
Fusion von Zeitreihen- und Ereignisdaten zur Optimierung der Zustandsdiagnose technischer Systeme

Enhancing the diagnostic performance of Condition Based Maintenance through fusion of event data and sensor data

Zur Erlangung des akademischen Grades Doktor-Ingenieur (Dr.-Ing.)

Genehmigte Dissertation von Henrik Simon

Tag der Einreichung: 28. Juni 2024, Tag der Prüfung: 19. November 2024

1. Gutachten: Prof. Dr.-Ing. Uwe Klingauf

2. Gutachten: Prof. Dr.-Ing. Joachim Metternich
Darmstadt, Technische Universität Darmstadt



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Fachbereich
Maschinenbau

Institut für Flugsysteme
und Regelungstechnik

Fusion von Zeitreihen- und Ereignisdaten zur Optimierung der Zustandsdiagnose technischer Systeme

Enhancing the diagnostic performance of Condition Based Maintenance through fusion of event data and sensor data

Genehmigte Dissertation von Henrik Simon

Tag der Einreichung: 28. Juni 2024

Tag der Prüfung: 19. November 2024

Darmstadt, Technische Universität Darmstadt

Bitte zitieren Sie dieses Dokument als:

URN: urn:nbn:de:tuda-tuprints-288015

URL: <http://tuprints.ulb.tu-darmstadt.de/28801>

Jahr der Veröffentlichung auf TUprints: 2024

Dieses Dokument wird bereitgestellt von tuprints,

E-Publishing-Service der TU Darmstadt

<http://tuprints.ulb.tu-darmstadt.de>

tuprints@ulb.tu-darmstadt.de

Die Veröffentlichung steht unter folgender Creative Commons Lizenz:

Namensnennung – Weitergabe unter gleichen Bedingungen 4.0 International

<https://creativecommons.org/licenses/by-sa/4.0/>

Erklärungen laut Promotionsordnung

§ 8 Abs. 1 lit. c PromO

Ich versichere hiermit, dass die elektronische Version meiner Dissertation mit der schriftlichen Version übereinstimmt.

§ 8 Abs. 1 lit. d PromO

Ich versichere hiermit, dass zu einem vorherigen Zeitpunkt noch keine Promotion versucht wurde. In diesem Fall sind nähere Angaben über Zeitpunkt, Hochschule, Dissertationsthema und Ergebnis dieses Versuchs mitzuteilen.

§ 9 Abs. 1 PromO

Ich versichere hiermit, dass die vorliegende Dissertation selbstständig und nur unter Verwendung der angegebenen Quellen verfasst wurde.

§ 9 Abs. 2 PromO

Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Darmstadt, 28. Juni 2024

Henrik Simon

Kurzfassung

Ein wichtiger Bestandteil der digitalen Transformation der Industrie (Industrie 4.0) ist die zustandsbasierte und vorausschauende Instandhaltung von technischen Systemen. Neben der intelligenten Vernetzung von Maschinen und Abläufen mit Hilfe von Informations- und Kommunikationstechnologien werden dafür mathematische Modelle benötigt, welche die akkurate Diagnose des Systemzustandes und die Prognose der verbleibenden Lebensdauer ermöglichen. Das vorliegende Dissertationsvorhaben setzt in diesem Themengebiet an und erweitert traditionelle zeitreihenbasierte Modellansätze durch die Fusion und Integration von zeit-diskreten Ereignisdaten.

Die aktuelle Forschung in der zustandsbasierten Instandhaltung und Systemdiagnose technischer Systeme, wie bei Flugzeugtriebwerken, Produktionsmaschinen oder Windkraftanlagen beschäftigt sich vor allem mit der Analyse von Zeitreihendaten basierend auf Sensoren. Kenngrößen wie Temperaturen, Drehzahlen oder Vibrationen werden dabei von Modellen verarbeitet, der aktuelle Gesundheitszustand herausgearbeitet und Restlebensdauern von Komponenten werden prognostiziert. Neben Sensordaten fallen bei technischen Systemen jedoch auch erhebliche Mengen an Ereignisdaten, wie Fehlermeldungen oder Wartungslogbücher an. Diese enthalten wertvolle Informationen über den aktuellen Zustand des Systems, werden nach dem aktuellen Stand der Forschung aber allenfalls für die Bestimmung der abhängigen Variablen (Label) und zur Evaluation der Modelle verwendet. In anderen Industriezweigen finden sich vereinzelt Ansätze zur Verarbeitung dieser Ereignisdaten. Die Anwendung im Maschinenbaukontext ist bisher jedoch unerforscht und stellt die Forschungslücke dar.

Zur Untersuchung der identifizierten Forschungslücke wird ein Konzept vorgeschlagen, um Ereignisdaten gewinnbringend in traditionelle Ansätze der Zustandsdiagnose einbinden zu können. Ziel ist es dadurch die Diagnosegenauigkeit zu erhöhen und Unsicherheiten zu reduzieren. Schlüssel zur Integration der Ereignisdaten und Kern des Dissertationsvorhabens ist die entsprechende Vorverarbeitung und zeitliche Anpassung der Ereignisdaten (*time-conversion*). Dabei werden Ansätze von der Überwachung von elektronischem Gerät (wie Festplatten, Geldautomaten oder Medizintechnik) als

Grundlage verwendet und für die Applikation in Maschinenbauanwendungen weiterentwickelt. Der Ansatz wird dabei modular konzipiert und möglichst weit abstrahiert: Das Gesamtkonzept und die Kernansätze zur *time-conversion* werden generalisiert entwickelt und sind somit unabhängig vom Anwendungsfall gültig. Einige Module der Datenvorverarbeitung erfordern jedoch Domänenwissen, sodass die Ereignisdaten vorab entsprechend korrekt kategorisiert und bereinigt werden. Der gesamte Ansatz ist dabei als Erweiterung (Add-On) zu verstehen und soll bei bestehenden zeitreihenbasierten Zustandsdiagnosen, bei entsprechender Verfügbarkeit von Ereignisdaten, als Ergänzung eingebaut werden können.

Im Rahmen der Dissertation wird das vorgeschlagene Konzept und die Methodik am Beispiel von Daten eines Open-Source-Datensatzes eines Windkraftparks evaluiert. Als Vergleichsansatz wird zunächst ein einfaches Zustandsüberwachungs- und Diagnosemodell basierend auf Sensordaten aufgebaut. Die Ereignisdaten werden mithilfe deskriptiver und explorativer Methoden analysiert und kategorisiert. Sodann lassen sich verschiedene Methoden der Datentransformation anwenden um die Ereignisdaten in den Zeitreihenbereich zu erweitern und mit dem Vergleichsansatz zu fusionieren. Der verwendete Datensatz erwies sich dabei als nur bedingt geeignet für die Evaluation des Konzepts. Generell zeigen sowohl der Vergleichsansatz als auch das vorgeschlagene Konzept keine hinreichend guten Ergebnisse für eine Anwendung im Bereich der zustandsbasierten Instandhaltung. Dieser Umstand wird einerseits dem verwendeten Klassifikationsansatz in Zusammenhang mit der hohen Klassenimbalance und unscharfen Labels zugewiesen, andererseits der fehlenden Dokumentation und Domänenwissen um zielgerichtet Merkmale für die Ereignisdaten entwickeln zu können. Die Evaluation der Ergebnisse anhand gängiger Bewertungsmetriken zeigt eine Verbesserung in Präzision und Genauigkeit bei einer Verschlechterung des Recall des vorgeschlagenen Konzepts gegenüber dem Vergleichsansatz. Statistisch signifikante Unterschiede können nicht festgestellt werden, was maßgeblich der hohen Ereignisdatenflut und fehlenden Interpretierbarkeit der Ereignisdaten zugewiesen wird. Eine Sensitivitätsanalyse beleuchtet dabei Einflussparameter auf die Diagnosegüte und ermöglicht die Identifikation von Grenzen des vorgeschlagenen Konzepts. Das abschließende Fazit mit der Thematisierung von Herausforderungen und einem Ausblick auf weitere Forschungspotenziale rundet die Arbeit ab.

Abstract

An important component of the digital transformation of industry (Industry 4.0) is the condition-based and predictive maintenance of technical systems. In addition to the intelligent networking of machines and processes using information and communication technologies, this requires mathematical models that enable an accurate diagnosis of the system status and a prediction of the remaining useful lifetime. The following dissertation project addresses this issue and extends traditional time series-based modeling approaches by fusing and integrating time-discrete event data.

Current research in condition-based maintenance and health diagnostics of technical systems, such as aircraft engines, production machines or wind turbines, is primarily concerned with analyzing time series data based on sensors. Parameters such as temperatures, rotational speeds or vibrations are processed by models, the current state of health is worked out and the remaining useful lifetime of components is predicted. However, in addition to sensor data, technical systems also generate considerable amounts of event data, such as error messages or maintenance logbooks. These contain valuable information about the current state of the system, but according to the current state of research, they are only used to determine the dependent variables (labels) and to evaluate the models. There are few approaches to processing this event data in other branches of industry, however, the application in the context of mechanical engineering has not yet been researched and represents the research gap for this work.

To investigate the identified research gap, a new concept is proposed to integrate event data gainful into traditional approaches to condition diagnosis. The aim is to increase diagnostic accuracy and reduce uncertainties. The key to the integration of event data and the core of the dissertation is the corresponding pre-processing and temporal adaptation of the event data (*time-conversion*). Approaches from the monitoring of electronic devices (such as hard disks, ATMs or medical technology) are used as a basis and further developed for application in mechanical engineering applications. The approach is designed to be generic and abstracted as far as possible: The overall

concept and the core approaches for *time-conversion* are developed in a generalized way and are therefore valid regardless of the actual application. However, some data pre-processing modules require domain knowledge so that the event data is correctly categorized and cleansed in advance. The entire approach is to be understood as an extension (add-on) and should be able to be integrated as a supplement to existing time-series-based diagnostic frameworks, provided that event data is available.

The proposed concept and methodology is evaluated using the example of data from an open source data set of a wind farm. As a reference approach, a simple condition monitoring and diagnostic model based on 10-minute sensor data (SCADA) is first built. The event data is analyzed and categorized using descriptive and explorative methods. Various data transformation methods can then be applied to extend the event data into the time series range and merge it with the reference approach. The given data set proved to be only partially suitable for evaluating the concept. In general, both the comparative approach and the proposed concept do not show sufficiently good results for an application in CBM. This circumstance is attributed to the classification approach used in connection with the high class imbalance and fuzzy labels on the one hand, and the lack of documentation and domain knowledge to be able to develop targeted features for the event data on the other. The evaluation of the results using common evaluation metrics shows an improvement in precision and accuracy with a deterioration in the recall of the proposed approach compared to the comparative approach. Statistically significant differences cannot be determined, which is mainly attributed to the high flood of event data and the lack of interpretability of the event data. A sensitivity analysis assesses the parameters influencing the diagnostic quality and enables a discussion of the limits of the proposed concept. The final conclusion with a discussion of challenges and an outlook on further research potential complete the work.

Danksagung

Die vorliegende Dissertation entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Flugsysteme und Regelungstechnik der TU Darmstadt unter Leitung von Herrn Prof. Dr.-Ing. Uwe Klingauf. Ihm, als meinem Doktorvater, gebührt mein besonderer Dank für die fachliche Unterstützung und Anregungen während der Anfertigung der Dissertation. Darüber hinaus möchte ich mich bei Ihm bedanken für die großen Freiheiten in der Gestaltung meiner Forschung sowie die vielen Möglichkeiten sich am Institut einzubringen und dessen Inhalte und Ausrichtung aktiv mitgestalten zu können. Für die Übernahme des Korreferats und das damit verbundene Interesse an meiner Arbeit danke ich Herrn Prof. Dr.-Ing. Joachim Metternich sehr herzlich.

Bei allen Kolleginnen und Kollegen des FSR möchte ich mich für die einzigartige Zeit, die tolle Atmosphäre und die schönen gemeinsamen Momente innerhalb und außerhalb der Universität bedanken. Ein besonderer Dank gilt dabei den Kollegen aus dem PHM-Team für die vielen freundschaftlichen und fachlichen Gespräche und die schöne Zusammenarbeit: den Kollegen Simon Mehringskötter, Henrik Heier und Lorenz Dingeldein danke ich vor allem für die Unterstützung bei der Themenfindung und -spezifikation, den Kollegen Florian Coors, Immo Schmidt und Max Weigert für die kritische Auseinandersetzung mit diesem Dokument und die ehrlichen Rückmeldungen. Auch bedanke ich mich bei den vielen Studierenden, die durch Abschlussarbeiten und Studienprojekte, zu dieser Dissertation beigetragen haben.

Abschließend möchte ich mich ganz besonders bei meiner Familie und meinen Freunden bedanken, die mir während dem gesamten Studium und der Zeit als wissenschaftlicher Mitarbeiter stets mit einem offenen Ohr zur Seite standen. Meinen Eltern bin ich zutiefst dankbar, dass sie mir diesen Weg ermöglicht und mich immer unterstützt haben. Ein ganz besonderer Dank gilt meiner Frau Jana für die stetige Motivation, Unterstützung, den nötigen Druck und vor allem, dass sie mir in den stressigen Phasen familiär den notwendigen Freiraum geschaffen hat.

Darmstadt, im Juni 2024

Henrik Simon

Inhaltsverzeichnis

Danksagung	ix
Symbol- und Abkürzungsverzeichnis	xv
1. Einleitung	1
1.1. Zielsetzung der Arbeit	3
1.2. Struktur der Arbeit	4
2. Stand der Forschung im Bereich der datenbasierten Schadensdiagnose	5
2.1. Grundlagen der Instandhaltung	5
2.1.1. Ziele der Instandhaltung	7
2.1.2. Unterschiedliche Arten der Instandhaltung	8
2.2. PHM in der Instandhaltung	10
2.2.1. Allgemeine Architektur eines PHM-Systems	11
2.2.2. Datenvorverarbeitung und Merkmalsentwicklung	13
2.2.3. Datenbasierte Schadensdiagnose	17
2.2.4. Metriken zur Evaluation der Fehlerdiagnose	21
2.3. Forschungslücke	24
2.3.1. Unterschiedliche Datentypen und -strukturen als Grundlage	24
2.3.2. Stand der Forschung in der Fusion unterschiedlicher Datenstrukturen	27
2.4. Beitrag dieser Arbeit	28
3. Konzept zur Schadensdiagnose unter zusätzlicher Berücksichtigung von Ereignisdaten	31
3.1. Überblick über das vorgeschlagene Konzept	31
3.1.1. Gesamtstruktur und Zielsetzung	31
3.1.2. Randbedingungen, Annahmen und Anforderungen	33
3.2. Methoden zur Umsetzung des Konzepts	36
3.2.1. Ereignisdaten als zusätzliche Datenquelle	36

3.2.2.	Datenvorverarbeitung der Ereignisdaten	37
3.2.3.	Merkmalsentwicklung der Ereignisdaten	41
3.3.	Fusion von Ereignisdaten und Zeitreihendaten im Diagnoseschritt . . .	48
3.4.	Vorgehen zur Evaluation	50
3.4.1.	Vergleichsansatz	51
3.4.2.	Metriken und Darstellungsform	52
4.	Evaluation des vorgeschlagenen Konzepts am Beispiel von Windkraftanlagen	53
4.1.	Identifikation und Beschreibung der Datengrundlage	53
4.1.1.	Anforderungen an die Datengrundlage	53
4.1.2.	Beschreibung der Datengrundlage	54
4.2.	Datenvorverarbeitung der Ereignisdaten	62
4.3.	Merkmalsentwicklung der Ereignisdaten	69
4.4.	Aufbau des Diagnosemodells	79
4.4.1.	Wahl der Trainings- und Testdatenbasis	80
4.4.2.	Gradient Boosting zur Klassifikation der Schadenszustände . .	81
4.5.	Ergebnisse der Diagnose	83
4.5.1.	Plausibilisierung der Merkmalsentwicklung: Diagnose auf Basis der Ereignisdaten	84
4.5.2.	Vergleichsansatz: Diagnose auf Basis der Sensordaten	87
4.5.3.	Vorgeschlagenes Konzept: Diagnose auf Basis der fusionierten Daten	88
4.6.	Sensitivitätsanalyse	94
4.6.1.	Variation des Ausfallfensters	95
4.6.2.	Variation des Fensters für statistisch basierte Merkmale	98
4.6.3.	Variation des Fensters für sequenzbasierte Merkmale	100
5.	Diskussion des vorgeschlagenen Konzepts und der verwendeten Methodik	103
5.1.	Zusammenfassung der Ergebnisse und Diskussion des Konzepts	103
5.2.	Grenzen des Konzepts	106
6.	Fazit und Ausblick	109
6.1.	Zusammenfassung und Fazit der Arbeit	109
6.2.	Ausblick für zukünftige Arbeiten	111
	Literaturverzeichnis	113
A.	Anhang Kapitel 4	127
A.1.	Ergänzende Informationen zum Datensatz	128

A.2. Ergänzende Informationen zur statistischen Analyse	130
A.3. Ergänzende Informationen zur Hyperparameteroptimierung	132
A.4. Ergänzende Darstellungen zur Sensitivitätsanalyse	136

Symbol- und Abkürzungsverzeichnis

Symbolverzeichnis

Lateinische Buchstaben

Symbol	Einheit	Bezeichnung
<i>E</i>	-	Menge an Ereignissen in einer Datenbank
<i>e</i>	-	Einzelereignis der Menge <i>E</i>
<i>G</i>	-	Statistisch basierte Merkmale
<i>H</i>	-	Sequenzbasierte Merkmale
<i>I</i>	-	Fehlerähnlichkeitsbasierte Merkmale
<i>J</i>	-	Profilbasierte Merkmale
<i>K</i>	-	Menge an Kategorien der Ereignisdaten
<i>k</i>	-	einzelne Kategorie der Ereignisdaten
<i>L</i>	-	Länge einer Ereignis-Episode (Anzahl an Ereignissen)
<i>T</i>	-	Menge an Zeitstempel
<i>t</i>	sec	Zeitstempel (eines Ereignisses bzw. Datenpunkts)
<i>W</i>	-	allg. Beobachtungsfenster zur Bildung von Merkmalen
W	-	Subsequenz, eine Ereignissequenz in einem Fenster
<i>w</i>	sec	Fensterlänge (in Zeit) einer Subsequenz W
<i>X</i>	-	Ausfallereignis in einer Ereignisdatenbank
<i>y</i>	-	Abhängige Variable (Label bzw. Target)
<i>Z</i>	-	Merkmalsraum der Ereignisdaten

Griechische Buchstaben

Symbol	Einheit	Bezeichnung
α	-	Bezeichnung einer seriellen Episode von Ereignissen oder des Signifikanzniveaus für statistische Tests.
Φ	-	parallele Episode
γ	-	komposit Episode an Ereignissen

Abkürzungsverzeichnis

Acc Accuracy

AUC Area Under Curve (in Bezug auf Receiver Operating Characteristic)

BN Bayesian Network

CBM Condition Based Maintenance

CRISP-DM Cross Industry Standard Process for Data Mining

DL Deep Learning

DT Decision Tree

EKF Extended Kalman Filter

FFT Fast Fourier Transformation

FN False Negative

FP False Positive

GB Gradient Boosting

HMM Hidden Markov Model

HsMM Hidden semi-Markov Model

HUMS Health and Usage Monitoring

IFD Intelligent Fault Diagnosis

KI Künstliche Intelligenz

kNN künstliches Neuronale Netz

LSTM Long Short-Term Memory

MAE Mean Absolute Error

MAPE Mean Absolute Percentage Error

MIL Mulit-Instance Learning

ML Machine Learning

MSE Mean Squared Error
mRMR minimum Redundancy Maximum Relevance
NLP Natural Language Processing
OSA-CBM Open System Architecture for Condiditon Based Maintenance
PCA Principal Component Analysis
PdM Predictive Maintenance
PHM Prognostics and Health Management
Prec Precision
Rec Recall
RF Random Forest
ROC Receiver Operating Characteristic
RMSE Root Mean Squared Error
RNN Rekurrentes Neuronales Netz
RUL Remaining Useful Life
STFT Short-time Fourier Transformation
SCADA Supervisory Control and Data Aquisition
SVM Support Vector Machine
SPMF Sequential Pattern Mining Framework
t-SNE t-distributed stochastic neighbor embedding
TN True Negative
TP True Positive
TVKF Time-Varying Kalman Filter
UKF Unscented Kalman Filter
WKA Windkraftanlage

1. Einleitung

Die Ära der Digitalen Transformation der Industrie (Industrie 4.0) ist geprägt durch eine stetig wachsende Vernetzung von Maschinen, der Verschmelzung von IT und physischen Prozessen, sowie zunehmender Komplexität industrieller Anlagen. Dazu erfordert der steigende Wettbewerbsdruck in einer dynamischen Produktionslandschaft höhere Verfügbarkeiten, Zuverlässigkeiten und Auslastungen bestehender Maschinen und Anlagen. Die daraus resultierenden höheren Verschleiß- und Abnutzungserscheinungen erhöhen die Anforderungen an die Instandhaltung und stellen herkömmliche Instandhaltungsansätze vor neue Herausforderungen. Auch neue Geschäftsmodelle im Betreiber-Hersteller Verhältnis, wie Leistungsverträge (engl. Performance-based Contracting), die dem Betreiber eine garantierte Verfügbarkeit zur Nutzung zugesichert werden, motivieren den Hersteller zu einer effizienten Instandhaltung und hohen Zuverlässigkeit. Neben diesen ökonomischen Herausforderungen besteht auch eine wachsende gesellschaftliche und politische Motivation, die Nachhaltigkeit zu steigern. Im Kontext der Instandhaltung bedeutet dies eine effizientere Nutzung der Ressourcen indem vorzeitige Komponentenwechsel vermieden und Bauteillebensdauern ausgenutzt werden. [Str12; Ape18]

All diese Aspekte motivieren eine Verbesserung bestehender Instandhaltungssysteme. Ein Kernaspekt der Digitalen Transformation der Industrie (Industrie 4.0) sind die zustandsbasierte und vorausschauende Instandhaltung (engl. Condition Based Maintenance (CBM) bzw. Predictive Maintenance (PdM)). Neben der intelligenten Vernetzung von Maschinen und Abläufen mit Hilfe von Informations- und Kommunikationstechnologien werden dafür mathematische Modelle benötigt, welche die akkurate Diagnose des Systemzustandes und die Prognose der verbleibenden Lebensdauer ermöglichen. Der Forschungsbereich des Prognostics and Health Management (PHM) widmet sich diesen Herausforderungen und forscht seit rund zwei Jahrzehnten an Methoden und Lösungen um die Verfügbarkeit, Zuverlässigkeit und Nachhaltigkeit von Anlagen zu erhöhen. Die zunehmende Anzahl an Sensoren, Datenmengen und Komplexität der Systeme erschwert die konventionelle physikalische Modellbildung und motiviert den

Einsatz von datenbasierten Methoden aus dem Bereich des Machine Learning (ML) und der Künstlichen Intelligenz (KI). Die dabei entwickelten Lösungen, Ansätze und Methoden werden industrieübergreifend, von der Luftfahrtindustrie über produzierende Unternehmen bis in die Energieerzeugung, eingesetzt. [Sch10]

Maßgebliche Hemmnisse bei der Umsetzung von CBM und PdM Strategien sind neben den hohen Implementierungsaufwänden und fehlenden Know-How im Ingenieurbereich insbesondere mangelnde Daten und zu große Unsicherheiten in der Modellbildung. [Ver22b; Ver22a] Gleichzeitig steigt die Vielfalt und das Volumen bei der Datenerfassung moderner Maschinen stetig an. Während der Fokus der meisten PHM-Ansätze und Methoden auf der Analyse von Zeitreihendaten liegt, wird ein Großteil der erfassten Daten nicht für die Modellbildung verwendet [KS22]. Im Rahmen der vorliegenden Dissertation werden dabei von allem die nicht verwendeten Ereignisdaten betrachtet.

Die Analyse von temporalen, ereignisbasierten Daten und Sequenzen ist in unterschiedlichsten Alltagsanwendungen bereits unausweichlich: Von vorgeschlagenen Beiträgen in sozialen Netzwerken und Kurznachrichtenplattformen wie X über Produktempfehlungen beim Online-Shopping bis hin zur Codevervollständigung beim Programmieren¹.

In der technische Disziplin des PHM und allgemein in der Zustandsdiagnose finden diese Daten bislang abgesehen von der essentiellen Bedeutung in Alarmmeldungen sehr wenig Aufmerksamkeit und sind unzureichend erforscht. Dieses Defizit hängt unter anderem mit der Komplexität der Ereignisdaten zusammen. Die Ereignisse sind textueller Natur, haben mehrere Attribute, stammen von mehreren Geräten, treten unregelmäßig auf und weisen parallele sowie sequentielle Korrelationen auf. Viele Beobachtungen in Ereignisdaten spiegeln dabei typische PdM-Herausforderungen wider: Es sind hochdimensionale Daten und sie weisen räumlich-zeitliche Muster auf. Sie zeigen Variationen sowohl vor dem Ausfall als auch im normalen Betrieb und Ausfälle bilden eine Minderheitenklasse mit extremem Ungleichgewicht zwischen den Klassen. Zuletzt sind die Zeitstempel der Ereignisse sind unregelmäßig. [BKS22]

Diese Herausforderungen und Randbedingungen bilden die Ausgangslage der vorliegenden Dissertation.

¹siehe z. B. Github Copilot.

1.1. Zielsetzung der Arbeit

Im zuvor beschriebenen Themengebiet setzt die vorliegende Dissertation an und erweitert traditionelle zeitreihenbasierte Modellansätze durch die Fusion und Integration von zeit-diskreten Ereignisdaten. Zielsetzung der Arbeit ist die Untersuchung und Bewertung der Potenziale dieser Ereignisdaten im Kontext der Zustandsüberwachung und vorausschauenden Wartung. Der Schwerpunkt liegt dabei auf der Entwicklung von Vorverarbeitungsstrategien zur Fusion unterschiedlicher Datenstrukturen, namentlich Zeitreihendaten und temporale Ereignisdaten. Dafür werden bekannte Methoden aufgegriffen, adaptiert und weiterentwickelt.

Ausgangsbasis für die Arbeit ist der aktuelle Stand der Forschung, also bestehende PHM-Diagnosesysteme auf Basis von Zeitreihendaten (Sensoraten), wie sie bei Flugzeugtriebwerken, Produktionsmaschinen oder Windkraftanlagen anzufinden sind. Kenngrößen wie Temperaturen, Drehzahlen oder Vibrationen werden dabei von Modellen verarbeitet, der aktuelle Gesundheitszustand herausgearbeitet und Restlebensdauern von Komponenten prognostiziert.

Ereignisdaten, wie Fehlernachrichten oder Wartungslogbücher, enthalten wertvolle Informationen über den aktuellen Zustand des Systems, werden nach dem aktuellen Stand der Forschung aber allenfalls für die Bestimmung der abhängigen Variablen (Label) und zur Evaluation der Modelle verwendet. In anderen Industriezweigen finden sich vereinzelt Ansätze zur Verarbeitung dieser Ereignisdaten. Die Anwendung im Maschinenbaukontext ist bisher jedoch unerforscht und stellt die Forschungslücke dar.

Im Rahmen der Arbeit wird ein Konzept vorgeschlagen, um Ereignisdaten gewinnbringend in traditionelle Ansätze der Zustandsdiagnose einbinden zu können. Ziel ist es dadurch die Diagnosegenauigkeit zu erhöhen und Unsicherheiten zu reduzieren. Schlüssel zur Integration der Ereignisdaten und Kern des Dissertationsvorhabens ist die entsprechende Vorverarbeitung und zeitlichen Anpassung der Ereignisdaten (*time conversion*). Dabei werden Ansätze von der Überwachung von elektronischem Gerät (wie Festplatten, Geldautomaten oder Medizintechnik) als Grundlage verwendet und für die Applikation in Maschinenbauanwendungen weiterentwickelt. Der Ansatz wird dabei modular konzipiert und möglichst weit abstrahiert: Das Gesamtkonzept und die Kernansätze zur *time-conversion* werden generalisiert entwickelt und sind somit unabhängig des Anwendungsfalls gültig. Einige Module der Datenvorverarbeitung erfordern jedoch Domänenwissen, sodass die Ereignisdaten vorab entsprechend korrekt

kategorisiert und bereinigt werden. Der gesamte Ansatz ist dabei als Erweiterung (Add-On) zu verstehen und soll bei bestehenden zeitreihenbasierten Zustandsdiagnosen, bei entsprechender Verfügbarkeit von Ereignisdaten, als Ergänzung eingebaut werden können.

Insgesamt fokussiert das Vorhaben auf den Diagnoseschritt und schafft damit eine solidere Basis für die Umsetzung von CBM. Die Prognose des Gesundheitszustands in die Zukunft und damit die Ermöglichung von PdM ist explizit nicht Teil des Konzepts.

1.2. Struktur der Arbeit

Die Arbeit ist in sechs Kapitel strukturiert. Nach der Einführung in die Thematik und Darstellung der Zielsetzung wird in Kapitel 2 der aktuelle Stand der Forschung im Bereich der datenbasierten Schadensdiagnose beschrieben. Dabei werden sowohl die Grundlagen der Instandhaltung und des PHM eingeführt, als auch die gängigen Ansätze und Methoden dargestellt. Ausgehend von der Forschungslücke wird in Kapitel 2 ebenfalls der Beitrag der Arbeit beschrieben. In Kapitel 3 wird das zentrale Konzept zur Schadensdiagnose unter der Berücksichtigung von Ereignisdaten vorgestellt. Die Methodik wird in diesem Kapitel allgemeingültig formuliert und das Vorgehen zur Evaluation beschrieben. In Kapitel 4 wird das vorgeschlagene Konzept am Beispiel von Windkraftanlagen evaluiert. Ausgehend von der Darstellung der Datengrundlage wird darin die Datenvorverarbeitung und Merkmalsentwicklung am spezifischen Anwendungsfall dargestellt und die Ergebnisse präsentiert sowie mit dem Vergleichsansatz gegenübergestellt. Das Kapitel schließt mit einer Sensitivitätsanalyse der Variationsparameter bei der Erstellung der Merkmale. Kapitel 5 diskutiert das Konzept anhand der zusammengefassten Ergebnissen. Dabei werden weiterhin die Grenzen des Konzepts aufgezeigt. Die Arbeit schließt mit einer Zusammenfassung und einem Ausblick für künftige Arbeiten in Kapitel 6.

2. Stand der Forschung im Bereich der datenbasierten Schadensdiagnose

Das folgende Kapitel gibt einen Überblick über den aktuellen Stand der Forschung im Bereich der Systemdiagnose und des Prognostics and Health Managements (PHM). Hierfür wird zunächst eine Einleitung in den Bereich der technischen Wartung und Instandhaltung mit Beschreibung der verschiedenen Instandhaltungsarten sowie den grundlegenden Begriffen und Definitionen gegeben. Ausgehend davon werden klassische Diagnoseverfahren und aktuelle Trends aus der Forschung sowie dazugehörige Bewertungsmetriken präsentiert. Im speziellen wird dabei auf die Anforderungen hinsichtlich der Datengrundlage eingegangen. Abschließend wird auf die Besonderheiten der Vorverarbeitung von ereignisbasierten Daten eingegangen und die daraus resultierende Forschungslücke abgeleitet.

2.1. Grundlagen der Instandhaltung

Die Begriffe der Wartung und Instandhaltung werden häufig vermischt und das Verständnis ist nicht immer einheitlich definiert. Daher wird in diesem Abschnitt eine Einführung in den Bereich der Wartung und Instandhaltung gegeben und die in dieser Arbeit verwendeten Begriffe definiert.

Instandhaltung ist nach [DIN19] als

„Kombination aller technischen und administrativen Maßnahmen sowie Maßnahmen des Managements während des Lebenszyklus eines Objekts, die dem Erhalt oder der Wiederherstellung ihres funktionsfähigen Zustands dient, sodass es die geforderte Funktion erfüllen kann“

definiert. Die Instandhaltung wird nach [DIN19] weiterhin in die Grundmaßnahmen Wartung, Inspektion, Instandsetzung und Verbesserung untergliedert (vgl. Abbildung 2.1).

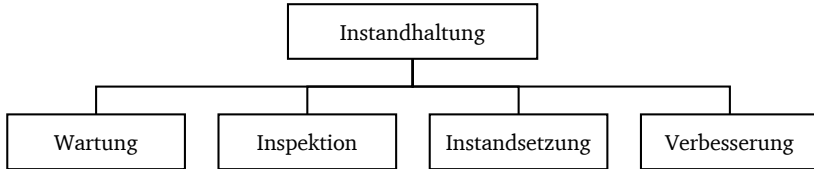


Abbildung 2.1.: Untergliederung der Instandhaltung nach [DIN19]

Die Wartung beschreibt dabei alle Maßnahmen zur Verzögerung der Abnahme des so genannten *Abnutzungsvorrats*. Der Abnutzungsvorrat ist dabei definiert als der Vorrat zur Funktionserfüllung unter bestimmten Bedingungen. Abbildung 2.2 stellt den Abnutzungsvorrat über die Zeit dar. Die Inspektion ist in der Instandhaltung die Prüfung der maßgeblichen Merkmale eines Objekts auf Konformität. Sie kann durch Messung, Beobachtung oder Funktionsprüfung vollzogen werden. Die Instandsetzung beschreibt die durchgeführte, physische Maßnahme zur Wiederherstellung der Funktion eines defekten Objekts. Die Verbesserung fasst alle technischen, administrativen und Managementmaßnahmen zusammen, die zur Steigerung der Zuverlässigkeit, Sicherheit oder Instandhaltbarkeit beitragen.

In Abbildung 2.2 sind die entsprechenden Ereignisse der Instandhaltung in Abhängigkeit einer idealisierten Abnutzungskurve dargestellt. Ausgehend vom Neuzustand wird der Abnutzungsvorrat durch chemische und/oder physikalische Prozesse abgebaut. Ohne weitere Eingriffe wird der Abnutzungsvorrat bis zur Ausfallgrenze aufgebraucht und das System ist nicht mehr funktionsfähig. Folge ist eine korrektive Instandsetzung nach dem Ausfall des Systems mit erhöhtem Zeitaufwand. Die Inspektion ist eine Prüfung des Zustands, die jedoch keinen Einfluss auf den Verlauf der Abnutzungskurve hat. Die Wartung ist hingegen ein Ereignis, das die Abnutzung verzögern kann. Alternativ kann die präventive Instandhaltungsmaßnahme angewendet werden. Ausführlichere Beschreibungen der Instandhaltungsarten finden sich in Abschnitt 2.1.2.

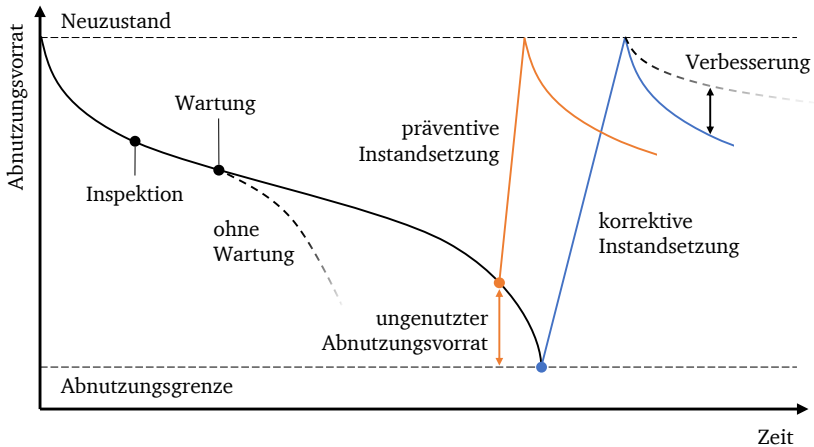


Abbildung 2.2.: Darstellung des Abnutzungsvorrats über die Zeit in Anlehnung an [DIN19; Ape18]

2.1.1. Ziele der Instandhaltung

Das primäre Ziel der Instandhaltung ist die Verzögerung der Abnutzungsgeschwindigkeit, bzw. die Verhinderung von Ausfällen und Schäden an Objekten [Str12]. Darunter werden die Ziele und gleichermaßen Aufgaben der Instandhaltung nach Leidinger wie folgt beschrieben:

- Erhöhung der Sicherheit
- Erhöhung der Verfügbarkeit
- Erhöhung der Zuverlässigkeit
- Der Wert der Anlage/ des Systems soll erhalten bleiben

Häufig können dabei die Ziele nicht gleichermaßen optimiert werden, und je nach Anwendung erfolgt eine Priorisierung.

2.1.2. Unterschiedliche Arten der Instandhaltung

Die Instandhaltungsart beschreibt die grundsätzliche Herangehensweise an das Instandhaltungsproblem und legt fest, mit welcher Methodik, in welcher zeitlichen Abfolge und in welchem Umfang die zuvor definierten Instandhaltungsmaßnahmen durchgeführt werden.

Die grundlegende Unterteilung der Instandhaltungsarten ist in Abbildung 2.3 dargestellt. Zunächst wird dabei zwischen korrektiver und präventiver Instandhaltung unterschieden.

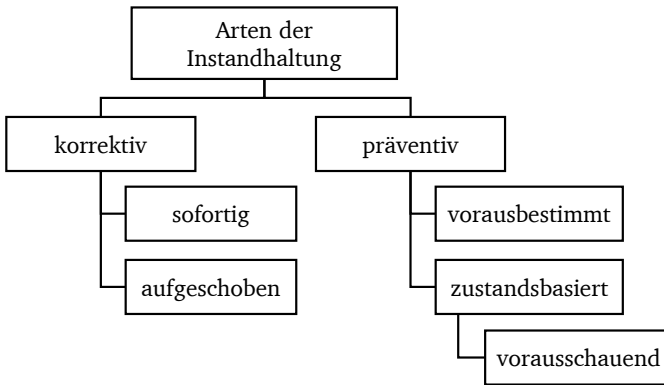


Abbildung 2.3.: Instandhaltungsarten in Anlehnung an [DIN18]

Die korrektive (oft auch reaktive) Instandhaltung ist der einfachste und älteste Ansatz. Eine Instandhaltung wird erst dann durchgeführt, wenn ein Objekt seine Funktion nicht mehr erfüllen kann - der daraus resultierende Ausfall des Objekts wird also bewusst in Kauf genommen. Vorteile sind hierbei die Ausnutzung der maximalen Lebensdauer und somit die Vermeidung ungenutzten Abnutzungsvorrats sowie der geringe Planungsaufwand. Nachteile sind jedoch der möglicherweise höhere Aufwand bei der Beschaffung notwendiger Ersatzteile und potenziell längere Stillstandszeiten, da Ausfälle unvorhergesehen auftreten. Diese Art der Instandhaltung kann insbesondere bei schnell verfügbaren Ersatzteilen, nicht sicherheitskritischen Objekten oder bei redundanten Systemen sinnvoll sein. Ein alltägliches Beispiel ist der Wechsel von Leuchtmitteln im Haushalt. Die korrektive Instandhaltung kann weiter unterteilt werden in sofortige und aufgeschobene Maßnahmen. [Lei14]

Ziel der vorbeugenden Instandhaltung ist die Vermeidung von Ausfällen und damit die Instandsetzung vor Erreichen der Abnutzungsgrenze. Der Vorteil liegt in der Vermeidung von unerwarteten Ausfällen und teuren Stillstandszeiten sowie einer insgesamt höheren Zuverlässigkeit und Verfügbarkeit. Entsprechende Instandhaltungsmaßnahmen sind besser planbar und können gut in den operativen Betrieb integriert werden. Diese Art der Instandhaltung wird insbesondere bei kostenintensiven Objekten oder sicherheitskritischen Systemen, wie z. B. in der Luftfahrt, eingesetzt.

