Appl*, vol. 82, no. 22, pp. 34183–34198, 2023, doi: 10.1007/s11042-023-16349-y.

[34] D. P. F. Möller, “Machine Learning and Deep Learning,” in *Guide to Cybersecurity in Digital Transformation: Trends, Methods, Technologies, Applications and Best Practices*, D. P. F. Möller, Ed., Cham: Springer Nature Switzerland, 2023, pp. 347–384. doi: 10.1007/978-3-031-26845-8_8.

[35] L. Leoni, A. Bahoo Toroody, M. M. Abaei, A. Cantini, F. Bahoo Toroody, and F. De Carlo, “Machine learning and deep learning for safety applications: Investigating the intellectual structure and the temporal evolution,” *Saf Sci*, vol. 170, p. 106363, 2024, doi: 10.1016/j.ssci.2023.106363.

[36] J. Zhu, N. Chen, and W. Peng, “Estimation of Bearing Remaining Useful Life Based on Multiscale Convolutional Neural Network,” *IEEE Transactions on Industrial Electronics*, vol. 66, no. 4, pp. 3208–3216, 2019, doi: 10.1109/TIE.2018.2844856.

[37] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, “A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects,” *IEEE Trans Neural Netw Learn Syst*, vol. 33, no. 12, pp. 6999–7019, 2022, doi: 10.1109/TNNLS.2021.3084827.

[38] S. Geman, E. Bienenstock, and R. Doursat, “Neural Networks and the Bias/Variance Dilemma,” *Neural Comput*, vol. 4, no. 1, pp. 1–58, Jan. 1992, doi: 10.1162/neco.1992.4.1.1.

[39] K. R. Varshney, “Engineering safety in machine learning,” in *2016 Information Theory and Applications Workshop (ITA)*, 2016, pp. 1–5. doi: 10.1109/ITA.2016.7888195.

[40] P. Goodarzi, A. Schütze, and T. Schneider, “Comparison of different ML methods concerning prediction quality, domain adaptation and robustness,” *tm - Technisches Messen*, vol. 89, no. 4, pp. 224–239, 2022, doi: 10.1515/teme-2021-0129.

[41] P. Goodarzi, A. Schütze, and T. Schneider, “Comparing AutoML and Deep Learning Methods for Condition Monitoring using Realistic Validation Scenarios,” *arXiv preprint arXiv:2308.14632*, 2023.

[42] C. Janiesch, P. Zschech, and K. Heinrich, “Machine learning and deep learning,” *Electronic Markets*, vol. 31, no. 3, pp. 685–695, 2021, doi: 10.1007/s12525-021-00475-2.

[43] E. Trunzer *et al.*, “Big Data trifft Produktion,” *atp magazin*, vol. 61, pp. 90–98, 2019, doi: 10.17560/atp.v61i1-2.2394.

[44] A. Weckenmann and M. Knauer, “Causes and consequences of measurement uncertainty in production metrology,” *e&i Elektrotechnik und Informationstechnik*, vol. 116, no. 4, pp. 234–239, 1999, doi: 10.1007/BF03159510.

[45] T. Dorst, T. Schneider, S. Eichstädt, and A. Schütze, “Influence of measurement uncertainty on machine learning results demonstrated for a smart gas sensor,” *Journal of Sensors and Sensor Systems*, vol. 12, no. 1, pp. 45–60, 2023, doi: 10.5194/jsss-12-45-2023.

[46] G. Sartorius, “Messtechnik und maschinelles Lernen,” in *Erfassen, Verarbeiten und Zuordnen multivariater Messgrößen: Neue Rahmenbedingungen für das Nächste-Nachbarn-Verfahren*, Wiesbaden: Springer Fachmedien, 2019, pp. 53–71. doi: 10.1007/978-3-658-23576-5_5.

[47] T. Mühl, *elektrische Messtechnik*. Springer Vieweg Wiesbaden, 2020. doi: 10.1007/978-3-658-29116-7.

[48] BIPM *et al.*, “Evaluation of measurement data — Guide to the expression of uncertainty in measurement.” [Online]. Available: https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf/cb0ef43f-baa5-11cf-3f85-4dcd86f77bd6

[49] BIPM *et al.*, “Evaluation of measurement data — Supplement 1 to the ``Guide to the expression of uncertainty in measurement'' — Propagation of distributions using a Monte Carlo method.” [Online]. Available: https://www.bipm.org/documents/20126/2071204/JCGM_101_2008_E.pdf/325dcaad-c15a-407c-1105-8b7f322d651c

[50] BIPM *et al.*, “Evaluation of measurement data — Supplement 2 to the ``Guide to the expression of uncertainty in measurement'' — Extension to any number of output quantities.” [Online]. Available: https://www.bipm.org/documents/20126/2071204/JCGM_102_2011_E.pdf/6a3281aa-1397-d703-d7a1-a8d58c9bf2a5

[51] K.-D. Sommer and B. R. L. Siebert, “Praxisgerechtes Bestimmen der Messunsicherheit nach GUM (Practical Determination of the Measurement Uncertainty under GUM),” vol. 71, no. 2, pp. 52–66, 2004, doi: 10.1524/teme.71.2.52.27068.

[52] T. Dorst, T. Schneider, S. Eichstädt, and A. Schütze, “Uncertainty-aware automated machine learning toolbox,” vol. 90, no. 3, pp. 141–153, 2023, doi: 10.1515/teme-2022-0042.

[53] T. Dorst *et al.*, “Uncertainty-aware data pipeline of calibrated MEMS sensors used for machine learning,” *Measurement: Sensors*, vol. 22, p. 100376, 2022, doi: 10.1016/j.measen.2022.100376.

[54] T. Lange and K. Mosler, “Normalverteilung und zentraler Grenzwertsatz,” in *Statistik kompakt: Basiswissen für Ökonomen und Ingenieure*, T. Lange and K. Mosler, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2017, pp. 55–62. doi: 10.1007/978-3-662-53467-0_6.

[55] P. Driscoll, F. Lecky, and M. Crosby, “An introduction to everyday statistics—1,” *J Accid Emerg Med*, vol. 17, no. 3, p. 205, 2000, doi: 10.1136/emj.17.3.205-a.

[56] N. Mohd Razali and B. Yap, “Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests,” *J. Stat. Model. Analytics*, vol. 2, Jan. 2011.

[57] D. Öztuna, A. Elhan, and E. Tüccar, “Investigation of four different normality tests in terms of type 1 error rate and power under different distributions,” *Turk J Med Sci*, vol. 36, pp. 171–176, Jan. 2006.

[58] J. I. Marden, “Positions and QQ Plots,” *Statistical Science*, vol. 19, no. 4, pp. 606–614, 2004, [Online]. Available: http://www.jstor.org/stable/4144431

[59] S. DEMİR, “Comparison of Normality Tests in Terms of Sample Sizes under Different Skewness and Kurtosis Coefficients,” *International Journal of Assessment Tools in Education*, vol. 9, no. 2, pp. 397–409, 2022, doi: 10.21449/ijate.1101295.

[60] V. W. Berger and Y. Zhou, “Kolmogorov–Smirnov Test: Overview,” in *Wiley StatsRef: Statistics Reference Online*, 2014. doi: 10.1002/9781118445112.stat06558.

[61] K. Klauenberg and C. Elster, “Testing normality: An introduction with sample size calculation in legal metrology,” *tm - Technisches Messen*, vol. 86, no. 12, pp. 773–783, 2019, doi: doi:10.1515/teme-2019-0148.

[62] S. S. Shapiro and M. B. Wilk, “An Analysis of Variance Test for Normality (Complete Samples),” *Biometrika*, vol. 52, pp. 591–611, 1965, doi: 10.2307/2333709.

[63] T. Schneider, T. Dorst, C. Schnur, P. Goodarzi, and A. Schütze, “Einfluss von Datenqualität, Domain Shift und Messunsicherheit auf die Vorhersagequalität smarter Sensorsysteme: Influence of data quality, domain shift, and measurement uncertainty on the prediction quality of smart sensor systems,” *tm - Technisches Messen*, vol. 90, no. s1, pp. 33–36, 2023, doi: 10.1515/teme-2023-0087.

[64] L. Budach *et al.*, “The Effects of Data Quality on Machine Learning Performance,” 2022. [Online]. Available: https://arxiv.org/abs/2207.14529

[65] Y. Wilhelm, U. Schreier, P. Reimann, B. Mitschang, and H. Ziekow, “Data Science Approaches to Quality Control in Manufacturing: A Review of Problems, Challenges and Architecture,” in *Service-Oriented Computing*, S. Dustdar, Ed., Cham: Springer International Publishing, 2020, pp. 45–65.

[66] J. P. Rohweder, G. Kasten, D. Malzahn, A. Piro, and J. Schmid, “Informationsqualität – Definitionen, Dimensionen und Begriffe,” in *Daten- und Informationsqualität: Die Grundlage der Digitalisierung*, K. Hildebrand, M. Gebauer, and M. Mielke, Eds., Wiesbaden: Springer Fachmedien Wiesbaden, 2021, pp. 23–43. doi: 10.1007/978-3-658-30991-6_2.

[67] V. Gudivada, A. Apon, and J. Ding, “Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations,” *International Journal on Advances in Software*, vol. 10, pp. 1–20, Jul. 2017.

[68] C. Schnur, D. Schmidt, A. Blum, Schütze Andreas, and T. Schneider, “KI-Projekte erfolgreich umsetzen: Eine Checkliste für den Mittelstand,” 2024. doi: doi: 10.5281/zenodo.10069539.

[69] B. Heinrich and M. Klier, “Datenqualitätsmetriken für ein ökonomisch orientiertes Qualitätsmanagement,” in *Daten- und Informationsqualität: Die Grundlage der Digitalisierung*, K. Hildebrand, M. Gebauer, and M. Mielke, Eds., Wiesbaden: Springer Fachmedien Wiesbaden, 2021, pp. 47–65. doi: 10.1007/978-3-658-30991-6_3.

[70] M. D. Wilkinson *et al.*, “The FAIR Guiding Principles for scientific data management and stewardship,” *Sci Data*, vol. 3, no. 1, p. 160018, 2016, doi: 10.1038/sdata.2016.18.

[71] P. Rocca-Serra *et al.*, “The FAIR Cookbook - the essential resource for and by FAIR doers,” *Sci Data*, vol. 10, no. 1, p. 292, 2023, doi: 10.1038/s41597-023-02166-3.

[72] X. Wang and R. Sun, “Domain Adaptation of Time Series Classification,” *IEEE Access*, vol. 12, pp. 38456–38472, 2024, doi: 10.1109/ACCESS.2024.3375334.