Um den Zeitpunkt einer präventiven Maßnahme zu bestimmen, wird weiter zwischen der vorausbestimmten und zustandsbasierten Instandhaltung unterschieden. Bei der vorausbestimmten Instandhaltung werden auf Basis von Erfahrungswerten feste Intervalle definiert, nach denen entsprechende Maßnahmen durchgeführt werden, z. B. Zeitintervalle oder Nutzungszyklen [Ape18]. Wichtig bei der vorausbestimmten Instandhaltung ist die richtige Wahl des Intervalls: Wird die Maßnahme zu früh geplant, bleibt ein ungenutzter Abnutzungsvorrat bestehen. Wird die Maßnahmen jedoch zu spät geplant, besteht das Risiko eines ungeplanten Ausfalls. [Eic+13; DBT19]

Um diesen Nachteilen entgegenzuwirken und gleichzeitig die Vorteile der präventiven Instandhaltung zu nutzen, ist die zustandsbasierte Instandhaltung (engl. Condition Based Maintenance (CBM)) entstanden. Diese kann weiter unterschieden werden in die vorausschauende Instandhaltung, auch prädiktive Instandhaltung (engl. Predictive Maintenance (PdM)) und nicht vorausschauende Instandhaltung (CBM) [DIN18]. Bei der nicht prädiktiven Instandhaltung werden Maßnahmen auf Basis einer Abschätzung des *aktuellen* Objektzustands ergriffen. Unter dem Begriff CBM wird eine Vielzahl von Ansätzen und teilweise auch Instandhaltungsstrategien zusammengefasst, mit denen der aktuelle (Gesundheits-) Zustand bestimmt werden kann. Regelmäßige Inspektionen und genaue Modelle bilden dabei die Grundlage um den Zustand abzubilden und den verbleibenden Abnutzungsvorrat zu quantifizieren. [Ape18]

Die konsequente Weiterentwicklung der CBM ist die zustandsbasierte vorausschauende Instandhaltung, besser bekannt als PdM. Dabei wird der zu erwartende Zustand des Systems in die Zukunft prognostiziert. Das Ergebnis ist eine erwartete Restlebensdauer (engl. Remaining Useful Life (RUL)). [Sch10]. Kernelemente von CBM und PdM sind entsprechende Algorithmen und Modelle. Für deren Entwicklung hat sich in den letzten Jahrzehnten die Forschungsdisziplin Prognostics and Health Management (PHM) herausgebildet, die in Abschnitt 2.2 näher beschrieben wird.

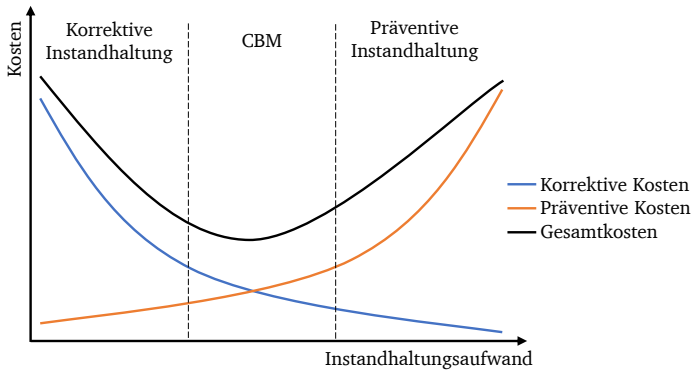


Abbildung 2.4.: Schematische Darstellung des Optimierungsproblems bei der Wahl der Instandhaltungsart nach [KAC17]

2.2. PHM in der Instandhaltung

Grundlage für die Umsetzung von zustandsbasierter Instandhaltung sowie prädiktiver Instandhaltung ist die Kenntnis des aktuellen Objektzustands. PHM Methoden bieten hier wesentliche Werkzeuge, um das in Abbildung 2.4 dargestellte Kostenoptimum zu finden. Zu diesem Zweck wird in diesem Abschnitt eine Einführung in die wesentlichen Aspekte von Diagnose- und Prognoseverfahren für technische Systeme gegeben.

Die Disziplin des PHM kann als Methodik oder technischer Ansatz verstanden werden, um CBM und PdM zu ermöglichen. In der Praxis werden die Begriffe manchmal auch als Synonyme betrachtet und häufig miteinander vermischt. Im Rahmen der Dissertation wird unter PHM die Entwicklung von Modellen und Algorithmen verstanden, die den aktuellen Gesundheitszustand eines Objekts (in Echtzeit) in Abhängigkeit der Betriebsbedingungen abbilden und in die Zukunft prognostizieren. Die Disziplin des PHM hat ihren Ursprung in der Entwicklung des Health and Usage Monitoring (HUMS), das Ende der 1980er Jahre zur Überwachung von Hubschraubergetrieben in der Luftfahrtindustrie eingeführt wurde [KAC17]. Ursprünglich mit dem Ziel entwickelt, die Sicherheit von Hubschraubern zu erhöhen, finden sich heute Anwendungen in Fertigungslinien, dem Energiesektor und vielen anderen Industriezweigen. Heute verfolgt PHM in der Instandhaltung vor allem wirtschaftliche Ziele mit der Umsetzung von

CBM und damit der Kostenoptimierung zwischen korrektiver und präventiver Instandhaltung. [Lan01; Zak+06] Der Begriff vereint dabei mehrere Forschungsdisziplinen, unter anderem Sensorik, *Physics of Failure*, Machine Learning (ML), Statistik und *Reliability Engineering*. Die Grundidee besteht darin, aus den Betriebsdaten eines Objekts wertvolle Informationen über seinen Gesundheitszustand abzuleiten. [KAC17]

2.2.1. Allgemeine Architektur eines PHM-Systems

Um einen Rahmen für ein einheitliches Vorgehen bei der Entwicklung und Implementierung von PHM-Systemen zu schaffen, werden verschiedene Prozessmodelle genutzt. Die beiden bekanntesten sind der Cross Industry Standard Process for Data Mining (CRISP-DM) und der Open System Architecture for Condition Based Maintenance (OSA-CBM), von denen es je nach Anwendung weitere Varianten gibt [Hub+19; VWD14]. Der CRISP-DM zielt generell auf datengetriebene Geschäftsmodelle ab. Die Daten stehen im Zentrum der iterativen Entwicklung und sowohl das Business Understanding, als auch das abschließende Deployment der Modelle sind Teil des Gesamtprozesses. Der Kern der Entwicklung deckt sich mit dem auf technische Anwendungen fokussierten OSA-CBM. Letzterer wurde explizit für die Anwendung in der Instandhaltung entwickelt und bildet die Grundlage für die Entwicklung von PHM-Systemen in den verschiedensten Industriezweigen.

Die ISO 13374 beschreibt den grundsätzlichen Aufbau eines PHM-Systems und die Vorgehensweise bei der Implementierung in sechs Schritten. Die Schritte laufen sequentiell ab und sind in Abbildung 2.5 dargestellt. Das Prozessmodell des OSA-CBM ergänzt das Modell um den siebten Schritt der Visualisierung (ebenfalls in der Abbildung enthalten). Da das OSA-CBM-Modell die Vorgehensweise für die weitere Arbeit beschreibt, werden die einzelnen Schritte im Folgenden kurz erläutert. [CLS06; Int03]

Das Prozessmodell beginnt mit der physikalischen Ebene der Datenerfassung. Diese kann in Sensorik und Messhardware unterteilt werden. Mit Hilfe der Sensorik werden objektive Messgrößen am beobachteten Objekt erfasst und gespeichert. Je nach Art der Problemstellung ist die physikalische Datenerfassung nicht Bestandteil der Entwicklung eines PHM-Systems oder kann nicht beeinflusst werden. In der vorliegenden Arbeit werden keine Daten erfasst, sondern der vorgeschlagene Ansatz wird anhand eines Anwendungsbeispiels aus dem Energiesektor evaluiert (vgl. Kapitel 4). In diesem Fall stellt die Identifikation und Aufbereitung einer geeigneten Open-Source-Datenbasis den Schritt der Datenaquisition dar.



Abbildung 2.5.: OSA-CBM Prozessmodell in Anlehnung an [Int03]

Im nächsten Schritt werden die erfassten Daten im Rahmen der Datenaufbereitung für die weitere Analyse vorbereitet. Hierzu werden vorrangig Filtermethoden zur Erhöhung des Signal-Rausch-Verhältnisses, zur Eliminierung von Ausreißern oder zur Synchronisation der Sensordaten eingesetzt. Datenpunkte und Variablen mit geringem Informationsgehalt in Bezug auf die Zielgröße werden ebenfalls entfernt. Ziel dieses Prozessschrittes ist es, eine sinnvolle Vorauswahl der Daten zu treffen (Selektion) und die Qualität der Daten für die Zielanwendung zu erhöhen. Der Prozessschritt der Datenvorbereitung steht in engem Zusammenhang mit der Zustandsdetektion.

Ziel der Zustandsdetektion ist die Generierung einer aussagekräftigen Datenbasis, auf deren Grundlage im nächsten Schritt die Diagnose des Gesundheitszustands erfolgt. In diesem Schritt werden die zuvor aufbereiteten Daten in Informationen überführt. Wesentlich hierfür sind die Ansätze der Merkmalsentwicklung (engl. Feature Engineering). Der Begriff vereint zum einen die Merkmalsextraktion und -generierung (engl. Feature Extraction and Generation) und zum anderen die Merkmalsreduktion und -selektion (engl. Feature Reduction and Selection) [Agg15; Guy08]. Für die Merkmalsentwicklung können bereits Ansätze aus dem Bereich des ML verwendet werden. Da die Datenvorbereitung sowie die Zustandsdetektion den Kern der vorliegenden Dissertation bilden, wird dieser Themenkomplex in den folgenden Abschnitten ausführlicher behandelt (siehe Abschnitt 2.2.2 und Abschnitt 2.2.3).

Im Schritt der Zustandsbewertung erfolgt eine erste Abstraktion der generierten Merkmale, um eine quantifizierte Aussage über den Gesundheitszustand treffen zu können. Damit wird die Grundlage für CBM gelegt. Als Synonym wird häufig auch der Begriff Diagnose verwendet. Generell lassen sich die dazu notwendigen Modelle und Ansätze in modellbasierte (oder physikalische) Ansätze, datenbasierte Ansätze und hybride Ansätze unterteilen, die in Abschnitt 2.2.3 näher beschrieben werden. [Lee+14]

Ausgehend von der quantifizierten Bewertung des aktuellen Ist-Zustands des Objekts wird im Schritt der Prognose die weitere Entwicklung des Abnutzungsvorrats bzw. der Restlebensdauer RUL in die Zukunft prognostiziert. Der Schritt der Prognose ist auch Schlüssel zur Implementierung von PdM.

Im letzten Schritt des Prozessmodells wird schließlich eine Handlungsempfehlung abgeleitet. Dazu werden die Ergebnisse der vorangegangenen Schritte zusammengefasst und so aufbereitet, dass daraus operative Handlungsempfehlungen abgeleitet werden können. So können beispielsweise Wartungs- und Instandhaltungsmaßnahmen frühzeitig in den operativen Betrieb integriert, Ressourcen rechtzeitig disponiert oder die operativen Betriebsbedingungen so angepasst werden, dass die nächste Routinemaßnahme erreicht werden kann.

2.2.2. Datenvorverarbeitung und Merkmalsentwicklung

Im Rahmen der Datenvorverarbeitung werden die noch unverarbeiteten Rohdaten zunächst auf Plausibilität geprüft, ergänzt und bereinigt. Die Datenvorverarbeitung spielt damit eine essentielle Bedeutung in der Datenverarbeitung und kann als Qualitätskontrolle verstanden werden. Dazu gehört ebenfalls die korrekte Formatierung der Daten, sodass IDs beispielsweise als ganze Zahlen definiert werden, numerische Werte nicht als Text sowie Zeitstempel in adäquaten *Datetime*-Formaten. Auch das Angleichen verschiedener Datenquellen mit unterschiedlichen Zeitstempeln kann Teil einer Vorverarbeitung sein. Daneben beschäftigt sich die Vorverarbeitung mit dem Filtern (Bereinigen) und der Detektion von Ausreißern. In der Literatur gibt es eine Vielzahl von Methoden und Ansätzen, die sich mit Methoden hierfür beschäftigen. Die Methoden sind untereinander kombinierbar und müssen grundsätzlich auf die vorliegende Problemstellung angepasst und zugeschnitten werden, so dass keine allgemeingültige Vorgehensweise festgelegt werden kann. Umfassende Übersichten zu Methoden zur Bereinigung von Daten finden sich unter anderem in Meyer; Chandola et al.; Aggarwal.

Im Rahmen der Merkmalsentwicklung werden die vorhandenen (bzw. erfassten) Daten in aussagekräftige Merkmale überführt. Dabei bestehen direkte Abhängigkeiten zwischen der Wahl der Methoden zu Merkmalsentwicklung und der nachgelagerten Schadensdiagnose oder allgemein der verwendeten Modellierungsmethodik. Entsprechend müssen die Methoden auf die Zielsetzung der Problematik angepasst werden und erhöhen dann direkt die Aussagekraft der Daten. Allgemein erfolgt dies in zwei Schritten: Im ersten Schritt werden im Rahmen der Merkmalsextraktion und -generierung Merkmale generiert, die einen direkten Bezug zum Gesundheitszustand des Systems haben. Im zweiten Schritt erfolgt die Merkmalsreduktion und -selektion, bei der redundante Merkmale herausgefiltert werden und nur noch die für die Diagnose aussagekräftigsten Merkmale übrig bleiben. Beide Schritte werden im Folgenden näher erläutert.

Merkmalsextraktion und -generierung

Die Merkmalsextraktion und -generierung kann prinzipiell nach der *Art* der Daten und bei technischen Anwendungen nach der *Beschreibungsform* der Merkmalsextraktion (Zeitbereich, Frequenzbereich, ...) unterteilt werden. Die Art der Daten kann bei der übergeordneten Betrachtung der Daten für ML vielfältig sein und ist nicht einheitlich definiert: Aggarwal unterscheidet zwischen unabhängig orientierten Daten (z. B. unstrukturierte Textdaten) und abhängig orientierten Daten (z. B. Bilddaten oder Zeitreihendaten). Weitere Einordnungen verschiedener Datenstrukturen in Bezug auf Anwendungen der Anomaliedetektion finden sich in Chandola et al. und Gupta et al. Relevant im Anwendungsbereich des CBM und im Fokus dieser Arbeit sind zunächst Zeitreihendaten, die nach Jardine et al. in drei Kategorien unterteilt werden:

- einzelne Messwerte zu einem bestimmten Zeitpunkt (bspw. Temperatursignale oder eine elektrische Leistung),
- wellenförmige Messwerte in einer bestimmten Zeitspanne (bspw. Vibrationsdaten) und
- mehrdimensionale Messwerte (bspw. Bilddaten).

Von besonderem Interesse für CBM sind einzelne und wellenförmige Messwerte. Die Ausprägung der Werte kann dabei binär, diskret oder kontinuierlich sein [Guy08]. Auch Transformationen zwischen den einzelnen Kategorien sind möglich: Bei Anwendungen von rekurrenten Neuronalen Netzen (RNN) werden z. B. Ausgangsdaten aus einzelnen oder wellenförmigen Messwerten als Bilder in den mehrdimensionalen Bereich transformiert.

Für die Überführung in Merkmale existieren unterschiedliche Ansätze. In Abbildung 2.6 ist eine Auswahl der dafür gängigen Methoden nach Weizhong Yan et al. und Lei et al. in Abhängigkeit der Beschreibungsform dargestellt. Eine aufgezeichnete Messgröße kann allerdings auch direkt als Merkmal verwendet werden. Insbesondere bei der Anwendung von künstlichen Neuronalen Netzen (kNN) wird kontrovers diskutiert, ob die Verwendung von Rohdaten nicht vorteilhaft ist. Gerade im technischen Kontext des CBM und zur Verbesserung der Nachvollziehbarkeit der Ergebnisse wird eine Extraktion von Merkmalen jedoch bevorzugt. [WHN08; Lei+20]

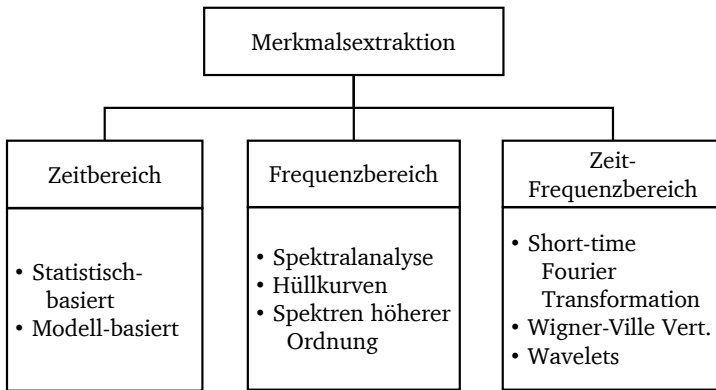


Abbildung 2.6.: Verfahren der Merkmalsextraktion in Anlehnung an [WHN08] und [Lei+20]

Bei der Merkmalsextraktion im Zeitbereich kann weiter zwischen statistisch basierter und modellbasierter Merkmalsextraktion unterschieden werden. Als deskriptive statistische Merkmale können neben den statistischen Momenten Mittelwert, Standardabweichung, Schiefe und Kurtosis auch Parameter zur Bestimmung des Energieinhalts (z. B. Root Mean Square (RMS)) oder der Stärke (z. B. Minimal- und Maximalwerte) gebildet werden. Diese Merkmale können sowohl aus den Rohdaten, als auch aus den bereits gefilterten Daten gebildet werden. Bei der modellbasierten Merkmalsbildung kann ein parametrisierbares Zeitreihenmodell, z. B. ein physikalisches Modell, angepasst werden. Neben der direkten Ableitung von Merkmalen aus den physikalischen Größen, ist es auch üblich Messwerte oder Merkmale so zu kombinieren, dass kein direkter Bezug zur physikalischen Messgröße besteht. Merkmale aus dem Zeitbereich eignen sich gut zur Fehlerdetektion, weniger zur Fehlerisolierung. [WHN08]

Im Frequenzbereich können hingegen Merkmale extrahiert werden, die für die Fehlerisolierung besser geeignet sind. Die am weitesten verbreitete Methode hierfür ist die Spektralanalyse, die mithilfe der Fourier-Transformation (engl. Fast Fourier Transformation (FFT)) vorgenommen wird. Dabei wird typischerweise nicht das gesamte Spektrum betrachtet, sondern nur diejenigen Frequenzanteile, die für den zu erwartenden Fehler ausschlaggebend sind. Analysen im Frequenzbereich sind im Bereich der Fehlerisolierung der Standard moderner technischer Systeme, wie z. B. Windkraftanlage (WKA) [Rom+16]. Neben der Spektralanalyse kommen hierbei auch Hüllkurven oder Spektren höherer Ordnung (z. B. die Spektralanalyse statistischer Momente) zum Einsatz. [WHN08]

Die Methoden aus dem Zeit-Frequenzbereich eignen sich vor allem für nicht-stationäre Messreihen, bei denen wechselnde Betriebsbedingungen Veränderungen in den gemessenen Daten hervorrufen. Dabei werden die Daten simultan sowohl im Zeitbereich, als auch im Frequenzbereich analysiert. Die meist verwendete Methode hierfür ist die Short-time Fourier Transformation (STFT), bei der die FFT für kurze Zeitabschnitte berechnet wird. Weitere Methoden sind die Wigner-Ville Verteilung und Wavelets. [WHN08]

Merkmalsreduktion und -selektion

Das Resultat der zuvor beschriebenen Verfahren zur Merkmalsextraktion bzw. -generierung ist eine Vielzahl an Merkmalen. So gibt es bspw. Python-Bibliotheken zur Generierung von Merkmalen, die aus einer gemessenen Zeitreihengröße bis zu 200 Merkmale erstellen¹. Aber auch eine FFT erzeugt bei ausreichend hoher Abtastrate mehrere Tausend Merkmale. Viele der Merkmale werden jedoch für den Diagnose- oder Prognoseschritt keine Aussagekraft haben. Häufig werden auch konstante Größen gemessen, wie bspw. die Drehzahl eines Generators, die keinen Mehrwert für die Diagnose liefern. Ziel der Merkmalsselektion und -reduktion ist daher, den Merkmalsraum (engl. feature space) im nächsten Schritt auf ein aussagekräftiges Parameterset einzugrenzen. Neben der verbesserten Übersichtlichkeit und Nachvollziehbarkeit der Eingaben (Inputs) in die Algorithmen, reduziert sich gleichzeitig die Komplexität und damit die Trainingszeit der Modelle. [Are+18; Rom+16]

Bei der Merkmalsreduktion werden alle oder nur ein Teil der Merkmale in einen neuen Merkmalsraum mit niedrigerer Dimension transformiert. Die bei PHM Systemen beliebteste Methode hierfür ist die Hauptkomponentenanalyse (engl. Principal Component

¹z. B. *tsfresh*: <https://tsfresh.readthedocs.io/en/latest/index.html>, aufgerufen am 21.03.2023

Analysis (PCA)). Die PCA ist eine orthogonale Lineartransformation, bei der die erklärende Varianz in den Hauptkomponenten maximiert wird, die erste Hauptkomponente also die größte Varianz repräsentiert, die zweite Hauptkomponente die zweitgrößte Varianz, u.s.w.. Die Reduktion der Parameter geschieht allerdings erst im zweiten Schritt, bei dem alle Hauptkomponenten unterhalb einer zuvor definierten Varianzschwelle (z. B. 95%) abgeschnitten werden. Alternativen zur PCA sind die Faktorenanalyse oder auch das t-distributed stochastic neighbor embedding (t-SNE). [BNS20; Cer19]

Bei der Merkmalsselektion wird der Merkmalsraum hingegen nicht durch Transformation sondern durch Selektion einzelner oder mehrerer Merkmale reduziert. Dies kann einerseits basierend auf Expertenwissen geschehen, bspw. durch die Berücksichtigung der ausgewählten Frequenzbereiche einer FFT für ein bestimmtes Schadensbild. Andererseits können auch hierfür algorithmische Ansätze genutzt werden, die sich in Filter [LH03], Wrapper [KJ97] und eingebettete Methoden [Lal+06] aufteilen lassen. Filter beurteilen die Relevanz von Merkmalen durch uni- oder multivariate Kennwerte, wie bspw. das Bayessche Informationskriterium oder das Minimum Redundanz Maximum Relevanz Verfahren, auch bekannt als minimum Redundancy Maximum Relevance (mRMR) [PLD05]. Filtermethoden haben den Nachteil, dass sie Redundanzen innerhalb der Merkmale nicht unbedingt erkennen, weswegen häufig auch Wrapper zum Einsatz kommen. Dabei wird analysiert, welchen Einfluss einzelne Merkmale innerhalb einer Klassifizierungsaufgabe auf die Performanz (meistens Genauigkeit, vgl. Abschnitt 2.2.3) haben. Resultat ist eine sortierte Liste der Wichtigkeit einzelner Merkmale im Hinblick auf die gewünschte Zielgröße [Lei+20]. Eingebettete Methoden nutzen die inhärente Selektion der Merkmale während des Trainings eines Algorithmus. So enthält der Entscheidungsbaum in den obersten Knoten typischerweise die wichtigsten Merkmale und gibt damit automatisch ein Ranking der wichtigsten Merkmale mit aus.

Ausgehend von den generierten und reduzierten bzw. selektierten Merkmalen, werden diese im nächsten Schritt in der Diagnose in Zustandsinformationen transformiert.

2.2.3. Datenbasierte Schadensdiagnose

Die Schadensdiagnose beschäftigt sich mit der Bewertung des Objekt- oder Systemzustands. Im Falle der hier beschriebenen datenbasierten Schadensdiagnose werden die Prozesseingangs- und Messgrößen gemäß den zuvor beschriebenen Schritten in Merkmale überführt, sodass eine Zustandsbewertung ermöglicht wird.

Die recht allgemeine Aufgabe der Fehlerdiagnose kann weiter untergliedert werden in Fehlerdetektion, -isolation und -identifikation, wie in Abbildung 2.7 dargestellt. Auf der rechten Seite des Diagnoseausgangs können verschiedene Weiterverarbeitungsschritte stehen. Bei der Fehlerdetektion wird zunächst nur festgestellt, ob ein Fehler im System vorhanden ist oder nicht. Die Fehlerisolation lokalisiert den Fehler im System und die Fehleridentifikation liefert quantitative oder qualitative Informationen zum Fehler und bildet Grundlage für eine Rekonfiguration [Oss14]. Während die Detektion z. B. eine Detailanalyse von Systemexperten erfordert, können die Phasen der Isolation und Identifikation gemäß dem Leitbild des CBM direkt zur Planung von Instandhaltungsmaßnahmen genutzt werden oder die weitere Verarbeitung in Prognosealgorithmen ermöglichen. Für die einzelnen Schritte bieten sich mitunter verschiedene Herangehensweisen und Ansätze der Merkmalsextraktion an. Der Gesamtansatz kann so auch hierarchisch aufgebaut werden. Die Aufgabe der Fehlerdetektion kann bspw. mit Regressionsansätzen und Merkmalsextraktionsansätzen im Zeitbereich bewerkstelligt werden, während die Fehlerisolation meistens Aufgabe von Klassifikationsalgorithmen ist.

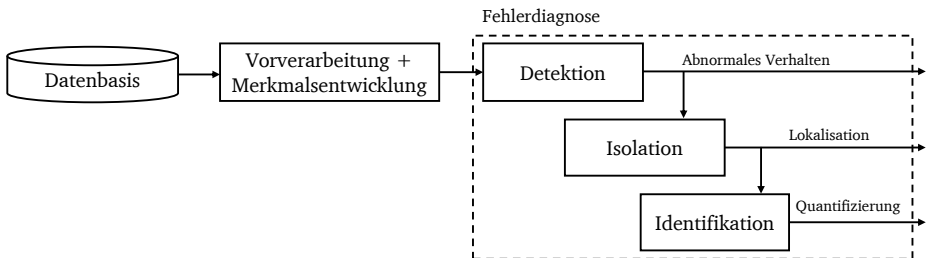


Abbildung 2.7.: Ablauf einer Fehlerdiagnose in Anlehnung an [RCB00]

Kern der Entwicklung ist demnach der Aufbau des entsprechenden Modells. Dabei kommen neben einfachen statistischen Verfahren vor allem Ansätze aus dem Bereich des ML und der Künstlichen Intelligenz (KI) zum Einsatz. Die Methoden des ML können in der Anwendung der Fehlerdiagnose und des CBM generell in *überwachtes* (engl. *supervised*) und *unüberwachtes* (engl. *unsupervised*) Lernen untergliedert werden. Beim überwachten Lernen kann jedem Datenpunkt in der Trainingsdatensatz ein bekannter Ausgabewert (engl. Label) zugewiesen werden. Ist der Ausgabewert binär oder diskret, spricht man von einer Klassifikation, ist er kontinuierlich, so spricht man von einer Regression. Dem gegenüber steht das unüberwachte Lernen, bei dem keine Labels vorhanden sind. Auch in diesem Fall ist eine Diagnose möglich, jedoch wird die

Interpretierbarkeit und Bewertung erschwert. Meist beschränkt sich das unüberwachte Lernen auf Dimensionsreduktion (siehe Abschnitt 2.2.2), grafische Interpretation und Gruppierungsaufgaben (engl. Clustering).

Bei der Fehlerdiagnose kann weiter unterschieden werden zwischen qualitativer und quantitativer Betrachtung, für die jeweils verschiedene methodische Herangehensweisen infrage kommen (siehe Abbildung 2.8). Im Falle der quantifizierten Betrachtung ist es das Ziel einen Gesundheitsindikator (engl. Health Indicator, HI) zu bilden, der üblicherweise das Verhältnis des aktuellen Zustands im Vergleich zur Abnutzungsgrenze prozentual darstellt (bspw. zwischen 0% und 100%). Da dieser die Grundlage für Prognoseanwendungen bildet, wird dieser Weg vor allem im Bereich des PdM gewählt. Hierfür kommen vor allem Regressionsansätze zum Einsatz, da sie den Indikator kontinuierlich abbilden. Daneben ist es ebenfalls möglich - wenn auch weniger gebräuchlich - eine Mehrklassen-Klassifikation zu nutzen. [Sus+15; Lee+14]

In praktischen Anwendungen der datenbasierten Fehlerdiagnose (bei Verwendung von ML-Algorithmen häufig auch Intelligent Fault Diagnosis (IFD)) wird der Gesundheitsindikator jedoch häufiger qualitativ (binär) klassifiziert: Ein Fehler ist vorhanden oder nicht vorhanden [Lei+20]. Entsprechend wird auch vereinfacht von gesunden (*healthy*) und schadhaften (*faulty*) Zuständen gesprochen. Auch hierbei ist die Verwendung von Regressionsansätzen möglich, wobei in diesem Fall keine Labels der Zielgröße notwendig sind, sondern Messgrößen angelernt werden, die beobachtbar sind. Der Vergleich mit einem Grenzwert liefert dann den binären Zustand. Dieser Ansatz ist vergleichbar mit einem regelungstechnischen Beobachter sowie klassischen physikalischen Verfahren, findet jedoch besonders bei der Fehlerdetektion in Zusammenhang mit ML Anwendung. So bauen Jannis Weinert et al. und Schlechtingen et al. z. B. so genannte *Normal Behaviour Modelling* Ansätze für die Überwachung der Betriebsvariablen von Windkraftanlagen auf. [JS16; SSA13]

Typische Ansätze und Algorithmen aus der Literatur

In der Literatur kann eine Vielzahl von datenbasierten Ansätzen zur Schadensdiagnose gefunden werden. Die verbreitetsten Ansätze für die algorithmische Verknüpfung von Eingangs- und Labeldaten der vergangenen Jahre sind dabei künstliches Neuronales Netz (kNN), die Stützvektormethode (engl. Support Vector Machine (SVM)), Bayes'sche Netze (engl. Bayesian Network (BN)) sowie Entscheidungsbäume (engl. Decision Tree (DT)) bzw. deren Ensembles Random Forest (RF). Aktuelle Trends gehen

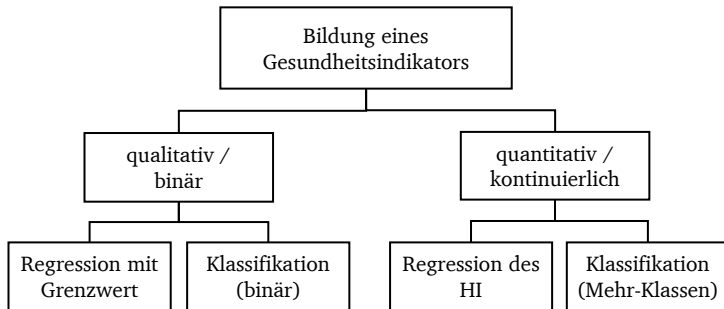


Abbildung 2.8.: Ansätze zur Erstellung eines Gesundheitsindikators (eigene Darstellung)

vor allem in das Themengebiet des *tiefen Lernens* (engl. Deep Learning (DL)) und weiteren speziellen Formen von kNN, wie z. B. Long Short-Term Memory (LSTM)-Netze. Auch das Themengebiet des *Transfer Learning* ist derzeit im Fokus der Forschung, bei dem bereits vortrainierte Netzwerke mit kleinen Anpassungen auf das gegebene Problem adaptiert werden. Die genannten Verfahren sind nur ein Auszug aus einer Vielzahl an möglichen Ansätzen. Ein umfassender Überblick über Anwendungsfälle würden den Rahmen dieses Grundlagenkapitels jedoch sprengen. Eine kurze Beschreibung der bereits genannten Verfahren sowie umfangreiche Anwendungsbeispiele im Kontext des CBM sowie PdM von RF, kNN und SVM können in Lei et al., Liu et al. und Carvalho et al. sowie für BN in Weber et al. gefunden werden.

Als Alternative zu datenbasierten Ansätzen, können auch modellbasierte Ansätze zur Schadensdiagnose verwendet werden. Modellbasierte Ansätze setzen ein physikalisches Abbild des betrachteten Objekts in Form eines Sets aus mathematischen Gleichungen voraus. Der Schlüssel der Diagnose liegt im Abgleich zwischen modellierten und gemessenen Parametern (Betrachtung der Residuen) [Ise84]. Durch geeignete Filteransätze, wie Kalman-Filter oder Partikelfilter, können die internen Parameter geschätzt werden, was einen großen Vorteil des Ansatzes darstellt. Somit ist insbesondere die Interpretierbarkeit und Nachvollziehbarkeit gegenüber datenbasierten Ansätzen besser. Da jedoch auch der physikalische Ursprung des Schadens korrekt abgebildet werden muss, erfordern diese Modelle ein tiefgehendes Systemverständnis [Ise06]. Besonders für komplexe Systeme steigt der Aufwand für den Aufbau eines physikalischen Modells

sehr schnell an oder es wird gar unmöglich die Wirklichkeit exakt abzubilden, insbesondere bei der Modellierung von Degradierungsprozessen.

2.2.4. Metriken zur Evaluation der Fehlerdiagnose

Abschließend erfolgt die Bewertung der Diagnosealgorithmen anhand von objektiven Metriken. Ziel ist dabei, die Güte der Diagnose und die Performanz der Algorithmen objektiv beurteilen und vergleichen zu können. Im Bereich der Diagnose wird dabei zwischen Metriken für Regression und Klassifikation unterschieden.

Für die Regression kann das Bestimmtheitsmaß R^2 betrachtet werden, das die Anpassung des Modells an die Daten beschreibt. Da R^2 dimensionslos und auf das Intervall $[0,1]$ beschränkt ist, ist die Interpretierbarkeit des Bestimmtheitsmaßes verhältnismäßig einfach. Es beschreibt dabei den Anteil an Varianz in der Zielvariablen, der durch das Modell beschrieben wird. Allerdings hat die Verteilung der Daten einen großen Einfluss auf die Aussagekraft dieser Metrik: Wenn die Anzahl der Datenpunkte hoch ist und die Varianz der Eingangsdaten im Vergleich zur Varianz der Zielvariablen klein ist, führt das zu einem hohen R^2 , obwohl das Modell die Daten mitunter schlecht beschreibt.

Neben R^2 werden verschiedene Metriken verwendet, die auf die Akkuratheit der Regression zurückzuführen sind. Die Akkuratheit (auch Regressionsresiduum) ist definiert als die Abweichung u_i des tatsächlich gemessenen Wertes y_i^* und des modellierten Wertes y_i : [Eck19]

$$u_i = y_i^* - y_i \quad (2.1)$$

Die perfekte Regression hat einen Fehler von $u_i = 0$. Hiervon ausgehend lassen sich mehrere Metriken und Berücksichtigung der Gesamtvorhersagen n herleiten:

- Der mittlere absolute Fehler (engl. Mean Absolute Error (MAE)) definiert als:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |u_i| \quad (2.2)$$

Tabelle 2.1.: Aufbau einer Konfusionsmatrix bei einem
Zwei-Klassen-Problem
vorhergesagte Klasse

		vorhergesagte Klasse	
		schadhaft	gesund
wahre Klasse	schadhaft	TP (richtig positiv)	FN (falsch negativ)
	gesund	FP (falsch positiv)	TN (richtig negativ)

- Der mittlere absolute relative Fehler (engl. Mean Absolute Percentage Error (MAPE))

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{100 \cdot u_i}{y_i^*} \right| \quad (2.3)$$

- Der Mean Squared Error (MSE)

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n u_i^2 \quad (2.4)$$

- Der Root Mean Squared Error (RMSE)

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (2.5)$$

Bei der Klassifikation wird beurteilt, wie gut der Algorithmus Datenpunkte den einzelnen Klassen zuordnet. Bei einem binären (zwei Klassen) Klassifikationsproblem werden alle weiteren Metriken anhand der Konfusionsmatrix (siehe Tabelle 2.1) gebildet. [SJS06]

In der Matrix entsprechen die Zeilen der wahren Klasse (Labels) und die Spalten der durch den Algorithmus vorhergesagten Klasse (Prediction). Stimmt die Vorhersage des Algorithmus mit den wahren Labels überein, spricht man von einer korrekten (richtigen) Klassifikation, andernfalls von einer falschen. Entsprechend finden sich auf der Hauptdiagonalen die korrekten Klassifikationen und auf der Nebendiagonalen die falschen Klassifikationen. Entscheidend für die Bildung der Metriken und Interpretation der Ergebnisse ist die Definition der positiven und negativen Klasse auf den Diagonalen. Im Rahmen dieser Arbeit wird die in der Literatur übliche Notation für

asymmetrische, binäre Attribute übernommen: Die häufiger vorkommende negative Klasse (0) entspricht dem gesunden bzw. nominalen Zustand, die seltener vorkommende positive Klasse (1) entspricht dem schadhafte[n] Zustand [HPK12]. Entsprechend gibt es lediglich vier Fälle, anhand derer die Klassifikation beurteilt wird: Richtig positiv (engl. True Positive (TP)), richtig negativ (engl. True Negative (TN)), falsch positiv (engl. False Positive (FP)) sowie falsch negativ (engl. False Negative (FN)). Durch die mathematische Kombination der vier Zellen werden im Folgenden die Metriken zur Evaluation gebildet: [Alp14]

- Genauigkeit (engl. Accuracy (Acc)):

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.6)$$

- Präzision (engl. Precision (Prec)):

$$Prec = \frac{TP}{TP + FP} \quad (2.7)$$

- Sensitivität (engl. Recall (Rec)):

$$Rec = \frac{TP}{TP + FN} \quad (2.8)$$

- f1-Wert (engl. f1-score):

$$f1 = 2 \cdot \frac{Prec \cdot Rec}{Prec + Rec} \quad (2.9)$$

Im weiteren Verlauf der Dissertation werden die englischen Fachbegriffe der Metriken verwendet. Die Metriken verfolgen dabei unterschiedliche Ziele der Bewertung. Während die Genauigkeit eine recht allgemeine Metrik ist, verliert sie gerade bei sehr unbalancierten Klassenverteilungen (high class imbalance) enorm an Aussagekraft. Gerade im Kontext der Fehlerdiagnose werden daher bevorzugt Recall, Precision und f1-Score betrachtet. Gerade Recall und Precision stehen dabei meist im Konflikt und nur eine der beiden Metriken kann zur Last der anderen optimiert werden. Eine Analogie zur Optimierung nach Sicherheit oder Kosten kann dabei aufgebaut werden: Recall wird herangezogen bzw. optimiert, wenn der Preis für FN hoch ist. Zur Erinnerung: Ein False Negative beschreibt einen unentdeckten Schaden und ist daher sicherheitskritisch. Entsprechend wird Recall für die Optimierung der Sicherheit betrachtet, Precision bei

Betrachtung der Instandhaltungseffizienz (engl. maintenance efficiency), respektive Kosten.

Die Receiver Operating Characteristic (ROC) Kurve gibt einen guten Überblick über den Kompromiss zwischen Recall und Precision, wobei die Fläche unterhalb der Kurve (engl. Area Under Curve AUC) als entsprechende Metrik dient und maximiert werden soll. [Sax+08b; SJS06]

2.3. Forschungslücke

Ein entscheidendes Merkmal bei den in der Literatur dargestellten und diskutierten Methoden ist die Verfügbarkeit von Zeitreihendaten. Die weitgehende Mehrheit der Publikationen entwickelt Methoden, die auf den zuvor dargestellten Ansätzen der Merkmalsentwicklung (vgl. Abschnitt 2.2.2) und ML-Algorithmen beruhen. Es wird davon ausgegangen, dass entsprechende Sensorik am System appliziert ist, bzw. schlichtweg ausreichend Zeitreihendaten verfügbar sind. Daneben liegen jedoch noch weitere Datenquellen mit abweichender Datenstruktur vor. Dieser Abschnitt führt in das Thema der Ereignisdaten (*Event Logs*) ein und beschreibt bereits existierende Ansätze zur Datenvorverarbeitung und Fusion von diesen.

2.3.1. Unterschiedliche Datentypen und -strukturen als Grundlage

Alternative Datenquellen, wie operationelle Informationen, Wartungs- und Instandhaltungslogbücher oder Ereignisdaten (*Event Logs*) werden derzeit vor allem zum Labeln von Zeitreihendaten und damit zur Erklärung und Evaluation der Modelle verwendet. Abseits von typischen Maschinenbauanwendungen aus der Produktion, Luftfahrt- oder Energiebranche finden sich vereinzelt Ansätze zur Verarbeitung von Ereignisdaten, die in diesem Abschnitt beleuchtet werden.

Ereignisdaten können vereinfacht als temporale Daten beschrieben werden, deren zeitliche Abfolge entgegen klassischen Zeitreihendaten keinen äquidistanten Zeitstempeln folgt und aufeinanderfolgende Einträge keine vorgeschriebene Beziehung haben. Sie sind typischerweise textbasiert, haben mehrere Attribute, kommen von unterschiedlichen Bauteilen, treten unregelmäßig auf und zeigen sowohl sequenzielle als auch parallele Korrelationen [BKS22]. Eine ausführlichere Einführung und Definition wird in Abschnitt 3.2.1 gegeben.

In Alam et al. wird einer der bekanntesten PHM-Vergleichsdatensätze (siehe [Sax+08a]) adaptiert und als Ereignisdatensatz interpretiert. Dabei werden verschiedene Ansätze aus dem DL verwendet um die RUL zu präzidieren. Die Datengrundlage entstammt jedoch einerseits aus einer Simulationsumgebung und andererseits werden in dieser Veröffentlichung lediglich Zeitreihendaten mit Filterverfahren sequenziert, sodass keine wahre Ereignisdatenbasis entsteht. [Ala+21]

Bonnevay et al. untersucht die Möglichkeiten der Fehlerdetektion basierend auf Ereignisdaten in der Elektroindustrie. Dabei werden aus den Ereignisdaten Wavelet-Merkmale generiert und diese mit einem *Random Forest* als Ensemble-Ansatz verarbeitet. Entscheidend ist hierbei, dass die Ereignisse in eine Funktion der Zeit überführt werden, ein häufig verwendeter Ansatz zur Vorverarbeitung von Ereignisdaten. Betrachtet wurde hierbei vor allem, wie früh ein Fehler vor dem Ausfall detektiert werden kann und wie sich der Fehler der Klassifikation (engl. False Negative Rate bzw. False Positive Rate) über die Zeit entwickelt: je weiter entfernt der Ausfall liegt, desto größer ist der Klassifikationsfehler. [BCG20]

In Gutschi et al. wird ein Ansatz zur Fehlerdetektion auf Basis von Ereignisdaten in einer Fertigungsstraße formuliert. Die Methode der Datenvorverarbeitung mit einer Segmentierung in Fenster und gleitende Fenster liegt im Fokus der Veröffentlichung. Die Klassifikation der Restlebensdauer erfolgt ebenfalls mit einem *Random Forest* Algorithmus. Eine Untersuchung verschiedener Ansätze der Merkmalsselektion zeigt einen hohen Einfluss auf die Ergebnisgüte der RUL-Prognose. [Gut+19]

Einen weiteren Ansatz der Segmentierung und Vorverarbeitung von Ereignisdaten in gleitenden Fenstern wird in Korvesis et al. vorgeschlagen. Bei der Anwendung auf so genannte *Post flight reports* aus der Luftfahrtindustrie werden die Ereignisse in einen hoch-dimensionalen Vektorraum projiziert, der im nächsten Schritt als Regression der Ausfallwahrscheinlichkeit interpretiert wird. Als Algorithmus wird ein *Random Forest* Regressor verwendet. Der Ansatz kann auch als Weiterentwicklung der von Sipos et al. vorgeschlagenen Methode verstanden werden. Hier wird ebenfalls ein *Bag-of-Events*-Ansatz entwickelt, vergleichbar mit dem *Bag-of-Words*-Ansatz aus der Verarbeitung natürlicher Sprache [Joa98]. Eingebettet in einen *Mult-Instance Learning (MIL)*²-Ansatz wird eine SVM zu Klassifikation der Fehler angelernt und auf das Anwendungsgebiet von medizinischem Equipment angewendet.

Auch Bakdi et al. verwendet das Verfahren der Zeit- und Fensteraggregation von Ereignissen zur Fehlerdiagnose bzw. sehr kurzfristigen Prognose in einem Zeitfenster von

²Bei MIL-Ansätzen erhält der Lernalgorithmus ein Set aus gelabelten *Bags* (mehrere/viele Instanzen), an Stelle von einzeln gelabelten Instanzen. [DLL97]

maximal drei Tagen. Am Beispiel von Fehlernachrichten eines elektrischen Schiffantriebs wird mithilfe von RF und des temporalen Verarbeitungsansatzes für natürliche Sprache *Random Indexing*³ dabei ein Modell zur Fehler- und Ausfallklassifikation aufgebaut. Schlüssel ist die Transformation der textbasierten Nachrichten in einen numerischen Raum um Zeitreihenanalysen zu ermöglichen. Die Besonderheit ist weiterhin die Berücksichtigung von so genannten „schwachen“ Labels, bei denen lediglich der Ausfallzeitpunkt bekannt ist, ohne weitere Informationen zu vorangegangenen Anzeichen oder Auslösern. [BKS22]

Guillaume et al. präzisiert den Ausfall von Geldautomaten in einer Flotte basierend auf Ereignisdaten (*Event Logs*). Dabei werden einzelne Fehlermeldungen betrachtet und mit einer *One-Hot*-Encodierung innerhalb gleitender Fenster als Merkmale interpretiert. Methodisch werden verschiedene *change point detection* Ansätze implementiert und untersucht. [GVW20]

Ein allgemeines Vorgehen zur Vorverarbeitung von Ereignisdaten wird von Wang et al. vorgeschlagen. Dabei werden übergeordnet vier Merkmalskategorien eingeführt: statistisch-basierte, Muster-basierte, Ähnlichkeits-basierte sowie Profil-basierte Merkmale. Das allgemeine Konzept wird ebenfalls am Beispiel verschiedener Ereignisdaten von Geldautomaten im Rahmen einer Klassifikation und unter Verwendung verschiedener baumbasierter Lernverfahren untersucht. [Wan+17]

Die zusammengestellte Literaturübersicht zeigt, dass es durchaus einige Forschungsarbeiten zur Verarbeitung von Ereignisdaten mit dem Ziel der Fehlerdetektion bzw. Prognose der RUL gibt. Der Fokus liegt dabei vor allem auf elektronischen Geräten und IT-Systemen. Diese Systeme erfassen üblicherweise keine oder nur sehr wenige Zeitreihendaten (meistens lediglich Temperaturwerte). Entsprechend besteht bei diesen Systemen die Notwendigkeit, auf die Ereignisdatenbasis zurückzugreifen, wobei Zeitreihendaten dabei vernachlässigt werden. Der potentielle Mehrwert von Ereignisdaten im Maschinenbaukontext ist bisher jedoch unerforscht. Weiterhin zeigt die Literaturübersicht in Abschnitt 2.2.3, dass der Schwerpunkt der Forschung bei den meisten Maschinenbauanwendungen (rotierenden Bauteilen und Maschinen) auf der Verarbeitung und Nutzung von Zeitreihendaten in klassischen PHM-Ansätzen liegt. Dabei werden Ereignisdaten allenfalls für die Bestimmung der abhängigen Variablen (*Label*) und zur Evaluation der Modelle verwendet.

Abseits des technischen Kontext der Zustandsdiagnose finden sich bereits Ansätze zur Fusion von unterschiedlichen Datenstrukturen, die im nächsten Abschnitt kurz

³Random Indexing wird in Kombination mit SVM von [Fro+13] zur Fehlerdiagnose in Software-Systemen vorgeschlagen.

erläutert werden sollen.

2.3.2. Stand der Forschung in der Fusion unterschiedlicher Datenstrukturen

Für die Fusion unterschiedlicher Datenstrukturen existieren bereits einige Ansätze, die diese Funktion inhärent ermöglichen. Der folgende Abschnitt gibt einen kurzen Überblick über diesen Themenbereich.

Im Kontext des *Reliability Engineering* finden sich grundlegende Ansätze zur Fusion von verschiedenen Datenstrukturen. Das Reliability Engineering wird z. B. in der Luftfahrtindustrie verwendet um Wartungsintervalle festzulegen. Grundlage hierfür sind Ereignisdaten, wie Ausfall- oder Wartungsdaten, die zur Anpassung von Wahrscheinlichkeitsverteilungen herangezogen werden (üblicherweise Weibull-Verteilungen). Eine logische Weiterentwicklung des Reliability Engineering ist dabei die Kombination mit sensorbasierten Diagnosemodellen. Eine solche Kombinationen sind dabei so genannte *Time Dependent Proportional Hazard* Modelle, bei denen Ereignisse Beobachtungspunkte für die Modelloptimierung darstellen. [JLB06; JAM87; VWZ04]

Auch Kalman-Filter, die insbesondere in der Zustandsschätzung und Multi-Sensordaten Fusion eine wichtige Rolle spielen, können mithilfe von Modifikationen der zeitlich-variablen Matrizen (engl. Time-Varying Kalman Filter (TVKF)) unterschiedliche Sensorraten verarbeiten [AYK23; Dei+23]. Insbesondere die Sonderformen des Extended Kalman Filter (EKF) sowie Unscented Kalman Filter (UKF) sind Erweiterungen des klassischen Kalman-Filter und können mit unvollständigen, zeitlich versetzten bzw. asynchronen Messdaten umgehen. Beispiele solcher Ausführungen finden sich in V. Fathabadi et al., Fatehi et al. und Wang et al.

Neben den bereits genannten Ansätzen bieten sich ebenfalls probabilistische Modelle an, um mit unterschiedlich häufig auftretenden Zeitstempeln umzugehen. Das bekannteste Beispiel hierfür sind BN, die z. B. in Cai et al. verwendet werden, um sensordatenbasierte Fehlerdiagnosemodelle durch Evidenzen von menschlichen Beobachtungen zu ergänzen. Dabei können beim untersuchten Anwendungsfall der Erdwärmepumpen signifikante Verbesserungen in der Fehlerdetektion erreicht werden. Dynamische BN werden in Li et al. verwendet, um das Risswachstum in Strukturrissen von Flugzeugflächen zu diagnostizieren und prognostizieren. Die Struktur der Netze erlaubt dabei auch die Integration heterogener Datenstrukturen bzw. -quellen.

Auch Hidden Markov Models (HMM) lassen sich in die Kategorie der probabilistischen Modelle einordnen und ermöglichen die gleichzeitige Verarbeitung von Event- und Zeitreihendaten. Dabei werden Schadensklassen als *Hidden States* interpretiert und Sensorsignale als Realisierung der Beobachtungen (siehe z. B. [BMA00; Esm+13]). Im Bereich der Instandhaltung kommt häufig auch deren Erweiterung als Hidden semi-Markov Modell zum Einsatz, wie bei Dong et al. für technische Diagnose- und Prognoseanwendungen sowie Salfner, der eine generelle mathematische Formulierung für temporale Sequenzen in Hidden semi-Markov Models (HsMM) vorschlägt.

Beispiele aus anderen Industriebereichen finden sich unter anderem im Gesundheitswesen, wo Liu et al. Continuous-Time Hidden Markov Model (HMM) mit dem Ziel der Analyse von unterschiedlicher Krankheitsverläufe anwendet. Im Finanzwesen wird in den vergangenen zehn Jahren ebenfalls an ähnlichen Problemstellungen geforscht. So fing Jianfeng Si et al. z. B. damit an, Sentiments (Meinungsbilder) aus Kurznachrichten von Plattformen wie X (ehemals Twitter) zu analysieren und daraus Aktienkursprognosen abzuleiten. Ni et al. erweitert den Ansatz um Zusammenhänge zwischen verschiedenen Tweets zu bilden, sowie diese mit klassischen zeitreihenbasierten Indikatoren (Tageshoch, -tief, Eröffnungs- und Schlusspreis) zu fusionieren. In den meisten Veröffentlichungen geht es dabei weniger um quantitative Prognosen als vielmehr darum, eine Indikation des Trends zu erhalten: steigende oder fallende Kurse.

Zusammenfassend existiert ein breiter Forschungsbereich, in dem die Fusion unterschiedlicher Datenstrukturen eine Rolle spielt. In einigen Fällen findet die Fusion allerdings vielmehr in Form eines Angleichens der Zeitreihendaten statt, wie beim TVKF. Teilweise wird die Fusion auch als sinnvolle Kombination von abhängigen und beschreibenden Variablen gesehen. Je nach Ansatz ist dabei das Ereignis entweder eine Evidenz (BN) oder ein verdeckter Zustand (HMM).

2.4. Beitrag dieser Arbeit

Die vorliegende Arbeit widmet sich der Aufgabe, die im vorherigen Abschnitt herausgestellte Forschungslücke zu schließen. Ziel der Arbeit ist die Erweiterung eines klassischen PHM-Ansatzes basierend auf Zeitreihendaten durch die Fusion und Integration von zeit-diskreten Ereignisdaten, wie Event-Logs oder Instandhaltungs- und Wartungsdaten. Bereits bestehende Ansätze sollen dabei unverändert weiterverwendet

werden können und lediglich durch die zusätzliche Datenbasis inklusive der entsprechenden Vorverarbeitung erweitert werden. Dafür wird ein generischer Ansatz vorgeschlagen, der Methoden aus der Verarbeitung von Ereignisdaten (vgl. Abschnitt 2.3.1) aufgreift, weiterentwickelt und als optionale Erweiterung (*Add-On*) in ein bestehendes PHM-Framework integriert. Während die in Abschnitt 2.3 vorgestellten Ansätze anwendungsspezifisch formuliert sind, d.h. einen speziellen Ansatz für ein bestimmtes Problem vorstellen, wird das im Rahmen dieser Arbeit vorgeschlagene Konzept allgemeingültig formuliert und anhand eines Anwendungsfalls evaluiert.

Die Arbeit verfolgt das übergeordnete Ziel die Diagnosegüte zu steigern. Insbesondere soll dabei die Genauigkeit der Fehlerklassifikation bei komplexen Systemen verbessert und Unsicherheiten reduziert werden. Schlüssel zur Integration der Ereignisdaten und Kern des Dissertationsvorhabens ist die entsprechende Vorverarbeitung und zeitliche Anpassung der Ereignisdaten (*time-conversion*). Die Evaluation des Forschungsbeitrags erfolgt anhand der Bewertungsmetriken der Klassifikation, die den Einfluss der zusätzlichen Datenbasis aufzeigt. Der Forschungsbeitrag ist in Abbildung 2.9 visualisiert.

Nachdem im vergangenen Kapitel eine Übersicht über den Stand der Technik gegeben und daraus der Forschungsbeitrag extrahiert wurde, wird im nachfolgenden Verlauf der Arbeit der Ansatz zur Fusion von Zeitreihendaten mit zeit-diskreten Ereignisdaten präsentiert und anhand eines Anwendungsbeispiels evaluiert.

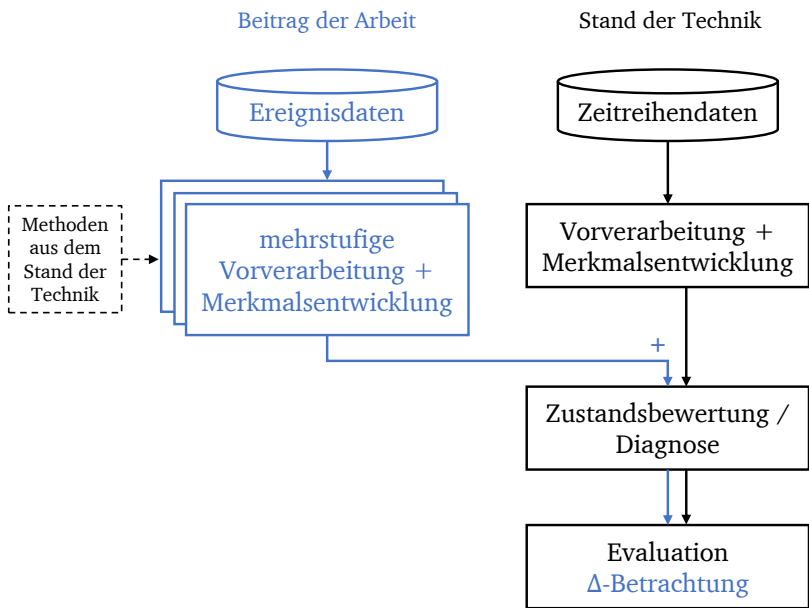


Abbildung 2.9.: Forschungsbeitrag der Dissertation

3. Konzept zur Schadensdiagnose unter zusätzlicher Berücksichtigung von Ereignisdaten

Aufbauend auf den zuvor dargestellten Grundlagen und der identifizierten Forschungslücke, wird in diesem Kapitel das vorgeschlagene Konzept zur Fusion von Zeitreihendaten und Ereignisdaten vorgestellt. Entsprechend stellt Abschnitt 3.1 die Gesamtstruktur dar und beschreibt die notwendigen Randbedingungen sowie dessen getroffenen Annahmen. In Abschnitt 3.2 wird auf den Kern der Entwicklung, die Methoden zur Umsetzung der Datenvorverarbeitung und -fusion eingegangen. Abschließend werden in Abschnitt 3.3 und Abschnitt 3.4 die Algorithmen zur Fehlerdiagnose und das Vorgehen der Evaluation vorgestellt.

3.1. Überblick über das vorgeschlagene Konzept

Der vorliegende Abschnitt gibt einen Überblick über das Konzept der Fehlerdiagnose auf Basis der fusionierten Zeitreihendaten mit zeit-diskreten Ereignisdaten. Hierfür werden zunächst die Gesamtstruktur dargestellt und die Zielsetzung erläutert und daraufhin die geltenden Randbedingungen, getroffenen Annahmen und Anforderungen (bspw. an die Datengrundlage) formuliert.

3.1.1. Gesamtstruktur und Zielsetzung

Die Gesamtstruktur des Konzepts ist aus Abbildung 2.9 ersichtlich. Als Ausgangslage und Basis wird ein bestehendes PHM-System zur Fehlerdiagnose angenommen. Entsprechend liegt eine Zeitreihendatenbasis vor, für die bereits ein Ansatz zur Detektion

und Klassifikation von Fehlerfällen existiert. Zusätzlich zur Zeitreihendatenbasis sind als zweite Datenbasis Ereignisdaten vorhanden.

Für beide Pfade (und somit auch Datengrundlagen) werden getrennte, eigene Methoden zur Datenvorverarbeitung und Merkmalsentwicklung verwendet. Für die Zeitreihenbasis bieten sich die in Abschnitt 2.2.2 dargestellten Ansätze an, wobei diese nicht im Fokus der Arbeit stehen. Im Rahmen der Arbeit wird daher mit einfachen, statistischen Extraktionen im Zeitbereich gearbeitet, die für das Konzept als gegeben angenommen werden. Die Vorverarbeitung und Merkmalsentwicklung der Ereignisdaten stellt den Kern der Arbeit dar und ist vereinfacht in Abbildung 3.1 dargestellt. Der Prozess ist dabei mehrstufig aufgebaut: zunächst findet eine Kategorisierung der Ereignisdaten statt. Für die Kategorisierung ist Expertenwissen notwendig. Ausgehend davon werden die Daten je nach Kategorie mehr oder weniger intensiv gefiltert und reduziert (siehe Abschnitt 3.2.2). Im nachgelagerten Schritt der Merkmalsentwicklung werden die Ereignisdaten in Merkmale überführt (siehe Abschnitt 3.2.3), die im abschließenden Schritt mit den Zeitreihendaten fusioniert werden. Der Schritt der Merkmalsentwicklung ist generisch formuliert und daher ohne Expertenwissen anwendbar. Der gesamte Vorverarbeitungsprozess der Ereignisdaten zur Fusion mit Zeitreihendaten wird als *time-conversion* bezeichnet [BCG20].

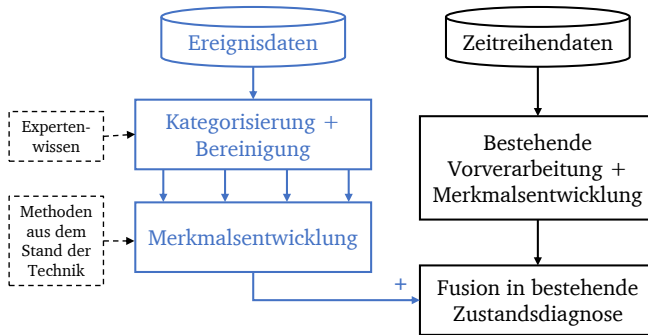


Abbildung 3.1.: Vorverarbeitungsschritte der Ereignisdaten

Im nachfolgenden Schritt werden die Ereignisdaten in den Diagnoseschritt eingebracht (siehe Abschnitt 3.3). Zur Evaluation des Konzepts (siehe Kapitel 4) werden hierfür mehrere Klassifikationsalgorithmen aufgebaut, deren Ansatz und Struktur stets gleich ist. Zunächst wird eine Plausibilisierung der vom Konzept vorgeschlagenen Merkmalsentwicklung vorgenommen, indem der Klassifikationsalgorithmus ausschließlich die

Merkmale der Ereignisdaten als Input erhält. Ist dabei eine generelle Klassifikation möglich, ist die Merkmalsentwicklung erfolgreich. Der Vergleichsansatz als zweiter Klassifikator ausschließlich basierend auf Zeitreihendaten stellt den Stand der Technik dar. Zuletzt werden beim vorgeschlagenen Konzept die Merkmale aus den Ereignisdaten als zusätzlicher Input dem Vergleichsansatz übergeben. Insofern ändert sich der Eingaberaum (Input-Werte) in den Algorithmus zwar, der algorithmische Ansatz bleibt jedoch gleich.

Damit werden zwei unterschiedliche Ergebnisse aus der Klassifikation generiert, die direkt gegenüber gestellt werden können. Anhand der Metriken zur Evaluation wird der quantitative Mehrwert durch die zusätzlichen Ereignisdaten beurteilt (siehe Abschnitt 3.4). Daneben lässt sich das Konzept auch anhand qualitativer Analysen, wie dem Gültigkeitsbereich in Abhängigkeit der Baugruppen oder Komponenten, diskutieren.

Nachdem in diesem Abschnitt ein kurzer Gesamtüberblick über das Konzept und das Vorgehen zur Evaluation gegeben wurde, soll im folgenden auf die übergeordnete Zielsetzung, Anforderungen und Randbedingungen eingegangen werden.

3.1.2. Randbedingungen, Annahmen und Anforderungen

Der vorliegende Abschnitt gibt eine Einführung in die Zielsetzung der Arbeit und die daraus resultierenden Randbedingungen, Annahmen und spezifischen Anforderungen an die Datengrundlage. Abschließend werden auch die erwarteten Grenzen der Arbeit kurz aufgezeigt. Eine Zusammenfassung der funktionalen Anforderungen, Annahmen und Anforderungen an die Datengrundlage ist in Tabelle 3.1 gegeben.

Funktionale Anforderungen beschreiben die Leistung und den Zweck oder auch den Funktionsumfang des Konzepts. Das grundlegende Ziel des Konzepts ist es, durch die Fusion der Ereignisdaten in einen bestehenden Prozess einen Informationsgewinn zu realisieren und damit die Diagnosegenauigkeit zu verbessern (FR1). Eine der Anforderungen ist dabei, dass das vorgeschlagene Konzept generisch ist (FR2). Daher wird es in diesem Kapitel anwendungsunabhängig formuliert. Zunächst werden die Ereignisdaten demnach ohne bezogenen Kontext analysiert und für verschiedene Kategorien unterschiedliche Vorschläge der Vorverarbeitung und Merkmalsentwicklung ausgearbeitet. Ziel ist somit die Übertragbarkeit des Konzepts auf unterschiedliche Anwendungsfälle zu gewährleisten (FR3). Des Weiteren wird davon ausgegangen, dass Sensordaten entsprechend der Dominanz im Stand der Forschung gegenüber Ereignisdaten die bessere Datenqualität darstellen. Daher soll das Konzept als Erweiterung in bereits

bestehenden PHM-Ansätzen bei entsprechender Verfügbarkeit von Ereignisdaten dienen können und ist insofern kein zwingender Bestandteil der (Erst-) Entwicklung von PHM-Systemen (FR4).

Neben den Anforderungen werden an dieser Stelle auch getroffene Annahmen formuliert. Generell wird davon ausgegangen, dass vorliegende bzw. verfügbare Ereignisdaten nicht nur aus den bereits verwendeten Sensordaten abgeleitet werden, sondern einen Informationsgewinn darstellen (A1). Diese Annahme kann auch als funktionale Anforderung formuliert werden, indem diese Unabhängigkeit gefordert wird. Weiterhin ist eine Annahme, dass die Ereignisdatenbasis prinzipiell im Zusammenhang mit Ausfällen generiert wird und damit eine Frühindikation zulässt (A2). Eine Randbedingung ist, dass das hier vorgestellte Konzept nicht auf die absolute Optimierung der Zustandsdiagnose abzielt, sondern nur den Einfluss der zusätzlichen Ereignisdaten beleuchtet (A3). Es wird angenommen, dass die Ereignisdaten prinzipiell in unterschiedliche Kategorien aufgeteilt werden können (A4). Da die Erfassung der Ereignisdaten häufig vom Hersteller einer Anlage festgelegt wird, kann diese als bekannt angenommen werden. Entsprechend wird im Rahmen der Arbeit angenommen, dass die Kategorien bereits bekannt und festgelegt sind (A5). D.h. die Kategorisierung ist an dieser Stelle eine Domänenaufgabe und wird nicht in aller Tiefe im Rahmen der Arbeit ausgeführt.

Bezüglich der Datengrundlage müssen einige Anforderungen formuliert werden, um eine erfolgreiche Evaluation zu ermöglichen. So muss sichergestellt sein, dass Informationen zu Ausfällen vorhanden sind und die Daten damit gelabelt sind (DR1). Weiterhin müssen sowohl Zeitreihendaten als auch Ereignisdaten vorhanden sein (DR2 & DR3). Abschließend muss für die erfolgreiche Evaluation und Beurteilung des Konzepts im Rahmen der Dissertation eine ausreichend große Datenbasis zum Trainieren der Modelle vorhanden sein (DR4).

An dieser Stelle sollen weiterhin einige erwartete Grenzen des Konzepts aufgeführt werden. Prinzipiell wird erwartet, dass eine Klassifikation der Zustände ausschließlich auf Basis der Ereignisdaten möglich ist, deren Genauigkeit jedoch dem Stand der Technik (auf Basis von Zeitreihendaten) unterliegt. Weiterhin wird erwartet, dass die Integration von Informationen aus Ereignisdaten nicht bei allen Baugruppen oder Komponenten einen signifikanten Einfluss auf die Diagnosegenauigkeit hat. So könnte das Konzept mehr Vorteile bei elektronischen Baugruppen liefern, als z. B. bei klassischen mechanischen Baugruppen wie Getrieben oder Lagern, die gut durch Schwingungssensorik überwacht werden können. Generell gilt hierbei jedoch zu beachten, dass eine Fehlerprognose bei elektronischen Bauteilen aufgrund des häufig zufälligen oder spontanen Ausfallverhaltens generell herausfordernd ist. An diesem Punkt liefert das vorgeschlagene Konzept keinen Mehrwert, da es ausschließlich auf die Diagnose des

aktuellen Zustands abzielt. Weiterhin werden Randbedingungen einer operationell produktiven Umgebung, wie z. B. Latenzen bei der Datenübertragung oder Laufzeiten von Datenverarbeitungspipelines, nicht betrachtet.

Tabelle 3.1.: Funktionale Anforderungen, Annahmen und Randbedingungen an das vorgeschlagene Konzept

Kategorie	ID	Definition
Funktionale Anforderungen	FR1	Das vorgeschlagene Konzept soll die Genauigkeit der Diagnose verbessern.
	FR2	Das vorgeschlagene Konzept soll Unsicherheiten (in Form von Falschklassifikationen) reduzieren.
	FR3	Das vorgeschlagene Konzept soll generisch formuliert und damit auf unterschiedliche Anwendungsfälle anwendbar sein.
	FR4	Das vorgeschlagene Konzept soll auf bestehenden Systemen aufbauen und diese erweitern.
Annahmen	A1	Die Ereignisdaten stellen einen Mehrwert gegenüber den Zeitreihendaten dar (sie sind unabhängig von den Sensordaten, keine Multikollinearität liegt vor).
	A2	Die vorliegenden Ereignisdaten stehen prinzipiell im Zusammenhang mit Ausfallereignissen.
	A3	Die Arbeit fokussiert nicht auf die Optimierung des Algorithmus, sondern auf der Betrachtung des Einflusses der zusätzlichen Datenbasis.
	A4	Die Logs lassen sich prinzipiell kategorisieren.
	A5	Kategorien werden durch Experten festgelegt und werden als vorhanden angenommen (Klassifikation der Logs in Kategorien vorhanden)
Anforderungen an die Datengrundlage	DR1	Ausfälle sind enthalten und bekannt (Labels vorhanden).
	DR2	Zeitreihendaten sind vorhanden.
	DR3	Ereignisdaten sind vorhanden.
	DR4	Eine ausreichend große Datenbasis zum Trainieren der Modelle ist vorhanden.

Mit den in diesem Abschnitt formulierten Zielsetzungen, Annahmen und Randbedingungen werden im Folgenden die Methoden zur Umsetzung des Konzepts dargelegt.

3.2. Methoden zur Umsetzung des Konzepts

In diesem Abschnitt werden die zur Umsetzung des Konzepts erforderlichen Methoden beschrieben. Eingangs wird dafür das Verständnis des Begriffs der *Ereignisdaten* und der entsprechenden Datenbasis selbst erläutert. Anschließend werden die relevanten Methoden zur Vorverarbeitung der Ereignisdaten und zur Merkmalsentwicklung beschrieben.

3.2.1. Ereignisdaten als zusätzliche Datenquelle

Ereignisdaten (engl. event data), diskrete temporale Daten oder sequenzielle Daten (engl. sequential data) werden meistens in so genannten Ereignisprotokollen (engl. event logs oder event log data) abgelegt. Im Rahmen der Arbeit wird der Begriff Ereignislog verwendet. Ereignislogs sind in mehrfacher Hinsicht einzigartig. Sie sind temporal und können sowohl als symbolische Sequenzen (über Ereigniscodes) als auch als numerische Zeitreihen betrachtet werden. Im letzteren Fall werden Variablen aus Meldungen extrahiert oder mit Ereignishäufigkeiten über ein bestimmtes Zeitfenster, z. B. Tage, definiert. Ein Log kann zusätzlich kategorische Merkmale (Ereigniscodes und kategorische Variablen im Text) enthalten und typischerweise sind Logs textbasiert und enthalten damit unstrukturierte Daten. Wesentliche Eigenschaften sind dabei, dass die Daten keine äquidistanten Zeitstempel aufweisen und nachfolgende Einträge keine Beziehung haben (müssen), d.h. sie treten unregelmäßig auf und können dabei sequenzielle und parallele Korrelationen aufweisen. [Sip+14; Agg15; CBK09; BKS22]

Mathematisch formuliert sind Ereignislogs eine Menge E von Ereignissen und eine Menge T von Zeitstempeln. Eine Ereignisdatenquelle erfasst eine Menge $D = \{e, t\} | e \in E, t \in T$, wobei jedes Element (e, t) in D ein Ereignis e mit einem zugehörigen Zeitstempel t darstellt. Jedes Ereignis e kann einen Satz von Eigenschaften besitzen, die als zusätzliche Informationen in D gespeichert werden können. Die Menge D kann als eine Relation zwischen E und T aufgefasst werden, die die zeitliche Abfolge von Ereignissen in E beschreibt.

In Tabelle 3.2 ist ein exemplarischer Auszug eines Ereignislogs für mehrere Geräte gezeigt. Die ersten drei Spalten der Darstellung entsprechen dem Minimalbeispiel, in dem lediglich Zeitstempel, die Geräteinformation und eine Bemerkung des Ereignisses gegeben sind. Optionale zusätzliche Informationen, wie eine Kategorisierung oder die Schwere des Ereignisses sind ausgegraut in zwei weiteren Spalten enthalten. Des

Weiteren können weitere Zeitstempel (z. B. wann eine Fehlermeldung aufgehoben wird), Identifier der Ereignisse oder dazugehörige ausgeführte (Instandhaltungs-) Maßnahmen enthalten sein. [KBV18]

Tabelle 3.2.: Beispielhafter Aufbau eines Ereignislogs

Timestamp	ID	Remark	Cat.	Severity
2018-01-17 04:38:29	dev01	gear 2, temp: 57 C	Info	low
2018-01-17 14:05:43	dev06	Key switch: Local control	Service	low
2018-01-17 14:28:45	dev04	Pause 1.2 kW 52 RPM	Ops.	low
2018-01-17 15:47:39	dev06	Running up timeout	Fault	mid

Ereignislogs finden sich häufig in der IT-Branche. Diese definiert das Ereignislog als eine Aufzeichnung der Ereignisse, die in den Systemen und Netzwerken eines Unternehmens auftreten. Logs bestehen aus Log-Einträgen; jeder Eintrag enthält Informationen zu einem bestimmten Ereignis, das in einem System oder Netzwerk aufgetreten ist. Ursprünglich wurden Logs in erster Linie für die Fehlerbehebung verwendet, aber mittlerweile erfüllen sie in den meisten Unternehmen viele Funktionen, wie z. B. die Optimierung der System- und Netzwerkleistung, die Aufzeichnung der Aktionen von Benutzern und die Bereitstellung von Daten, die für die Untersuchung bössartiger Aktivitäten nützlich sind. [Nat]

Einer der wesentlichen Unterschiede zu Zeitreihendaten ist die Unabhängigkeit aufeinanderfolgender Einträge. Bei Zeitreihendaten sind implizit Abhängigkeiten der Sensorwerte, also aufeinander folgende Einträge in den Daten, über die Zeit gegeben. Zum Beispiel wird eine Temperaturmessung über einen gewissen Zeitraum nur langsame Veränderungen aufzeigen. Im Falle der Ereignisdaten, wie in Tabelle 3.2 exemplarisch dargestellt, sind aufeinander folgende Einträge mehrheitlich unabhängig voneinander. [Agg15]

3.2.2. Datenvorverarbeitung der Ereignisdaten

Die zuvor adressierte Unabhängigkeit der Datenpunkte, bzw. Log-Einträge, erfordert eine Anpassung der Datenvorverarbeitung gegenüber des klassischen Vorgehens bei Zeitreihendaten. Ereignisdaten enthalten unterschiedlichste Arten der Informationen, daher sind die grundlegenden Aufgaben der Vorverarbeitung eine Kategorisierung sowie Filterung bzw. Bereinigung der Daten. Die Vorverarbeitung der Ereignisdaten ist

dabei domänen- und anwendungsspezifisch. Der nachfolgende Abschnitt gibt daher Anhaltspunkte zu der grundlegenden Herangehensweise, die Ausführung bleibt jedoch abhängig vom Anwendungsfall und wird in Abschnitt 4.2 detailliert vorgestellt.

Die Kategorisierung hat das Ziel, den Datensatz in verschiedene Informationsklassen zu unterteilen. Die jeweiligen Klassen lassen sich dann im nächsten Schritt bereinigen. Dazu liegen unterschiedliche Zielsetzungen vor: Zum einen werden die Ereignislogs vereinfacht und Einträge ohne direkte Aussagekraft eliminiert und zum anderen Datenpunkte, die hohe Korrelationen mit den Zeitreihendaten aufweisen und damit wenig Informationsgewinn für die Diagnose darstellen, entfernt. Für die einzelnen Kategorien der Ereignislogs kommen im darauffolgenden Schritt der Merkmalsentwicklung jeweils ein oder mehrere Ansätze zum Einsatz.