[73] P. Singhal, R. Walambe, S. Ramanna, and K. Kotecha, “Domain Adaptation: Challenges, Methods, Datasets, and Applications,” *IEEE Access*, vol. 11, pp. 6973–7020, 2023, doi: 10.1109/ACCESS.2023.3237025.

[74] Y. Feng, J. Chen, S. He, T. Pan, and Z. Zhou, “Globally Localized Multisource Domain Adaptation for Cross-Domain Fault Diagnosis With Category Shift,” *IEEE Trans Neural Netw Learn Syst*, vol. 34, no. 6, pp. 3082–3096, 2023, doi: 10.1109/TNNLS.2021.3111732.

[75] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, “Deep Model Based Domain Adaptation for Fault Diagnosis,” *IEEE Transactions on Industrial Electronics*, vol. 64, no. 3, pp. 2296–2305, 2017, doi: 10.1109/TIE.2016.2627020.

[76] R. Tanabe *et al.*, “MIMII Due: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection with Domain Shifts Due to Changes in Operational and Environmental Conditions,” *2021 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 21–25, 2021, [Online]. Available: https://api.semanticscholar.org/CorpusID:233864645

[77] I. Gao, S. Sagawa, P. W. Koh, T. Hashimoto, and P. Liang, “Out-of-Domain Robustness via Targeted Augmentations,” *ArXiv*, vol. abs/2302.11861, 2023, Accessed: Sep. 17, 2024. [Online]. Available: https://arxiv.org/abs/2302.11861

[78] J. Antony, “2 - Fundamentals of Design of Experiments,” in *Design of Experiments for Engineers and Scientists (Second Edition)*, J. Antony, Ed., Oxford: Elsevier, 2014, pp. 7–17. doi: 10.1016/B978-0-08-099417-8.00002-X.

[79] R. Fontana, A. Molena, L. Pegoraro, and L. Salmaso, “Design of experiments and machine learning with application to industrial experiments,” *Statistical Papers*, vol. 64, no. 4, pp. 1251–1274, 2023, doi: 10.1007/s00362-023-01437-w.

[80] H. Schiefer and F. Schiefer, “Statistische Versuchsplanung, Design of Experiments (DoE),” in *Statistik für Ingenieure: Eine Einführung mit Beispielen aus der Praxis*, H. Schiefer and F. Schiefer, Eds., Wiesbaden: Springer Fachmedien Wiesbaden, 2018, pp. 1–20. doi: 10.1007/978-3-658-20640-6_1.

[81] D. E. Huntington and C. S. Lyrintzis, “Improvements to and limitations of Latin hypercube sampling,” *Probabilistic Engineering Mechanics*, vol. 13, no. 4, pp. 245–253, 1998, doi: 10.1016/S0266-8920(97)00013-1.

[82] M. Liefvendahl and R. Stocki, “A study on algorithms for optimization of Latin hypercubes,” *J Stat Plan Inference*, vol. 136, no. 9, pp. 3231–3247, 2006, doi: 10.1016/j.jspi.2005.01.007.

[83] T. M. Cioppa and T. W. Lucas, “Efficient Nearly Orthogonal and Space-Filling Latin Hypercubes,” *Technometrics*, vol. 49, no. 1, pp. 45–55, Feb. 2007, doi: 10.1198/004017006000000453.

[84] T. Schneider, S. Klein, and A. Schütze, “Machine learning in industrial measurement technology for detection of known and unknown faults of equipment and sensors,” *tm - Technisches Messen*, vol. 86, no. 11, pp. 706–718, Sep. 2019, doi: 10.1515/teme-2019-0086.

[85] T. Schneider, “Methoden der automatisierten Merkmalextraktion und -selektion von Sensorsignalen,” Master Thesis, Lehrstuhl für Messtechnik, Universität des Saarlandes, Saarbrücken, 2015.

[86] D. H. Wolpert and W. G. Macready, “No free lunch theorems for optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, 1997, doi: 10.1109/4235.585893.

[87] S. Klein, T. Schneider, and A. Schütze, “Zustandsüberwachung in der Automatisierungstechnik mittels maschinellem Lernen,” in *20. GMA/ITG-Fachtagung Sensoren und Messsysteme 2019*, Nürnberg: AMA Service GmbH, 2019, pp. 366–372. doi: 10.22028/D291-37635.

[88] T. Dorst, T. Schneider, S. Klein, S. Eichstätdt, and A. Schütze, “Synchronisationsprobleme innerhalb eines Sensorsystems und deren Auswirkungen auf Ergebnisse des maschinellen Lernens,” in *20. GMA/ITG-Fachtagung Sensoren und Messsysteme 2019*, Nürnberg, 2019, pp. 444–450. doi: 10.5162/sensoren2019/5.4.1.

[89] J. Huang, Y.-F. Li, and M. Xie, “An empirical analysis of data preprocessing for machine learning-based software cost estimation,” *Inf Softw Technol*, vol. 67, pp. 108–127, 2015, doi: 10.1016/j.infsof.2015.07.004.

[90] A. Kawala-Sterniuk *et al.*, “Comparison of Smoothing Filters in Analysis of EEG Data for the Medical Diagnostics Purposes,” *Sensors*, vol. 20, no. 3, 2020, doi: 10.3390/s20030807.

[91] S. García, J. Luengo, and F. Herrera, “Data Preparation Basic Models,” in *Data Preprocessing in Data Mining*, S. García, J. Luengo, and F.

Herrera, Eds., Cham: Springer International Publishing, 2015, pp. 39–57. doi: 10.1007/978-3-319-10247-4_3.

[92] I. Bin Mohamad and D. Usman, “Standardization and its effects on k-means clustering algorithm,” *Res J Appl Sci Eng Technol*, vol. 6, no. 17, pp. 3299–3303, 2013, doi: 10.19026/rjaset.6.3638.

[93] O. Masmoudi, M. Jaoua, A. Jaoua, and S. Yacout, “Data Preparation in Machine Learning for Condition-based Maintenance,” *Journal of Computer Science*, vol. 17, no. 6, pp. 525–538, Jun. 2021, doi: 10.3844/jcssp.2021.525.538.

[94] A. Schutze, A. Gramm, and T. Ruhl, “Identification of organic solvents by a virtual multisensor system with hierarchical classification,” *IEEE Sens J*, vol. 4, no. 6, pp. 857–863, 2004, doi: 10.1109/JSEN.2004.833514.

[95] R. Bogue, “Sensors for condition monitoring: A review of technologies and applications,” *Sensor Review*, vol. 33, no. 4, pp. 295–299, Sep. 2013, doi: 10.1108/SR-05-2013-675.

[96] I. Guyon and A. Elisseeff, “An Introduction to Feature Extraction,” in *Feature Extraction: Foundations and Applications*, I. Guyon, M. Nikravesh, S. Gunn, and L. A. Zadeh, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 1–25. doi: 10.1007/978-3-540-35488-8_1.

[97] W. Caesarendra and T. Tjahjowidodo, “A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing,” Dec. 01, 2017, *MDPI AG*. doi: 10.3390/machines5040021.

[98] F. Reinhart, K. Neumann, W. Aswolinskiy, J. Steil, and B. Hammer, “Maschinelles Lernen in technischen Systemen,” in *Steigerung der Intelligenz mechatronischer Systeme*, A. Trächtler and J. Gausemeier, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2018, pp. 73–118. doi: 10.1007/978-3-662-56392-2_4.

[99] O. Aimiyekagbon, A. Bender, and W. Sextro, “Extraktion und Selektion geeigneter Merkmale für die Restlebensdauerprognose von technischen Systemen trotz aleatorischen Unsicherheiten,” *VDI-Berichte 2391*, pp. 197–210, Jan. 2021, doi: 10.51202/9783181023914-197.

[100] G. E. Kondhalkar and G. Diwakar, “Crest Factor Measurement by Experimental Vibration Analysis for Preventive Maintenance of

Bearing," in *ICRRM 2019 – System Reliability, Quality Control, Safety, Maintenance and Management*, Singapore: Springer Singapore, 2020, pp. 133–138.

[101] O. Robert, "Generalized Feature Extraction for Structural Pattern Recognition in Time-Series Data," PhD-Thesis, Carnegie Mellon University, Pittsburgh, 2001.

[102] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1, pp. 37–52, 1987, doi: 10.1016/0169-7439(87)80084-9.

[103] H. Ulrich and H. Weber, *Laplace-, Fourier- und z-Transformation: Grundlagen und Anwendungen*. Springer Fachmedien Wiesbaden, 2017. doi: 10.1007/978-3-658-03450-4.

[104] A. Neubauer, "Definition der DFT," in *DFT - Diskrete Fourier-Transformation: Elementare Einführung*, A. Neubauer, Ed., Wiesbaden: Vieweg+Teubner Verlag, 2012, pp. 17–46. doi: 10.1007/978-3-8348-1997-0_3.

[105] H. Rafiq, "Vibration Signal-Based Analysis for Gear Faults," in *Condition Monitoring and Nonlinear Frequency Analysis Based Fault Detection of Mechanical Vibration Systems*, H. Rafiq, Ed., Wiesbaden: Springer Fachmedien Wiesbaden, 2023, pp. 41–70. doi: 10.1007/978-3-658-42480-0_3.

[106] K. Uchtmann and R. Wirth, "Maschinendiagnose an drehzahlveraenderlichen Antrieben mittels Ordnungsanalyse," *Antriebstechnik*, vol. 38, pp. 44–49, 1999.

[107] D. Franke, "Signalanalyse und Anwendung schmalbandiger Kennwerte in komplexen Wälzlagerungen," in *Wälzlagerdiagnose an Maschinensätzen: Diagnose und Überwachung von Wälzlagerfehlern und -schäden*, D. Franke, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2022, pp. 145–173. doi: 10.1007/978-3-662-62620-7_8.

[108] A. Mertins, "Kurzzeit-Fourier-Transformation," in *Signaltheorie: Grundlagen der Signalbeschreibung, Filterbänke, Wavelets, Zeit-Frequenz-Analyse, Parameter- und Signalschätzung*, A. Mertins, Ed., Wiesbaden: Vieweg+Teubner, 2010, pp. 227–241. doi: 10.1007/978-3-8348-9381-9_8.

[109] A. A. Ogaili, M. N. Hamzah, A. Ab. Jaber, and E. Ghane, “Application of Discrete Wavelet Transform for Condition Monitoring and Fault Detection in Wind Turbine Blades: An Experimental Study,” *Engineering and Technology Journal*, vol. 42, no. 1, pp. 104–116, 2024, doi: 10.30684/etj.2023.142023.1516.