Prinzipiell können Ereignislogs nach verschiedenen Kriterien unterteilt werden:

- Kategorisierung nach Komponenten oder Subkomponenten eines Systems
- Kategorisierung nach unterschiedlichen Informationsklassen, bzw. Inhalten und Anwendungsbereichen der Logs (z. B. Operative Bedingungen, Instandhaltung, Leistungsdaten, Grenzwertüberschreitungen, sonstige IT-Ereignisse wie Sensor-Reset, ...)
- Kategorisierung nach der Schwere eines Ereignisses (z. B. Hinweis, Warnung, Fehler, Ausfall)

Da Ereignislogs mitunter sehr heterogene Einträge aus unterschiedlichsten Quellen beinhalten können, ist zunächst die Kategorisierung nach Informationsklassen, also den Inhalten und Anwendungsbereichen, von großer Bedeutung. Gemeinsam mit einer Bewertung der Schwere eines Ereignisses kann damit auch eine Priorisierung durchgeführt werden. Der Schritt der Kategorisierung nach Komponenten oder Subkomponenten eines Systems spielt zunächst eine untergeordnete Rolle, da in Ereignisdaten – genauso wie in Zeitreihendaten – auch Muster von Interaktionen zwischen Komponenten sichtbar werden können und sie damit wertvolle Informationen liefern. Für einzelne Ansätze der Merkmalsentwicklung macht es jedoch Sinn, die Ereignisse zunächst nach Komponenten zu filtern, um unabhängige Merkmale zu generieren.

Um Ereignisdaten entsprechend verschiedener Kategorien zu klassifizieren, wird der Ansatz der multiplen Klassifikation oder multi-label Klassifikation gewählt. Dabei wird jeder Eintrag in den Ereignisdaten mit einer oder mehreren Kategorien versehen. Gegeben seien eine Menge von Ereignissen $E = e_1, e_2, \dots, e_n$ und eine Menge von Kategorien $K = k_1, k_2, \dots, k_m$. Es wird eine Klassifikationsfunktion f definiert, die jedem Ereignis e_i eine Menge von Kategorien aus K zuordnet:

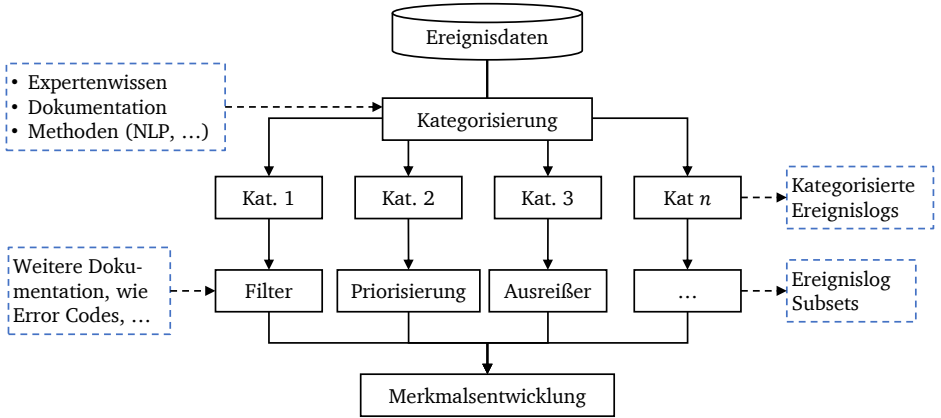


Abbildung 3.2.: Generische Vorverarbeitung der Ereignisdaten

$$f : E \rightarrow 2^K. \quad (3.1)$$

Dabei steht 2^K für die Potenzmenge von K , also die Menge aller möglichen Untermengen von K . Jedes Element in 2^K repräsentiert eine Kombination von Kategorien für ein Ereignis. Die Klassifikationsfunktion f kann beispielsweise als binärer Vektor dargestellt werden, in dem jedes Element des Vektors einer Kategorie entspricht. Ein Wert von 1 in einem Element des Vektors bedeutet, dass die entsprechende Kategorie dem Ereignis zugeordnet ist, während ein Wert von 0 bedeutet, dass die Kategorie nicht zugeordnet ist.

Für die initiale Festlegung der Kategorien bestehen mehrere Herangehensweisen. Experten einer Domäne oder Systemingenieure eines Systems kennen die Kategorien der Ereignislogs und können diese manuell definieren. Ist kein Domänenwissen über die Ereignislogs vorhanden, können Dokumentationen hinzugezogen oder alternativ Methoden aus der Kategorisierung und Gruppierung von Textdaten genutzt werden. Das Feld der Computerlinguistik (engl. Natural Language Processing (NLP)) liefert dabei Ansätze und Methoden zur Gruppierung unstrukturierter Texte, wie er in Ereignis-

nislogs häufig zu finden ist. Methoden wie Tf-idf¹ oder Doc2Vec² transformieren den Textkörper (Korpus) in einen Vektorraum und helfen dabei Wortähnlichkeiten zu identifizieren und zu gruppieren. Dieses Vorgehen kann insbesondere dann helfen, wenn kein Domänenwissen vorhanden bzw. keine Dokumentation verfügbar ist. Allerdings stoßen die Ansätze gerade im Kontext der technischen Anwendungen und bei sehr kurzen Dokumenten (in diesem Fall einzelne Logeinträge) schnell an ihre Grenzen. [Bon+]

Im Rahmen der Arbeit wird angenommen, dass eine entsprechende Kategorisierung gegeben ist, sodass der Prozess der Kategorisierung, sowie die Definition der Kategorien nicht Teil der Implementierung ist.

Auf die Kategorisierung der Ereignislogs folgt der Schritt der Bereinigung. Ähnlich wie bei der Vorverarbeitung von Zeitreihendaten, ist hierbei das Ziel das Signal-Rausch-Verhältnis zu erhöhen und die Aussagekraft der Daten für die Zustandsdiagnose zu verbessern. Aufgaben der Ereignisdatenbereinigung sind dabei: [Zhe+09; KBV18]

- Entfernen von sehr selten auftretenden Ereignissen.³
- Zusammenfassen von gleichen Einträgen, die in Eventlogs im selben Beobachtungsfenster sehr eng aufeinander folgen. Dies hat Folgen auf die Merkmalsentwicklung, weil statistisch basierte Merkmale verändert werden.
- Reduktion der Dimensionsvielfalt, indem ähnliche Ereignisse innerhalb einer Kategorie weiter zusammengefasst bzw. reduziert werden.
- Häufig auftretende Ereignisse haben oft wenig Aussagekraft und können somit entfernt werden.
- Entfernen von Ereignissen in aufeinanderfolgenden Beobachtungsfenstern. Dieser Schritt kann helfen, die statistische Signifikanz von Merkmalen zu erhöhen.
- Entfernen von mit den Zeitreihendaten hoch-korrelierten Einträgen (insbesondere dann relevant, wenn die Ereignislogs zusätzliche Sensorinformationen, wie z. B. Temperaturen oder Leistungen enthalten).

¹Term frequency - inverse document frequency (Tf-idf) ist ein Ansatz bei dem die Häufigkeit eines Wortes in einem Satz im Verhältnis zur Worthäufigkeit im gesamten Text analysiert wird. Darüber lässt sich schnell und einfach die Relevanz einzelner Wörter bestimmen.

²Doc2Vec (je nach Anwendung auch Word2Vec) sind flache Neuronale Netze, die als unüberwachte Lerner Wortassoziationen (bzw. Satzassoziationen bei Doc2Vec) innerhalb eines Textes lernen und als Vektorraum darstellen. [LM14; Mik+13]

³Seltene Ereignisse können ebenso gute Indikatoren für Anomalien und auftretende Schäden sein. Daher muss der Schritt situativ abgewogen werden.

Ergebnis der Vorverarbeitung der Ereignisdaten ist eine sortierte, kategorisierte und bereinigte Datenbasis. Dieser Schritt ist durchaus analog zur Vorverarbeitung von Zeitreihendaten zu interpretieren, lediglich mit anderen Methoden. Durch die sorgfältige Vorverarbeitung wird das Datenverständnis verbessert und weitere explorative Einblicke ermöglicht. Gleichzeitig sind die Daten für den folgenden Schritt der Merkmalsentwicklung aufbereitet.

3.2.3. Merkmalsentwicklung der Ereignisdaten

Im Schritt der Merkmalsentwicklung werden aufbauend auf den zuvor definierten und unterteilten Kategorien Merkmale generiert. Ziel ist es in diesem Schritt, die Ereignisdaten in Zeitreihendaten zu konvertieren (*time conversion*) und damit die Grundlage für die Fusion mit dem klassischen Ansatz zu schaffen. Bei dem hier vorgestellten datenbasierten Vorgehen werden zunächst temporale Muster in den Ereignisdaten identifiziert und im nächsten Schritt deren zeitliche Korrelation mit Ausfällen analysiert. Die Identifikation bzw. Kenntlichmachung temporaler Muster ist dabei Teil der Merkmalsentwicklung, während die Detektion von Korrelationen die Aufgabe des nachgelagerten Klassifikationsalgorithmus ist. Ein vereinfachtes Beispiel für ein solches Muster besteht aus den Ereignissen e_1 , e_2 und e^* :

Wenn die Ereignisse e_1 und e_2 im System auftreten
dann folgt darauf des Ausfallereignis e^*
mit einer (zeitlichen) Wahrscheinlichkeit P .

Für jede zuvor definierte Kategorie kommen dabei eine oder mehrere spezifische Extraktionstechniken zur Anwendung, um relevante Informationen zu extrahieren und die Daten in eine geeignete Form zu überführen. Im Gegensatz zum zuvor beschriebenen Schritt der Datenvorverarbeitung ist für den Schritt der Merkmalsentwicklung kein Expertenwissen notwendig. Die Ansätze werden generisch formuliert und sind prinzipiell ohne tieferes Systemverständnis anwendbar. Generell kann jede der im folgenden präsentierten Techniken auf jede Ereigniskategorie angewendet werden, allerdings können einzelne Kombinationen zu vorteilhaften Ergebnissen führen, während andere keinen Mehrwert liefern.

Prinzipiell werden vier Arten der Merkmalsentwicklung unterschieden: statistisch basierte Merkmale, sequenzbasierte Merkmale, fehlerähnlichkeitsbasierte Merkmale und profilbasierte Merkmale.

Statistisch basierte Merkmale

Bei statistisch basierten Merkmalen werden Ereignisse in einem (gleitenden) Beobachtungsfenster betrachtet und anhand von Häufigkeiten einzelner Ereignissen, Gruppen oder Distanzen statistische Merkmale gebildet. Nach Wang et al. wird unterschieden zwischen den einfachen statistischen Merkmalen (Anzahl der einzigartigen Ereignisse in einem Beobachtungsfenster) und aufwendigeren statistischen Merkmalen, wie der zeitliche Abstand eines Fehlers zum Ausfall- oder Vorhersagezeitpunkt. [Wan+17]

Bei der einfachen Form der Merkmale wird ein *Zeitfenster* betrachtet, innerhalb dessen Ereignisse oder Ereigniseigenschaften mit einer gewissen Häufigkeit (Frequenz) auftreten. Formal werden die Ereignisse $e_i \in E$, Ereigniskategorien $k_i \in K$ oder Distanzen zwischen Ereignissen/Kategorien d_j^i innerhalb des (gleitenden) Fensters W mit der Breite $w = t_e - t_s$ betrachtet. In Abbildung 3.3 ist exemplarisch eine Darstellung des gleitenden Fensters über drei Ereignisse gegeben.

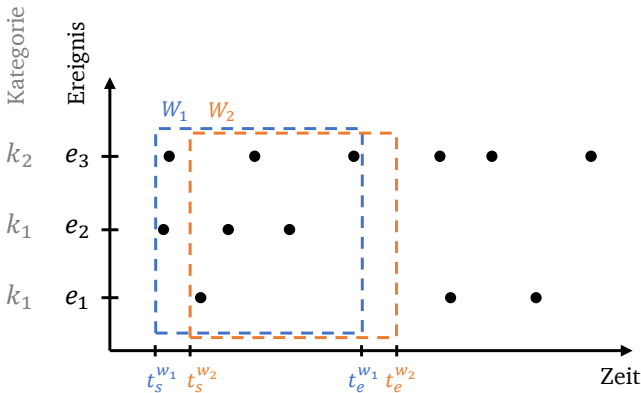


Abbildung 3.3.: Statistisch basierte Merkmale in einem gleitenden Fenster für Einzelereignisse

Dabei können einerseits die Ereignisse individuell oder Ereigniskategorien untersucht werden. Das einfachste Merkmal bildet dabei die absolute Häufigkeit des Ereignisses e_i im Zeitfenster w_j ab:

$$x_{stat_1}^{e_i} = \sum_{t=t_s}^{t_e} (e_i, t). \quad (3.2)$$

Äquivalent können die Häufigkeiten aller Ereignisse E im Zeitfenster W sowie die Häufigkeit aller Ereignisse innerhalb einer Kategorie k_j betrachtet werden:

$$x_{stat_2}^{k_j} = \sum_{t=t_s}^{t_e} (e_i, t), \quad e_i \in k_j. \quad (3.3)$$

Ergänzung der einfachen statistischen Merkmale ist die Betrachtung des zeitlichen Abstands zwischen Ereignissen. Dabei wird der mittlere Abstand zwischen den Ereignissen innerhalb eines Fenster inklusive den jeweils auf diese folgenden Ereignissen betrachtet. Es können ebenfalls einzelne Ereignisse, alle Ereignisse oder Ereignisse einer Kategorie betrachtet werden.

$$x_{stat_3}^{e_i} = \frac{1}{n} \sum_{e_i=e(t_s)}^{t_e} (t_{e_{i+1}} - t_{e_i}), \quad \text{mit } n = \sum e_i \text{ in } W. \quad (3.4)$$

Exemplarisch sind die auf Basis von Abbildung 3.3 generierten häufigkeitsbasierten, statistischen Merkmale tabellarisch in Tabelle 3.3 zusammengefasst.

Tabelle 3.3.: Beispiele statistisch basierter Merkmale auf Basis von Abbildung 3.3

	$x_{stat_1}^{e_1}$	$x_{stat_1}^{e_2}$	$x_{stat_1}^{e_3}$	$x_{stat_2}^{k_1}$	$x_{stat_2}^{k_2}$
W_1	3	3	1	4	3
W_2	2	2	1	3	2

Die Festlegung der Fensterbreite w obliegt dabei dem Systemexperten. Die Fensterbreite ist daher Gegenstand für spätere Parameterstudien der statistisch basierten Merkmale.

Sequenzbasierte Merkmale

Sequenzbasierte Merkmale stellen kausale Verbindungen zwischen einer Sequenz von auftretenden Ereignissen in Ereignislogs und einem Fehler bzw. Ausfall her. Eine Sequenz wird dabei als Folge von Ereignissen verstanden. In dieser Arbeit entspricht eine Sequenz den Ereignissen in einem Beobachtungsfenster. Durch die Methode des gleitenden Fensters entsteht eine Sequenzliste (oder -datenbank) mit einzelnen Sequenzen variierender Länge. Die Aufgabe bei der Erstellung von sequenzbasierten Merkmalen ist es, wiederholende bzw. relevante Subsequenzen in der Sequenzliste (bzw. -datenbank) zu finden. Diese werden häufig auch als *Episoden* bezeichnet [MTI97].

Formal betrachtet ist eine Sequenz eine Folge von Ereignissen in einem Zeitfenster, wobei jedem Ereignis ein Zeitpunkt des Auftretens zugeordnet ist. Bei einer Menge E von Ereignissen ist ein Ereignis ein Paar (e, t) , wobei $e \in E$ das Ereignis und t den Zeitpunkt seines Auftretens darstellt. Eine Ereignisfolge bzw. eine Sequenz s auf E ist ein Tripel (s, T_s, T_e) , wobei

$$s = \langle (e_1, t_1), (e_2, t_2), \dots, (e_n, t_n) \rangle$$

eine geordnete Folge von Ereignissen ist, so dass $e_i \in E$ für alle $i = 1, \dots, n$, und $t_i \leq t_{i+1}$ für alle $i = 1, \dots, n - 1$. Weiterhin sind T_s und T_e die Startzeit und die Endzeit, mit $T_s \leq t_i < T_e$ für alle $i = 1, \dots, n$. [MTI97]

Im Falle der Ereignisdaten kann die Länge einer Sequenz auf verschiedene Art und Weisen interpretiert und ausgelegt werden. Eine übliche Ausprägung ist dabei, dass eine Sequenz auf Tages, Wochen oder Monatsbasis gebildet wird, für eine komplette Anlage, Komponenten oder zuvor definierte Ereigniskategorien.

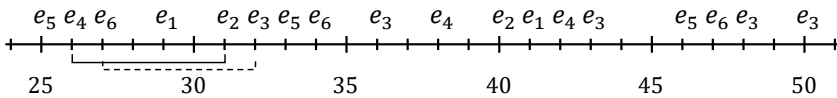


Abbildung 3.4.: Visuelle Darstellung einer Sequenz mit zwei Fenstern der Breite 5 in Anlehnung an [MTI97].

Bei der Analyse von Sequenzen sind relevante Subsequenzen (Episoden) von Interesse. Dabei wird erneut ein *Zeitfenster* betrachtet, innerhalb dessen einzelne Subsequenzen mit einer gewissen Häufigkeit (Frequenz) auftreten. Das Fenster auf einer Sequenz

$\mathbf{s} = (s, T_s, T_e)$ ist formal eine Subsequenz $\mathbf{W} = (s, t_s, t_e)$, die mit $t_s < T_e$ und $t_e > T_s$ und \mathbf{W} aus denjenigen Ereignissen von (e, t) aus s besteht, bei denen $t_s \leq t < t_e$. Die Zeitspanne $t_e - t_s$ wird als Fensterbreite w bezeichnet. Die Analyse von Episoden in definierten Fenstern limitiert somit die Anzahl und Länge der relevanten Subsequenzen.

Informell ist eine Episode eine Subsequenz einer Sequenz, die entweder als parallele (unter Vernachlässigung der Ereignisreihenfolge) oder serielle (unter Berücksichtigung der Ereignisreihenfolge) Episode bezeichnet wird. Episoden können als gerichtete azyklische Graphen beschrieben werden. Im Fall dieser Arbeit wird eine Episode auch als Ausfallsignatur bezeichnet, also eine Sequenz, die direkten Bezug zu einem Fehler oder Ausfall hat.

Sei X ein Ausfallereignis und T_r eine Subsequenz von Auslöseereignissen. Eine Episode (Subsequenz) $\mathbf{W} = (T_r, X, T_r < X)$ ist die Menge der Ereignisse T_r, X , die innerhalb von definierten Zeitintervallen auftreten und eine Gesamtreihenfolge $T_r < X$, die darstellt, dass die Auslöseereignisse T_r zeitlich vor X auftreten. Für die Ereignisse in $T_r : (T_r, \leq)$ kann eine partielle Ordnung \leq bestehen.

Die Fehlersignatur ist *injektiv*, wenn kein Ereignis zweimal oder öfter in der Signatur vorkommt. Die Länge L einer Episode \mathbf{W} ist die Anzahl der Ereignisse, die in der Ereignismenge T_r, X enthalten sind. Aus der Definition der Episode ergibt sich $L \geq 2$.

Basierend auf der partiellen Ordnung der Auslöseereignisse gibt es drei Arten von Episoden, die in Abbildung 3.5 dargestellt sind. Die Episode $\alpha = (\{e_1, e_2, X\}, \{e_1 < e_2 < X\})$ wird als *seriell* bezeichnet, wenn die Ordnungsrelation $e_1 < e_2$ der Auslöseereignisse e_1 und e_2 eine totale Ordnung ist. Die Episode $\Phi = (\{e_1, e_2, X\}, \{e_1, e_2 < X\})$ ist *parallel*, da die Ordnung von e_1 und e_2 trivial ist für alle $e_1 \neq e_2$. Bei den Fehlersignaturen α und Φ können wir sehen, dass die Fehlersignatur $\lambda = (\{e_1, X\}, \{e_1 < X\})$ eine Untersignatur ist, da sie in α und Φ enthalten ist. Die dritte Art von Fehlersignatur γ ist zusammengesetzt: sie wird als *komposit* bezeichnet und kann rekursiv aus den ersten beiden Typen konstruiert werden. [Li+07]

Im Rahmen der Arbeit sind vor allem serielle Sequenzen von Interesse, da parallele Sequenzen im Falle der kategoriebasierten Generierung schon durch die statistisch basierten Merkmale abgedeckt werden.

Bei einer festgelegten Fensterbreite w ist es möglich, viele verschiedene Episoden aus der Sequenz zu extrahieren. Neben der manuellen Bestimmung durch Systemexperten [HC94], zielen die meisten in der Literatur präsentierten Ansätze darauf ab Episoden zu identifizieren, die mit einer gewissen *Häufigkeit* auftreten. Formal gilt demnach, dass Sequenzen deren Support $\text{sup}(s)$ größer ist als eine Mindestschwelle

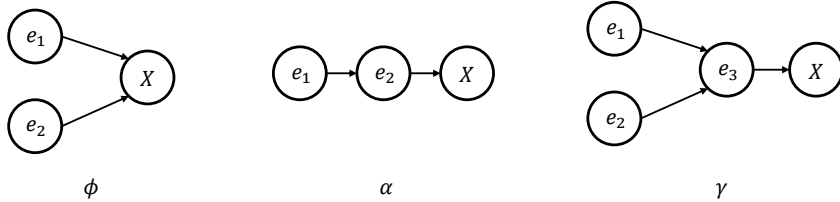


Abbildung 3.5.: Darstellung der drei Sequenzarten in Anlehnung an [Li+07]

$sup(s) \geq minsup$ von Bedeutung für die jeweilige Aufgabe sind. Bezogen auf die Detektion von Ausfallereignissen, werden häufig auch die Anzahl an Episoden mit einem folgenden Schadensereignis ins Verhältnis gesetzt zu jenen Episoden ohne Schadensereignis (siehe z. B. [Li+07]). Das übergeordnete Themengebiet ist dabei dem Forschungsgebiet des *Sequential Pattern Mining* zuzuordnen, das seit den 1990er Jahren insbesondere in der Analyse von Kaufverhalten im Retail-Bereich entsprechende Methoden liefert [AS94]. Die Aufgabe des Sequential Pattern Mining ist das Auflisten von allen möglichen Mustern, die mit einer vom Nutzer definierten Mindesthäufigkeit auftreten. Auch wenn es eine eindeutige Lösung für das Problem gibt, ist die Aufgabe der Identifikation mitunter rechenintensiv. Bei q Ereignissen in der Datenbank können bis zu $2^q - 1$ unterschiedliche Subsequenzen entstehen. Daher ist es notwendig effiziente Algorithmen zu entwerfen. In Philippe Fournier Viger et al. wird eine umfassende Übersicht über mögliche Ansätze und Algorithmen für das Problem gegeben.

Im Rahmen der Arbeit wird der „Prefix Span“ Algorithmus aus der Sequential Pattern Mining Framework (SPMF)⁴-Bibliothek von Philippe Fournier-Viger verwendet, der für das Problem des *Sequential Pattern Mining* den Benchmark darstellt. Durch die Möglichkeit den absoluten minimalen Support $minsup()$ sowie die maximale Länge der betrachteten Sequenzen zu definieren, lässt sich der Lösungsraum bereits effizient einschränken. [PEI+04; Fou+16]

Das Ergebnis der sequenzbasierten Merkmale ist eine Liste an relevanten Episoden, mit deren Häufigkeit im Beobachtungsfenster. Die Umsetzung im Merkmalsraum bzw. Input in den Algorithmus geschieht als One-Hot-Coding: jede relevante Episode wird als Merkmal in den Eingaberaum aufgenommen und bekommt im zutreffenden Beobachtungsfenster den Eintrag „1“, andernfalls den Eintrag „0“.

⁴SPMF ist eine in Java implementierte Open-Source Software und Data Mining Bibliothek. Sie ist spezialisiert auf Pattern Mining Probleme.

Fehlerähnlichkeitsbasierte Merkmale

Fehlerähnlichkeitsbasierten Merkmalen liegt die Annahme einer Beziehung zwischen sich wiederholenden Fehlern bzw. Ausfällen zugrunde. Es wird angenommen, dass diese Beziehung beschrieben werden kann. Dafür werden Einträge der Ereignislogs in fehlerbehafteten Beobachtungsfenstern und im aktuellen Beobachtungsfenster betrachtet. Es ergeben sich zwei Mengen/Kollektionen an Ereignissen. Das daraus gebildete Merkmal ist die Jaccard-Distanz, die die Ähnlichkeit der beiden Mengen und damit die Ähnlichkeit der Ereignisse in beiden Fenstern beschreibt. Formal sind zwei Kollektionen s_i, s^* an Ereignissen gegeben, wobei s^* eine schadhafte Sequenz darstellt. Bei beiden Sequenzen werden nur die einzigartigen Ereignisse berücksichtigt, alle Dopplungen werden entfernt. Die Jaccard-Distanz wird definiert als die Anzahl der gemeinsamen Elemente (Schnittmenge) geteilt durch die Größe der Vereinigungsmenge

$$J(s_i, s^*) = \frac{|s_i \cap s^*|}{|s_i \cup s^*|}$$

Für jedes vorhandene Schadensbild wird dabei ein eigenes Merkmal eingeführt. Eine Variation des Merkmals ist die Betrachtung von Kategorien der Ereignislogs anstelle einzelner Ereignisse. Die Herausforderung bei dieser Art der Merkmale ist es, die korrekten und relevanten Kollektionen an Ereignissen zu definieren bzw. zu finden. Diese Aufgabe kann - je nach untersuchtem Anwendungsfall oder Komplexität des Problems - zum Teil nur durch Experten umgesetzt werden.

Profilbasierte Merkmale

Die zuvor beschriebenen Merkmale sind stark vom aktuellen Diagnosezeitpunkt abhängig und verändern sich über die Zeit. Es sind damit so genannte temporale Merkmale [Zhe+09]. Daneben lassen sich übergeordnete, systemspezifische oder bei Mobilitätssystemen auch routenspezifische Merkmale ableiten. Für stehende Systeme sind das z. B. das Systemmodell oder Installationsdatum, bei Mobilitätssystemen sind das z. B. eine Flugroute oder Jahreszeit. Profilbasierte Merkmale entsprechen dabei nur bedingt der Definition von Ereignisdaten. Insbesondere bei stationären Maschinen sind diese Merkmale konstant und die Relevanz auf die Diagnose wird als vernachlässigbar eingestuft. Bei Mobilitätssystemen variiert dieses Merkmal jedoch häufig. So gibt es Untersuchungen der Korrelationen von Verschleißerscheinungen am Flugzeug mit

bestimmten Flugrouten, Flugzeugmustern oder Jahreszeiten in [CS18; TL] und von Umwelteinflüssen auf die Degradation von WKA in [RYM18].

3.3. Fusion von Ereignisdaten und Zeitreihendaten im Diagnoseschritt

Nachdem im vorherigen Abschnitt die aus den Ereignisdaten generierten Merkmale eingeführt wurden, beschreibt der folgende Abschnitt die Fusion mit Zeitreihendaten und die Zustandsdiagnose. Wie in Abbildung 2.9 dargestellt, erfolgt die Fusion erst im Schritt der Zustandsbewertung und -diagnose, nach der unabhängigen Vorverarbeitung und Merkmalsentwicklung.

Fusion von Ereignis- und Zeitreihendaten

Die Herausforderung bei der Fusion von Zeitreihendaten mit Ereignisdaten besteht im Vorhandensein von unterschiedlichen Datenstrukturen. Während Zeitreihendaten neue Messwerte für (alle) Variablen zu festgelegten, äquidistanten Zeitschritten liefern, folgen Ereignisdaten als unstrukturierte Daten keiner solchen Struktur. Prinzipiell gibt es Möglichkeiten mit dieser Herausforderung umzugehen: Die Datenstrukturen werden aneinander angepasst, oder algorithmische Ansätze werden gewählt, die mit unterschiedlichen Datenstrukturen umgehen können. Abschnitt 2.3.2 gibt einen Überblick über algorithmische Ansätze für diese Problemstellung.

Die Anpassung der Datenstrukturen ermöglicht die Verwendung von bereits bestehenden Diagnoseumgebungen und erfüllt damit die Anforderung FR4 in Tabelle 3.1, dass der vorgestellte Ansatz als Erweiterung (Add-On) nutzbar sein soll. Ein weiterer Vorteil ergibt sich darin, dass die Vielfalt der möglichen implementierbaren ML-Algorithmen und -Toolboxen bei einer Anpassung der Datenstruktur erhalten bleibt.

Aufgrund der Vorteile der Anpassung der Datenstrukturen, wird der letztere Ansatz gewählt. Entsprechend ist es notwendig die Ereignisdaten so zu adaptieren, dass sie mit den Zeitreihendaten fusionierbar sind. Dieser Schritt wird als *time-conversion* bezeichnet. Die Ereignisdaten, bzw. die daraus generierten Merkmale müssen so formuliert und adaptiert werden, dass sie sich nahtlos in die Zeitreihendaten einfügen lassen. Dabei muss zu jedem Zeitschritt der Zeitreihendaten ein neuer Wert für die Ereignisdaten generiert oder interpoliert werden. Diese Randbedingung wird erfüllt, indem die

Ereignisdaten in gleitenden Fenstern auf Basis des Zeitstempels der Zeitreihendaten betrachtet und Merkmale in diesen Fenstern generiert werden. Der Abstand zwischen den gleitenden Fenstern bzw. die Updaterate der Merkmale entspricht dabei genau der Abtastfrequenz der Zeitreihendaten. Entsprechend bildet jeder Zeitschritt der Zeitreihendaten ein neues Beobachtungsfenster für die Generierung der Merkmale aus den Ereignisdaten.

Die Ausgangsbasis und Grundlage für den Vergleichsansatz ist dabei der aus Zeitreihendaten bestehende Eingaberaum $\mathbf{X} = X(t)$ mit $X(t) = x_1(t), x_2(t), \dots, x_n(t)$ Sensordaten und einer abhängigen Variable (*Target*) $\mathbf{y} = y(t)$ mit $y \in \{0, 1\}$, die das Label darstellt. Der durch die Ereignisdaten aufgespannte Merkmalsraum wird als Vektor $\mathbf{Z} = Z(t)$ bezeichnet. Er beinhaltet die Merkmale basierend auf Ereignisdaten $\mathbf{Z} = [G \ H \ I \ J]$, mit statistisch basierten Merkmalen $G(t)$, sequenzbasierten Merkmalen $H(t)$, fehlerähnlichkeitsbasierten Merkmalen $I(t)$ und profilbasierten Merkmalen $J(t)$. Ein Merkmal G wird formal innerhalb eines Fensters W mit dem Startzeitpunkt Zeitpunkt t_s gebildet und lautet demnach $G(W(t_s))$. Im Rahmen der Arbeit wird das Merkmal vereinfacht als $G(t)$ beschrieben.

Tabelle 3.4 stellt die Erweiterung des Eingaberaums bzw. Zusammenführung der beiden Datenstrukturen anhand eines Beispiels dar. Die Label in der letzten Spalte der Tabelle bleiben unverändert, sodass der Trainings- und Testprozess für den Algorithmus bestehen bleibt.

Tabelle 3.4.: Darstellung des Fusionsschritts anhand des Eingaberaums

					Ereignismerkmale				y
	x_1	x_2	...	x_n	G	H	I	J	
t_1	$G(W_1)$	$H(W_1)$	$I(W_1)$	$J(W_1)$	0
t_2	$G(W_2)$	$H(W_2)$	$I(W_2)$	$J(W_2)$	0
...
t_n	$G(W_n)$	$H(W_n)$	$I(W_n)$	$J(W_n)$	1

Der Schritt der Datenfusion geschieht somit nach der unabhängigen Entwicklung der Merkmale und vor der Zustandsdiagnose, direkt beim Einlesen der Daten in das Diagnosemodell. Die Vorverarbeitung der Zeitreihendaten und die eigentliche Modellbildung im Diagnoseansatz bleiben unabhängig von der Merkmalsentwicklung der Ereignisdaten und ermöglichen damit eine akkurate Beurteilung des Ansatzes.

Zustandsbewertung und Diagnose

Der fusionierte Eingaberaum, bestehend aus Zeitreihendaten und Merkmalen der Ereignisdaten, ist Input für das Diagnosemodell. Die Problemstellung wird dabei als Klassifikationsansatz mit binärer Zustandsdiagnose des Gesundheitszustands interpretiert (vgl. Abschnitt 2.2.3). Mit einem ML-Klassifizierungsalgorithmus lässt sich die binäre Zielgröße (Label) *gesund* oder *schadhaft* vorhersagen, bzw. diagnostizieren. Bei mehreren möglichen Fehlern oder Komponenten werden mehrere Modelle verwendet.

Generell sind für diesen Schritt verschiedene Klassifikationsalgorithmen einsetzbar. Die Auswahl hängt von den spezifischen Anforderungen und vom Anwendungsfall ab. Im Allgemeinen ist es von Vorteil, verschiedene Algorithmen miteinander zu vergleichen. Zum Beispiel erfordern einfache unüberwachte Cluster-Algorithmen wie der kNN-Algorithmus die gesamte zugrunde liegende Datenbasis, um die Klassenzugehörigkeit basierend auf der Distanz zu den k nächsten Punkten zu bestimmen. Im Gegensatz dazu verwenden Entscheidungsbäume einen trainierten Regelsatz, der die Trainingsdatenbasis widerspiegelt. Das bedeutet, dass nicht die gesamte Datenbasis benötigt wird, um Klassifikationen durchzuführen, was den Rechenaufwand erleichtert. Entscheidungsbäume und kNN sind nur zwei von vielen Beispielen. In der Literatur gibt es zahlreiche weitere Algorithmen (vgl. auch Abschnitt 2.2.3), die je nach Komplexität des Systems, der Anzahl der Klassen, dem Umfang der Daten und der gewünschten Art der Ausgabe - zum Beispiel der reinen Klassenausgabe im Vergleich zur Ausgabe der Wahrscheinlichkeiten für jede Klasse - gegeneinander abgewogen werden müssen. [Agg15; Bis09]

3.4. Vorgehen zur Evaluation

Nachdem das Gesamtkonzept der Dissertation und die dafür notwendigen Methoden in diesem Kapitel dargestellt wurden, beschreibt der vorliegende Abschnitt das Vorgehen zur Evaluation. Eine grafische Darstellung des Prozesses ist in Abbildung 3.6 gegeben. Dabei bestehen mehrere Stränge, mit den beiden Datenquellen als Ausgangsbasis. Die Ereignisdaten werden wie beschrieben vorverarbeitet und Merkmale generiert. Auf dieser Basis findet ebenfalls eine Merkmalselektion sowie eine Hyperparameteroptimierung des verwendeten Algorithmus statt. Die Plausibilisierung der Merkmalsentwicklung geschieht durch die Klassifikation auf Basis der Ereignisdaten. Parallel dazu wird ein Vergleichsansatz (Benchmark) auf Basis der Zeitreihendaten aufgebaut. Die Ergebnisse der dabei durchgeführten Hyperparameteroptimierung werden

mit den bereits gefundenen Hyperparametern abgeglichen und beide Merkmalsräume fusioniert. Der Vergleich der Metriken von Vergleichsansatz und fusioniertem Ansatz (vorgeschlagenes Konzept) ermöglicht dabei abschließend eine Quantifizierung der Güteveränderung durch die zusätzliche Berücksichtigung von Ereignisdaten. Der nachfolgende Abschnitt beschreibt das Vorgehen zur Evaluation und führt die notwendigen Metriken und Darstellungsformen ein.

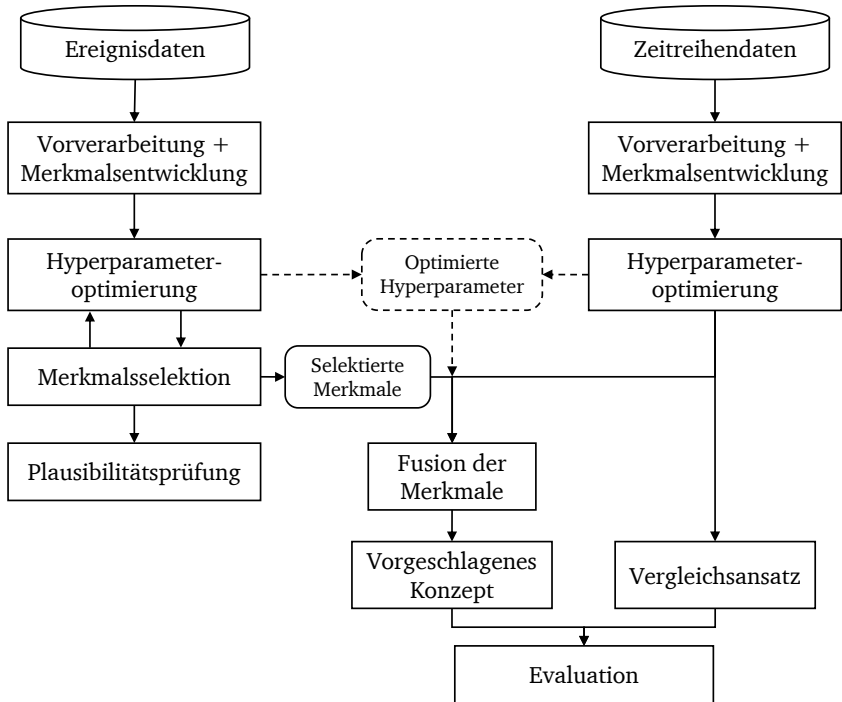


Abbildung 3.6.: Vorgehen zur Evaluation des vorgeschlagenen Konzepts

3.4.1. Vergleichsansatz

Der Vergleichsansatz stellt den aktuellen Stand der Technik in der Zustandsdiagnose dar. Der Modellaufbau entspricht dem in Abschnitt 3.3 dargestellten Klassifikationsan-

satz und ausschließlich dem rechten Pfad in Abbildung 2.9. Der Unterschied besteht vor allem in der Datenbasis, auf der das Modell trainiert und getestet wird. Eingabe in das Modell sind dabei ausschließlich Zeitreihendaten. Die genauen Ansätze zur Datenvorverarbeitung und Merkmalsentwicklung sind dabei abhängig vom exakten Anwendungsfall und werden für den betrachteten Anwendungsfall in Kapitel 4 ausgeführt. Um eine Vergleichbarkeit der Ergebnisse zu erzielen, wird die Aufteilung von Trainings- und Testdaten beim Vergleichsansatz und fusionierten Ansatz identisch durchgeführt. D.h. der Gesamtdatensatz wird für beide Fälle gleichermaßen zeitlich, oder nach Komponenten, separiert. Ebenfalls bleibt der algorithmische Klassifikationsansatz gleich. Zwar erfordert jede Trainingsdatenbasis einen individuellen Trainingsprozess, wodurch unterschiedliche Modellparameter (z. B. Gewichte eines kNN) entstehen, das Grundgerüst inklusive der Hyperparameter soll jedoch weitgehend gleich bleiben. Damit wird eine differenzierte und bezüglich der Bewertung der Eingabedaten akkurate Grundlage zur Evaluation gewährleistet.