[110] T. N. Kruglova, “Wavelet analysis for fault diagnosis of electrical machines using current signals,” in *2016 2nd International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*, 2016, pp. 1–5. doi: 10.1109/ICIEAM.2016.7911652.

[111] M. Vetterli and C. Herley, “Wavelets and filter banks: theory and design,” *IEEE Transactions on Signal Processing*, vol. 40, no. 9, pp. 2207–2232, 1992, doi: 10.1109/78.157221.

[112] D. Theng and K. K. Bhoyar, “Feature selection techniques for machine learning: a survey of more than two decades of research,” *Knowl Inf Syst*, vol. 66, no. 3, pp. 1575–1637, 2024, doi: 10.1007/s10115-023-02010-5.

[113] A. Nowak-Brzezińska, “Feature Selection Approach for Rule-Based Knowledge Bases,” in *Advances in Feature Selection for Data and Pattern Recognition*, U. Stańczyk, B. Zielosko, and L. C. Jain, Eds., Cham: Springer International Publishing, 2018, pp. 163–182. doi: 10.1007/978-3-319-67588-6_9.

[114] P. Dhal and C. Azad, “A comprehensive survey on feature selection in the various fields of machine learning,” *Applied Intelligence*, vol. 52, no. 4, pp. 4543–4581, 2022, doi: 10.1007/s10489-021-02550-9.

[115] J. Li *et al.*, “Feature Selection: A Data Perspective,” *ACM Comput. Surv.*, vol. 50, no. 6, Dec. 2017, doi: 10.1145/3136625.

[116] B. Venkatesh and J. Anuradha, “A Review of Feature Selection and Its Methods,” *Cybernetics and Information Technologies*, vol. 19, no. 1, pp. 3–26, 2019, doi: 10.2478/cait-2019-0001.

[117] A. Jović, K. Brkić, and N. Bogunović, “A review of feature selection methods with applications,” in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2015, pp. 1200–1205. doi: 10.1109/MIPRO.2015.7160458.

[118] J. M. Sutter and J. H. Kalivas, “Comparison of Forward Selection, Backward Elimination, and Generalized Simulated Annealing for

Variable Selection," *Microchemical Journal*, vol. 47, no. 1, pp. 60–66, 1993, doi: 10.1006/mchj.1993.1012.

[119] Z. Yan and Y. Yao, "Variable selection method for fault isolation using least absolute shrinkage and selection operator (LASSO)," *Chemometrics and Intelligent Laboratory Systems*, vol. 146, pp. 136–146, 2015, doi: 10.1016/j.chemolab.2015.05.019.

[120] G. Wang, L. Sun, W. Wang, T. Chen, M. Guo, and P. Zhang, "A feature selection method combined with ridge regression and recursive feature elimination in quantitative analysis of laser induced breakdown spectroscopy," *Plasma Science and Technology*, vol. 22, no. 7, p. 074002, 2020, doi: 10.1088/2058-6272/ab76b4.

[121] N. Spolaôr, E. A. Cherman, M. C. Monard, and H. D. Lee, "ReliefF for Multi-label Feature Selection," in *2013 Brazilian Conference on Intelligent Systems*, 2013, pp. 6–11. doi: 10.1109/BRACIS.2013.10.

[122] X. Lin *et al.*, "A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information," *Journal of Chromatography B*, vol. 910, pp. 149–155, 2012, doi: 10.1016/j.jchromb.2012.05.020.

[123] P. Bugata and P. Drotar, "On some aspects of minimum redundancy maximum relevance feature selection," *Science China Information Sciences*, vol. 63, no. 1, p. 112103, 2019, doi: 10.1007/s11432-019-2633-y.

[124] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," in *Computational Systems Bioinformatics. CSB2003. Proceedings of the 2003 IEEE Bioinformatics Conference. CSB2003*, 2003, pp. 523–528. doi: 10.1109/CSB.2003.1227396.

[125] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural Comput Appl*, vol. 24, no. 1, pp. 175–186, 2014, doi: 10.1007/s00521-013-1368-0.

[126] G. A. Darbellay and I. Vajda, "Estimation of the information by an adaptive partitioning of the observation space," *IEEE Trans Inf Theory*, vol. 45, no. 4, pp. 1315–1321, 1999, doi: 10.1109/18.761290.

[127] MathWorks Inc., "Dokumentation fscmrmr." Accessed: Jul. 13, 2024. [Online]. Available:

https://de.mathworks.com/help/stats/fscmrmr.html#mw_4bc5d3f4-bef8-4dd3-bcc9-877d3ea65840

[128] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction," *Journal of Applied Science and Technology Trends*, vol. 1, no. 1, pp. 56–70, May 2020, doi: 10.38094/jastt1224.

[129] S. Velliangiri, S. Alagumuthukrishnan, and S. I. Thankumar joseph, "A Review of Dimensionality Reduction Techniques for Efficient Computation," *Procedia Comput Sci*, vol. 165, pp. 104–111, 2019, doi: 10.1016/j.procs.2020.01.079.

[130] A. Handl, "Hauptkomponentenanalyse," in *Multivariate Analysemethoden: Theorie und Praxis multivariater Verfahren unter besonderer Berücksichtigung von S-PLUS*, A. Handl, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 115–147. doi: 10.1007/978-3-642-14987-0_5.

[131] M. Greenacre, P. Groenen, T. Hastie, A. I. D'Enza, A. Markos, and E. Tuzhilina, "Principal component analysis," *Nature Reviews Methods Primers*, vol. 2, no. 1, p. 100, 2022, doi: 10.1038/s43586-022-00184-w.

[132] K. Backhaus, B. Erichson, W. Plinke, and R. Weiber, "Faktorenanalyse," in *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*, Springer Gabler, Berlin, Heidelberg, 2018, pp. 365–433. doi: 10.1007/978-3-662-56655-8_8.

[133] I. T. Jolliffe, "Mathematical and Statistical Properties of Population Principal Components," in *Principal Component Analysis. Springer Series in Statistics*, Springer, New York, 1986, pp. 8–22. doi: 10.1007/978-1-4757-1904-8_2.

[134] J. L. Lastovicka, "Book Review: A User's Guide to Principal Components," *Journal of Marketing Research*, vol. 29, no. 4, pp. 485–486, Nov. 1992, doi: 10.1177/002224379202900412.

[135] R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *Ann Eugen*, vol. 7, no. 2, pp. 179–188, Sep. 1936, doi: 10.1111/j.1469-1809.1936.tb02137.x.

[136] R. D. Decker, S. Rašković, and K. Brunsiek, "Diskriminanzanalyse," in *Handbuch der sozialwissenschaftlichen Datenanalyse*, C. Wolf and H.

Best, Eds., Wiesbaden: VS Verlag für Sozialwissenschaften, 2010, pp. 495–523. doi: 10.1007/978-3-531-92038-2_20.

[137] A. Handl, “Diskriminanzanalyse,” in *Multivariate Analysemethoden. Statistik und ihre Anwendungen*, A. Handl, Ed., Springer, Berlin, Heidelberg, 2002, pp. 315–361. doi: 10.1007/978-3-662-08887-6_12.

[138] J. Janssen and W. Laatz, “Diskriminanzanalyse,” in *Statistische Datenanalyse mit SPSS für Windows: Eine anwendungsorientierte Einführung in das Basissystem und das Modul Exakte Tests*, Springer, Berlin, Heidelberg, 2003, pp. 439–456. doi: 10.1007/978-3-662-10038-7_20.

[139] M. Feilmeier, I. Fergel, and G. Segerer, “Lineare Diskriminanzanalyse und Clusteranalyseverfahren bei Kreditscoringsystemen,” *Zeitschrift für Operations Research*, vol. 25, no. 2, pp. B25–B38, 1981, doi: 10.1007/BF01919910.

[140] P. A. Lachenbruch, C. Sneeringer, and L. T. Revo, “Robustness of the linear and quadratic discriminant function to certain types of non-normality,” *Communications in Statistics*, vol. 1, no. 1, pp. 39–56, Jan. 1973, doi: 10.1080/03610927308827006.

[141] K. Bian and R. Priyadarshi, “Machine Learning Optimization Techniques: A Survey, Classification, Challenges, and Future Research Issues,” *Archives of Computational Methods in Engineering*, 2024, doi: 10.1007/s11831-024-10110-w.

[142] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, “Bearing Fault Detection by a Novel Condition-Monitoring Scheme Based on Statistical-Time Features and Neural Networks,” *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3398–3407, 2013, doi: 10.1109/TIE.2012.2219838.

[143] X. Chen, P. Wang, Y. Hao, and M. Zhao, “Evidential KNN-based condition monitoring and early warning method with applications in power plant,” *Neurocomputing*, vol. 315, pp. 18–32, 2018, doi: 10.1016/j.neucom.2018.05.018.

[144] J. Zhong, Z. Yang, and S. F. Wong, “Machine condition monitoring and fault diagnosis based on support vector machine,” in *2010 IEEE International Conference on Industrial Engineering and Engineering Management*, 2010, pp. 2228–2233. doi: 10.1109/IEEM.2010.5674594.

[145] M. Gerdes, “Decision trees and genetic algorithms for condition monitoring forecasting of aircraft air conditioning,” *Expert Syst Appl*, vol. 40, no. 12, pp. 5021–5026, 2013, doi: 10.1016/j.eswa.2013.03.025.

[146] P. C. Sen, M. Hajra, and M. Ghosh, “Supervised Classification Algorithms in Machine Learning: A Survey and Review,” in *Emerging Technology in Modelling and Graphics. Advances in Intelligent Systems and Computing*, J. K. Mandal and D. Bhattacharya, Eds., Springer, Singapore, 2020, pp. 99–111. doi: 10.1007/978-981-13-7403-6_11.

[147] P. Potočnik and E. Govekar, “Semi-supervised vibration-based classification and condition monitoring of compressors,” *Mech Syst Signal Process*, vol. 93, pp. 51–65, 2017, doi: 10.1016/j.ymssp.2017.01.048.

[148] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, “Machine learning: a review of classification and combining techniques,” *Artif Intell Rev*, vol. 26, no. 3, pp. 159–190, 2006, doi: 10.1007/s10462-007-9052-3.

[149] S. Abe, “Kernel-Based Methods,” in *Support Vector Machines for Pattern Classification. Advances in Pattern Recognition*, Springer, London, 2010, pp. 305–329. doi: 10.1007/978-1-84996-098-4_6.

[150] S. Abe, “Multiclass Support Vector Machines,” in *Support Vector Machines for Pattern Classification. Advances in Pattern Recognition*, Springer, London, 2010, pp. 113–161. doi: 10.1007/978-1-84996-098-4_3.

[151] M. Awad and R. Khanna, “Support Vector Regression,” in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, Apress, Berkeley, 2015, pp. 67–80. doi: 10.1007/978-1-4302-5990-9_4.

[152] A. Deris, A. M. Zain, and R. Sallehuddin, “Overview of Support Vector Machine in Modeling Machining Performances,” *Procedia Eng*, vol. 24, pp. 308–312, 2011, doi: 10.1016/j.proeng.2011.11.2647.

[153] H.-F. Eckey, R. Kosfeld, and M. Rengers, “Clusteranalyse,” in *Multivariate Statistik: Grundlagen — Methoden — Beispiele*, Gabler Verlag, Wiesbaden, 2002, pp. 203–288. doi: 10.1007/978-3-322-84476-7_4.

[154] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, “The Mahalanobis distance,” *Chemometrics and Intelligent Laboratory*

Systems, vol. 50, no. 1, pp. 1–18, 2000, doi: 10.1016/S0169-7439(99)00047-7.

[155] I. Kononenko and M. Kukar, “Chapter 3 - Machine Learning Basics,” in *Machine Learning and Data Mining*, Woodhead Publishing, 2007, pp. 59–105. doi: 10.1533/9780857099440.59.

[156] R. Gutierrez-Osuna, “Pattern analysis for machine olfaction: a review,” *IEEE Sens J*, vol. 2, no. 3, pp. 189–202, 2002, doi: 10.1109/JSEN.2002.800688.

[157] S. Wold *et al.*, “Multivariate Data Analysis in Chemistry,” in *Chemometrics: Mathematics and Statistics in Chemistry*, Springer Netherlands, Dordrecht, 1984, pp. 17–95. doi: 10.1007/978-94-017-1026-8_2.

[158] P. Geladi and B. R. Kowalski, “Partial least-squares regression: a tutorial,” *Anal Chim Acta*, vol. 185, pp. 1–17, 1986, doi: 10.1016/0003-2670(86)80028-9.

[159] I. S. Helland, “Partial Least Squares Regression and Statistical Models,” *Scandinavian Journal of Statistics*, vol. 17, no. 2, pp. 97–114, 1990, [Online]. Available: http://www.jstor.org/stable/4616159

[160] D. A. Clifton, L. A. Clifton, P. R. Bannister, and L. Tarassenko, “Automated Novelty Detection in Industrial Systems,” in *Advances of Computational Intelligence in Industrial Systems*, Springer, Berlin, Heidelberg, 2008, pp. 269–296. doi: 10.1007/978-3-540-78297-1_13.

[161] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, “A review of novelty detection,” *Signal Processing*, vol. 99, pp. 215–249, Jun. 2014, doi: 10.1016/j.sigpro.2013.12.026.

[162] M. Zhao and V. Saligrama, “Anomaly detection with score functions based on nearest neighbor graphs,” in *Proceedings of the 22nd International Conference on Neural Information Processing Systems*, in NIPS’09. Red Hook, New York, USA: Curran Associates Inc., 2009, pp. 2250–2258.

[163] P. Drews, P. Núñez, R. P. Rocha, M. Campos, and J. Dias, “Novelty detection and segmentation based on Gaussian mixture models: A case study in 3D robotic laser mapping,” *Rob Auton Syst*, vol. 61, no. 12, pp. 1696–1709, 2013, doi: 10.1016/j.robot.2013.06.004.

[164] J. Ma and S. Perkins, “Time-series novelty detection using one-class support vector machines,” in *Proceedings of the International Joint Conference on Neural Networks, 2003.*, 2003, pp. 1741–1745 vol.3. doi: 10.1109/IJCNN.2003.1223670.

[165] J. Shawe-Taylor and B. Žličar, “Novelty Detection with One-Class Support Vector Machines,” in *Advances in Statistical Models for Data Analysis. Studies in Classification, Data Analysis, and Knowledge Organization*, Springer, Cham, 2015, pp. 231–257.

[166] F. Del Buono, F. Calabrese, A. Baraldi, M. Paganelli, and F. Guerra, “Novelty Detection with Autoencoders for System Health Monitoring in Industrial Environments,” *Applied Sciences*, vol. 12, no. 10, 2022, doi: 10.3390/app12104931.

[167] X. Ding, Y. Li, A. Belatreche, and L. P. Maguire, “An experimental evaluation of novelty detection methods,” *Neurocomputing*, vol. 135, pp. 313–327, 2014, doi: 10.1016/j.neucom.2013.12.002.

[168] S. Klein, Y. Wilhelm, A. Schütze, and T. Schneider, “Combination of generic novelty detection and supervised classification pipelines for industrial condition monitoring,” *tm - Technisches Messen*, vol. 91, no. 9, pp. 454–465, 2024, doi: doi:10.1515/teme-2024-0016.

[169] C. C. Aggarwal, “Proximity-Based Outlier Detection,” in *Outlier Analysis*, C. C. Aggarwal, Ed., Springer, Cham, 2017, pp. 111–147. doi: 10.1007/978-3-319-47578-3_4.

[170] P. Refaeilzadeh, L. Tang, and H. Liu, “Cross-Validation,” in *Encyclopedia of Database Systems*, Springer, US, Boston, 2009, pp. 532–538. doi: 10.1007/978-0-387-39940-9_565.

[171] X. Zeng and T. R. Martinez, “Distribution-balanced stratified cross-validation for accuracy estimation,” *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 12, no. 1, pp. 1–12, Jan. 2000, doi: 10.1080/095281300146272.

[172] S. Arlot and A. Celisse, “A survey of cross-validation procedures for model selection,” *Stat Surv*, vol. 4, no. none, pp. 40–79, Jan. 2010, doi: 10.1214/09-SS054.

[173] B. G. Marcot and A. M. Hanea, “What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?,” *Comput Stat*, vol. 36, no. 3, pp. 2009–2031, 2021, doi: 10.1007/s00180-020-00999-9.

[174] V. K. Verma, K. Saxena, and U. Banodha, “Analysis Effect of K Values Used in K Fold Cross Validation for Enhancing Performance of Machine Learning Model with Decision Tree,” in *Advanced Computing. IACC 2023. Communications in Computer and Information Science, vol 2053*, Springer, Cham, 2024, pp. 374–396. doi: 10.1007/978-3-031-56700-1_30.

[175] M. W. Browne, “Cross-Validation Methods,” *J Math Psychol*, vol. 44, no. 1, pp. 108–132, 2000, doi: 10.1006/jmps.1999.1279.

[176] S. C. Smithson, G. Yang, W. J. Gross, and B. H. Meyer, “Neural networks designing neural networks: Multi-objective hyper-parameter optimization,” in *2016 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2016, pp. 1–8. doi: 10.1145/2966986.2967058.

[177] M. Hoffmann and R. Bergmann, “Improving Automated Hyperparameter Optimization with Case-Based Reasoning,” in *Case-Based Reasoning Research and Development. ICCBR 2022. Lecture Notes in Computer Science, vol 13405*, Springer, Cham, 2022, pp. 273–288. doi: 10.1007/978-3-031-14923-8_18.

[178] J. Nievergelt, “Exhaustive Search, Combinatorial Optimization and Enumeration: Exploring the Potential of Raw Computing Power,” in *SOFSEM 2000: Theory and Practice of Informatics. SOFSEM 2000. Lecture Notes in Computer Science, vol 1963*, Springer, Berlin, Heidelberg, 2000, pp. 18–35. doi: 10.1007/3-540-44411-4_2.

[179] J. Snoek, H. Larochelle, and R. P. Adams, “Practical Bayesian Optimization of Machine Learning Algorithms,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf

[180] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, and S.-H. Deng, “Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization,” *Journal of Electronic Science and Technology*, vol. 17, no. 1, pp. 26–40, 2019, doi: 10.11989/JEST.1674-862X.80904120.

[181] MathWorks Inc., “Dokumentation Bayesian Optimization Algorithm.” Accessed: Jul. 20, 2024. [Online]. Available:

https://de.mathworks.com/help/stats/bayesian-optimization-algorithm.html

[182] F. Archetti and A. Candelieri, “Automated Machine Learning and Bayesian Optimization,” in *Bayesian Optimization and Data Science. Springer Briefs in Optimization*, Springer, Cham, 2019, pp. 1–18. doi: 10.1007/978-3-030-24494-1_1.

[183] Z. Liu, R. Nötzel, and K.-W. Bonfig, “Abtastung schmalbandiger Signale und ihre Anwendung zur Hüllkurvenanalyse bei rechnergestützten schwingungsdiagnostischen Systemen (Sampling of Narrow Band Signals and its Application for Envelope Analysis in Computer-aided Diagnosis Systems),” *tm - Technisches Messen*, vol. 74, no. 2, pp. 63–73, 2007, doi: 10.1524/teme.2007.74.2.63.

[184] P. Goodarzi, S. Klein, A. Schütze, and T. Schneider, “Comparing Different Feature Extraction Methods in Condition Monitoring Applications,” in *2023 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2023, pp. 1–6. doi: 10.1109/I2MTC53148.2023.10176106.

[185] E. Behrends, “Die Exponentialverteilung,” in *Elementare Stochastik*, Wiesbaden: Vieweg+Teubner Verlag, Wiesbaden, 2013, pp. 187–204. doi: 10.1007/978-3-8348-2331-1_6.

[186] N. J. Helwig, “Zustandsbewertung industrieller Prozesse mittels multivariater Sensordatenanalyse am Beispiel hydraulischer und elektromechanischer Antriebssysteme,” Dissertation, Lehrstuhl für Messtechnik, Universität des Saarlandes, Saarbrücken, 2018. doi: 10.22028/D291-27896.

[187] N. Helwig, E. Pignanelli, and A. Schütze, “Detecting and Compensating Sensor Faults in a Hydraulic Condition Monitoring System,” in *Proceedings SENSOR 2015*, Nürnberg, 2015, pp. 641–646. doi: 10.5162/sensor2015/d8.1.

[188] T. Schneider, S. Klein, and M. Bastuck, “Condition monitoring of hydraulic systems Data Set at ZeMA,” Datensatz, 2018. doi: 10.5281/ZENODO.1323611.

[189] J. Weber and T. Radermacher, “Einführung in die Hydraulik,” in *Hydraulik – Fluid-Mechatronik: Grundlagen, Komponenten, Systeme, Messtechnik und virtuelles Engineering*, Springer, Berlin, Heidelberg, 2020, pp. 1–14. doi: 10.1007/978-3-662-60664-3_1.