3.4.2. Metriken und Darstellungsform

Die Evaluation des beschriebenen Ansatzes erfolgt anhand gängiger Metriken für die Beurteilung von Klassifikationsalgorithmen. Dabei werden die Metriken aus Abschnitt 2.2.4 verwendet. Aufbauend auf zwei Konfusionsmatrizen, je eine für den Vergleichsansatz und den fusionierten Ansatz, lassen sich immer die Metriken der beiden Modelle vorgeschlagenes Konzept und Vergleichsansatz direkt miteinander vergleichen. Genutzt werden dabei die üblichen Metriken der Accuracy, Recall, Precision und f1-Score. Bei Parametervariationen lassen sich ebenfalls Mittelwerte, Mediane und Streuungen mehrerer Modellergebnisse vergleichen. Die Ausführung ist in Abschnitt 4.5 ersichtlich.

4. Evaluation des vorgeschlagenen Konzepts am Beispiel von Windkraftanlagen

Im vorliegenden Kapitel wird das vorgeschlagene Konzept auf einen realen Datensatz angewendet und daran evaluiert. Der Aufbau des Kapitels entspricht dabei dem in Abbildung 3.6 dargestelltem Vorgehen. Ausgehend von der Identifikation geeigneter Datengrundlagen in Abschnitt 4.1.1 wird der genutzte Datensatz in Abschnitt 4.1.2 eingeführt und beschrieben. Die Abschnitte 4.2 und 4.3 beschreiben daraufhin die Ausführung der in Abschnitten 3.2.2 und 3.2.3 eingeführten Methoden zur Vorverarbeitung und Merkmalsentwicklung. Abschließend werden in Abschnitt 4.5 die Ergebnisse präsentiert und das vorgeschlagene Konzept daran evaluiert.

4.1. Identifikation und Beschreibung der Datengrundlage

4.1.1. Anforderungen an die Datengrundlage

Um das vorgeschlagene Konzept prinzipiell untersuchen und evaluieren zu können, werden Anforderungen an die Datengrundlage formuliert. Einige der übergeordneten Anforderungen im Hinblick auf die Daten sind bereits in Tabelle 3.1 in Abschnitt 3.1.2 beschrieben und müssen erfüllt sein, um das Konzept prinzipiell umsetzen zu können. Weiterhin sind die dabei genannten Anforderungen relevant um eine adäquate Evaluation des Konzepts durchführen zu können. Die wichtigsten Anforderungen dafür sind das Vorhandensein von Zeitreihen- und Ereignisdaten sowie bekannte Ausfälle oder Fehler um damit ein Labeling der Daten zu ermöglichen. Die Ausfälle sollten dabei ungeplant sein und die Gesamtdatenbasis hinsichtlich Volumen und Anzahl der

Ausfälle ausreichend für den Aufbau von datenbasierten Modellen sein. Erfahrungen aus der Forschungspraxis zeigen, dass ein Zeitraum von mindestens einem Jahr bei gegebenen Ausfallereignissen als ausreichend angesehen wird.

Aufgrund der Beschaffenheit der Daten und der gegebenen Anforderungen, insbesondere der Verfügbarkeit von Ereignisdaten und Ausfällen über einen Zeitraum von mehreren Jahren, kann eine geeignete Datengrundlage praktisch nur im operativen Betrieb realer Anlagen identifiziert werden. Die Erzeugung einer geeigneten Realdatenbasis im Rahmen der Promotion an der Universität ist dabei nicht realistisch. In dieser Hinsicht besteht eine Abhängigkeit von Projektpartnern, die jedoch zum Teil erheblichen Restriktionen in der Verwendung und Weitergabe von Daten auferlegen. Aus diesem Grund wird für die Evaluation auf Opensource Datensätze zurückgegriffen. Gängige Online-Communities¹ stellen viele Datensätze für die Entwicklung unterschiedlichster datenbasierter Methoden zur Verfügung. Auf Grundlage einer umfangreichen Online-Recherche konnte ein öffentlicher Datensatz identifiziert werden, der die Anforderungen erfüllt und unter CC BY-SA 4.0 frei zugänglich ist. [Ene18]

Im Folgenden wird der Datensatz vorgestellt und für die Evaluation des Konzepts aufbereitet.

4.1.2. Beschreibung der Datengrundlage

Die Datenbasis für die Evaluation des Konzepts wurde von *Energias de Portugal* erstmalig im Rahmen des Hackathons *Hack the Wind 2018* zum *Global Wind Summit* in Hamburg veröffentlicht und enthält mehrere Datensätze von WKA. Seitdem ist der Datensatz über das Open Data Portal² von *Energias de Portugal* frei verfügbar.

Im Datensatz sind insgesamt Daten von fünf Windkraftlagen eines Onshore-Windparks mit 2.000 Kilowatt Nennleistung über einen Zeitraum von zwei Jahren (2016-2017) enthalten. Ausführlichere Angaben zu der technischen Spezifikation sind in Tabelle A.2 in Abschnitt A.1 im Anhang gegeben. Der genaue Standort der Anlagen ist abstrahiert, der relative Abstand der Anlagen und einer Wetterstation zueinander ist bekannt. Neben den Daten der Wetterstation sind folgende Kerndatenbanken im Datensatz enthalten: operationelle Betriebsdaten (engl. Supervisory Control and Data Aquisition (SCADA)), Ereignisdaten/Eventlogs und Ausfalldaten. Die enthaltenen Daten sind in nachfolgender Tabelle 4.1 kurz beschrieben.

¹z. B. <https://www.kaggle.com>, aufgerufen am 14.07.2023

²<https://www.edp.com/en/innovation/open-data>, aufgerufen am 14.07.2023

Tabelle 4.1.: Informationen zum verwendeten Evaluationsdatensatz

Datensatz	Charakteristika
Wetterdaten	Windgeschwindigkeit und -richtung; Umgebungstemperatur und Luftdruck; Feuchtigkeit
Sensordaten (SCADA)	Generatordrehzahl, -temperatur und -leistungen; Getriebetemperaturen (Öl und Lager); Zellentemperatur; Aktive und Reaktive Leistungen; Pitchwinkel, ...
Ereignisdaten	Ereignisse aus verschiedenen Kategorien, Statusinformationen, Sensordaten, operationelle Zustände, und viele weitere.
Ausfalldaten	Zeitpunkte und kurze Informationen der 28 Ausfälle und Fehler an verschiedenen Komponenten (Generator, Generatorlager, Getriebe, Hydraulik und Transformator)

Neben den technischen Spezifikationen und einer Leistungskurve der WKA ist keine detailliertere Information oder Dokumentation gegeben. Während die Betriebsdaten mit Sensornamen und dem typischen Aufbau einer Windkraftanlage eine gute Interpretationsbasis ermöglichen, ist dies insbesondere bei den Ereignisdaten nicht gegeben. Basierend auf den gegebenen Spezifikationen der Anlage (vgl. Tabelle A.2) wird vermutet, dass es sich bei den Anlagen um *Vestas V90* handelt. Auf dieser Grundlage lassen sich im Internet einige nicht vertrauliche Datenblätter finden, die zum Teil hilfreiche Informationen für die Interpretation der Ereignisdaten liefern.

Die nachfolgende kurze deskriptive und explorative Beschreibung gemäß des Schritts *Datenerfassung* des OSA-CBM in Abbildung 2.5 ist in die drei Datenbanken Betriebsdaten, Ereignisdaten und Ausfalldaten untergliedert.

Betriebsdaten - SCADA Daten

Die aufgezeichneten Betriebsdaten der Windkraftanlagen kommen aus dem SCADA-System. Dabei werden neben Zeitstempel und Anlagen-ID Sensorwerte in zehnmütigen Intervallen abgespeichert. Bei fünf Anlagen und einem Zeitraum von zwei Jahren entspricht dies theoretisch 525.600 Datenpunkten. Für jeden Sensor fallen im entsprechenden Intervall Mittelwerte (Average) an, für einige zusätzlich auch Minimal- und Maximalwerte sowie die Standardabweichung. Insbesondere Temperaturen, aber auch einige der Leistungsdaten enthalten nur Mittelwerte. Es sind 54 einzelne Signale

enthalten, die den Kategorien Generator, Hydraulik, Rotor, Rotorblätter, Zelle, Umwelt, Transformator, Netz, Regelung und Produktion zugeordnet werden können. Aus den 54 Signalen werden insgesamt 81 Messwerte (durch Minimal-, Maximalwert und Standardabweichung) gebildet. Temperaturen sind als ganzzahlige Werte aufgenommen, der Rest als Gleitkommazahlen. Einige Messwerte aus der Netzkategorie (z. B. Netzspannung und Netzfrequenz) können als quasi-statisch angenommen werden. [Boy10]

Um eine spätere Übertragbarkeit der Diagnosemodelle zwischen verschiedenen Anlagen durchführen zu können und die Gesamtdatenbasis als Trainingsdatenbasis verwenden zu können, wird die Ähnlichkeit der Datengrundlagen der einzelnen WKA untersucht. In Abbildung 4.2 werden die Verteilungen der Generatorlagertemperatur aller Anlagen gezeigt. Es werden dabei die Verteilungen der Mittelwerte dargestellt, um eine mögliche Verzerrung zu reduzieren.

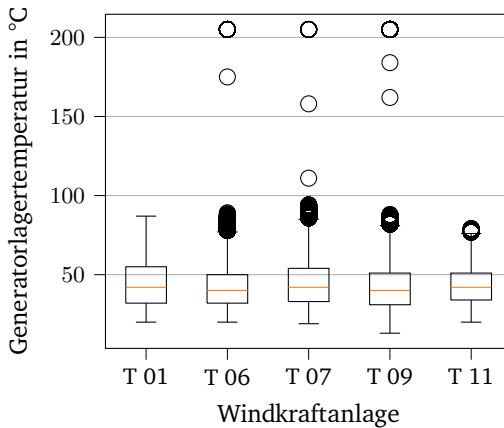


Abbildung 4.1.: Verteilung der Generatorlagertemperatur

Für die meisten der 54 Sensorwerte zeigen sich dabei keine signifikanten Abweichung der einzelnen Anlagen. Nur die Durchschnittswerte der Temperaturen des Hochvolttransformators, des Hydrauliköls und des Getriebelagers zeigen Auffälligkeiten. Bei den Verteilungen der durchschnittlichen Getriebelagertemperatur zeigt sich, dass die Verteilung von WKA T11 höhere Maximalwerte aufweisen und die gesamte Temperaturverteilung verglichen mit den restlichen Anlagen nach oben verschoben ist. Eine Visualisierung in Form von Boxplots ist in Abbildung 4.2 gegeben. Das spiegelt sich

auch in den statistischen Analysen wider: Der Mittelwert über den gesamten Datensatz liegt um $4,1^{\circ}\text{C}$ höher als bei den restlichen WKA.

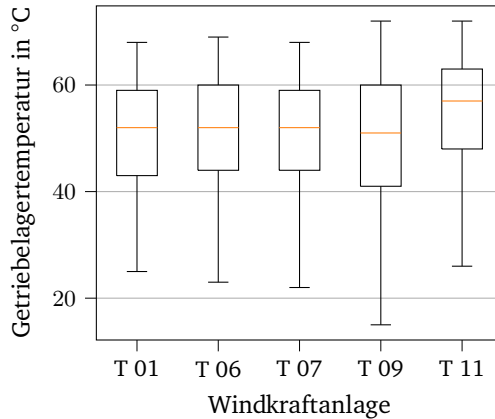


Abbildung 4.2.: Verteilung der Getriebelagertemperatur

Bei der Betrachtung der Datenreihen über die Zeit, ergibt sich für WKA 11 kein zeitlicher Trend oder Ausschläge der gemessenen Temperaturwerte. Vielmehr ist hier ein konstanter Offset von Anlage 11 gegenüber den anderen Anlagen ersichtlich.

Ein ähnliches Bild zeigt sich für die durchschnittlichen Getriebeöltemperaturen von WKA 9, die im Mittel $10,7^{\circ}\text{C}$ höher liegen als der Rest. Bei den Phasentemperaturen des Hochvolttransformators zeigt sich hingegen, dass die Temperaturen der Anlagen 6 und 11 im Durchschnitt über alle drei Phasen um $16,7^{\circ}\text{C}$ niedriger sind als die der restlichen Anlagen. Eine Korrelation der Differenzen mit auftretenden Schäden ist nicht ersichtlich.

Die zum Teil signifikanten Differenzen bei einzelnen Sensoren zwischen den Anlagen motivieren die Modellbildung von unabhängigen Modellen für jede WKA. Durch den vorliegenden Datensatz und die Verfügbarkeit mehrerer „Prüflinge“ (WKA) können mit moderater Belastbarkeit Aussagen zur Robustheit des Ansatzes getätigt werden. Dabei kann aus diesem Sachverhalt eine mögliche Untersuchungshypothese abgeleitet werden: Die Vergleichsansätze, die auf SCADA-Daten aufgebaut werden, müssen aufgrund unterschiedlicher Ausprägungen der Sensoren bei den Anlagen, jeweils pro WKA, aufgebaut werden. Dabei wird vermutet, dass die Modelle nur eingeschränkt

untereinander übertragbar sind. Das vorgeschlagene Konzept könnte beim Aufbau eines Gesamtmodells hingegen für alle Anlagen robuster sein als der Vergleichsansatz.

Ereignisdaten

Die Ereignisdaten enthalten neben einer Spalte mit dem eigentlichen Ereignisseintrag einen Zeitstempel, Anlagen-ID und optionalen Reset-Zeitstempel. Zu den Ereignisdaten ist keine Dokumentation verfügbar, sodass die Interpretation der Einträge auf Basis der deskriptiven Analyse und Online-Recherchen geschieht.

Alle fünf Windkraftanlagen umfassen im genannten Zeitraum von zwei Jahren insgesamt 318.835 Ereignisseinträge. Die Aufteilung zwischen den Anlagen ist dabei relativ gleichmäßig bei rund 61.000 Einträgen für WKA 7, 9 und 11. Die Anlagen 1 (76.485 Einträge) und 6 (54.068 Einträge) weisen etwas mehr bzw. weniger Einträge auf. Weiterhin lassen sich alle Einträge 11.751 *einzigartigen* Ereignissen zuordnen. Die hohe Vielfalt entsteht dabei durch numerische Werte in den Texteinträgen. Einen Großteil der Gesamtereignisse nehmen dabei mit 138.672 Vorkommen die Meldung „External Power Ref.:2000kW“ und mit 29.288 Vorkommen die Meldung „Generator 1 in“ ein. Da beide Meldungen für den Sachverhalt der Gesundheitsdiagnose keinen Mehrwert liefern, werden sie im Rahmen der Datenbereinigung (siehe Abschnitt 4.2) entfernt. Ein exemplarischer Auszug der Ereignisdatenbank für WKA 1 ist in Tabelle 4.2 gegeben.

Im Beispiel zeigen die ersten vier Einträge Meldungen mit Sensor- bzw. Leistungsinformationen. Insbesondere der zweite Eintrag mit Leistungsinformationen ist dabei alle 30 Minuten zu finden. In Abschnitt 4.2 werden Korrelationen dieser Einträge mit den SCADA-Daten untersucht und hoch korrelierte Einträge entfernt um Redundanzen zu vermeiden und so den Eingaberaum in das Diagnosemodell möglichst klein zu halten. Die Zeilen 6 bis 12 lassen ohne weitere Recherche nur Spekulationen des Sinngehalts zu. Nach intensiver Onlinerecherche lässt sich hierfür eine Interpretation finden. So deuten Zeilen 6 bis 8 auf (manuelle) Eingaben und die Übernahme der manuellen Kontrolle hin. Für die Zeilen 10 und 11 konnten im Internet Verweise auf Wartungsereignisse gefunden werden. Der Eintrag „Pause pressed on keyboard“ ist dabei der Hinweis auf das Wartungsereignis, der darauffolgende Eintrag referenziert mit dem Code „900“ auf „Pause pressed on keyboard“ und gibt eine Leistungsinformation. Ebenso ist ein Reset-Zeitstempel gegeben, zu dem die Anlage aus dem Wartungsmodus in den Betrieb startet. Die dazugehörige Sequenz ist in Zeile 14 bis 18 zu sehen. Sequenzen wie in diesem Beispiel können durch explorative Analyse identifiziert und als Teil der Datenbereinigung vereinfacht bzw. entfernt werden. Die relevante Information in diesem

Tabelle 4.2.: Auszug der EDP Ereignisdaten

Zeile	Zeitstempel	Reset	ID	Ereignis
1	2016-01-02T04:38:29		T01	GearoilCooler 2, gear: 57°C
2	2016-01-02T06:38:11		T01	External power ref.:2000kW
3	2016-01-02T06:40:24		T01	Yaw Speed Exc: 0° Rate:10sec
4	2016-01-02T06:56:43		T01	GearoilCooler 1, gear: 49°C
5
6	2016-01-20T12:48:53		T01	Key switch: Local/Remote
7	2016-01-20T12:48:53		T01	User 0 primary access
8	2016-01-20T12:48:54		T01	Key switch: Local control
9	2016-01-20T12:48:55	13:40	T01	Pause pressed on keyboard
10	2016-01-20T12:48:56		T01	Extra info. Err:900 P: -8.3kW
11	2016-01-20T12:48:56		T01	Pause -8.3kW 131RPM
12	2016-01-20T12:48:59		T01	New SERVICE state: 1, 0
13
14	2016-01-20 13:40:26		T01	New SERVICE state: 0, 0
15	2016-01-20 13:40:32		T01	Run
16	2016-01-20T13:42:13		T01	Key switch: Local/Remote
17	2016-01-20T13:42:16		T01	Key switch: Remote control
18	2016-01-20T12:56:46		T01	Generator 1 in

Fall ist, dass die Zeilen 6 bis 18 auf ein Wartungsereignis von rund einer Stunde Dauer hinweisen. Allerdings kommt der Eintrag „Pause pressed on keyboard“ auch in anderen Fällen vor: Nur 402 von insgesamt 808 Vorkommen haben ebenfalls einen Eintrag in der ResetTime. Das resultiert daraus, dass in einigen Fällen der Eintrag mehrfach direkt hintereinander mit nur wenigen Sekunden Abstand geloggt wird.

In Abbildung 4.3 ist eine exemplarische Darstellung der nicht bereinigten Ereignisse über die Zeit für WKA T01 gegeben. Die Darstellung zeigt die hohe Dimensionalität der Daten, sowie das unregelmäßige Auftreten. Gleichzeitig ist eine Häufung der Logs zum Ende des beobachteten Zeitraums ersichtlich, die sich neben Anlage T01 auch bei T06 und T09 zeigt. Laut Expertenangaben können Softwareupdates der Firmware solche drastischen Änderung im Loggingverhalten bedeuten [SKS22]. Da für den Datensatz keine solche Information gegeben ist und sich die Charakteristika der Daten deutlich verändert, wird im weiteren Verlauf der Datensatz nur bis zum 12.12.2017 betrachtet (vgl. Abschnitt 4.2).

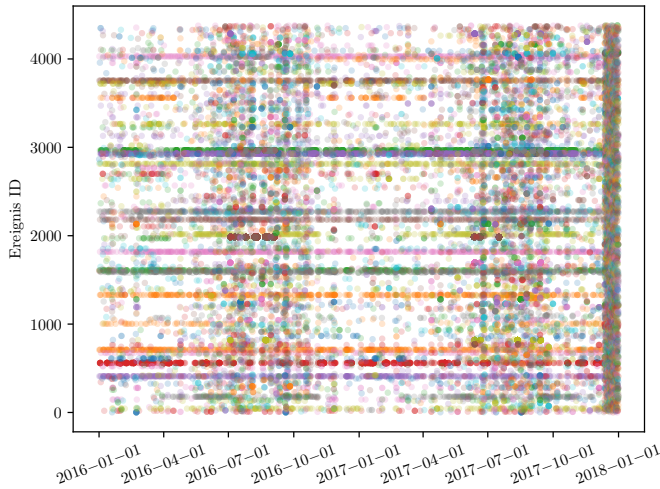


Abbildung 4.3.: Darstellung der Ereignisdaten am Beispiel von T01

Ausfalldaten

In den Ausfalldaten liegen insgesamt 28 Einträge von Fehlern, Ausfällen und Komponentenwechsel vor. Die Einträge teilen sich dabei auf die Komponenten Generator, Generatorlager, Getriebe, Transformator und Hydrauliksystem auf. Neben der Information, welche Anlage und Komponente betroffen ist, wird der Zeitstempel und eine kurze Beschreibung gegeben. Während es in der Luftfahrtindustrie üblich ist, bei Wartungseinträgen zwischen der Beschreibung des Problems (engl. „Description“) und der daraufhin ausgeführten Handlung (engl. „Action“) zu differenzieren, ist diese Unterscheidung in den verwendeten Ausfalldaten nicht gegeben.

Die Aufteilung von Vorkommnissen zwischen Komponenten und WKA ist in Tabelle 4.3 gegeben, sowie grafisch über die Zeit in Abbildung 4.4 dargestellt. Die Gesamtübersicht der Einträge mit Zeitstempel ist in Tabelle A.3 im Anhang dargestellt. Auffällig ist hierbei, dass Anlage 01 nur zwei Einträge aufweist (Ölpumpe des Getriebes und Trafolüfter defekt), während Anlage 06 alleine fünf Einträge am Generator aufweist. Dieser zeigt im Zeitraum von Juli bis Oktober 2016 mehrere Auffälligkeiten und wurde

zwei mal getauscht.

Tabelle 4.3.: Übersicht der aufgetretenen Fehler und Anlagen

Komp.	WKA					Σ
	T01	T06	T07	T09	T11	
Generatorlager	-	-	2x	4x	-	6
Generator	-	5x	1x	-	1x	7
Getriebe	1x	1x	-	2x	-	4
Transformator	1x	-	2x	-	-	3
Hydraulik	-	2x	2x	1x	3x	8
Σ	2	8	7	7	4	28

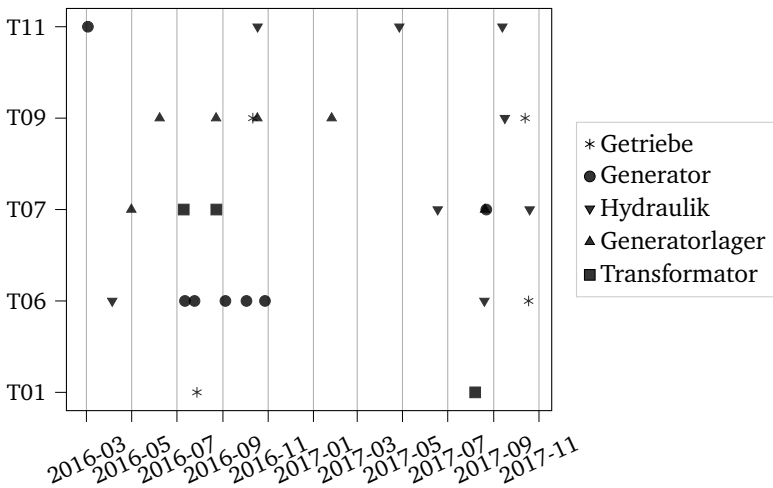


Abbildung 4.4.: Darstellung der aufgetretenen Fehler nach Komponente und Anlage

Die Spezifität der gegebenen Informationen in den Ausfalldaten variiert dabei: Neben spezifischen Angaben zu Wechseln von Komponenten (z. B. Generator(-lager) oder Temperatursensor getauscht) sind auch viele unspezifische Einträge vorhanden: „Getriebepumpe beschädigt“ ist bspw. zunächst nur eine Diagnose und gibt keinen

Aufschluss über die daraus resultierende Handlung. Zwar ist damit der Zeitpunkt der Feststellung gegeben, jedoch nicht der Zeitpunkt der Reparatur. Es findet sich hierzu kein passender bzw. dazugehöriger Eintrag. Ein gegensätzliches Beispiel ist der Eintrag „Getriebe repariert“, der keinen Aufschluss über den Zeitpunkt des ersten Feststellens gibt. Andere Beispiele sind die Einträge „Getriebelager hohe Temperatur“, „Fehler in der Pitch-Regelung“, oder „Hydraulikfehler im Bremskreis“, die alle in die Kategorie der Beschreibung des Problem einzuordnen sind (ohne dazugehörige Handlung). Die Ausfalldaten sind für die Arbeit die Grundlage zum Labeln der Daten, d.h. zum Definieren von Zeitabschnitten die später als fehlerhaft klassifiziert werden sollen. Dieser Zusammenhang wird in der Literatur auch als schwach gelabelte (engl. *weakly labeled*) Daten bezeichnet. Damit sind Label gemeint, bei denen zwar der Ausfallzeitpunkt bekannt ist, nicht aber vorgehende Symptome oder Ursachen. Aus diesem Grund wird folgende Annahme ergänzt. [BKS22]

Annahme A5 Um der starken Klassenimbalance entgegenzuwirken wird angenommen, dass alle enthaltenen Ausfallereignisse einen definierten Zeitraum *vor* dem eigentlichen Ereignis vorhanden und erkennbar sind. D.h. es wird angenommen, dass bereits x Tage vor dem eigentlich Fehlerereignis ein schadhafter Zustand vorliegt und damit ein positives Label gegeben ist. Das entsprechende Zeitfenster wird als w_{fail} definiert und wird im Rahmen der Sensitivitätsanalyse variiert.

Die detaillierte Betrachtung der Einträge inklusive die kritische Auseinandersetzung mit den entsprechenden Texteinträgen wird im Rahmen der Diskussion der Ergebnisse in Kapitel 5 vorgenommen.

4.2. Datenvorverarbeitung der Ereignisdaten

Gemäß dem in Abschnitt 3.2.2 dargestellten Vorgehen ist der erste Schritt bei der Datenvorverarbeitung die Kategorisierung der Ereignisdaten. Da für die vorliegende Datenbasis neben einer Liste mit einigen Alarm- und Fehlercodes keine weiteren Informationen oder Dokumentationen vorliegen, wird die Kategorisierung auf Basis explorativer Methoden durchgeführt. Dazu werden die Daten unter anderem gemeinsam mit den Ausfalldaten betrachtet und erste Relationen untersucht. Das Vorgehen erhöht dabei gleichzeitig das Datenverständnis und kann ebenfalls als Ergänzung zum vorherigen Abschnitt 4.1.2 genutzt werden.

Die Kategorisierung der Ereignisdaten erfolgt in diesem Fall zunächst am Beispiel der in Tabelle 4.2 gezeigten Einträge. Der erste und vierte Eintrag ist dabei zusätzlichen Temperaturdaten zuzuschreiben, während Einträge #2 Statusinformationen und #3 Grenzwertüberschreitungen zugeschrieben werden können. Die Sequenz der Einträge #6 bis #10 wird Wartungsereignissen zugeordnet, der darauffolgende Eintrag #11 ist sowohl der Kategorie Statusinformation als auch „Pause“ zugeordnet. Die nachfolgenden Einträge #12 bis #17 gehören wiederum der Wartungssequenz an und der letzte Eintrag ist ebenfalls eine Statusinformation.

Eine Kategorisierung nach Komponenten oder Subkomponenten des Systems ist dabei auf Basis der vorliegenden Informationen schwierig zu realisieren, bzw. aufgrund einiger übergeordneter Einträge (z. B. „Run“, „Pause“, „Key Switch“) nicht möglich und wird verworfen. Diese Vereinfachung wird im späteren Verlauf Einfluss auf die Modellbildung haben und bei der Darstellung der Ergebnisse erneut aufgegriffen. Die im Folgenden gezeigte Kategorisierung orientiert sich an verschiedenen Informationsklassen und Anwendungsbereichen:

- Zusätzliche Temperaturdaten verschiedener Systeme, die teils in regelmäßigen Abständen, teils sporadisch geloggt werden.
- Wartungsereignisse: Eine Kombination aus mehreren Ereignissen, siehe Abschnitt 4.1.2.
- Umweltbedingungen.
- Operationelle Bedingungen und Zustände, wie Abschaltung aufgrund von Turbulenz, Abschaltung aufgrund von Einspeisemanagement (externe Bedingungen) oder Statusupdates wie Systemstart.
- Statusinformationen (z. B. Sensor-Resets, Akkutests, Time-Outs oder ähnliches).
- Grenzwertüberschreitungen.
- Fehler- und Alarmlogmeldungen.

Daneben können mit den in [Ves07] beschriebenen Logtexten einige Alarm und Loginformationen in die Kategorien

- Run,
- Pause,
- Stop sowie

- Emergency

einsortiert werden und erlauben so eine Einordnung der Schwere eines Ereignisses. Auffällig ist hierbei, dass alle Ereignisse, die in eine der zuvor genannten Schwerekategorien eingruppiert werden können, ebenfalls einen *Reset* Zeitstempel aufweisen. Ein Restanteil von 0,022 % der Logdaten lässt sich keiner der oben genannten Kategorien zuweisen. Die entsprechenden Ereignisse treten zumeist singulär auf, stehen zeitlich nicht im Zusammenhang mit Ausfallereignissen und werden daher vernachlässigt.

Um eine Vorselektion der Daten zu vereinfachen, werden zunächst die in Abschnitt 4.1.2 bereits beschriebenen Meldungen „External Power Ref.: 2000kW“ sowie „Generator 1 in“ eliminiert sowie der Datensatz vom 01.01.2016 bis 12.12.2017 betrachtet. Die danach resultierende Verteilung der Gruppen ist in Abbildung 4.5 dargestellt.

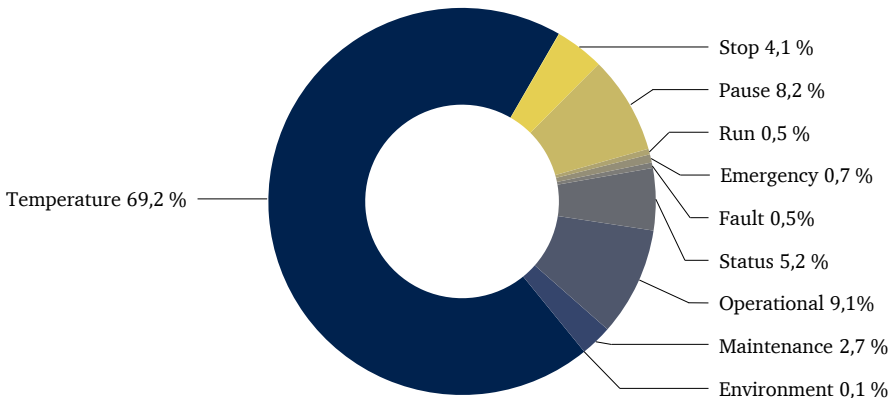


Abbildung 4.5.: Verteilungen der definierten Kategorien im Ereignisdatsatz.

Im nächsten Schritt können alle Einträge, die redundante Informationen mit den SCADA-Daten aufweisen, eliminiert werden.

Bereinigung redundanter Informationen in den Logdaten

Bei der Sichtung der Daten konnte festgestellt werden, dass die Logdaten ebenfalls Informationen von Sensoren enthalten (vgl. Zeilen 1-4 in Tabelle 4.2). Im Wesentlichen

sind dabei Temperaturdaten enthalten. Dabei liegt der Verdacht nahe, dass diese Informationen mit den SCADA-Daten redundant vorliegen. In einer ersten Analyse werden aus den Logdaten entsprechende Temperaturen herausgelöst und tabellarisch mit SCADA-Daten verglichen, sowie eine Korrelationsanalyse durchgeführt. Ziel ist es, redundante Parameter, die in den Logdaten und SCADA-Daten enthalten sind, zu identifizieren und zu reduzieren.

Hierfür liegt der Fokus zunächst auf Temperaturen, da sie einen Großteil der zusätzlichen Sensordaten in den Ereignislogs ausmachen. Mit einer Stichwortsuche nach dem Begriff "°C" werden alle Einträge, in denen numerische Temperaturwerte vorkommen, identifiziert und isoliert. Die Einträge lassen sich zusammenfassen und es können bis zu 37 einzigartige Einträge/Variablen gefunden werden, abhängig der betrachteten WKA.

Um mögliche Korrelationen mit den SCADA-Daten zu analysieren, werden beide Tabellen (SCADA und Ereignislogs) zusammengefasst. Die Fusion der Tabellen wird anhand des nächstmöglichen Zeitstempels mit einer maximalen Toleranz von zwei Minuten vorgenommen. Da in diesem Fall Temperaturdaten betrachtet werden und die SCADA-Daten eine zehnmütige Auflösung aufweisen, wird der erwartete Fehler als gering eingeschätzt. Für die zusammengefassten Daten werden Korrelationskoeffizienten nach Pearson berechnet. Eine Darstellung der Ergebnisse für Korrelationen von Logdaten zu SCADA Daten werden in Form einer Heatmap in Abbildung 4.6 dargestellt.

Von den insgesamt 37 untersuchten Ereignissen lassen sich nicht für alle Korrelationen mit den Sensordaten analysieren. Zwölf Einträge treten nur sporadisch auf und zeigen keinerlei Korrelationen mit den Sensordaten. Auf Basis eines Schwellwerts der Korrelationskoeffizienten von über 0,9 werden 14 der 25 verbleibenden Einträge identifiziert und reduziert. Insgesamt betrifft das 57.604 Einträge in den Logdaten. Hochkorrelierte Einträge enthalten vor allem Informationen, die sich ebenfalls der Kategorie *Statusinformation* zuordnen ließen. Jene Logeinträge, die niedrigere Korrelationen erreichen, sind tendenziell auffälligeren Einträgen wie „High temp. gen bearing“ oder „High temperature“ zuzuordnen. Die verbleibenden elf niedrigkorrelierten Einträge (insgesamt 21.235 Ereignisse) verbleiben in der Datenbank unter der Kategorie „Temperaturdaten“.

Untersuchung operationeller Aspekte

Neben der Elimination von redundanten Informationen werden im Rahmen der Vorverarbeitung operationelle Aspekte untersucht. So ist z. B. ein Instandhaltungsereignis

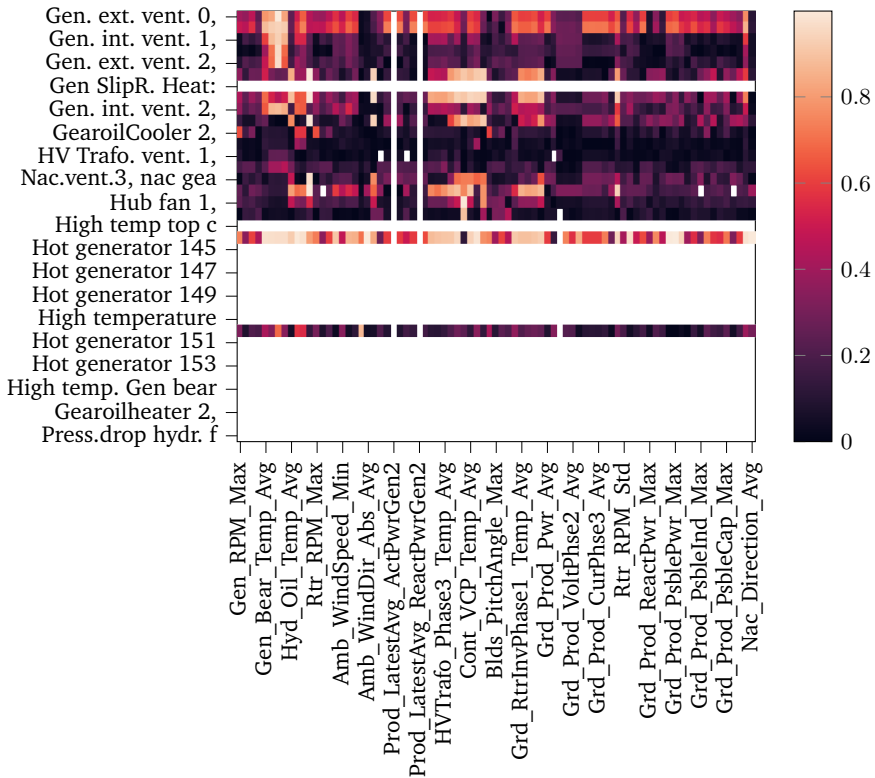


Abbildung 4.6.: Korrelationen von Temperaturen aus Logdaten mit SCADA-Daten am Beispiel von WKA 06

durch eine Sequenz von Ereignissen definiert, von der Übernahme der manuellen Kontrolle bis zum Einschalten des Generators. Untersucht werden soll, ob Auffälligkeiten von Instandhaltungsereignissen hinsichtlich der Ausfallereignisse feststellbar sind. Eine Darstellung der Instandhaltungsereignisse je WKA über der Zeit mit Markierung von Ausfällen ist in Abbildung 4.7 gegeben. Die Breite der Balken entspricht dabei der Dauer des Ereignisses. Um eine adäquate Visualisierung zu gewährleisten, ist für den Plot eine Mindestdauer von sechs Stunden definiert, auch wenn die durchschnittliche Länge eines Instandhaltungsereignisses 2h40min beträgt, bzw. der Median bei

45 Minuten liegt. Insbesondere bei den Anlagen T01, T07 und T09 ist dabei eine Häufung von Instandhaltungsereignissen rund um Ausfallereignisse erkennbar. Die Gesamtstillstandszeit je Anlage beträgt im Mittel 8 Tage und 02h51min über zwei Jahre, bzw. 1,11% der Zeit. Auffällig ist hierbei die große Streuung von knapp über 3 Tagen gesamte Stillstandszeit an WKA 11 bis zu 12,5 Tagen an WKA 07.

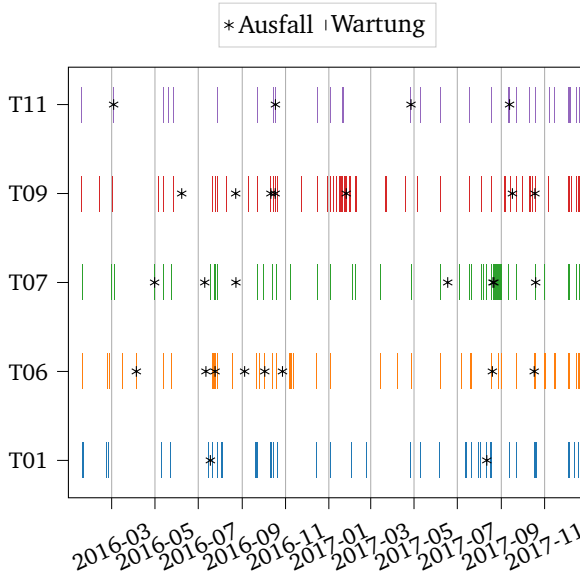


Abbildung 4.7.: Wartungsereignisse und Ausfälle aller WKA

Neben Instandhaltungsereignissen unterliegen die Windkraftanlagen außerdem netzgesteuerten Abschaltvorgängen, z. B. wenn insgesamt zu viel Strom in das Netz geliefert wird. Auch diese Zustände, in denen die Anlage aufgrund des so genannten Einspeisemanagements still steht, sind in den Ereignislogs ersichtlich. Auffällig ist hierbei, dass die Stillstandszeiten immer alle Anlagen betreffen. Insgesamt stehen die Anlagen aufgrund des Einspeisemanagements über den Zeitraum von zwei Jahren im Mittel 3 Tage 02h43min still, bzw. 0,43% der Zeit. Eine Visualisierung der Zeiträume ist in Abbildung 4.8 gegeben.