[190] I. Effenberger, H. Eigenbrod, A. Frommknecht, C. Jauch, J. Denecke, and M. Huber, “Qualitätssicherung in der Produktion,” in *Handbuch Industrie 4.0: Band 1: Produktion*, Springer, Berlin, Heidelberg, 2023, pp. 53–69. doi: 10.1007/978-3-662-58532-0_166.

[191] ZEISS, “Integrierte Qualitätssicherung durch optische 3D-Messtechnik, Expertenforum Maschinenautomation mit Mehrwert,” Dec. 2022. Accessed: Feb. 12, 2024. [Online]. Available: https://automationspraxis.industrie.de/wp-content/uploads/A/P/AP_Reader_Fanuc.pdf

[192] Festo SE & Co. KG, “Anwendungswissen und Lösungen für die Automobil- und Zulieferindustrie,” Esslingen, Aug. 2012. Accessed: Feb. 12, 2024. [Online]. Available: https://www.festo.com/PDF_Flip/industry/ami/DE/files/assets/basic-html/index.html#page2

[193] Y. Wilhelm, P. Reimann, W. Gauchel, S. Klein, and B. Mitschang, “Pusion - A Generic and Automated Framework for Decision Fusion,” in *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, 2023, pp. 3282–3295. doi: 10.1109/ICDE55515.2023.00252.

[194] M. Wuensch *et al.*, “Datenbasis zur Bewertung von Energieeffizienzmaßnahmen 2008,” 2012, *Dessau-Roßlau*. Accessed: Jul. 31, 2024. [Online]. Available: https://www.umweltbundesamt.de/publikationen/datenbasis-zur-bewertung-von-0

[195] Universität Stuttgart Institut für Energiewirtschaft und Rationelle Energieanwendung (IER) and Fraunhofer System und Innovationsforschung, “Fakten zur Druckluft, Druckluft effizient,” Karlsruhe, Aug. 2003. Accessed: Apr. 13, 2023. [Online]. Available: http://druckluft-effizient.de/wp-content/uploads/2020/09/fakten-00-09.pdf

[196] Festo SE & Co. KG, “Projektinterne Präsentation,” 2020.

[197] STMicroelectronics, “MP23ABS1 low-power microphone Datenblatt,” Oct. 2020. Accessed: Apr. 13, 2023. [Online]. Available: https://www.st.com/resource/en/datasheet/mp23abs1.pdf

[198] STMicroelectronics, “IMP23ABSU low-power microphone Datenblatt,” Sep. 2020. Accessed: Apr. 13, 2023. [Online]. Available: https://www.st.com/resource/en/datasheet/imp23absu.pdf

[199] E. NAGATA *et al.*, “Gear Skiving for Mass Production,” *The Proceedings of the JSME international conference on motion and power transmissions*, vol. 2017, pp. 2–13, Nov. 2017, doi: 10.1299/jsmeimpt.2017.02-13.

[200] T. Arndt, J. Klose, M. Gerstenmeyer, and V. Schulze, “Tool wear development in gear skiving process of quenched and tempered internal gears,” *Forsch Ingenieurwes*, vol. 86, no. 3, pp. 587–594, 2022, doi: 10.1007/s10010-021-00544-0.

[201] X. Lin, Y. Liu, S. Sun, G. Jin, and R. Hong, “Prediction and optimization of gear skiving parameters and geometric deviations,” *The International Journal of Advanced Manufacturing Technology*, vol. 121, no. 5, pp. 4169–4185, 2022, doi: 10.1007/s00170-022-09639-6.

[202] H. I. Faulstich, “Wälzschälen von Innenverzahnungen,” *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, vol. 72, no. 3, pp. 115–119, 1977, doi: 10.1515/zwf-1977-720303.

[203] Schaeffler Technologies AG & Co. KG, “Projektinterne Präsentation,” 2020.

[204] PCB PIEZOTRONICS, “PCB-356A45 Datenblatt.” Accessed: Dec. 22, 2023. [Online]. Available: https://www.pcbpiezotronics.de/produkte_skript/downloads/specs/J356A45_specs.pdf

[205] PCB PIEZOTRONICS, “PCB-356A15 Datenblatt.” Accessed: Dec. 22, 2023. [Online]. Available: https://www.pcbpiezotronics.de/produkte_skript/downloads/specs/356A15_specs.pdf

[206] KISTLER, “AE-8152C0050500 Datenblatt.” Accessed: Dec. 22, 2023. [Online]. Available: https://kistler.cdn.celum.cloud/SAPCommerce_Download_original/003-120e.pdf

[207] R. Wirth, “Maschinendiagnose an Industriegetrieben Teil II: Signalidentifikation in der Praxis,” *Antriebstechnik*, vol. 37, pp. 77–81, 1988, Accessed: Sep. 17, 2024. [Online]. Available: https://maschinendiagnose.de/mosaic/_M_userfiles/PDF/Downloads_DE/Fachbeitraege/Maschinendiagnose_Teil_2.pdf

[208] National Instruments, “cRIO-9040 Datenblatt.” Accessed: Dec. 22, 2023. [Online]. Available: https://www.ni.com/docs/de-DE/bundle/crio-9040-specs/page/specs.html

[209] National Instruments, “NI-9232 Datenblatt.” Accessed: Dec. 22, 2023. [Online]. Available: https://www.ni.com/docs/en-US/bundle/ni-9232-specs/page/specs.html

[210] National Instruments, “NI-9223 Datenblatt.” Accessed: Dec. 22, 2023. [Online]. Available: https://www.ni.com/docs/en-US/bundle/ni-9223-specs/page/specs.html

[211] KISTLER, “AE-Piezotron Kuppler Datenblatt.” Accessed: Dec. 22, 2023. [Online]. Available: https://kistler.cdn.celum.cloud/SAPCommerce_Download_original/003-119e.pdf

[212] National Instruments, “NI-9775 Datenblatt.” Accessed: Dec. 22, 2023. [Online]. Available: https://www.ni.com/docs/de-DE/bundle/ni-9775-specs/page/specs.html

[213] D. F. Plöger, “Modulation der Zahneingriffsschwingungen von Planetengetrieben,” Dissertation, Institut für Mechatronische Systeme im Maschinenbau (IMS), Technische Universität Darmstadt, Darmstadt, 2020. doi: 10.25534/tuprints-00011329.

Abbildungsverzeichnis

Abbildung 2-1: Übersicht über verwendete Sensoren, Methoden und Ziele der Zustandsüberwachung (angepasst nach [15]). 9
Abbildung 2-2: Übersicht über verfügbare Methodenfamilien im Bereich des maschinellen Lernens (angepasst nach [18]). 10
Abbildung 2-3: Entwicklung des Abnutzungsvorrats an einer Maschine mit den zugehörigen Detektionsschwellen für überwachtes und unüberwachtes maschinelles Lernen sowie den Handlungszeiträumen für verschiedene Instandhaltungsstrategien. Bei den Verläufen von Abnutzungsvorrat, Kosten und Verschleißfortschritt handelt es sich um schematische Darstellungen, welche in der Realität stark vom individuellen Prozess abhängig sind (angepasst nach [31]). 11
Abbildung 2-4: Unterschiedliche Herangehensweisen von klassischem maschinellem Lernen und Deep Learning (angepasst nach [42]). 13
Abbildung 2-5: Einfluss von Schiefe (3. statistisches Moment) und Kurtosis (4. Statistisches Moment) auf die Dichtefunktion einer Normalverteilung sowie die zugehörigen Q-Q-Plots zum Überprüfen der Daten auf Normalverteilung. 16
Abbildung 2-6: Vollautomatisierte Toolbox für die überwachte Klassifizierung (angepasst nach [9]). 23
Abbildung 2-7: In der Toolbox verfügbare Methoden zur Signalvorverarbeitung (Standardisierung). 25
Abbildung 2-8: In der Toolbox verfügbare Methoden zur Merkmalsextraktion. 26
Abbildung 2-9: Grundsätzlicher Ablauf der Merkmalsselektion für die verfügbaren Typen Filter, Wrapper und Embedded, angepasst nach [116]. 32
Abbildung 2-10: In der Toolbox verfügbare Methoden zur Merkmalsselektion. 33
Abbildung 2-11: In der Toolbox verfügbare Methoden zur Dimensionsreduktion. 36
Abbildung 2-12: Gegenüberstellung der Projektionsrichtungen der beiden vorgestellten dimensionsreduzierenden Verfahren PCA und LDA anhand eines Testdatensatzes mit zwei Clustern bzw. Gruppen. 39
Abbildung 2-13: In der Toolbox verfügbare Methoden zur Klassifizierung und Regression. 40
Abbildung 2-14: Gegenüberstellung der euklidischen und der Mahalanobis-Distanz. 41
Abbildung 2-15: In der Toolbox verfügbare Methoden zur Novelty Detection. 43
Abbildung 2-16: Eingriffspunkte der Novelty Detection in Kombination mit überwachten maschinellen Lernverfahren (angepasst nach [84]). 45
Abbildung 2-17: a) Vorgehen bei der systematischen x-fachen Kreuzvalidierung. b) Visuelle Validierung durch die Projektion von Trainingsdaten, wie es

dimensionsreduzierende Verfahren wie in diesem Fall die LDA ermöglichen[168]. 47
Abbildung 2-18: In der Toolbox verfügbare Methoden zur Hyperparameteroptimierung. 48
Abbildung 2-19: Exemplarische Stacks für eine ND und überwachte Klassifizierung mit Domain Shift Detection. 49
Abbildung 3-1: Ablauf der Mess- und Datenplanung mit anschließender Datengenerierung nach DoE. 52
Abbildung 3-2: Szenario-Format mit klar definiertem Lernproblem. 53
Abbildung 3-3: Ablaufdiagramm für eine adaptive Zustandsüberwachung durch Kombination von ND und überwachter Lernverfahren. Nachdem beide Modelle zunächst trainiert wurden, können neue Zustände von der ND erkannt und schrittweise in das überwachte Modell übertragen werden. 56
Abbildung 3-4: Schematische Darstellung des allgemeinen Ablaufs und der Funktionsweise des TFEx. Segmentierung des Zeitsignals (links) und anschließende Berechnung des Amplitudenspektrums (Mitte) für jedes Zeitsegment. Aufzählung der zu berechnenden statistischen und formbeschreibenden Merkmale (rechts). 58
Abbildung 3-5: Stack für die im Rahmen der adaptiven Zustandsüberwachung verwendeten Novelty Detection. 59
Abbildung 3-6: Stacks für die im Rahmen der adaptiven Zustandsüberwachung verwendeten Klassifizierung und Regression mit Domain Shift Detection. 60
Abbildung 3-7: Entwicklung des Grenzwertes in Abhängigkeit der Anzahl der im Training verfügbaren Samples. 62
Abbildung 3-8: Entwicklung der zugrundeliegenden Verteilung der kNN basierten Novelty Scores in Abhängigkeit der Anzahl der Dimensionen des Merkmalsraums. 63
Abbildung 4-1: Hydraulischer Schaltplan der Prüfstände [186]. 68
Abbildung 4-2: Hydraulische Systeme, welche in [186] für die Generierung der Trainingsdaten verwendet wurden. Prüfstand 1 wurde dabei für die umfangreiche Charakterisierungsmessung mit einzelnen und überlagerten Fehlermechanismen verwendet, wobei Prüfstand 2 für Stabilitätsuntersuchungen im Dauerbetrieb verwendet wurde [186]. 69
Abbildung 4-3: Arbeitszyklus des hydraulischen Aggregats dargestellt anhand des Signals des Drucksensors P1. 70
Abbildung 4-4: Vom DoE vorgegebene, gesteuerte Zustände der Komponenten Kühler, Ventil, Pumpe und Speicher während der Charakterisierungsmessung. In den grauen Bereichen ist der Zustand aufgrund größerer Zeitkonstanten des Systems z.B. bei einer sich ändernden Öltemperatur durch Variieren der Kühlereffizienz noch nicht eingeschwungen, weshalb der Systemzustand nicht klar definiert ist. 71
Abbildung 4-5: Verwendeter Stack für die überwachte Klassifizierung im Hydraulik Use Case. 72

Abbildung 4-6: Ergebnisse der überwachten Modellbildung bei Nutzung des gesamten Datensatzes und einer randomisierten 10-fachen Kreuzvalidierung. 72
Abbildung 4-7: Herkunft der selektierten Merkmale für die Modelle der einzelnen Fehlermechanismen bezüglich der Position im Arbeitszyklus (links) und der zugrundeliegenden Sensorsignale (rechts). Die Darstellung der Segmente, aus denen die Merkmale extrahiert werden, erfolgt dabei anhand des Sensorsignals, welches den höchsten Anteil an der Gesamtkorrelation aufweist. 74
Abbildung 4-8: Einfluss einer sinkenden Kühlereffizienz bzw. der daraus resultierenden steigenden Öltemperatur auf die Vorhersagegenauigkeit der Modelle für die einzelnen Fehlermechanismen..... 75
Abbildung 4-9: Stack für die Domain Shift Detection..... 75
Abbildung 4-10: Domain Shift Scores für die überwachten Modelle der kombinierten Fehlermechanismen..... 76
Abbildung 4-11: Stack für die Novelty Detection für das Ventilschaltverhalten. 77
Abbildung 4-12: Novelty Detection Modell für das Ventilschaltverhalten und der Einfluss auf die Novelty Scores bei Eintreten einer verringerten Kühlereffizienz. 78
Abbildung 4-13: Reduzierter Datensatz für die Simulation isoliert auftretender Fehlermechanismen in steigender Intensität..... 79
Abbildung 4-14: a) Histogramm der Novelty Scores der Trainingsdaten im Normalzustand, b) QQ- Diagramm der Novelty Scores der Trainingsdaten im Normalzustand. 80
Abbildung 4-15: a) Reaktion der Novelty Detection bei Auftreten des verzögerten Schaltens von V10, b) Ergebnis der überwachten Klassifizierung anhand der ersten Diskriminanzfunktion (DF1) für Training (blau) und Test (grün) auf Basis der Ausgabe der Novelty Detection sowie der Vorhersage des Modells (orange). 81
Abbildung 4-16: Position des höchstkorrelierten, identifizierten Merkmals für die Erkennung eines fehlerhaften Schaltverhaltens von V10, berechnet mittels Kurtosis aus dem Signal des Sensors PS2, dem Drucksensor, der sich unmittelbar hinter dem Ventil befindet..... 81
Abbildung 4-17: a) Initialisierung der Novelty Detection a) und der Domain Shift Detection b) nach der Detektion eines verschlechterten Ventilschaltverhaltens von V10. 82
Abbildung 4-18: Reaktion der Novelty Detection auf eine interne Pumpenleckage. 83
Abbildung 4-19: a) Reaktion der überwachten Klassifizierung und b) der Domain Shift Detection auf eine interne Pumpenleckage..... 83
Abbildung 4-20: a) Herkunft des höchstkorrelierten Merkmals des überwachten Modells zur Trennung der bekannten Zustände vom neuen unbekannten Zustand. b) Ergebnis des überwachten Modells zur Trennung der bekannten Zustände des Hydraulikaggregats. 84

Abbildung 4-21: Initialisierung der Novelty Detection a) und der Domain Shift Detection b) nach der Detektion einer internen Pumpenleckage. 85
Abbildung 4-22: Reaktion der Novelty Detektion auf einen verringerten Vordruck des Hydrospeichers. 85
Abbildung 4-23: a) Reaktion der überwachten Klassifizierung und b) der Domain Shift Detektion auf einen verringerten Vordruck des Hydrospeichers. 86
Abbildung 4-24: Modell zur Trennung des unbekannten Zustands von den bereits bekannten. 87
Abbildung 4-25: Überwachtes Modell mit allen Fehlermechanismen des Hydraulikaggregats nach stufenweiser Erweiterung ausgehend vom Gutzustand. Die Umrandung der Marker (von Gelb zu Rot) beschreibt die Intensität der einzelnen Fehlermechanismen. 87
Abbildung 4-26: Initialisierung der Novelty Detection a) und der Domain Shift Detection b) nach der Detektion eines verringerten Speichervordrucks. 88
Abbildung 4-27: Reaktion der Novelty Detektion a) und der überwachten Klassifizierung b) auf eine erhöhte Öltemperatur. 89
Abbildung 4-28: Reaktion der Domain Shift Detection auf eine erhöhte Öltemperatur. 89
Abbildung 4-29: Reaktion der Novelty Detektion auf einen verringerten Speichervordruck bei T3 und die Abweichung der Arbeitszyklen bei Fehlfunktion von Ventil V11 aufgrund einer zu hohen Öltemperatur. 91
Abbildung 4-30: Vollständig angelerntes Modell zur Überwachung des Hydraulikaggregats (ausgenommen verringerter Vordruck des Speichers bei T3). 92
Abbildung 4-31: Schematische Abbildung einer Euro-X-Schweißzange mit integriertem servopneumatischem Antriebssystem, wie es beispielsweise im Karosserierohbau verwendet wird [13]. 96
Abbildung 4-32: Versuchsaufbau bei Festo zum Aufzeichnen der Daten im Labor [193]. 97
Abbildung 4-33: DoE für die aufgezeichneten Daten der Schweißzange. 98
Abbildung 4-34: a) Histogramm der Novelty Scores der Trainingsdaten im Normalzustand, b) QQ-Diagramm der Novelty Scores der Trainingsdaten im Normalzustand. 100
Abbildung 4-35: a) Reaktion der Novelty Detection bei Auftreten einer Leckage am HZ Kammer 1, b) Ergebnis der überwachten Klassifizierung anhand der ersten Diskriminanzfunktion (DF1) für Training (blau) und Test (grün) auf Basis der Ausgabe der Novelty Detection sowie der Vorhersage des Modells (orange). 100
Abbildung 4-36: a) QQ-Diagramms der Initialisierung der Domain Shift Detection nach Detektion der Leckage HZ Kammer 1, b) zugehörige Domain Shift Scores. 101
Abbildung 4-37: a) QQ-Diagramm der Initialisierung der Novelty Detection nach Detektion der Leckage HZ Kammer 1, b) die entsprechenden Novelty Scores. 102
Abbildung 4-38: a) Reaktion der Novelty Detektion, b) der überwachten Klassifizierung sowie c) der Domain Shift Detection auf die Leckage des HZ Kammer 2. 103

Abbildung 4-39: Überwachtes Modell für die Leckagedetektion beider Kammern des HZ. 104
Abbildung 4-40: a) Reaktion der Novelty Detektion, b) der überwachten Klassifizierung sowie c) der Domain Shift Detection auf ein erhöhtes Lagermoment. 105
Abbildung 4-41: a) Reaktion der Novelty Detektion, b) der überwachten Klassifizierung sowie c) der Domain Shift Detection auf die Leckage des AZ Kammer 1. 106
Abbildung 4-42: Überwachtes Modell für die Leckagen an HZ und AZ, jeweils für Kammer 1 und 2. 107
Abbildung 4-43: a) Reaktion der Novelty Detektion, b) der überwachten Klassifizierung sowie c) der Domain Shift Detection auf die erhöhte Lagerreibung. 108
Abbildung 4-44: Überwachtes Modell der Schweißzange inklusive des Fehlers erhöhte Lagerreibung. Die erhöhte Lagerreibung hebt sich erst in höheren Diskriminanzfunktionen (DF5) von den Normaldaten ab. 109
Abbildung 4-45: a) Reaktion der Novelty Detektion, b) der überwachten Klassifizierung sowie c) der Domain Shift Detection auf den Wechsel des Ventils MPYE. 110
Abbildung 4-46: Visualisierung aller verfügbaren Daten der Schweißzange mittels PCA. Der Ventilwechsel ist deutlich als Domain Shift in den Daten erkennbar. Die Daten mit erhöhter Reibung heben sich erst in höheren Komponenten ab und trennen sich in zwei Cluster, wobei ein Großteil der Daten sehr nah am Cluster der Normaldaten liegt. 112
Abbildung 4-47: Vollständig angelerntes Modell mit allen verfügbaren Fehlern. Die Fehler vor (MPYE 1) und nach (MPYE 2) Wechsel des Ventils werden als separate Klassen behandelt. 113
Abbildung 4-48: Zielvorstellung einer akustischen Leckagedetektion für pneumatische Antriebe. KI-Sensorsystem mit hardwarenaher Signalverarbeitung befestigt an einem pneumatischen Antrieb [196]. 116
Abbildung 4-49: Manipulierte Dichtung an der Kolbenstange. Die Dichtung wird manipuliert und anschließend in den Zylinder eingesetzt. Nach der Montage wird die Leckage bestimmt [196]. 117
Abbildung 4-50: Prüfstand zum Generieren von Daten für die Leckagedetektion an Pneumatikzylindern. Ein oder mehrere Prüflinge können zum Testen montiert und über eine SPS automatisiert betrieben werden. Parallel werden die Daten der verwendeten Sensoren und Steuersignale über einen Datenlogger erfasst (angepasst nach [196]). 118
Abbildung 4-51: Arbeitszyklus des Pneumatikzylinders anhand des Mikrofonsignals von MP23ABS1 (10 kHz) und der zusätzlich verfügbaren Steuersignale der beiden Endschalter und der Ventile zum Belüften der Kammern des doppelwirkenden Zylinders. 120