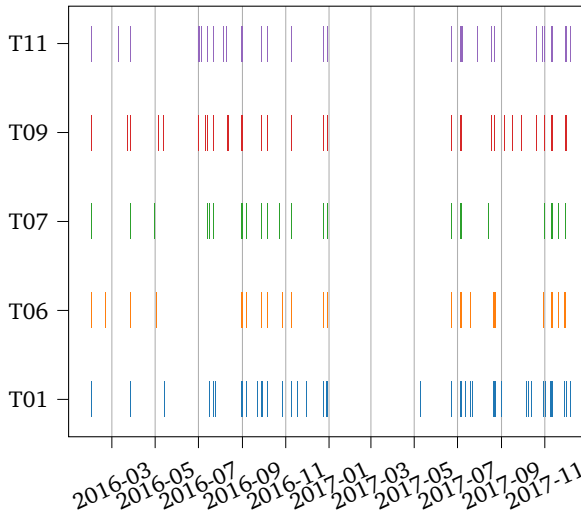


Abbildung 4.8.: Stillstandszeiten aufgrund des Einspeisemanagements aller WKA

Selektion und Reduktion der Ereignisdaten

Auf Basis der zuvor dargestellten Kategorien werden im folgenden die für eine Schadensdiagnose relevanten Ereignisdaten selektiert und unnötige Informationen reduziert. Dabei gilt es qualitativ zu beurteilen, welche Informationen für die Diagnose relevant sind. Das Hauptaugenmerk liegt dabei darin, die Vielfalt und Menge an Logdaten für den Schritt der Merkmalsentwicklung zu reduzieren. Wie zuvor beschrieben, werden die Logdaten um redundante Temperaturinformationen und die beiden Statusmeldungen „External Power Ref.:xxxxkW“ sowie „Generator 1 in“ reduziert. Dadurch lässt sich die Gesamtdatenbasis von 318.835 Ereignissen auf 134.217 Ereignisse, bzw. von 11.751 auf 5.790 einzigartige Ereignisse reduzieren. Da keine semantische Analyse der Ereignisse beabsichtigt ist, werden nach der Kategorisierung gemäß des üblichen Vorgehens beim *Natural Language Processing* alle Zahlenwerte eliminiert. Damit wird die Anzahl der einzigartigen Ereigniseinträge weiterhin auf 157 reduziert. Nur auf Basis deskriptiver Methoden können für die einzelnen Kategorien keine Aussage über die Relevanz für die Diagnose getroffen werden, sodass eine weitere Selektion und Reduktion erst im Rahmen der Merkmalsentwicklung vorgenommen wird. Es wird

jedoch davon ausgegangen, dass die Kategorien der Statusinformationen, Instandhaltungsereignisse und operationelle Ereignisse wenig Mehrwert für die Diagnose liefern. Den Kategorien um Fehler- und Alarmloginformation sowie allen Einträgen, die sich einer Schwere des Ereignisses (Run, Pause, Stop, Emergency) zuordnen lassen, wird eine höhere Relevanz zugeordnet.

Die Beurteilung der Kategorien wird im folgenden Abschnitt im Rahmen der Bewertung der extrahierten und entwickelten Merkmale vorgenommen.

4.3. Merkmalsentwicklung der Ereignisdaten

Dieser Abschnitt befasst sich mit der Beschreibung der Ausprägungen der in Abschnitt 3.2.3 beschriebenen Merkmalsentwicklung der Ereignisdaten. Ausgangsbasis für die Entwicklung aller Merkmale sind die kategorisierten und um redundante Temperaturinformationen bereinigten Ereignislogs. Ein wichtiger Schritt bei der Generierung von Merkmalen ist die gleichzeitige Selektion, da die Ansätze für sich auf Basis der vorliegenden Ereignisdaten mehrere hundert Merkmale generieren.

Die Ereignisdaten werden für die Merkmalsentwicklung in definierten Zeiträumen, oder auch Fenstern betrachtet. Dabei werden in diesem Schritt ebenfalls Fenster für das Labeln der Daten festgelegt. Basierend auf den Ausfalldaten wird damit ein Zeitraum w_{fail} von x Tagen vor dem Ausfallereignis als *schadhaft* bezeichnet. Der Rest des Datensatzes wird als *gesund* beschrieben. Weiterhin werden alle Ereignisse im Rahmen von gleitenden Fenstern analysiert um Merkmale zu generieren. Diese Fenster bewegen sich in Größenordnungen von zwölf Stunden bis eine Woche, je nach Merkmal und Wahl der Kategorie. Zusammenfassend sind demnach drei Fenster für die Generierung und Beurteilung der Merkmale von Bedeutung:

- das Ausfallfenster w_{fail} und
- das Beobachtungsfenster w_{obs} , das weiter unterteilt wird in
 - das Beobachtungsfenster für statistisch basierte Merkmale $w_{obs_{stats}}$ und
 - das Beobachtungsfenster für sequenzbasierte Merkmale $w_{obs_{seq}}$.

Die Fenstergrößen sind Variationsparameter und Gegenstand von Parameterstudien der Sensitivitätsanalyse in Abschnitt 4.6. Gemäß dem Aufbau von Abschnitt 3.2.3, werden im Folgenden zunächst die statistisch basierten Merkmale und sequenzbasierte Merkmale beschrieben.

Die Entwicklung von fehlerähnlichkeitsbasierten Merkmalen ist auf dieser Datenbasis ohne Expertenwissen nicht darstellbar. Das alternative, datenbasierte Vorgehen zur Identifikation von Schadenssignaturen lieferte keine eindeutigen Ergebnisse, weshalb fehlerähnlichkeitsbasierte Merkmale in diesem Fall nicht entwickelt werden können. Die Entwicklung von profilbasierten Merkmalen, die im Zusammenhang Systemspezifika stehen, wird im Rahmen der Arbeit durch den Aufbau von individuellen Modellen gewährleistet. So wird für jede Anlage bzw. für jeden Schadensfall ein eigenständiges Modell aufgebaut, wodurch keine entsprechenden Merkmale für diese Eigenschaften generiert werden müssen.

Entwicklung von statistisch basierten Merkmalen

Auf Basis der kategorisierten Ereignislogs können mehrere statistisch basierte Merkmale entwickelt werden. Im Folgenden werden die Merkmale auf Basis der Häufigkeit von einzigartigen Ereignissen im Beobachtungszeitfenster $w_{obs_{stats}}$ und der Häufigkeit aller Ereignisse innerhalb einer Kategorie im Fenster $w_{obs_{stats}}$ beschrieben.

Im ersten Schritt werden einzelne Ereignisse betrachtet. Um dem Einfluss von dimensionsbehafteten Größen entgegenzuwirken, werden dabei Ziffern aus den Ereignissen entfernt, was in 157 einzigartigen Ereignissen mündet. Daraus resultieren 157 Merkmale, deren Signifikanz für die Fehlerdiagnose zu beurteilen ist. Dafür werden die Merkmale zur deskriptiven Analyse gemeinsam mit den Ausfallzeitpunkten zunächst über die Zeit analysiert. Abbildung 4.9 zeigt diese zeitliche Darstellung am Beispiel von WKA T01 und den Merkmalen der Ereignisse *heating slipring time* und *pause kw rpm* bei einem Beobachtungszeitfenster $w_{obs_{stats}} = 3d$. Interessant ist hierbei der zusätzlich deskriptive Charakter, der sich durch die Merkmalsentwicklung und gewählte Darstellungsform ergibt. So kann an diesem Beispiel gezeigt werden, dass das Ereignis *heating slipring time* saisonal in den Wintermonaten auftritt und keine Korrelation mit den Ausfallereignissen zeigt.

Die Darstellung der Merkmale über die Zeit gibt allerdings nicht in allen Fällen einen so eindeutigen Aufschluss über die Eignung des Merkmals zur Zustandsdiagnose wie das Ereignis bzw. Merkmal *heating slipring time*. Um die Interpretierbarkeit zu erhöhen, werden im Folgenden die Verteilungen der Häufigkeiten in gesunden Fenstern w_{health} und schadhaften Fenstern w_{fail} miteinander verglichen. Abbildung 4.10 stellt die beiden daraus resultierenden Verteilungen für die Merkmale basierend auf den Ereignissen *heating slipring time* und *pause kw rpm* als Boxplot dar. Aufgrund der hohen Schiefe der Verteilungen und Dominanz der Werte nahe Null, ist die Aussagekraft der

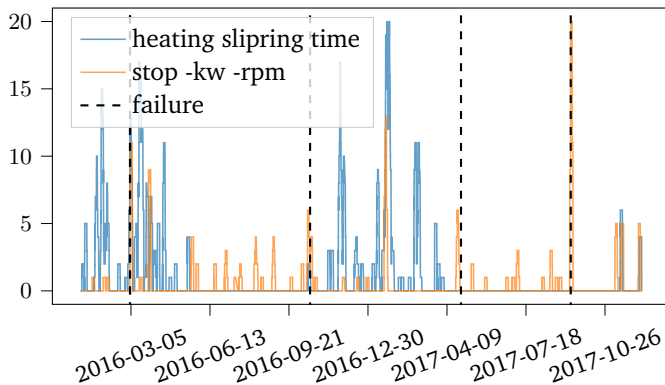


Abbildung 4.9.: Darstellung der statistisch basierten Merkmale für die Ereignisse *heating slipping time* und *stop -kw -rpm* über die Zeit an WKA T11 bei $w_{obs} = 3d$.

Boxplots begrenzt. Unter Vernachlässigung der dargestellten Ausreißer ergibt sich jedoch das Bild, dass für das Merkmal *heating slipping time* keine Differenzierung der Verteilungen ersichtlich ist, für das Merkmal *stop -kw -rpm* tendenziell eher.

Dabei ist zu beachten, dass die Größe des Ausfallfensters w_{fail} Einfluss auf die Verteilungen nimmt. Weiterhin bietet sich eine differenzierte Betrachtung der Verteilungen nach WKA sowie nach Schadensfall bzw. Komponente an, um mögliche Effekte der Merkmale bei gewissen Schadensbildern bzw. WKA identifizieren zu können. Auch die zeitliche Lage der Ausfallereignisse spielt eine Rolle. Bei WKA T11 liegen mehrere Ausfälle in den Wintermonaten, was beim Merkmal des Ereignisses *heating slipping time* zu einer Scheinkorrelation führt: Hier liegen die Ausfallereignisse in den Zeiträumen, in denen das Ereignis saisonal bedingt häufiger auftritt.

Aus der Interpretation von Zeitverlauf und Verteilungen wird ersichtlich, dass nicht alle Merkmale geeignet sind, um eine statistische Differenzierung der beiden Klassen (*schadhaft* und *gesund*) zu gewährleisten. Ein Ansatz zur Merkmalsreduktion ist es entsprechend nur jene Merkmale zu verwenden, bei deren die Verteilungen einen signifikanten Unterschied zeigen. Um die Signifikanz objektiv und quantitativ zu beurteilen, wird ein Mann-Whitney U-Test herangezogen für unabhängige und nicht normal verteilte Stichproben mit unterschiedlicher Varianz. Der Test prüft die Mittelwerthomogenitätshypothese $H_0 : \mu_1 = \mu_2$, also ob beide Verteilungen den gleichen Mittelwert

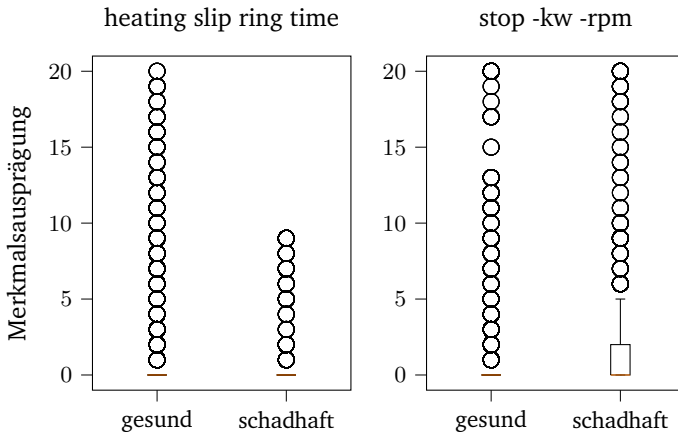


Abbildung 4.10.: Darstellung der statistisch basierten Merkmale für die Ereignisse *heating slipring time* und *stop -kw -rpm* über die Zeit an WKA T11 bei $w_{obs} = 3d$.

haben [Eck19]. Dabei werden beide Verteilungen als Input und ein Signifikanzniveau von $\alpha = 0,05$ verwendet.

Die entsprechenden p-Werte liegen für die in Abbildung 4.10 gezeigten Verteilungen bei 0,999 für das Merkmal *heating slipring time* und bei 0,0007 für das Merkmal *stop -kw -rpm*. Entsprechend hat nur das letztere Merkmal statistisch gesehen eine Relevanz für die Diagnose an WKA T01. Eine Darstellung aller Merkmale basierend auf einzigartigen Ereignissen in Boxplots oder deren p-Werte ist nicht sinnvoll. Ausschnittsweise werden die p-Werte der Signifikanztests in Form einer Heatmap für die ersten 30 Merkmale der Ereignisse in Abbildung 4.11 dargestellt. Werte unter 0,05 (besonders dunkle Bereiche) bedeuten einen signifikanten Unterschied der Verteilungen und damit eine sinnvolle Eignung der Merkmale für die Klassifikation. Auffällig sind die Unterschiede, die sich bei einigen Merkmalen in Abhängigkeit der WKA ergeben. Demnach weisen die Anlagen einzigartige Charakteristika auf, was die Modellbildung je Anlage motiviert.

Neben der Betrachtung von einzigartigen Ereignissen werden weiterhin Kategorien von Ereignissen genutzt um statistisch basierte Merkmale zu generieren. Anstelle der Häufigkeit einzigartiger Ereignisse wird dabei die Häufigkeit aller Ereignisse in der jeweiligen Kategorie innerhalb des Beobachtungsfensters betrachtet. Beispielhaft ist

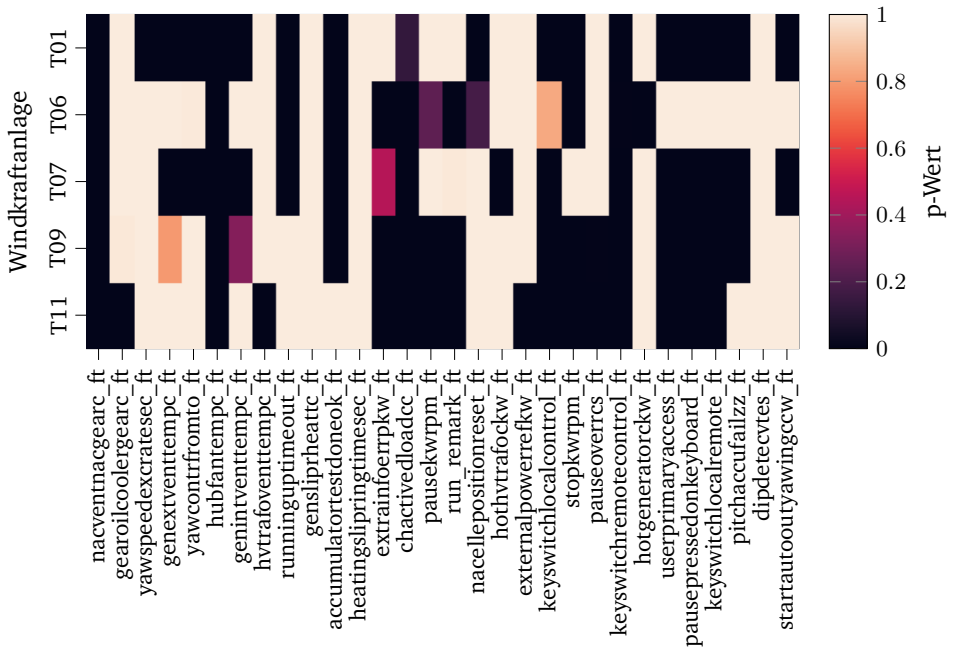


Abbildung 4.11.: Darstellung der p-Werte für die Merkmale der ersten 30 Ereignisse für alle Komponenten, bei $w_{obs} = 3d, w_{fail} = 30d$.

eine solche Darstellung in Abbildung 4.12 für WKA T01 und die Kategorien *Pause*, *Stop* und *Emergency* gegeben. Eine Häufung der Merkmale ist dabei in Bereichen von Ausfällen ersichtlich, allerdings nicht eindeutig vor dem Ausfallereignis.

Abbildung 4.13 stellt die beiden daraus resultierenden Verteilungen für die Merkmale basierend auf der Gruppe „Pause“ und „Stop“ dar.

Bei der Betrachtung aller Komponenten bzw. Schadensbilder zeigt sich hingegen das Bild, dass nahezu alle oben gezeigten Merkmale eine statistische Signifikanz aufweisen. Die Darstellung der Signifikanz der zuvor gezeigten Merkmale auf Basis der Kategorien über alle Komponenten ist in Abbildung 4.14 dargestellt. Zusätzlich sind hierbei die

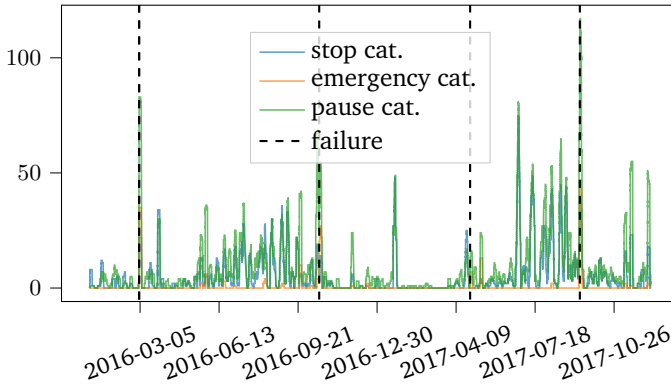


Abbildung 4.12.: Darstellung der statistisch basierten Merkmale für die Kategorien *Stop*, *Emergency* und *Pause* über die Zeit an WKA T11 bei $w_{obs} = 3d$.

Merkmale der Kategorien *Temperatur*, *Umwelt*, *Wartung*, *Operations* und *Status* ergänzt. Die Darstellung bestätigt den Verdacht, dass die Kategorien *Umwelt* und *Operations* keinen signifikanten Mehrwert für die Diagnose bieten und die Kategorien *Temperatur*, *Umwelt* und *Wartung* nur für einige WKA einen Mehrwert liefern. Das zuvor gezeigte einzelne Merkmal *heating slipping time* ist in die Kategorie *Temperatur* einsortiert.

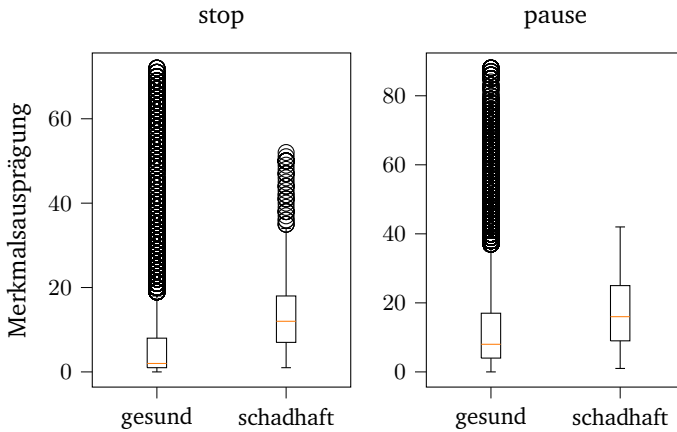


Abbildung 4.13.: Darstellung der statistisch basierten Merkmale für die Kategorien Stop und Pause an WKA T01 und Komponente Getriebe, bei $w_{obs} = 3d, w_{fail} = 30d$.

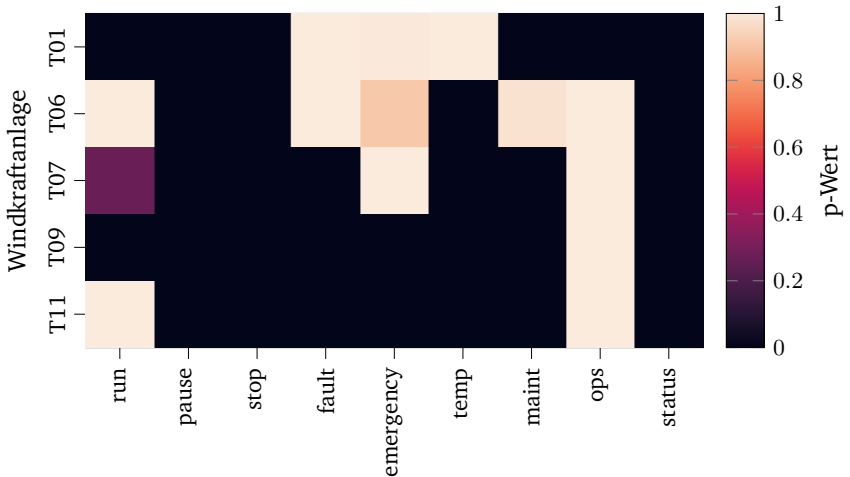


Abbildung 4.14.: Darstellung der p-Werte für die Merkmale mehrerer Kategorien für alle Komponenten, bei $w_{obs} = 3d, w_{fail} = 30d$.

Allgemein wird die Aussagekraft der statistischen Tests zur Merkmalselektion für diesen Anwendungsfall aufgrund der teilweise robusten Mittelwerte (vgl. Abbildung 4.10) nicht sehr hoch eingeschätzt. Infolge dessen werden alle Merkmale für die Modellbildung genutzt und die Merkmalsrelevanz im Nachhinein durch den Algorithmus erneut beurteilt.

Entwicklung von sequenzbasierten Merkmalen

Die Entwicklung der sequenzbasierten Merkmale basiert ebenfalls auf der Fenstermethode. Da für sequenzbasierte Merkmale andere Fenstergrößen von Interesse sind, wird hierbei ein neues Fenster $w_{obs_{seq}}$ betrachtet. Der gesamte Ereignisdatensatz wird je WKA gruppiert bzw. partitioniert und mit gleitenden Fenstern betrachtet. Die Ereignisse werden wie zuvor um Zahlenwerte bereinigt, sodass insgesamt 157 einzigartige Ereignisse übrig bleiben. Innerhalb eines Fensters wird dann, wie in Abschnitt 3.2.2 beschreiben, jede mögliche serielle Sequenz betrachtet und deren Häufigkeit (*Support*) im Gesamtdatensatz gespeichert. Als Randbedingungen werden nur Sequenzen bis zu einer Länge von vier Ereignissen und mit einer relativen Mindesthäufigkeit (min_{sup}) von 10% aller auftretenden Sequenzen berücksichtigt. Weiterhin soll die absolute Sequenzlänge mindestens zwei betragen, da Sequenzen der Länge eins formal einzelne Ereignisse sind und damit bereits als statistische Merkmale erfasst sind. Wie in Abschnitt 3.2.3 beschrieben, wird zur Identifikation der Sequenzen der „Prefix Span“ Algorithmus verwendet.

Die generierten Merkmale entsprechen dabei direkt den gefundenen Episoden (Sequenzen, die die oben genannten Randbedingungen erfüllen) und können gemäß dem Prinzip des One-Hot Coding interpretiert werden. Dabei wird jedem Zeitschritt eine „Eins“ zugewiesen, in dem die entsprechende Episode vorhanden ist, anderenfalls eine „Null“. Die Anzahl der Episoden/Merkmale ist abhängig von der Länge des Beobachtungsfensters. Um Komplexität und Zahl der Dimensionen des Merkmalraums für das ML-Modell zu begrenzen, wird für den Vergleichsfall ein Beobachtungsfenster von $w_{obs_{seq}} = 12$ Stunden gewählt. Weiterhin werden zunächst nur die Ereignisse der Kategorien *run*, *pause*, *stop*, *fault* und *emergency* berücksichtigt, die insgesamt 105 der 157 einzigartigen Ereignisse abdecken. Bei 20.467 ($min_{sup} = 2.047$) möglichen Sequenzen im Gesamtdatensatz können daraus 198 Episoden bzw. Merkmale generiert werden.

Die Anzahl der verwendeten Ereigniskategorien nimmt dabei Einfluss auf die Anzahl der Episoden. Bei Verwendung von zehn Kategorien, wie in Abbildung 4.14 dargestellt

Tabelle 4.4.: Beispiele für identifizierte Sequenzen

Sequenz ID	Support	Sequenz / Ereignisse
#85	5753	{pause -kw -rpm}, {run}
#145	2350	{running up timeout}, {running up timeout}

gegenüber fünf Kategorien, steigt die Anzahl an möglichen Sequenzen auf 147.550. Die Anzahl der gefundenen Episoden, respektive die Anzahl der Merkmale, bei festem $min_{sup} = 2.047$ auf rund 50.000. In jenem Fall wäre die Anzahl der Merkmale bei der limitierten Anzahl an Messwerten zu hoch für das Training eines ML-Modells. Der min_{sup} nimmt hier allerdings auch Einfluss. Wird dieser auf 10% erhöht (jede Sequenz muss mindestens 14.755 mal vorkommen), sinkt die Anzahl der Episoden auf 148. Auch die Wahl des Beobachtungsfensters nimmt Einfluss auf die Zahl der generierten Merkmale. So steigt die Anzahl der generierten Merkmale für den Referenzfall mit Ereignissen aus fünf Kategorien bei einem Beobachtungsfenster von 24 Stunden auf 570 und bei zwei Tagen auf 1.672.

Die visuelle Darstellung der Merkmale orientiert sich an den zuvor dargestellten Beispielen. Zunächst wird in Abbildung 4.15 eine Darstellung einiger ausgewählter sequenzbasierter Merkmale über die Zeit dargestellt. Die Aussagekraft der Darstellung bleibt aufgrund des binären Charakters jedoch begrenzt. Als Beispielsequenzen werden die Sequenzen #85 und #145 verwendet. Beide sind in Tabelle 4.4 beschrieben. Der Support ist hierbei über den gesamten Ereignisdatensatz berechnet, also unabhängig von Anlagen, Komponenten und Ausfallfenstern.

Die Darstellung von Verteilungen in Form von Boxplots wie im vorherigen Abschnitt ist aufgrund der binären Ausprägung der Merkmale nicht sinnvoll. Alternativ können jedoch die Mittelwerte der jeweiligen Merkmale in den beiden Klassen *schadhaft* und *gesund* verglichen werden, die im Prinzip eine Reduktion der Verteilungen darstellen. Entsprechend sind die Mittelwerte für die zuvor gezeigten Merkmale der Sequenzen #85 und #145 in Abbildung 4.16 für die Anlagen T01 und T11 dargestellt.

Die Betrachtung des Signifikanztests zur Selektion relevanter Merkmale wie im Abschnitt zuvor ist auf Basis der Binärdaten nicht sinnvoll. Relevante Merkmale können auf Basis des Vergleichs der beiden Mittelwerte ausgewählt oder ausgelassen werden. Allerdings zeigen die Mittelwerte der Merkmale hierbei für die einzelnen WKA unterschiedliche Trends. Allgemein sind die Niveaus bei den Anlagen T01, T06 und T07 deutlich niedriger als bei den Anlagen T09 und T11, was eine geringere absolute Häu-

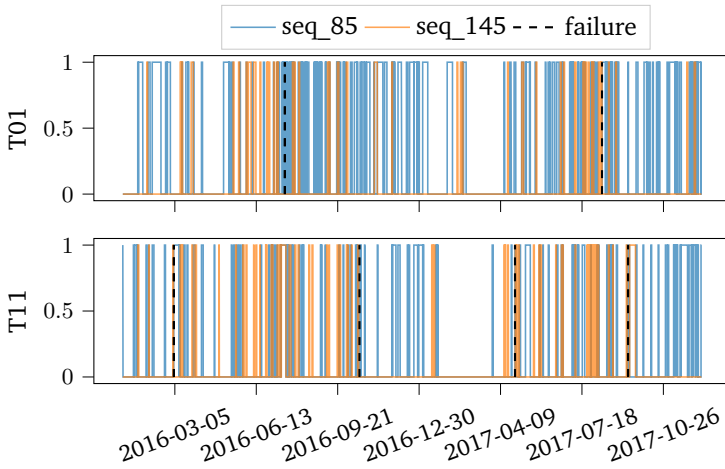


Abbildung 4.15.: Darstellung der sequenzbasierten Merkmale für die Sequenzen #85 und #145 über die Zeit an WKA T01 und T11 bei $w_{obs_{seq}} = 2d$.

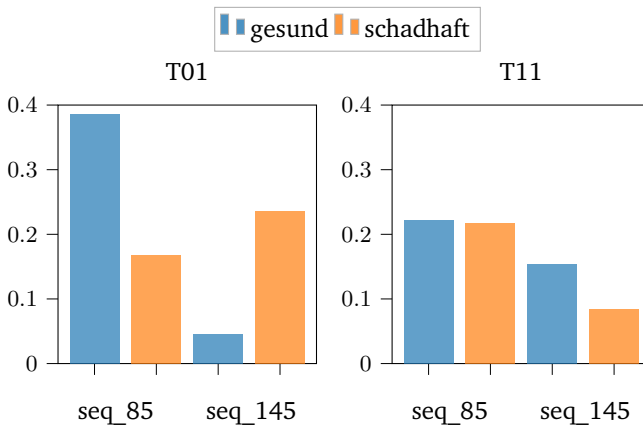


Abbildung 4.16.: Darstellung der Mittelwerte der sequenzbasierten Merkmale für die Sequenzen #85 und #145 an WKA T01 und T11 bei $w_{obs_{seq}} = 2d$.

figkeit von Sequenzen über den gesamten Zeitraum bedeutet. Ebenso verhält es sich mit den Überhöhungen der Mittelwerte in schadhafte Fenstern im Vergleich zu gesunden Fenstern. Auch hier zeigen die sequenzbasierten Merkmale bei den Anlagen T09 und T11 im Mittel deutlichere Überhöhungen beim schadhafte Fenster gegenüber den Anlagen T01, T06 und T07. Insgesamt kann festgehalten werden, dass die Ausprägung der sequenzbasierten Merkmale WKA-abhängig ist, was eine übergreifende Selektion der Merkmale an dieser Stelle erschwert. Entsprechend wird die Merkmalselektion ebenfalls im Rahmen der Modellentwicklung durchgeführt.

Zusammenfassung der Merkmalsentwicklung

Zusammenfassend lässt sich festhalten, dass auf Basis der statistischen und sequenzbasierten Ansätze eine Vielzahl an Merkmalen generiert werden kann. Die Reduktion der Merkmale, insbesondere bei Berücksichtigung einzelner Anlagen und Schadensbilder ist aufgrund der Vielzahl der Dimensionen und auf Basis der untersuchten Verteilungen schwer möglich. Soll weiterhin die Abhängigkeit der Variationsparameter bei der Merkmalsgenerierung berücksichtigt werden, erschwert dies eine allgemeine Merkmalsselektion weiter. Der statistische Vergleich der beiden Verteilungen in gesunden und schadhafte Fenster erlaubt zwar die Eliminierung von gänzlich unbrauchbaren untauglichen Merkmalen, eine zuverlässige Selektion von aussagekräftigen Merkmalen für den Diagnoseschritt ist jedoch auf dieser Basis nicht möglich. Aus diesem Grund wird der Aspekt bei der Diskussion der Ergebnisse und mit Hilfe der Klassifikationsmodellparameter erneut aufgegriffen.

4.4. Aufbau des Diagnosemodells

Die zuvor beschriebenen Merkmale werden im Diagnoseschritt in ML-Modellen verarbeitet. Der nachfolgende Abschnitt geht dabei zunächst auf die Wahl der Trainings- und Testdatenbasis und die daraus resultierenden Möglichkeiten bei der Modellbildung ein. Abschließend wird das verwendete Klassifikationsmodell vorgestellt und die Datenfusion beschrieben.

4.4.1. Wahl der Trainings- und Testdatenbasis

Die Wahl der Trainings- und Testdaten spielt eine wichtige Rolle bei der Verwendung von ML-Algorithmen. Bei der vorliegenden, abgeschlossenen Datenbasis, mit der begrenzten Verfügbarkeit von Daten innerhalb von zwei Jahren, ist zu berücksichtigen, dass die zugrunde liegenden Datenverteilungen in Trainings- und Testzeitraum ähnliche Charakteristika aufweisen. In der vorliegenden Arbeit orientiert sich die Wahl dieser Zeiträume an den im Hackathon verwendeten Zeiträumen. Dabei wird der Datensatz in folgende Subdatensätze geteilt:

- Trainingsdaten: 01.01.2016-31.08.2017
- Testdaten: 01.09.2017-12.12.2017

Zu beachten ist, dass der Trainingszeitraum gegenüber dem des Hackathons gekürzt ist, da ab dem 12.12.2017 bei den Anlagen T01, T06 und T09 Ausreißer und fehlerhafte Daten in den Logdaten enthalten sind (vgl. Abschnitt 4.1.2). Da in diesem Zeitraum jedoch keine Schäden mehr auftreten, wird die Auswirkung auf die Modellbildung als gering eingeschätzt.

Zur weiteren Evaluation der Ergebnisse, sowie zur Untersuchung und Bewertung der Robustheit des Modells wird eine 5-fache Kreuzvalidierung herangezogen. Dabei wird der Datensatz (je WKA oder je Komponente) in fünf gleich große Teile unterteilt. Vier der Teile werden zum Training genutzt, der fünfte Teil zum Testen. Nach fünf Durchläufen ist der gesamte Datensatz in unterschiedlichen Kombinationen zum Trainieren und Testen verwendet worden. Stimmen die dabei erzeugten Ergebnisse mit den zuvor gezeigten Ergebnissen überein, kann das Modell als robust angesehen werden.

Daneben muss die Art bzw. Zielsetzung der Modellbildung bei der Wahl der Trainings- und Testdaten berücksichtigt werden. Für ein Klassifikationsmodell je WKA und für alle Schäden werden die Daten nach WKA gefiltert. Entsprechend verkleinert sich die verwendete Datenbasis je Klassifikationsmodell auf rund ein Fünftel des Gesamtdatensatzes. Bei einem Modell je Baugruppe (Komponente) wird hingegen der gesamte Datensatz verwendet und nicht reduziert. Das beeinflusst insbesondere den relativen Anteil der Schadensvorkommnisse bzw. positiven Labels und damit die Klassenimbalance. So steigt die mittlere Klassenimbalance von 21% positive Labels für die Modellbildung je WKA auf 4% positive Labels für die Modellbildung für jede Komponente. Neben der Klassenimbalance nehmen auch die Zeitpunkte der Ausfälle bzw. die Wahl der Trainings- und Testdatenzeiträume Einfluss auf die Ergebnisse. So sind im oben genannten Testzeitraum nur Schäden an den Baugruppen Hydraulik und

Getriebe vorhanden. Für je ein Modell pro Schaden und Anlage ist das Volumen der Datengrundlage nicht ausreichend.

4.4.2. Gradient Boosting zur Klassifikation der Schadenszustände

An dieser Stelle wird der verwendete Klassifikationsalgorithmus kurz beschrieben und seine Funktionsweise erläutert. Im Allgemeinen bieten sich eine Vielzahl an Algorithmen an, um die Klassifikationsaufgabe zu erbringen. Im Rahmen einer Vorabuntersuchung, wurden neben dem final ausgewählten Gradient Boosting (GB), auch Entscheidungsbäume, RF, SVM sowie kNN betrachtet. Der GB Algorithmus lieferte dabei die besten Ergebnisse bei einer akzeptablen Rechenzeit.

Im Allgemeinen ist die Idee des so genannten *Boosting*, mehrere schwache Lerner zu einem Komitee zusammen zu bringen um eine leistungsfähige und robuste Klassifikation bzw. Regression zu ermöglichen [HTF09]. GB liefert dabei ein Vorhersagemodell auf Basis eines Ensembles an Entscheidungsbäumen (DT). Der Algorithmus basiert auf dem Prinzip des Gradientenabstiegs, wobei durch die Minimierung einer Kostenfunktion schrittweise ein Modell aufgebaut wird, das die bestmöglichen Vorhersagen treffen kann.

Zur Initialisierung beginnt der Prozess mit einem einfachen Modell, das als Startpunkt dient. Dies kann ein konstantes Vorhersagemodell sein, das beispielsweise den Durchschnitt der Zielvariablen verwendet. Darauf werden die Residuen (Differenzen zwischen den tatsächlichen und den vorhergesagten Werten) für dieses Modell berechnet. Das Ziel besteht darin, ein neues Modell zu erstellen, das diese Residuen minimiert. GB verwendet Entscheidungsbäume mit begrenzter Tiefe (flache Bäume), als Basis für den Aufbau des kombinierten Modells. Die Entscheidungsbäume werden in jeder Iteration trainiert, um die Residuen des vorherigen Modells besser zu approximieren.

Bei der Modellanpassung wird der Gradient der Kostenfunktion (z. B. der mittlere quadratische Fehler bei einer Regression oder der Log-Loss bei einer Klassifikation) bezüglich der Vorhersagen des aktuellen Modells berechnet. Das nächste Modell wird daraufhin so trainiert, dass es entlang des negativen Gradienten dieser Kostenfunktion arbeitet, um die Residuen des vorherigen Modells zu reduzieren. Im Schritt des „Boosting“ werden die neuen schwachen Modelle nach jeder Iteration kombiniert (auch „Aggregation“), indem ihre Vorhersagen gewichtet werden. Die Gewichtung erfolgt anhand der Lernrate (engl. „learning rate“), die steuert, wie stark die beitragenden Modelle in die Gesamtvorhersage einfließen. Dieser Prozess der Iteration, bei dem immer ein neues Modell hinzugefügt wird, das sich auf die Fehler des vorherigen

Tabelle 4.5.: Verwendete Hyperparameter für den GB-Algorithmus

SK-learn Parameter	Ausprägung	Beschreibung
Estimators	200	Anzahl der Bäume/Stufen
max_depth	5	Tiefe der Regressionsbäume
max_features	5	Anzahl der Merkmale für besten Split
subsample	0, 1	Anteil der Samples zum Anpassen der jeweiligen Basislerner
min_samples_split	30	Mindestanzahl an Samples um einen internen Knoten zu teilen
loss	exponential	Verlustfunktion

Modells konzentriert, wird wiederholt, bis eine vordefinierte Anzahl von Iterationen erreicht ist oder bis ein bestimmtes Kriterium für die Fehlerreduktion erfüllt ist. [Fri01; Fri02; HTF09; Ke+17]

Der Algorithmus wird im Rahmen des Trainings einer Hyperparameteroptimierung unterzogen. Hierfür wird ein Gridsearch angewandt und ein Parameterraum nach der besten Kombination abgesucht. Dabei wird der gesamte Datensatz herangezogen und mit einer 5-fachen Kreuzvalidierung untersucht. Die Optimierung geschieht dabei sowohl für das auf den Ereignisdaten basierende Modell (vgl. Abschnitt 4.5.1), als auch für das Vergleichsmodell basierend auf Zeitreihendaten (vgl. Abschnitt 4.5.2). Optimierte Metriken sind dabei die Gesamtgenauigkeit, sowie der f1-Score. Insgesamt zeigt sich, dass die Hyperparameteroptimierung vor allem die Varianz der Ergebnisse reduziert, der Bias ändert sich hingegen wenig. Eine Gegenüberstellung der Ergebnisse der Hyperparameteroptimierung ist in Abbildung A.2 im Anhang gegeben. Abschließend kann eine Parameterkonfiguration gefunden werden, die für beide Datenbasen optimale Ergebnisse findet.