Abbildung 4-52: Datenvorverarbeitung, um die Mikrofondaten in den einzelnen Ruhephasen der Endlage 2 in ein Szenario-Format zu bringen. 121
Abbildung 4-53: Stack für die Auswertung der Szenarien der einzelnen Zylinder. 122
Abbildung 4-54: a) Modell für Zylinder 4 (DSBC-32-100) auf Basis von IMP23ABSU, b) Frequenzintervalle der für die Modellbildung verwendeten Merkmale. 123
Abbildung 4-55: a) Frequenzbereich, aus dem das höchstkorrelierte Merkmal extrahiert wird, b) Verlauf des höchstkorrelierten Merkmals, berechnet mittels RMS im Bereich von 2500 – 5000 Hz. 123
Abbildung 4-56: DoE für die Leckagedetektion für die untersuchten Baugrößen DSBC-32-100 und DSBC-63-100. 124
Abbildung 4-57: Übertragbarkeit der Modelle zwischen Zylindern anhand des Test RMSE. Das Modell von Zylinder 4 lässt sich vergleichsweise gut auf Zylinder 5 übertragen, wohingegen eine schlechte Übertragbarkeit zu Zylinder 6 gegeben ist. 125
Abbildung 4-58: Ergebnisse der überwachten Modellbildung für a) eine randomisierte Kreuzvalidierung und b) eine LOCO Kreuzvalidierung mit Angabe der Fehler (RMSE) für die einzelnen Zylinder. 126
Abbildung 4-59: Frequenzbereich aus dem eines der höchskorrelierten Merkmale im Bereich zwischen 12 und 16 kHz extrahiert wird. 128
Abbildung 4-60: a) Histogramm der Novelty Scores der Trainingsdaten im Normalzustand, b) QQ-Diagramm der Novelty Scores der Trainingsdaten im Normalzustand. 129
Abbildung 4-61: a) Reaktion der Novelty Detection auf das Auftreten einer Leckage, b) Training des überwachten Modells mit der erkannten Leckage von Zylinder 4. 130
Abbildung 4-62: a) Reaktion der Novelty Detection auf das Auftreten der Leckage von Zylinder 5, b) Vorhersage der Leckage auf Basis des überwachten Modells für Zylinder 5. 131
Abbildung 4-63: a) Reaktion der Novelty Detection auf das Auftreten einer Leckage, b) Vorhersage der Leckage auf Basis des überwachten Modells für Zylinder 3. 131
Abbildung 4-64: Grundlegender Aufbau des Wälzschälprozesses mit relevanten Achsen für translatorische und rotatorische Bewegungen (angepasst nach [203]). 135
Abbildung 4-65: Ausstattung der Wälzschälmaschine mit zusätzlichen Beschleunigungs- und Acoustic Emission Sensoren mittels Klemmringen an den Spindelgehäusen (angepasst nach [203]). 137
Abbildung 4-66: Verwendete Messkoffer zum Aufzeichnen der Daten der Beschleunigungs- und Acoustic Emission Sensoren. 138
Abbildung 4-67: Übersicht über das Messsystem zum Aufzeichnen der Daten von Beschleunigungs- und Acoustic Emission Sensoren sowie der Prozessdaten, die direkt von der Anlage erfasst werden. Die Daten der verschiedenen Datenquellen

werden in separaten Dateien abgelegt und im Nachhinein softwareseitig mit Hilfe der Triggersignale synchronisiert und segmentiert. 139
Abbildung 4-68: Exemplarisches Signal der X-Achse des Beschleunigungssensors an der Werkzeugspindel und der Trigger zum Anzeigen der Start- und Endzeitpunkte der Einzelschnitte. 140
Abbildung 4-69: Vorverarbeitung des Triggersignals und Segmentierung der Sensordaten in einzelne Schnitte..... 141
Abbildung 4-70: Erzeugen eines künstlichen Triggers anhand der Werkzeugposition und die daraus abgeleiteten Start- und Stopppunkte der einzelnen Schnitte. 142
Abbildung 4-71: Zeitsignal der Y-Achse des Beschleunigungssensors an der Werkzeugspindel C1 mit original Abtastrate von 100 kHz (oben), Signal mit Downsampling um Faktor 10 (Mitte), und das Spektrogramm der Originaldaten im Bereich 1 - 5 kHz (unten)..... 143
Abbildung 4-72: a) Spektrogramm des ersten Schruppschnitts der Y-Achse des Beschleunigungssensors an der Werkstückspindel. Bei Werkzeugeingriff ist deutlich die Entstehung der Zahneingriffsfrequenz zu erkennen, b) Zahneingriffsfrequenz mit Seitenbändern des Sensors ACCYC4, c) Zahneingriffsfrequenz mit Seitenbändern des Sensors ACCYC1. 144
Abbildung 4-73: Von der Werkzeugdrehzahl abhängige Zahneingriffsfrequenz. Die Abweichung zwischen berechneter (Soll) und gemessener (Ist) Frequenz wird in der Regelung vermutet, die abhängig von den Regelparametern die Schnittgeschwindigkeit anpasst. 145
Abbildung 4-74: Ordnungsspektren des Sensors ACCYC1 über den gesamten Ordnungsbereich. Im unteren Ordnungsbereich zeigen sich deutlich die Zahneingriffsfrequenz sowie deren erste und zweite Harmonische. Im höheren Ordnungsbereich dominieren Strukturresonanzen das Spektrum. 146
Abbildung 4-75: Zeitsignal des Acoustic Emission Sensors an der Werkzeugspindel C1 und das zugehörige Spektrogramm im Bereich von 100 bis 250 kHz. 147
Abbildung 4-76: Positions- und Stromsignale für einen gesamten Bearbeitungsprozess. 147
Abbildung 4-77: Messaufbau zum Bestimmen des Werkzeugverschleiß nach einer Variante, also 10 gefertigten Innenverzahnungen [203]. 150
Abbildung 4-78: Mikroskop Aufnahmen Zahn oben rechts und Zahn oben links des ersten Werkzeugs, jeweils zu Beginn der Messkampagne und jeweils nach 10 gefertigten Teilen [203]. 153
Abbildung 4-79: Mikroskop Aufnahmen Zahn oben rechts und Zahn oben links des zweiten Werkzeugs, jeweils zu Beginn der Messkampagne und jeweils nach 10 gefertigten Teilen. Am Zahn oben rechts ist nach Variante 3 ein deutlicher Ausbruch am Werkzeug erkennbar [203]. 154

Abbildung 4-80: Stack für die überwachte Modellbildung für Werkzeug 1. 155
Abbildung 4-81: Das Ergebnis einer 10-fachen randomisierten Kreuzvalidierung für Schnitt 2 und Schnitt 5 sowie die Herkunft der für die Modellbildung verwendeten Merkmale für Schnitt 2. 155
Abbildung 4-82: Zeit- und Frequenzbereiche, aus denen die selektierten Merkmale der Beschleunigungssensoren stammen. Die Plots sind exemplarisch anhand des Zeitsignals und Spektrums des Sensors ACCYC1 erstellt. 156
Abbildung 4-83: Zeit- und Frequenzbereich, aus dem das höchstkorrelierte Merkmal mittels RMS extrahiert wird. Es zeigt sich deutlich eine Frequenz bei 360 Hz, deren Amplitude gerade gegen Ende der Messreihe einen deutlichen Anstieg verzeichnet. 157
Abbildung 4-84: Ergebnis einer 6-fachen randomisierten Kreuzvalidierung auf Basis der Daten von Schruppschnitt 2 sowie die Herkunft der für die Modellbildung verwendeten Merkmale bezüglich der Sensoren. 158
Abbildung 4-85: Ergebnis einer Leave-One-Tool-Out Kreuzvalidierung auf Basis der Daten von Schruppschnitt 2 sowie die Herkunft der für die Modellbildung verwendeten Merkmale bezüglich der Sensoren. 159
Abbildung 4-86: Ergebnis einer Leave-One-Tool-Out Kreuzvalidierung auf Basis der Daten von Schruppschnitt 2, wobei für die Merkmalsextraktion der Beschleunigungssensoren die Ordnungsspektren verwendet wurden. Für die restlichen Sensoren bleibt die Merkmalsextraktion unverändert. 160

Tabellenverzeichnis

Tabelle 4-1: Liste der im hydraulischen System installierten Prozesssensoren. 68
Tabelle 4-2: Zusammenfassung der vorhandenen Zustände für die manipulierten Komponenten. 71
Tabelle 4-3: Durch Bayes Optimierung gefundene Hyperparameter für die Modellbildung im Hydraulik Use Case. 73
Tabelle 4-4: Zusammenfassung der Ergebnisse der adaptiven Zustandsüberwachung eines hydraulischen Systems. 90
Tabelle 4-5: Verfügbare Sensoren bzw. Messsysteme für die Überwachung der Schweißzange. 97
Tabelle 4-6: Zusammenfassung der Ergebnisse der adaptiven Zustandsüberwachung einer servopneumatischen Schweißzange. 111
Tabelle 4-7: Jährliche Energieverluste und -kosten durch Leckage an Druckluftanlagen. Basierend auf einem Preis von 0,06 € pro kWh und 8000 Betriebsstunden pro Jahr [195]. 115
Tabelle 4-8: Verfügbare Sensoren für die Leckagedetektion. 118
Tabelle 4-9: Zusammenfassung der Ergebnisse für die Quantifizierung der Leckage für beide Zylindertypen und Validierungsszenarien. 127
Tabelle 4-10: Verfügbare Sensoren zur Verschleißdetektion beim Wälzschälen. 136
Tabelle 4-11: Für die Bearbeitungsgeschwindigkeit, die Produktqualität, die Verschleißentwicklung relevanten Prozessparameter sowie die mit den Prozessexperten von Schaeffler festgelegten Prozessfenster. 148
Tabelle 4-12: DoE für die Datengenerierung von Versuch 1 - 6. Die gelb markierte Variante befindet sich mittig im Parameterraum und ist bei allen Versuchen vorhanden. 149
Tabelle 4-13: Ergebnis der Messkampagnen in Form der verfügbaren Daten für jede Variante sowie die dabei entstandenen Datenmengen. 152
Tabelle 4-14: Parameter der optimierten Stacks sowie die Ergebnisse der 10-fachen randomisierten Kreuzvalidierung für die einzelnen Schnitte von Werkzeug 6 bei Vorgabe eines linearen Verschleißfortschritts. 156
Tabelle 4-15: Parameter der optimierten Stacks sowie die Ergebnisse der 6-fachen randomisierten Kreuzvalidierung für die einzelnen Schnitte aller bei Vorgabe eines linearen Verschleißfortschritts. 158
Tabelle 4-16: Parameter der optimierten Stacks sowie die Ergebnisse der Leave-One-Tool-Out Kreuzvalidierung für die einzelnen Schnitte aller bei Vorgabe eines linearen Verschleißfortschritts. 159

Eigene Veröffentlichungen

Journal Veröffentlichungen

- S. Klein, Y. Wilhelm, A. Schütze, T. Schneider, "Combination of generic novelty detection and supervised classification pipelines for industrial condition monitoring", *tm - Technisches Messen*, vol. 91, no. 9, 2024, pp. 454-465. https://doi.org/10.1515/teme-2024-0016
- C. Schnur, S. Klein, A. Schütze, T. Schneider, A. Blum, „Steigerung der Datenqualität in der Montage“, WT Werkstattstechnik, 112 (2022) NR. 11-12, S. 783 - 787, doi: 10.37544/1436–4980–2022–11–12–57
- A. Blum, Y. Wilhelm, S. Klein, C. Schnur, P. Reimann, R. Müller, A. Schütze, „Ganzheitlicher Ablaufplan für wissensgetriebene Projekte des maschinellen Lernens in der Produktion (Holistic workflow for knowledge-driven machine learning projects in manufacturing)“, tm - Technisches Messen, 89(5), 363-383, doi: 10.1515/teme-2022-0027
- A. Blum, S. Klein, K. Kühn, T. Schneider, A. Schütze, R. Müller, „Inprozess-Dichtheitsprüfung in der Montage“, wt Werkstattstechnik online, 9-2021, S. 650, DOI 10.37544/1436-4980-2021-09-75
- T. Schneider, S. Klein, A. Schütze, „Machine learning in industrial measurement technology for detection of known and unknown faults of equipment and sensors“, tm - Technisches Messen (2019), 86 (11), 706–718, doi: 10.1515/teme-2019-0086

Konferenzbeiträge

- C. Fuchs, S. Klein, J. Fischer, A. Schütze, T. Schneider, „Eine Methode zur erklärbaren Merkmalsextraktion aus dem Zeit- und Frequenzbereich für Condition Monitoring“, 22. GMA/ITG Fachtagung Sensoren und Messsysteme 2024, 11. und 12. Juni 2024, Nürnberg, Tagungsband, S. 556 - 562, doi: 10.5162/sensoren2024/P36
- P. Goodarzi, S. Klein, A. Schütze, T. Schneider, „Comparing Different Feature Extraction Methods in Condition Monitoring Applications“, 22. GMA/ITG Fachtagung Sensoren und Messsysteme 2024, 11. und 12. Juni 2024, Nürnberg, Tagungsband, S. 556 - 562, doi: 10.5162/sensoren2024/P36

- C. Fuchs, S. Klein, P. Goodarzi, A. Schütze, T. Schneider, „Analyse zum Einfluss von Labeling-Fehlern im Kontext von Luftschall- und Vibrationsdatensätzen für maschinelles Lernen“, DAGA 2023 - 49. Jahrestagung für Akustik, Tagungsband, S. 80-83, Sitzung "Akustische Messtechnik und Sensorik 1", Hamburg, 06.-09. März 2023
- C. Schnur, S. Klein, A. Blum, T. Schneider, R. Müller, A. Schütze, „Mess- und Datenplanung für Modelle des maschinellen Lernens an Bestandsanlagen“, 16. Dresdner Sensor-Symposium, Posterbeitrag, Dresden, 5.-7.12.2022
- C. Fuchs, S. Klein, S. Saller, D. Spies, A. Schütze, T. Schneider, „Entwicklung akustischer Messungen für industrielles maschinelles Lernen“, DAGA 2022 - 48. Jahrestagung für Akustik, Vorkolloquium „Künstliche Intelligenz für akustische Sensorsysteme“, 21. - 24. März 2022
- T. Dorst, T. Schneider, S. Klein, S. Eichstädt, A. Schütze, „Influence of synchronization within a sensor system on machine learning results“, oral presentation, MATHMET 2019 International Workshop, Lissabon (Portugal), 20–22 November 2019
- A. Schütze, S. Klein, T. Dorst, T. Schneider, „Sensorik 4.0 – smarte Sensorsysteme ermöglichen Zustandsbewertung, Selbstüberwachung und Prozessoptimierung“, Vortrag, Jahrestreffen der ProcessNet-Fachgemeinschaften "Prozess-, Apparate- und Anlagentechnik" unterstützt durch „Sustainable Production, Energy and Resources", 4.-5. November 2019, Dortmund
- T. Schneider, S. Klein, A. Blum, L. Schirmer, R. Müller, A. Schütze, „Combination of Human and Machine Intelligence to Optimize Assembly“, Societal Automation - Technological & Architectural Frameworks, Krakow (Poland), 4-6 September 2019
- T. Dorst, T. Schneider, S. Klein, S. Eichstädt, A. Schütze, „Synchronisationsprobleme innerhalb eines Sensorsystems und deren Auswirkungen auf Ergebnisse des maschinellen Lernens“, 20. GMA/ITG Fachtagung Sensoren und Messsysteme 2019, Nürnberg, 25. und 26. Juni 2019
- S. Klein, T. Schneider, A. Schütze, „Zustandsüberwachung in der Automatisierungstechnik mittels maschinellem Lernen“, 20. GMA/ITG Fachtagung Sensoren und Messsysteme 2019, Nürnberg, 25. und 26. Juni 2019
- N. Immesberger, S. Klein, M. Jochum, A. Schütze, „Überprüfung und Bewertung des Schriftbildes sowie des Anlagenzustandes industrieller Signieranlagen mit maschinellem Lernen“, 20. GMA/ITG Fachtagung Sensoren und Messsysteme 2019, Nürnberg, 25. und 26. Juni 2019

- T. Schneider, N. Helwig, S. Klein, A. Schütze, „Influence of sensor network sampling rate on multivariate statistical condition monitoring of industrial machines and processes“, EUROSENSORS 2018, poster presentation, September 9-12, 2018, Graz, Austria
- T. Schneider, S. Klein, N. Helwig, A. Schütze, M. Selke, C. Nienhaus, D. Laumann, M. Siegwart, K. Kühn, „Big Data Analytik mit automatisierter Signalverarbeitung für Condition Monitoring“, Sensoren und Messsysteme 2018, 19. ITG/GMA-Fachtagung, Vortrag, Session Sensorik für die Industrie 4.0, 26. - 27. Juni 2018, Nürnberg, D. in: ITG-Fachbericht 281: Sensoren und Messsysteme, VDE-Verlag Berlin (2018), ISBN 978-3-8007-4683-5, S. 259-262
- N. Helwig, S. Klein, A. Schütze, „Identification and Quantification of Hydraulic System Faults based on Multivariate Statistics using Spectral Vibration Features“, EUROSENSORS 2015, XXIX edition of the conference series, poster presentation, Freiburg, Germany, September 6 to 9, 2015

Datensätze

- S. Klein, „Sensor data set, electromechanical cylinder at ZeMA testbed (2kHz)“, dataset published on Zenodo, August 2, 2018
- T. Schneider, S. Klein, M. Bastuck, „Condition monitoring of hydraulic systems Data Set at ZeMA“, dataset published on Zenodo, April 26, 2018
- T. Schneider, S. Klein, M. Bastuck, Condition monitoring of hydraulic systems Data Set, dataset published in UCI machine learning repository, April 26, 2018

Sonstige

- C. Schnur, S. Klein, A. Blum, „Checklist – Measurement and data planning for machine learning in assembly“, Checklist (engl. Version) published at Zenodo, doi: 10.5281/zenodo.7556875
- C. Schnur, S. Klein, A. Blum, „Checkliste – Mess- und Datenplanung für das maschinelle Lernen in der Montage“, Checkliste veröffentlicht bei Zenodo, doi: 10.5281/zenodo.6943476
- S. Klein, T. Schneider, A. Schütze, „Bessere Produkte Dank künstlicher Intelligenz - Wie maschinelles Lernen die Qualität voraussagen kann“, IM + io 12/2019, August-Wilhelm Scheer Institut für digitale Produkte und Prozesse, S. 62-67, 2019, ISSN: 1616-1017

- S. Klein, T. Schneider, A. Schütze, „Automatisierte Zustandsbewertung industrieller Anlagen mittels maschinellem Lernen“, in: Jahresmagazin Mess- und Sensortechnik 2019/2020, Institut für wissenschaftliche Veröffentlichungen (IWV), 2019, S. 72-78

Danksagung

An dieser Stelle möchte ich allen beteiligten Personen meinen Dank aussprechen, die mich bei der Anfertigung meiner Dissertation unterstützt haben.

Allen voran gilt mein herzlicher Dank Prof. Dr. Andreas Schütze, der mir die Möglichkeit zur Promotion gegeben hat und mich über den gesamten Entstehungsprozess dieser Arbeit begleitet und unterstützt hat. Dabei schätzte ich sowohl die hervorragenden Rahmenbedingungen am Lehrstuhl für Messtechnik, die eine angenehme und effektive Arbeit in einem spannenden und innovativen Forschungsfeld ermöglicht haben, als auch die inhaltlichen Diskussionen und Anregungen, die ganz wesentlich zum Gelingen dieser Arbeit beigetragen haben.

Auch Herrn Prof. Dr.-Ing. Dirk Bähre möchte ich an dieser Stelle für das Übernehmen des Zweitgutachtens danken.

Weiterer Dank gilt natürlich meinen Arbeits- und Bürokollegen am Zentrum für Mechatronik und Automatisierungstechnik sowie am Lehrstuhl für Messtechnik. Bei Euch möchte ich mich herzlich für die gute Zusammenarbeit und die schöne gemeinsame Zeit bedanken. Durch den ständigen Austausch aber auch den ein oder anderen Spaß ist die Zeit der Promotion verflogen und ich werde mich immer gerne an diese Zeit zurückerinnern.

Katharina, mein größter Dank gilt Dir. Ohne Deine Unterstützung und Dein Verständnis wäre die Fertigstellung dieser Arbeit neben dem beruflichen Alltag nicht möglich gewesen.

Abschließend danke ich meiner Familie und im besonderen Maße meinen Eltern für ihre immerwährende Unterstützung, bedingungslosen Rückhalt und positiven Zuspruch. Ihr habt diesen Werdegang erst ermöglicht.