Die im Rahmen der Arbeit verwendeten Hyperparameter des in Scikit Learn trainierten GB sind in Tabelle 4.5 zusammengefasst. Die in der Arbeit verwendeten Python-Versionen sind in Tabelle A.1 im Anhang aufgeführt.

4.5. Ergebnisse der Diagnose

Der nachfolgende Abschnitt stellt die Ergebnisse der Zustandsdiagnose dar. Entsprechend der Ausführungen in Abschnitt 4.4.1 werden die Algorithmen auf dem Trainingsdatensatz trainiert und der Testdatensatz zur Beurteilung der Modellgüte zurückgehalten. Mit den bekannten wahren Labels im Testdatensatz lassen sich die entwickelten Modelle hinsichtlich Accuracy, Recall, Precision und f1-Score bewerten.

Im Folgenden werden zunächst die Ergebnisse der Klassifikation ausschließlich auf Basis der Ereignisdaten dargestellt und damit die Merkmalsentwicklung plausibilisiert. Im weiteren Verlauf zeigen die Ergebnisse des Vergleichsansatzes die Modellgüte des Stands der Technik. Diese werden mit den Ergebnissen des vorgeschlagenen Konzepts gegenübergestellt, was die Beurteilung des entwickelten Ansatzes ermöglicht. Die Ergebnisse auf Basis der Ereignisdaten sowie des Vergleichsansatzes basieren dabei auf der Modellbildung je WKA anhand des manuellen Trainings- und Testdatensplits. Die Ergebnisse der Modelle je Baugruppe sowie die der Kreuzvalidierung werden im Rahmen der Gegenüberstellung von vorgeschlagenem Konzept und Vergleichsansatz in Abschnitt 4.5.3 aufgegriffen.

Die in diesem Kapitel vorgestellten Ergebnisse basieren auf dem Vergleichsparameterset mit folgenden Ausprägungen:

- Ausfallfenster $w_{fail} = 30d$,
- Beobachtungsfenster für statistisch basierte Merkmale $w_{obs_{stat}} = 3d$,
- Beobachtungsfenster für sequenzbasierte Merkmale $w_{obs_{seq}} = 12h$ sowie
- der Eingrenzung der Ereignisse auf die Kategorien *pause*, *run*, *stop*, *emergency*, *fault*, *status*, *operations*, *maintenance* und *temperatures*.

Die Einflussgrößen der genannten Parameter werden in Abschnitt 4.6 im Rahmen der Sensitivitätsanalyse weiter untersucht.

4.5.1. Plausibilisierung der Merkmalsentwicklung: Diagnose auf Basis der Ereignisdaten

Um die prinzipielle Eignung der Ereignisdaten für die Zustandsdiagnose zu beurteilen, wird die Klassifikation der Zustände in diesem Abschnitt zunächst ausschließlich auf Basis der Ereignisdaten vorgenommen. Die im Folgenden gezeigten Ergebnisse werden daher als Plausibilisierung der Merkmalsentwicklung verstanden. Gleichzeitig erlauben die Ergebnisse eine isolierte Betrachtung der Merkmalsrelevanz und können damit für die Selektion von Merkmalen herangezogen werden.

Für die nachfolgend vorgestellten Ergebnisse werden demnach nur Merkmale aus Ereignisdaten verwendet, mit Labels aus den Ausfalldaten. Für jede WKA wird ein eigenes Modell aufgebaut. Die Ergebnisse für die einzelnen Modelle werden im weiteren Verlauf des Abschnitts zusammengefasst und gemittelte Metriken betrachtet und diskutiert. Da die Testdaten in diesem Fall einen zusammenhängenden Zeitraum abdecken, ist es möglich die Vorhersageergebnisse über die Zeit darzustellen. In Abbildung 4.17 ist die Darstellung der Klassifikationsergebnisse von WKA T07 auf Basis der Testdaten ab dem 01.09.2017 über die Zeit gegeben. Die blaue Linie stellt dabei die wahre Klasse dar, orange ist die Vorhersage. Der GB-Algorithmus ermöglicht ebenfalls eine Ausgabe der Klassenwahrscheinlichkeit, die in grau auf der zweiten Y-Achse darstellt ist.

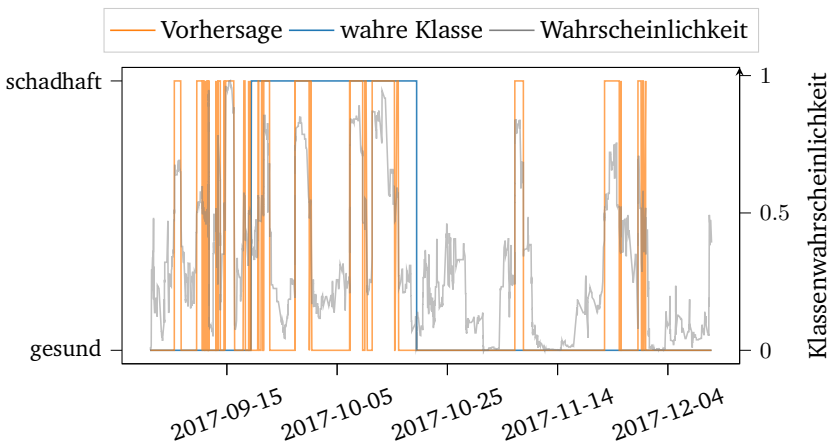


Abbildung 4.17.: Klassifikation der Testdaten von WKA T07 über die Zeit ausschließlich auf Basis von Ereignisdaten

Aus dieser Darstellung lässt sich die Konfusionsmatrix ableiten, die in Abbildung 4.18 dargestellt ist. Aus den Konfusionsmatrizen werden die Klassifikationsmetriken abgeleitet. Die Metriken für die dargestellte Konfusionsmatrix und die Ergebnisse der anderen WKA sind in Tabelle 4.6 zusammengefasst. Anzumerken ist dabei, dass die Ergebnisse von Anlage 01, die im Testzeitraum keine Ausfälle aufweist, keine Berechnung von Precision und Recall zulässt.

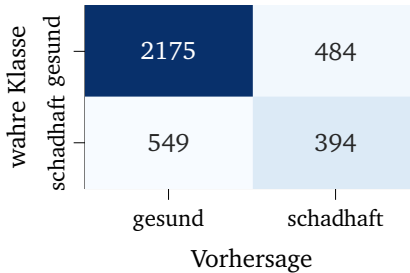


Abbildung 4.18.: Konfusionsmatrix der Testdaten von WKA T07 bei der Klassifikation ausschließlich auf Basis von Ereignisdaten

Tabelle 4.6.: Metriken zu der Konfusionsmatrix aus Abbildung 4.18

WKA	Accuracy	Recall	Precision	f1-Score
T01	0,92	-	-	-
T06	0,72	0,33	0,44	0,37
T07	0,71	0,42	0,45	0,43
T09	0,58	0,46	0,70	0,55
T11	0,74	0,50	0,38	0,44
Mittelwert	0,74	0,34	0,39	0,36
Std-abw.	0,12	0,20	0,25	0,21

Die Ergebnisse sind für alle Anlagen besser als eine Zufallsklassifikation und zeigen damit, dass im Allgemeinen eine Klassifikation auf Basis der Ereignisdaten möglich ist. Allerdings weisen sowohl Accuracy, als auch Precision und Recall (respektive f1-Score) Potenziale zur Verbesserung auf. Eine erfolgreiche Anwendung und Umsetzung eines CBM-Systems auf Basis dieser Modellergebnisse ist nicht empfohlen.

Selektion der Merkmale

Basierend auf den Ergebnissen dieser Modelle werden die generierten Merkmale im folgenden beurteilt und selektiert. Der vollständige Merkmalsraum basierend auf den Ereignisdaten enthält 364 Merkmale. Da explizit keine Vermischung der Merkmale gewünscht ist, um die Diskussion des Konzepts nicht zu beeinträchtigen, wird ausschließlich eine Selektion heran gezogen und bewusst auf eine Reduktion (im Sinne des Verschneidens mehrerer Merkmale) verzichtet. Durch die algorithmische Umsetzung im Form von Entscheidungsbäumen gibt der Algorithmus eine Rangfolge der Wichtigkeit der Merkmale (engl. *feature importance*) aus. Dabei wird für jedes Modell (jede WKA) ein eigenes Ranking erstellt. Die Rankings werden über alle fünf Modelle bzw. Anlagen gemittelt und absteigend sortiert. Auf dieser Basis lassen sich Aussagen über die kumulative Wichtigkeit der Merkmale treffen und eine Selektion vornehmen. Um 95% der kumulierten Wichtigkeit abzudecken (vergleichbar mit erklärter Varianz), werden 201 Merkmale benötigt. Mit 100 Merkmalen lassen sich 83% erreichen und mit 50 Merkmalen 69%. Dabei werden 100 Merkmale als Trade-Off zwischen einem akzeptablen Verlust an Information und einer handhabbaren Anzahl an zu trainierenden Parametern eingeschätzt. Entsprechend werden die Ergebnisse im Folgenden anhand der 100 relevantesten Merkmale, gemittelt über alle Modelle (WKA), gezeigt. Weitere Ausführungen zu den Einflussgrößen der Merkmale lassen sich den Abbildungen A.3 und A.4 im Anhang entnehmen.

Werden die Modelle nun mit dem reduzierten Merkmalsraum trainiert, verändern sich dadurch die Ergebnisse der Klassifikation. Diese sind in Tabelle 4.7 zusammengefasst. Allgemein sind die Auswirkungen der Merkmalsreduktion auf die Ergebnisse gering. Insgesamt ergibt sich sogar eine leichte Verbesserung der durchschnittlichen Accuracy sowie Precision, wobei auch die Streuung in den Ergebnissen leicht ansteigt.

Tabelle 4.7.: Klassifikationsmetriken mit reduziertem Merkmalsraum

WKA	Accuracy	Recall	Precision	f1-Score
T01	0,97	-	-	-
T06	0,71	0,43	0,43	0,43
T07	0,68	0,32	0,37	0,34
T09	0,64	0,47	0,80	0,59
T11	0,85	0,47	0,65	0,55
Mittelwert	0,77	0,34	0,45	0,38
Std-abw.	0,14	0,20	0,31	0,23

4.5.2. Vergleichsansatz: Diagnose auf Basis der Sensordaten

Der Vergleichsansatz stellt den aktuellen Stand der Technik dar und wird auf Basis der in 10-Minuten-Intervallen aufgezeichneten Sensordaten (SCADA-Daten) umgesetzt. Die Randbedingungen, also das Labeling der Daten mittels einer Zeitspanne von 30 Tagen vor dem Ausfall sowie die Festlegung von Trainings- und Testdaten bleiben dabei unverändert.

Als Merkmale werden 41 der insgesamt 81 Sensoren verwendet. Dabei wird insbesondere auf netzbezogene Variablen verzichtet, da sie nach Betrachtung der deskriptiven und explorativen Analyse keinen Mehrwert für die Zustandsdiagnose liefern. Die Liste der verwendeten Merkmale ist im Anhang in Tabelle A.5 zu finden.

Die Ergebnisse des Vergleichsansatzes sind in Tabelle 4.8 zusammengefasst. Insgesamt zeigt sich ein vergleichbares Bild mit denen der zuvor gezeigten Klassifikation auf Basis der Ereignisdaten in Abschnitt 4.5.1. Während die mittlere Accuracy und Recall leicht unterhalb der Klassifikation auf Basis der Ereignisdaten liegt, ist die Precision etwas höher. Insgesamt ist die Streuung der Metriken über alle Anlagen geringfügig höher.

Es zeigt sich damit, dass auch der Vergleichsansatz prinzipiell eine Klassifikation des Zustands ermöglicht, die absolute Leistung der Klassifikation aufgrund der zu hohen Ungenauigkeit jedoch für die tatsächliche Anwendung als nicht sinnvoll erachtet wird. Im Folgenden wird der Vergleichsansatz um die Merkmale der Ereignisdaten erweitert und die daraus resultierenden Ergebnisse verglichen.

Tabelle 4.8.: Klassifikationsergebnisse des Vergleichsansatzes

WKA	Accuracy	Recall	Precision	f1-Score
T01	0, 81	-	-	-
T06	0, 68	0, 48	0, 45	0, 46
T07	0, 67	0, 64	0, 46	0, 53
T09	0, 76	0, 67	0, 77	0, 71
T11	0, 51	0, 59	0, 13	0, 21
Mittelwert	0, 68	0, 48	0, 36	0, 39
Std-abw.	0, 12	0, 28	0, 30	0, 28

4.5.3. Vorgeschlagenes Konzept: Diagnose auf Basis der fusionierten Daten

Nach der Fusion der Ereignisdaten-basierenden Merkmale mit den Sensordaten-basierten Merkmalen (vgl. Abschnitt 3.3) ergeben sich daraus in einem dritten Modell die finalen Ergebnisse des vorgeschlagenen Konzepts. Diese sind für den Vergleichsfall mit den zuvor genannten Ausprägungen der Zeitfenster w_{fail} , $w_{obs_{stats}}$ und $w_{obs_{seq}}$ im Folgenden dargestellt und werden anschließend mit den Ergebnissen des Vergleichsansatzes gegenübergestellt.

Für das vorgeschlagene Konzept wird der Eingaberaum in das Modell ausgehend vom Vergleichsansatz mit den 100 reduzierten Merkmalen erweitert. Die Fusion erfolgt basierend auf dem nächstmöglichen Zeitstempel. Folglich basiert das Modell auf 141 Merkmalen und den gegebenen Labels des Vergleichsansatzes. Die funktionale Anforderung FR4 der einfachen Erweiterbarkeit von bestehenden Methoden aus Tabelle 3.1 ist damit erfüllt. Die Hyperparameter des Algorithmus werden weiterhin nicht verändert. Die Ergebnisse der Klassifikation für das vorgeschlagene Konzept sind in Tabelle 4.9 zusammengefasst.

Auf Basis dieser Ergebnisse kann nun die Gegenüberstellung von Vergleichsansatz und dem vorgeschlagenen Konzept vorgenommen werden und damit das Konzept evaluiert werden. Die Gegenüberstellung ist in Tabelle 4.10 zusammengefasst und zeigt, dass insbesondere die Accuracy sowohl beim Mittelwert als auch im Median verbessert werden kann. Während die falsch negativ Rate beim Konzept gegenüber dem Vergleichsansatz ansteigt (Recall sinkt), sinkt die falsch positiv Rate (Precision steigt). Der f1-Score ändert sich damit nicht signifikant. Ebenfalls kann die Varianz der Diagnose insgesamt

Tabelle 4.9.: Klassifikationsergebnisse des vorgeschlagenen Konzepts

WKA	Accuracy	Recall	Precision	f1-Score
T01	0,98	-	-	-
T06	0.69	0.47	0.47	0.47
T07	0.74	0.39	0.60	0.47
T09	0.69	0.51	0.71	0.59
T11	0.85	0.19	0.26	0.22
Mittelwert	0,78	0,31	0,41	0,35
Std-abw.	0,13	0,21	0,28	0,24

leicht reduziert werden, was in Form einer geringeren Standardabweichung bei den Metriken Recall, Precision und f1-Score ersichtlich wird.

Tabelle 4.10.: Klassifikationsergebnisse des vorgeschlagenen Konzepts

	Konzept				Vergleichsansatz			
	Acc.	Recall	Prec	f1	Acc.	Recall	Prec	f1
Mittelwert	0,78	0,31	0,41	0,35	0,68	0,48	0,36	0,39
Median	0,74	0,39	0,47	0,47	0,68	0,59	0,45	0,46
Std-Abw.	0,13	0,21	0,28	0,24	0,12	0,28	0,30	0,28

Die grafische Gegenüberstellung der Metriken Accuracy und f1-Score ist in Abbildung 4.19 für die einzelnen WKA gegeben. Aus der Darstellung wird ersichtlich, dass die Verbesserungen der Accuracy maßgeblich durch die Anlagen T01 und T11 herrühren, während der Vergleichsansatz z. B. bei T09 sowohl bei der Accuracy als auch beim f1-Score bessere Ergebnisse liefert als das vorgeschlagene Konzept.

Zur übersichtlichen Gegenüberstellung der Ergebnisse werden in Vergleichsplots in Abbildung 4.20 die Differenzen der Metriken der einzelnen Modelle gebildet und farblich markiert. Grün hinterlegt sind dabei die Fälle, in denen das vorgeschlagene Konzept bessere Ergebnisse als der Vergleichsansatz liefert, rot hinterlegt sind die Fälle in denen der Vergleichsansatz bessere Ergebnisse liefert. Die Darstellung bestätigt die bereits getroffenen allgemeinen Aussagen zu Accuracy und f1-score und führt die Ergebnisse zu Recall und Precision weiter aus. Ersichtlich wird daraus, dass das vorgeschlagene Konzept insbesondere bei Anlage T09 in allen Metriken schlechtere

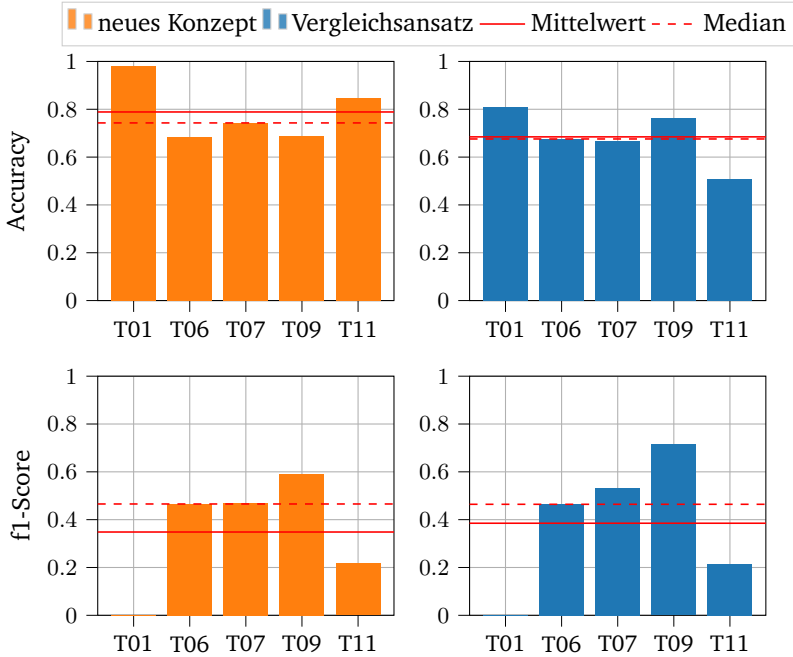


Abbildung 4.19.: Vergleich von vorgeschlagenem Konzept und Vergleichsansatz für das Vergleichsparameterset

Ergebnisse liefert als der Vergleichsansatz.

Abschließend wird mit einem statistischen Test die Signifikanz der Null-Hypothese getestet, die besagt, dass die beiden Verteilungen für die jeweiligen Metriken gleich sind. Aufgrund der unverbundenen Stichproben, einer angenommenen Normalverteilung innerhalb der Klassen, sowie Gleichheit der Varianz (zutreffend auf Accuracy) wird ein zweiseitiger Student-T-Test herangezogen. Mit einem p-Wert von 0,21 wird die Nullhypothese angenommen, was bedeutet, dass die Mittelwerte der beiden Verteilungen - statistisch gesehen - nicht signifikant unterschiedlich sind. Zu beachten ist hierbei, dass die Aussagekraft des t-Test bei einer so geringen Stichprobengröße eingeschränkt ist und deutliche Unterschiede zwischen den Verteilungen für eine statistisch signifikante Unterscheidung benötigt würden.

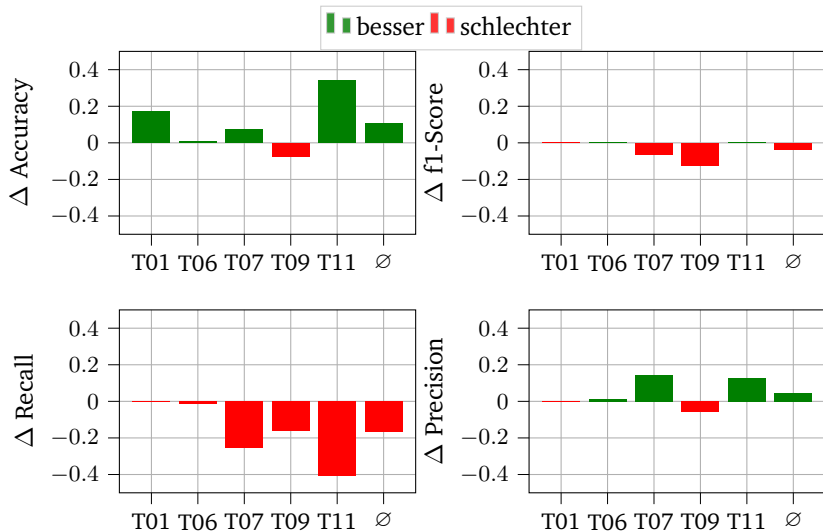


Abbildung 4.20.: Vergleich von vorgeschlagenem Konzept und Vergleichsansatz für das Vergleichsparameterset anhand der Differenzen

Ergebnisse der Kreuzvalidierung

Um die Robustheit der Ergebnisse hinsichtlich der gewählten Trainingsdatenbasis zu untersuchen, wird eine 5-fache Kreuzvalidierung angewendet. Dabei wird der Datensatz in fünf gleich große Teile (*Folds*) unterteilt, wobei immer vier Teile zum Training und der verbleibende Teil zum Testen verwendet wird. Daraus entstehen fünf Modellergebnisse die abschließend gemittelt werden. Generell wird bei dem Verfahren im Falle der vorliegenden Klassifikation versucht, in allen Teilen ähnliche Klassenverteilungen zu erreichen. Entsprechend sind die Daten zeitlich zumindest teilweise durchmischt. Das bedeutet, dass zwei aufeinander folgende Datenpunkte in Trainings- und Testdatenbasis geteilt werden können. Weiterhin verhindert die Kreuzvalidierung per sé kein Overfitting, soll jedoch aufgrund der Nutzung aller Daten für das Trainieren und Testen eine ungünstige Wahl der Trainingsdaten vermeiden. Die Ergebnisse der Kreuzvalidierung des vorgeschlagenen Konzepts sowie des Vergleichsansatzes sind in Abbildung 4.21 dargestellt.

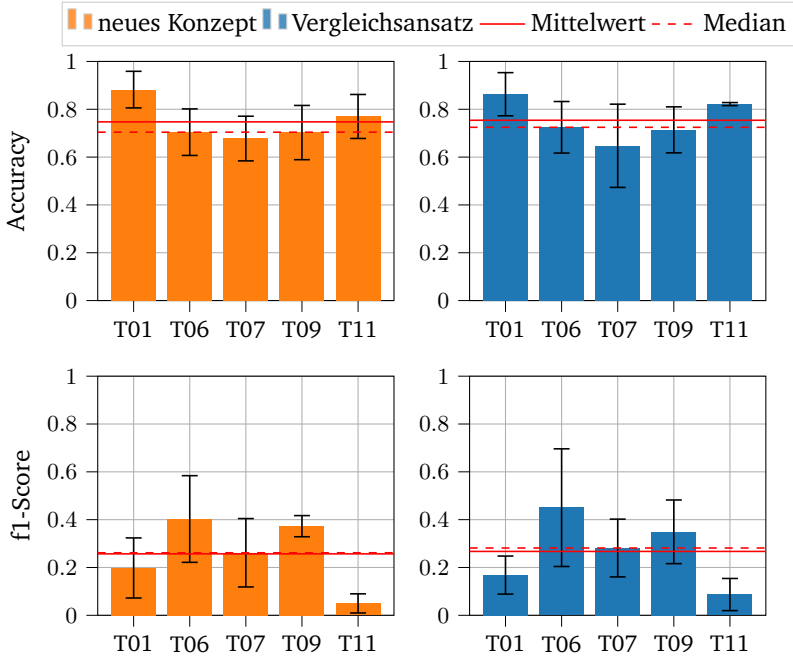


Abbildung 4.21.: Vergleich von vorgeschlagenem Konzept und Vergleichsansatz auf Basis der Kreuzvalidierung

Beim Vergleich mit den zuvor dargestellten Ergebnissen auf Basis des manuellen Trainings- und Testsplits (vgl. Abbildung 4.19) wird deutlich, dass der Vergleichsansatz bei der Kreuzvalidierung geringfügig besser abschneidet (rund 5% bei der Accuray) und das vorgeschlagene Konzept etwas schlechter (rund 4% bei der Accuray). Bei der Gegenüberstellung von Konzept und Vergleichsansatz auf Basis der Kreuzvalidierungsergebnisse zeigt sich demnach kein signifikanter Unterschied der Metriken. Die Ergebnisse des Konzepts zeigen tendenziell geringere Streuungen innerhalb der fünf Kreuzvalidierungsergebnisse, die Effekte sind jedoch nicht signifikant.

Zusammenfassend zeigt sich, dass die Ergebnisse der Kreuzvalidierung die Ergebnisse auf Basis des manuellen Splits nicht bestätigen. Insbesondere der f1-Score zeigt bei der Kreuzvalidierungen schlechtere Ergebnisse, sowohl für den Vergleichsansatz als auch für das neue Konzept.

Ergebnisse der Diagnose nach Komponenten

Neben dem Aufbau eines Diagnosemodells je WKA ist es ebenfalls möglich, ein Modell je Komponente bzw. Baugruppe aufzubauen. Die Ergebnisse der Modellbildung nach Baugruppen werden im Folgenden vorgestellt.

Die Herausforderung bei der Modellbildung je Baugruppe ist, dass sich einerseits die bestehende Klassenimbalance verschlechtert und andererseits die Trainingsdatenbasis ungleich verteilt ist. So liegen für die Baugruppe Transformator z. B. lediglich drei Ausfälle im gesamten Zeitraum vor. Im Testzeitraum des manuellen Trainings- und Testsplits finden sich nur Ausfälle an den Baugruppen Getriebe und Hydraulik. Die Ergebnisse der Klassifikation sind in Abbildung 4.22 dargestellt. Zwar zeigen die Ergebnisse allgemein höhere Accuracies als bei der Modellbildung je WKA, jedoch deutlich niedrigere f1-Scores, was durch die höhere Klassenimbalance begründet sein kann. Weiterhin zeigt das vorgeschlagene Konzept auch für diesen Fall durchschnittlich 12, 2% bessere Ergebnisse als der Vergleichsansatz bei Betrachtung der Accuracy. Der f1-Score des Konzepts fällt jedoch um 6% schlechter aus als beim Vergleichsansatz. Insbesondere bei den Baugruppen ohne einen Schaden im Testzeitraum (Generator, Generatorlager und Transformator) bedeuten die besseren Ergebnisse des vorgeschlagenen Konzepts eine geringere Anzahl an falsch positiven Klassifikationsergebnissen.

Auf eine Darstellung der Ergebnisse der Kreuzvalidierung bei der Modellbildung je Baugruppe wird an dieser Stelle verzichtet, da die Ergebnisse im Wesentlichen mit denen der Kreuzvalidierung der Modellbildung nach WKA übereinstimmen. Insgesamt zeigen das vorgeschlagene Konzept und der Vergleichsansatz dabei vergleichbare Accuracies, während das vorgeschlagene Konzept geringfügig bessere f1-scores als der Vergleichsansatz aufweist. Generell wird die Modellgüte mit f1-Scores unterhalb von 0, 2 jedoch als ungeeignet für eine CBM-Anwendung bewertet.

Die bisher dargestellten Ergebnisse zeigen die prinzipielle Eignung der Ereignisdaten für eine Zustandsdiagnose. Weiterhin wurde eine geringfügige Verbesserung der Diagnoseergebnisse durch die Fusion von Ereignisdaten und Sensordaten gegenüber dem Vergleichsansatz nachgewiesen. Da die Merkmale, die aus den Ereignisdaten generiert werden, jedoch verschiedenen Einflussparameter unterliegen, soll im Folgenden eine Sensitivitätsanalyse Aufschluss über den Einfluss dieser Parameter geben.

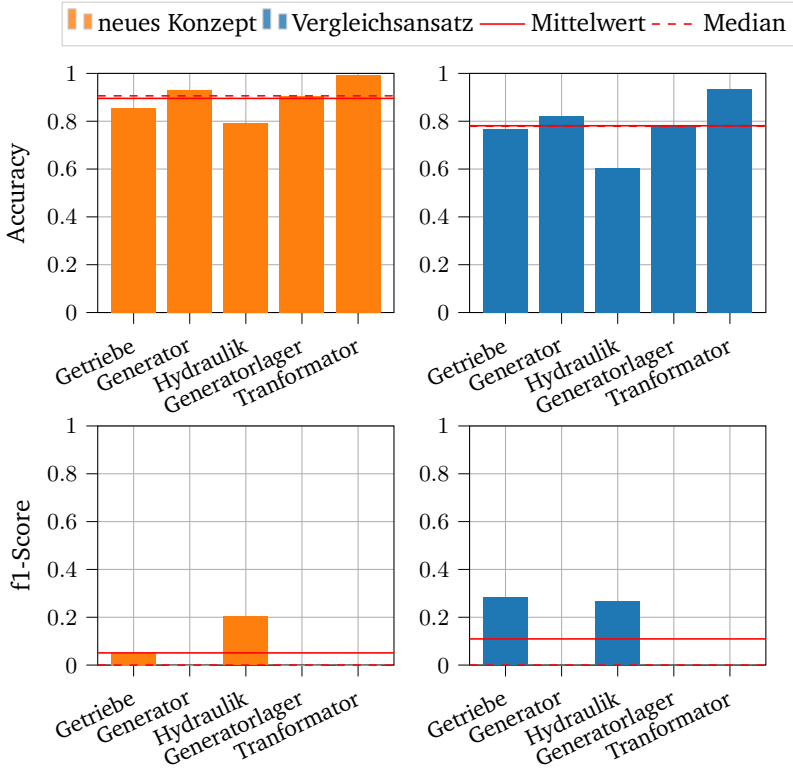


Abbildung 4.22.: Vergleich nach Komponenten

4.6. Sensitivitätsanalyse

Die zuvor gezeigten Ergebnisse stellen den Vergleichsfall der Ergebnisse dar und erlauben aufgrund des Grades an Annahmen, insbesondere bei der Generierung der Merkmale und Label der Daten, nur bedingt allgemeingültige Rückschlüsse auf den Vorteil des Konzepts. Die Sensitivitätsanalyse liefert hierfür Einblicke in potenzielle Einflussgrößen der Parameter, die während der Generierung der Merkmale variiert wurden.

Bei der folgenden Darstellung der Ergebnisse wird zwischen allgemeinen Variationsparametern und solchen Parametern, die nur die Ereignismerkmale betreffen, unterschieden. Als allgemeiner Parameter wird z. B. die Fensterbreite des Ausfallfensters w_{fail} betrachtet, da dieser sowohl für die Ereignisdaten, als auch für die Sensordaten einen Einfluss auf die Ergebnisse hat. Parameter, die die Ereignismerkmale betreffen sind hingegen jene Variationsparameter, die während der Generierung der Ereignismerkmale verwendet wurden, z. B. die Fenster $w_{obs_{stats}}$ und $w_{obs_{seq}}$ oder die verwendeten Kategorien. Die Ergebnisse des Vergleichsansatzes bleiben von diesen Veränderungen unberührt. Tabelle 4.11 gibt einen Überblick über die im Rahmen der Sensitivitätsanalyse untersuchten Parameter und deren Ausprägungen. Die Veränderung der Parameter wird dabei unabhängig vorgenommen. Damit werden keine Kreuzeffekte erfasst und untersucht. Die Wahl der Fenstergrößen und -schrittweiten der Veränderungen basiert auf Annahmen und Erfahrungswerten.

Tabelle 4.11.: Ausprägungen der Variationsparameter in der Sensitivitätsanalyse (Vergleichsfall unterstrichen)

Parameter	Ausprägungen
Ausfallfenster w_{fail}	3d, 7d, 15d, <u>30d</u> , 50d
Fenster für statistisch basierte Merkmale $w_{obs_{stats}}$	12h, 1d, <u>3d</u> , 5d, 7d
Fenster für sequenzbasierte Merkmale $w_{obs_{seq}}$	3h, 6h, <u>12h</u> , 1d, 2d

Die Ergebnisse der Sensitivitätsanalyse werden nachfolgend in der Ordnung der genannten Fenster anhand der Modellbildung nach WKA sowie für den manuellen Trainings- und Testdatensplit dargestellt.

4.6.1. Variation des Ausfallfensters

Die Größe des Ausfallfensters beeinflusst, aufgrund der Veränderung der Label, sowohl den Vergleichsansatz als auch das vorgeschlagene Konzept. Entsprechend werden bei der Analyse der Ergebnisse ebenfalls Veränderungen beim Vergleichsalgorithmus analysiert.

Die Entwicklung der Accuracy sowie des f1-Scores sind in Abbildung 4.23 über die verschiedenen Ausfallfenster exemplarisch für WKA T07 und gemittelt über alle WKA dargestellt. Aus der Abbildung ist der generelle Trend erkennbar, dass die Accuracy mit zunehmend längeren Ausfallfenstern abnimmt und der f1-Score gleichzeitig ansteigt.

Insbesondere bei sehr kurzen Ausfallfenstern von drei oder sieben Tagen ist der f1-Score dabei null oder nahezu null. Das bedeutet, dass hierbei keine korrekte Klassifikation vorgenommen wird und alle Datenpunkte der gesunden Klasse zugeteilt werden. Die Accuracy bleibt dabei aufgrund der großen Klassenimbalance, bzw. der sehr kleinen Anzahl an schadhafte Datenpunkten hoch. Folglich wird die Klassifikation beider Ansätze erst ab einem angenommenen Schadensfenster von 15 Tagen akzeptabel. Gleichzeitig zeigt der Vergleich der Metriken von Konzept und Vergleichsansatz über alle Fenster hinweg eine höhere Accuracy beim vorgeschlagenen Konzept, jedoch einen niedrigeren f1-Score.

Um die Vor- und Nachteile des vorgeschlagenen Konzeptes über die verschiedenen Ausfallfenster diskutieren zu können, wird die weitere Ergebnisdarstellung in Abbildung 4.24 auf die Differenzen zwischen Konzept und Vergleichsansatz reduziert. Dabei werden die vorher gezeigten Vergleichsplots genutzt und lediglich der Mittelwert der Metriken über alle WKA diskutiert sowie die Streuung in Form der Standardabweichung dargestellt. Über alle Ausfallfenster zeigt sich dabei bei der Accuracy eine gleichbleibende Verbesserung des Konzepts gegenüber des Vergleichsansatzes von rund 0,12 Prozentpunkten. Die Varianz ist dabei insbesondere bei längeren Ausfallfenstern größer, was vor allem in den Anlagen T01 und T11 begründet ist. Für den f1-Score zeigt sich, wie bereits in Abbildung 4.23 dargestellt, eine durchschnittliche Verschlechterung der Metrik über alle Ausfallfenster. Einzig bei WKA T11 schneidet das Konzept im Mittel geringfügig besser ab als der Vergleichsansatz. Insbesondere der Recall verschlechtert sich durch die Fusion mit Ereignisdaten. Bei der Precision zeigt sich keine eindeutige Veränderung über die Ausfallfenster.

Beim Vergleich der Streuung der fünf Modelle (je WKA) wird ersichtlich, dass das vorgeschlagene Konzept insbesondere bei kürzeren Ausfallfenstern über alle Metriken geringere Standardabweichungen aufweist als der Vergleichsansatz. Beim f1-Score sowie dem Recall zeigt das vorgeschlagene Konzept gegenüber dem Vergleichsansatz geringere Standardabweichungen, während die Accuracy und Precision keine signifikante Differenz aufweist. Über alle Metriken hinweg steigt dabei die Standardabweichung mit längeren Ausfallfenstern an. Die entsprechende Visualisierung der Standardabweichungen beider Ansätze über das Ausfallfenster ist in Abbildung A.5 im Anhang gegeben.

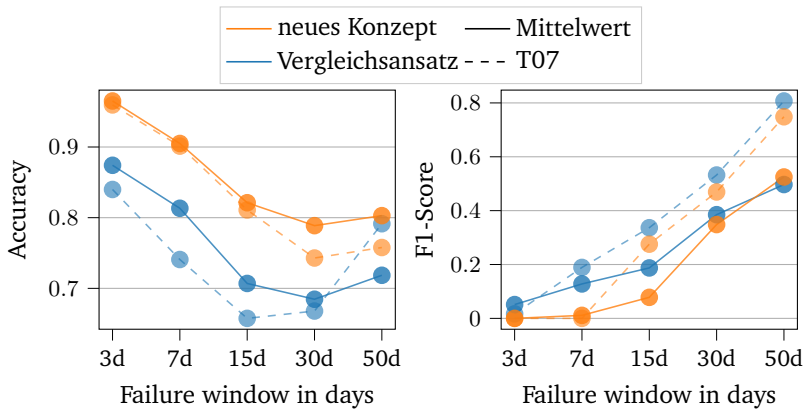


Abbildung 4.23.: Ergebnisse der Sensitivitätsanalyse des Ausfallfensters am Beispiel von WKA T07

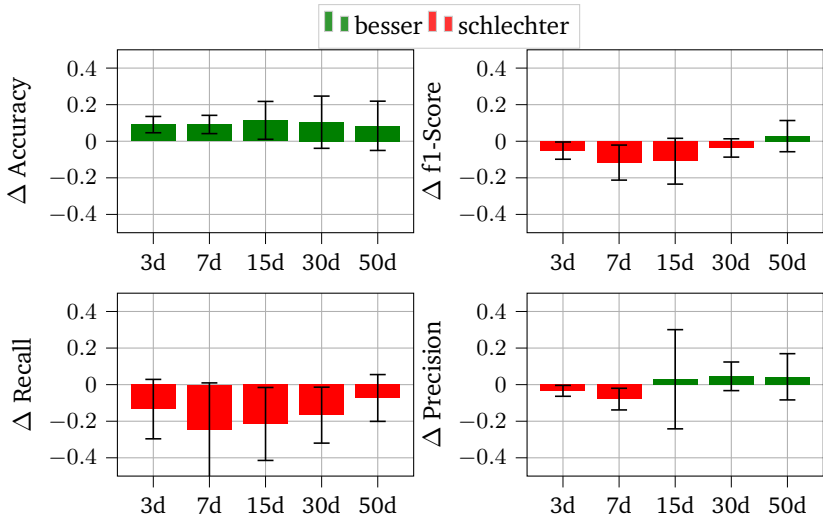


Abbildung 4.24.: Ergebnisse der Sensitivitätsanalyse des Ausfallfensters gemittelt über alle WKA

4.6.2. Variation des Fensters für statistisch basierte Merkmale

Nach der Darstellung des Einfluss des Ausfallfensters, wird in diesem Abschnitt die Sensitivität des Konzepts auf die Größe des Fensters für statistisch basierte Merkmale untersucht. Das Fenster nimmt nur Einfluss auf die Merkmale, die aus den Ereignisdaten generiert werden, weswegen die Ergebnisse des Vergleichsansatzes in diesem Abschnitt unverändert konstant bleiben und dem Vergleichsfall aus Abschnitt 4.5.2 entsprechen.

In Abbildung 4.25 ist analog zur vorherigen Darstellung der Verlauf der Metriken der Accuracy sowie des f1-Score über die verschiedenen Fenstergrößen dargestellt. Dabei fällt auf, dass die Merkmalsgenerierung über besonders kurze Zeitfenster von 12 Stunden keinen signifikanten Vorteil beim Konzept liefern. Generell steigt der Zugewinn in der Accuracy mit längeren Zeitfenstern an, während der f1-Score tendenziell abnimmt. Die Ausnahme bildet das längste Zeitfenster von sieben Tagen: hier steigt der f1-Score wieder auf nahezu das Niveau des Vergleichsansatzes an.

Die oben getroffenen Aussagen werden durch die Darstellung der Differenzen der Metriken in Abbildung 4.26 bestätigt. Die maximale Verbesserung in der Accuracy wird vom vorgeschlagenen Konzept gegenüber dem Vergleichsansatz bei einer Fenstergröße von sieben Tagen erreicht. Dabei verschlechtert sich der Recall und verbessert sich die Precision was in einem unveränderten f1-Score resultiert.

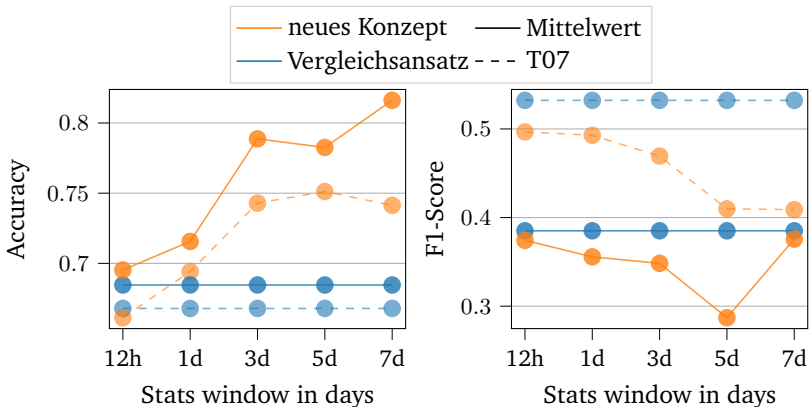


Abbildung 4.25.: Ergebnisse der Sensitivitätsanalyse des statistisch basierten Fensters am Beispiel von WKA T07

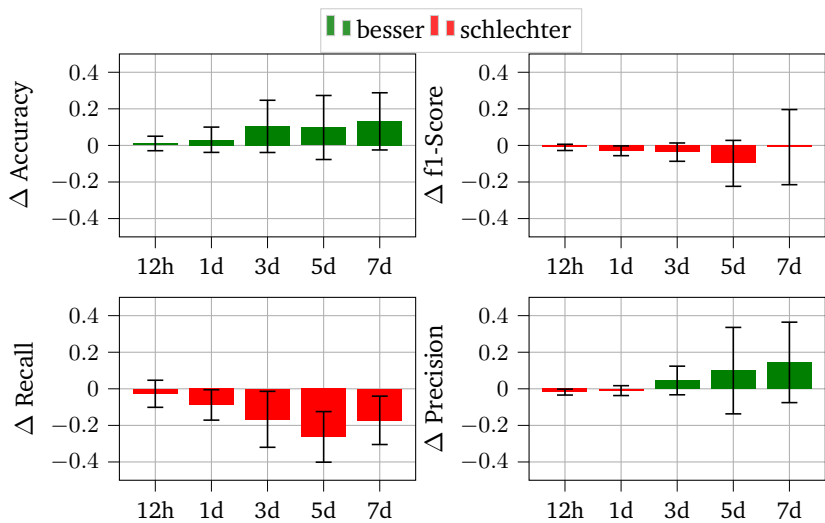


Abbildung 4.26.: Ergebnisse der Sensitivitätsanalyse des statistisch basierten Fensters gemittelt über alle WKA

Bei Betrachtung der Varianzen (vgl. Abbildung A.6 im Anhang) zeigt sich kein signifikanter Unterschied zwischen Konzept und Vergleichsansatz. Bei der Accuracy zeigt das vorgeschlagene Konzept bei kleinen Zeitfenstern zwar eine geringere Streuung (Standardabweichung), bei größeren Zeitfenstern ist sie jedoch höher als die des Vergleichsansatzes. Der f1-Score und Recall zeigen beim vorgeschlagenen Konzept insgesamt eine geringere Streuung, während die Precision ähnlich wie die Accuracy von geringerer Streuung auf größere Streuung mit längeren Zeitfenstern ansteigt.

4.6.3. Variation des Fensters für sequenzbasierte Merkmale

Zuletzt wird der Einfluss der Fenstergröße der sequenzbasierten Merkmale analysiert. Da die Länge des Fensters auch einen Einfluss auf die Anzahl der generierten Merkmale hat, werden hierbei entgegen der vorherigen Sensitivitätsanalysen auch der Merkmalsraum, sowie die Namen der Merkmale verändert. Zwangsweise ändern sich damit auch die Ergebnisse des Vergleichsfalls bei einer Fenstergröße von 12 Stunden für die sequenzbasierten Merkmale.

Analog zu den vorherigen Darstellungen ist die absolute Entwicklung der Accuracy sowie des f1-Scores in Abbildung 4.27 dargestellt. Generell fällt der Gewinn an Accuracy des vorgeschlagenen Konzepts gegenüber dem Vergleichsansatz geringer aus. Auch bei der Fenstergröße von 12 Stunden liegt die hier dargestellte Accuracy unterhalb der des Vergleichsfalls. Das kann unter anderem daran liegen, dass für diesen Fall keine Merkmale reduziert bzw. selektiert wurden und der Merkmalsraum damit eine höhere Dimension hat, als im Vergleichsfall. Generell sind die Auswirkungen der Fenstergröße auf die Ergebnisse vergleichsweise gering. Die höchsten Accuracies werden bei kürzeren Zeitfenstern für sequenzbasierte Merkmale erreicht. Die Veränderungen der Fenster hat keinen signifikanten Einfluss auf den f1-Score.

Das zuvor skizzierte Bild wird durch die Darstellung der Differenzen der Metriken von Vergleichsansatz und vorgeschlagenem Konzept in Abbildung 4.28 bestätigt. Dabei kann ein nahezu gleichbleibende Verbesserung der Accuracy durch das Konzept festgestellt werden, während der f1-Score geringfügig schlechter ausfällt, maßgeblich begründet durch den signifikant schlechteren Recall des Konzepts gegenüber dem Vergleichsansatz. Bei der Precision zeigt sich über alle Fenstergrößen hinweg kein erwähnenswerter Unterschied.

Die Analyse der Streuung über die einzelnen Modelle (Anlagen) zeigt, dass kürzere Zeitfenster zu einer höheren Standardabweichung bei der Accuracy und niedrigeren Standardabweichung für den f1-Score führen. Die Streuung aller vier Metriken nähert

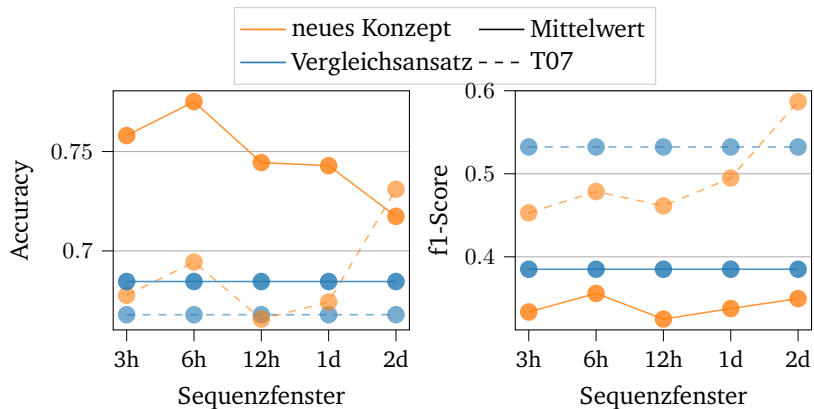


Abbildung 4.27.: Ergebnisse der Sensitivitätsanalyse des sequenzbasierten Fensters am Beispiel von WKA T07

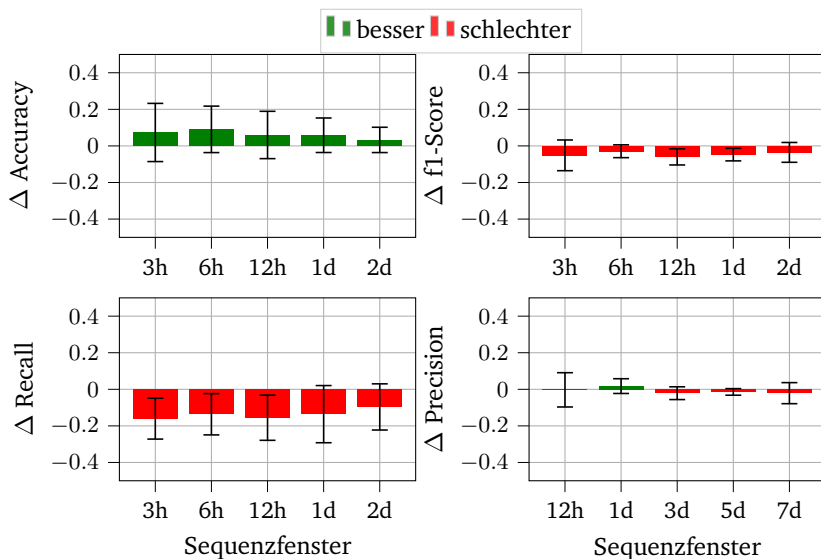


Abbildung 4.28.: Ergebnisse der Sensitivitätsanalyse des sequenzbasierten Fensters gemittelt über alle WKA

sich bei längeren Beobachtungsfenstern von einem bzw. zwei Tagen den Standardabweichungen des Vergleichsansatzes. Die entsprechende Visualisierung kann Abbildung A.7 im Anhang entnommen werden.

Zusammenfassung

Zusammenfassend zeigt die Sensitivitätsanalyse für jeden Variationsparameter isoliert die Einflussgrößen auf die Ergebnisse der Diagnose. Für die Variation des Ausfallfensters zeigt sich dabei ein nahezu identischer Einfluss auf das vorgeschlagene Konzept wie auf den Vergleichsansatz. Bei der Generierung der statistisch basierten Merkmale zeigen sich starke Effekte auf die Accuracy in Abhängigkeit der gewählten Fenstergröße. Mit längeren Beobachtungsfenstern bei der Generierung statistisch basierter Merkmale steigt dabei die Accuracy des vorgeschlagenen Konzepts. Die Auswirkungen der veränderten Fenstergröße auf sequenzbasierte Merkmale sind hingegen gering und zeigen keinen eindeutigen Trend. Die abschließende Diskussion und kritische Betrachtung der Ergebnisse mit Ableitung von Grenzen für das vorgeschlagene Konzept wird in Kapitel 5 beschrieben.

5. Diskussion des vorgeschlagenen Konzepts und der verwendeten Methodik

Nachdem die Ergebnisse zuvor in verschiedenen Konfigurationen dargestellt wurden, fasst das vorliegende Kapitel die Kernergebnisse hinsichtlich der gestellten Anforderungen und Hypothesen zusammen und beleuchtet die Grenzen des Konzepts.

5.1. Zusammenfassung der Ergebnisse und Diskussion des Konzepts

Die generelle Güte der Ergebnisse, sowohl des Vergleichsansatzes als des vorgeschlagenen Konzepts, wird als ungeeignet für eine reale Anwendung in der zustandsbasierten Instandhaltung beurteilt. Das ergibt sich maßgeblich durch die hohen entstehenden Ungenauigkeiten und hohe Anzahl an falsch-positiven bzw. -negativen Klassifikationsergebnissen. Die Gründe hierfür werden maßgeblich auf den verwendeten Klassifikationsansatz von Zeitreihendaten zurückgeführt. Im Rahmen der Modellbildung wurde eine Vielzahl an Klassifikationsalgorithmen untersucht. Neben dem final verwendeten Gradient Boosting Verfahren kamen auch einfache Entscheidungsbäume, SVM, kNN sowie Naive Bayes Klassifikatoren zum Einsatz. Trotz umfangreicher Hyperparameteroptimierungen konnten dabei keine Modelle identifiziert werden, die zu besseren Ergebnissen führten. Dies unterstreicht die zuvor getroffene Aussage, dass es der Ansatz der Klassifikation in diesem Fall nicht zielführend ist.

Auch die verwendete Evaluationsdatenbasis mit einer hohen Flut an Ereignisdaten und fehlenden Möglichkeiten der Interpretation erschweren die zielgerichtete Merkmalsentwicklung für die Ereignisdaten. Die mit der Datenbasis verbundenen Herausfor-

derungen werden in Abschnitt 5.2 tiefer beleuchtet. Dennoch können auf Basis der Gegenüberstellung der Ergebnisse die Grundgedanken des Konzepts diskutiert und potenzielle Grenzen beschrieben werden.

Die in Abschnitt 4.5.3 dargestellten Ergebnisse zeigen für den Vergleichsfall eine Verbesserung der Accuracy (Diagnosegenauigkeit) des vorgeschlagenen Konzepts gegenüber des Vergleichsansatzes. Die Ergebnisse beruhen dabei auf der Bildung von fünf einzelnen Modellen, eins für jede WKA. Im Mittel über die fünf Modelle kann die Diagnosegenauigkeit (Accuracy) durch das Konzept um 10% bei Mittelwert und 6% beim Median gesteigert werden. Ein statistisch signifikanter Unterschied kann auf Basis der geringen Stichprobenzahl nicht bestätigt werden. Die Anzahl der Falsch-Klassifikationen im Verhältnis zu Richtig-Positiv-Klassifikationen steigt mit dem vorgeschlagenen Konzept allerdings auch, was sich in einem geringfügig schlechteren f1-Score und Recall zeigt. Mit Blick auf die Definition der Metriken in Abschnitt 2.2.4 bedeutet die verbesserte Accuracy und der schlechtere f1-Score vereinfacht ausgedrückt, dass das vorgeschlagene Konzept zwar den gesunden Zustand besser detektiert (klassifiziert), aber nicht zwingend bei der Erkennung von schadhaften Zuständen hilft. Daneben zeigt sich jedoch auch, dass mit dem vorgeschlagenen Konzept die Varianz unter den Modellen geringfügig reduziert werden kann, was insgesamt eine höhere Robustheit und Verlässlichkeit der Modelle und Diagnoseergebnisse verspricht.

Bei Betrachtung der einzelnen Anlagen und dem Vergleich aller Ergebnisse lässt sich ein eindeutiger Trend erkennen. Unter Einbeziehung aller 15 Modellläufe der Sensitivitätsanalyse kann gezeigt werden, dass die mittlere Accuracy je WKA bei den Anlagen T01 sowie T11 durch das vorgeschlagene Konzept am meisten gegenüber dem Vergleichsansatz verbessert wird. Auch für Anlage T07 besteht eine Verbesserung der Accuracy, wenn auch weniger stark ausgeprägt. Bei den Anlagen T06 und T09 zeigen sich hingegen keine signifikanten Änderung oder eine Verschlechterung der Accuracy vom Konzept gegenüber des Vergleichsansatzes. Bei Betrachtung des f1-Scores zeigt sich ein weniger eindeutiges Bild: die generelle Verschlechterung der Metrik beim Konzept gegenüber des Vergleichsansatzes bleibt bestehen und kann vor allem Anlagen T06, T07 und T09 zugeschrieben werden, jedoch unterliegen die Mittelwerte einer hohen Streuung. Bei Anlage T11 wird in einigen Fällen auch eine Verbesserung festgestellt, sodass sich hier der f1-Score im Mittel nicht verändert. Eine Visualisierung des Zusammenhangs ist in Abbildung 5.1 gegeben.

Bei der Diskussion dieser Aussagen können auch die Schadensbilder der jeweiligen WKA hinzugezogen werden. Anlage T01 weist im Testzeitraum keine Schäden auf. Die dennoch höhere erreichbare Accuracy bedeutet, dass mit dem vorgeschlagenen Konzept demnach die falsch-positiv Klassifikationen reduziert werden. Diese Aussage

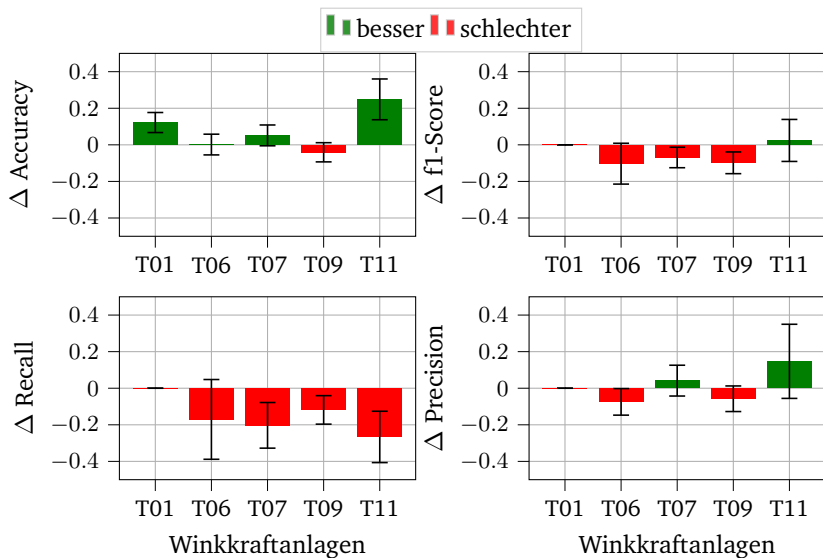


Abbildung 5.1.: Besser-schlechter Darstellung aller Ergebnisse gemittelt je WKA.

wird ebenfalls durch die baugruppenbasierte Klassifikation untermauert (vgl. Abbildung 4.22). Dabei zeigt sich eine geringfügige Verbesserung der Accuracy durch das Konzept bei allen Komponenten ohne Schäden oder Ausfälle im Testzeitraum. Weiterhin können die Vorteile in der Accuracy durch das Konzept bei WKA T07 und T11 mit der Baugruppe „Hydraulik“ korreliert werden. Beide Anlage weisen im Testzeitraum nur Schäden an dieser Baugruppe auf. Die Anlagen T06 und T09 zeigen im Testzeitraum hingegen Schäden an der Baugruppe „Getriebe“, die der Vergleichsansatz bereits gut detektiert. Beide Aussagen werden durch die Ergebnisse der baugruppenbasierten Klassifikation bestätigt. Abbildung 4.22 zeigt eine eindeutige Verbesserung der Accuracy bei annähernd gleichbleibendem f1-Score durch das Konzept. Bei der Klassifikation der Getriebeschäden ist hingegen eine signifikante Verschlechterung des f1-Scores bei annähernd gleichbleibender Accuracy ersichtlich.

Die 5-fache Kreuzvalidierung kann die Verbesserungen der Accuracy im Mittel nicht bestätigen. Für Anlage T01 und T07 ist zwar eine mittlere Verbesserung der Accuracy nachweisbar, der Effekt fällt jedoch deutlich geringer aus, als beim manuellen

Trainings-Test-Split. Weiterhin kann auch mit der Kreuzvalidierung eine Reduktion der Varianz gezeigt werden, aber es finden sich keine statistisch signifikanten Änderungen zwischen dem vorgeschlagenen Konzept und dem Vergleichsansatz. Es können einzelne Kreuzvalidierungspaare gefunden werden, bei denen das vorgeschlagene Konzept signifikante Verbesserungen bringt: Bei Fold 2 an WKA T07 zeigt das Konzept z. B. eine Verbesserung von 19% gegenüber dem Vergleichsansatz. Gleichzeitig finden sich jedoch auch Gegenbeispiele, wie an Fold 2 von WKA T11, in denen der Vergleichsansatz um 21% höhere Accuracies liefert als das Konzept. Die Verwendung der Kreuzvalidierung bei Zeitreihendaten ist jedoch im Allgemeinen zu hinterfragen, da durch die zeitliche Durchmischung der Daten üblicherweise auch Trainingsdaten verwendet werden, die zeitlich hinter den Testdaten liegen, was der Anwendung eines solchen Ansatzes im Realfall widerspricht.

Die funktionale Anforderung FR3, dass das Konzept anwendbar auf unterschiedliche Anwendungsfälle ist, kann im Rahmen der Arbeit nur bedingt getestet werden. Der Ansatz ist zwar generisch formuliert und wurde anhand eines passenden Problems evaluiert, dennoch zeigte die Ausarbeitung den großen Implementierungsaufwand, der damit verbunden ist. Dabei ist Expertenwissen, sowohl in der technischen Domäne der Anlage (Windkraft in diesem Fall), als auch im Aufbau entsprechender Analysemethoden und Monitoringansätze hilfreich bzw. erforderlich.

5.2. Grenzen des Konzepts

Der folgende Abschnitt fasst die Grenzen des Konzepts zusammen. Dabei wird zunächst auf die Grenzen hinsichtlich der vorliegenden Datenbasis und des konkreten Evaluationsfalls eingegangen. Im weiteren Verlauf des Abschnitts werden auch die generellen Gültigkeitsbereiche des Konzepts über den vorliegenden Datensatz hinaus adressiert.

Grenzen der verwendeten Evaluationsdatenbasis

Das Konzept zeigt auf Basis der Evaluationsdaten zwei wesentliche Grenzen: Einerseits kann die Klassifikation bei einer Datenbasis mit hoher Klassenimbalance nicht verbessert werden (vgl. in Abschnitt 4.5.3, Diagnose auf Basis der Baugruppen), andererseits zeigen die Ergebnisse nur Verbesserungseffekte bei gewissen Anlagen bzw. Schadensbildern.

Allgemein erfüllt die Evaluationsdatenbasis zwar die formulierten Anforderungen aus Abschnitt 3.1.2, jedoch zeigen sich bei der praktischen Umsetzen zwei wesentliche Herausforderungen hinsichtlich der Ereignisdaten. Generell zeigt die Evaluation am Beispiel der WKA jedoch die Herausforderungen bei der Anwendung eines solchen Konzepts ohne tiefes System- und Expertenwissen. Die Inhalte der Ereignismeldungen können nur ansatzweise interpretiert werden, wodurch keine zuverlässige Einschätzung der Relevanz einzelner Ereignis hinsichtlich eines Ausfalls vorgenommen werden kann. Weiterhin liegt insgesamt eine hohe Ereignisdatenflut vor. In Zusammenhang mit dem fehlenden Domänenwissen ist auch hier schwierig zu beurteilen, welche Ereignisse eine Aussagekraft hinsichtlich des Zustands der Anlagen enthalten und welche nicht.

Auch die zum Teil unscharfe Definition der Ausfalldaten (schwache Labels) erschwert eine belastbare Aussage des Konzepts. Gerade im Hinblick auf die im Testdatensatz enthaltenen Ausfälle zeigt sich, dass der bei WKA T09 beschriebene Getriebeschaden mit „Gearbox Noise“ relativ unspezifisch ist und möglicherweise keinen tatsächlichen Ausfall nach sich zieht. Die darauffolgenden Wartungsereignisse lassen zumindest keinen Rückschluss auf eine Reparatur oder sonstige aufwendigere Maßnahmen zu. Generell ist bei dieser Art der Problemstellung die Differenzierung zwischen einem beobachteten Zustand (in diesem Fall die Meldung „Gearbox Noise“) und dem tatsächlichen (wahren) Zustand des Systems eine große Herausforderung bzw. technisch nicht immer möglich. Die hier getroffene Annahme, dass jeder Eintrag einem Ausfall entspricht, ist in diesem Rahmen eine Vereinfachung, die zur Verfälschung der Ergebnisse führen kann und ohne Expertenwissen nicht zu prüfen ist.

Der präsentierten Evaluationsdatenbasis liegt die Annahme zugrunde, dass Ereignisdaten nicht (ausschließlich) auf Basis von Sensordaten generiert werden, sodass keine Multikollinearität zwischen Sensordaten und Ereignisdaten vorliegen. Diese Annahme kann ohne Experteninformationen nicht bestätigt oder widerlegt werden. Sollten allerdings ein Großteil der Ereignisdaten aus den bereits vorliegenden Sensordaten abgeleitet werden, wären Informationen im Diagnosemodell redundant vorhanden. Entsprechend gering fiel in diesem Fall der Mehrwert des vorgeschlagenen Konzeptes aus. Schlussfolgernd bringt das vorgeschlagene Konzept nur dann Vorteile und ist anwendbar, wenn die Ereignisdaten (nachweislich) unabhängig von bereits verwendeten Sensordaten generiert werden.

Generelle Grenzen des vorgeschlagenen Konzepts

Weitere Grenzen des Konzepts liegen in der häufig vorliegenden zeitlichen Unsicherheit von Ereignisdaten. Im Rahmen einer mehrjährigen Projektarbeit mit Lufthansa Technik wurde festgestellt, dass in der Luftfahrtindustrie eine Häufung von Einträgen und Ereignissen im Rahmen von Inspektionen (Checks) festgestellt werden kann [KS22]. Die Grenzen des vorgeschlagenen Konzepts zeigen sich in diesem Kontext, da nur Ereignisse berücksichtigt werden können, deren Zeitstempel auch dem tatsächlichen Auftreten entsprechen.

Daneben sollte die vorteilhafte Anwendbarkeit des Konzepts in Bezug auf Schadensbilder und -arten eingegrenzt werden. Bisherige Annahmen, dass das Konzept insbesondere bei Baugruppen einen großen Effekt erzielt, die durch rein sensorbasierte Monitoringansätze nicht gut überwacht werden können, wie z. B. elektrische Komponenten (Generator und Transformator im Anwendungsbeispiel), konnte auf der untersuchten Datenbasis nicht nachgewiesen werden. Die Ergebnisse bestätigen zwar, dass das Konzept beim Getriebe keinen großen Effekt zeigt, bei dem mechanische Schäden dominieren und schnell in Form von Temperaturanstiegen detektierbar werden. Allerdings konnte am Generator oder Transformator kein Vorteil erkannt werden. Anzumerken ist hierbei, dass der Generator im Datensatz lediglich drei Ausfälle aufweist, die auf das Kühlsystem zurückzuführen sind. Entsprechend kann in Bezug auf diese Annahme keine belastbare Aussage getroffen werden.

6. Fazit und Ausblick

Im letzten Kapitel wird die Arbeit zusammengefasst und ein abschließendes Fazit gezogen. Zuletzt wird ein Ausblick auf zukünftige Arbeiten gegeben.

6.1. Zusammenfassung und Fazit der Arbeit

Die vorliegende Dissertation verfolgt das Ziel, bestehende zeitreihenbasierte Modellansätze durch die Fusion von zeit-diskreten Ereignisdaten zu verbessern.

In Kapitel 2 wird dafür der aktuelle Stand der Technik im Bereich der datenbasierten Modellbildung für Zustandsüberwachungsmodelle dargestellt. Das Forschungsgebiet des PHM bildet dabei die Ausgangslage für diese Arbeit. Der aktuelle Stand der Forschung in der zustandsbasierten Instandhaltung und Systemdiagnose technischer Systeme beschäftigt sich nahezu ausschließlich mit der Analyse von Zeitreihendaten basierend auf Sensoren. Kenngrößen wie Temperaturen, Drehzahlen oder Vibrationen werden dabei von Modellen verarbeitet, der aktuelle Gesundheitszustand herausgearbeitet und Restlebensdauern von Komponenten werden prognostiziert. Für die diagnostische Verarbeitung von zeit-diskreten Ereignisdaten existieren zwar Ansätze und Methoden, diese fokussieren sich jedoch auf IT-Systeme und Computerlogs. Bestehende PHM-Ansätze verwenden die zusätzlich vorhandenen Informationen auf Basis zeit-diskreter Ereignisse nicht. Damit wurde in Abschnitt 2.3 die Forschungslücke dargelegt und die Aufgabe der Arbeit beschrieben.

Um die Genauigkeit von bestehenden zeitreihenbasierten Diagnose- und Monitoringmodellen zu erhöhen, wurde in Kapitel 3 dieser Arbeit ein Konzept vorgeschlagen, das die Fusion der zeit-diskreten Ereignisdaten in bestehende PHM-Systeme erlaubt. Schlüssel für die Integration ist die mehrstufige Vorverarbeitung und Merkmalsentwicklung der Ereignisdaten. Gemäß dem traditionellen Vorgehen des CRISP-DM müssen die

Ereignisdaten im Rahmen der Datenvorverarbeitung bereinigt, kategorisiert und aussagekräftige Merkmale generiert werden. Zur Entwicklung der Merkmale wurden vier Ansätze vorgeschlagen und ausführlich beschrieben. Die Ansätze gehen von einfachen Zählern von einzelnen Ereignissen und Ereigniskategorien in definierten Zeitfenstern (Häufigkeiten) über sequenzbasierte Analysen hin zu fehlerähnlichkeitsbasierten sowie profilbasierte Merkmale. Bis auf die letzte Kategorie haben alle Merkmale die Eigenschaft, dass sie eine Anpassung der temporalen Daten mit ungleichmäßigen Zeitstempeln auf die Zeitreihendaten ermöglichen. Diese Eigenschaft, im Rahmen der Arbeit als *time-conversion* bezeichnet, ermöglicht die einfache Fusion mit bestehenden Diagnosesystemen und erlaubt damit die Verwendung verschiedener Diagnosealgorithmen. Eine breite Anwendbarkeit über verschiedene Anwendungsfälle hinweg ist damit sichergestellt.

Die Evaluierung des Konzepts erfolgte anhand von Daten eines Windkraftparks und ist in Kapitel 4 dargestellt. Die vorliegende Datenbasis erfüllt die gestellten Anforderungen formal, wenngleich für die Ereignisdaten keine ausführliche Dokumentation vorhanden ist. Dieser Umstand, in Zusammenhang mit dem fehlenden Domänenwissen, erschwert eine zielgerichtete Entwicklung der Merkmale auf Basis der Ereignisdaten. Allgemein zeigen die Ergebnisse sowohl für das vorgeschlagene Konzept als auch für den Vergleichsansatz keine hinreichend hohe Klassifikationsgüte für die Anwendungen in der zustandsbasierten Instandhaltung. Dennoch erlaubte der Vergleich der Ergebnisse Rückschlüsse auf das vorgeschlagene Konzept. Insgesamt zeigten sich dabei vielversprechende Ergebnisse. Die Fusion von Ereignisdaten mit den vorhandenen Sensordaten führte zu einer verbesserten Diagnosegenauigkeit und Präzision unter spezifischen Randbedingungen. Dabei zeigten sich insbesondere bei einzelnen Anlagen deutliche Verbesserungen vom Konzept gegenüber dem Vergleichsansatz (vgl. Kapitel 5). Insgesamt zeigt sich beim Basisfall der Parameter eine Verbesserung der Genauigkeit von 10% im Mittelwert und 6% im Median. Generell ist somit das Ziel der Fusion unterschiedlicher Datenstrukturen gelungen und der Nachweis für eine Verbesserung der Modellgüte für Klassifikationsmodelle erbracht. Der Effekt der Verbesserung ist jedoch nicht für alle untersuchten Parameter statistisch signifikant. Die Sensitivitätsanalyse verdeutlichte die Bedeutung verschiedener Einflussfaktoren und identifizierte Grenzen des vorgeschlagenen Ansatzes, wodurch wichtige Erkenntnisse für zukünftige Entwicklungen gewonnen wurden.

Zusammenfassend lässt sich somit hinsichtlich der Evaluation sagen, dass die verwendete Evaluationsdatenbasis unter dem Hintergrund der Interpretierbarkeit der Daten nicht vollumfänglich geeignet ist um das Konzept zu beurteilen. Anders herum ist der Ansatz und die Ergebnisse dennoch vielversprechend. Die vorliegende Arbeit

legt somit den Grundstein für weiterführende Forschungsbemühungen im Bereich der zustandsbasierten Instandhaltung und Systemdiagnose. Trotz vielversprechender Ergebnisse sind noch weitere Untersuchungen erforderlich, um die Robustheit und Anwendbarkeit des Konzepts in verschiedenen industriellen Kontexten zu validieren und potenzielle Herausforderungen zu überwinden.

6.2. Ausblick für zukünftige Arbeiten

Abschließend soll der letzte Abschnitt dieser Arbeit einen Ausblick auf zukünftige Arbeiten geben, die im Rahmen der Anfertigung dieser Dissertation an verschiedensten Stellen erschienen sind. Dabei werden zunächst konkrete Vorschläge für die in der Evaluation aufgetretenen Herausforderungen gegeben und im weiteren Verlauf des Abschnitts Alternativen zum vorgeschlagenen Konzept präsentiert.

Eine der Herausforderungen beim konkreten Anwendungsbeispiel in der Evaluation war das hohe Klassenungleichgewicht, das in Kombination mit den getroffenen Annahmen zum Ausfallfenster zu schwachen Klassifikationsergebnissen geführt hat. Das verwendete Label ist dabei wegen zwei Gründen unzuverlässig: Erstens geht das Label des Systemzustands direkt von 0 (gesund) auf 1 (schadhaft) über. Zweitens ist die Annahme von festen Zeiträumen mit schadhaftem Label vor dem eigentlichen Ausfall der Komponente mit hoher Unsicherheit verbunden. Beides begründet möglicherweise die schwierige bzw. unzuverlässige Diagnose. Für letzteren Grund kann die Definition von individuellen Intervallen jeden Schadensfall, z. B. basierend auf Sensordaten, Abhilfe schaffen. Allerdings würde diese Festlegung auf Basis der Sensordaten gleichzeitig den Vergleichsansatz bevorzugen. Der ersten Vereinfachung (Sprung der Labels von 0 auf 1) kann durch die Verwendung einer Sigmoid-Funktion zur Darstellung der Schadenswahrscheinlichkeit begegnet werden, wie z. B. in Korvesis et al. Diese Art der Definition von Labels motiviert zugleich die Verwendung von Regressionsansätzen anstelle von Klassifikationsansätzen. Beobachter-Systeme und (Auto-) Encoder-Decoder Verfahren sind dabei vielversprechende Alternativen, die auch bei ungelabelten Daten Anomalien, respektive Ausfälle, detektieren können.

Das in dieser Arbeit vorgeschlagene Konzept führt eine zeitliche Transformation (*time-conversion*) der temporalen Ereignisdaten durch, um sie mit den Zeitreihendaten fusionieren zu können. Während dies ein nicht selten genutzter Weg in der Analyse von Ereignisdaten ist, besteht auch die Möglichkeit Zeitreihendaten zu transformieren, um sie in Sequenzen analysieren zu können. Der in Keogh et al. vorgeschlagene

HOT SAX Algorithmus (Heuristically Ordered Time series using Symbolic Aggregate ApproXimation) verwendet eine Kombination aus Dimensionsreduktionstechniken und statistischen Analysen, um effizient spezifische Zeitreihen-Teilsequenzen zu finden, die sich von den üblichen Mustern unterscheiden. Entsprechend werden Zeitreihen in Sequenzen transformiert, womit neben der in Keogh et al. beschriebenen Methodik auch andere Sequence Data Mining Ansätze anwendbar sind.

Bei der Analyse der Ergebnisse hat sich gezeigt, dass die Variation der Fensterlänge der Sequenzfenster keinen großen Einfluss auf die Güte des vorgeschlagenen Konzepts hat. Bei der Generierung von sequenzbasierten Merkmalen ist dabei anzumerken, dass gängigen Methoden aus dem Forschungsbereich des *Sequence Data Mining* darauf abzielen *häufige* Sequenzen bzw. Episoden zu detektieren. Dies ist z. B. bei der Prognose von Kaufverhalten im Konsumbereich von Bedeutung und ein gängiges Verfahren um Warenkorbvorschläge im Online-Shopping zu ermöglichen [SA96; FGT12]. Im Falle der Anwendung in technischen Bereichen ist häufig das Ziel, seltene Ereignisse (respektive Sequenzen) zu detektieren, so genannte *Black Swan Events*. Hierfür liefert die Literatur bisher jedoch keine Ansätze im Bereich des Sequence Data Mining. Eine Herausforderung dabei ist, dass die Identifikation von möglichst selteneren Sequenzen in einer nahezu unendlichen Vielfalt mündet.

Der Vergleich eines Künstliche Intelligenz (KI)- oder datenbasierten Diagnosesystems mit einem expertenbasierten System ist ebenfalls eine spannende Forschungsaufgabe. Bei der Deutschen Bahn werden etwa auf der Fernverkehrsflotte der Intercity Express (ICE) Züge regelmäßig so genannten Schadecodes (Ereignisdaten mit einer höheren Kritikalität bzw. Priorität) übermittelt. Mit Hilfe von Fahrzeugexperten wurden dabei über mehrere Jahre regelbasierte Methoden aufgebaut, die auf Basis der Schadcodes aktiv Ausfälle vermeiden. Beispiele sind logische bzw. kausale Kombinationen und Folgen von Ereignissen (Sequenzen) oder Häufungen von Ereignissen in Zeitfenstern (entsprechend der statistisch basierten Merkmale). Die Bahn kann damit rund ein bis zwei technische Ausfälle und signifikante Zugverspätungen pro Tag und zusätzlich unnötige Instandhaltungsereignisse vermeiden [Stö23]. Diese aktive Nutzung von Ereignisdaten bestätigt das große Potenzial des untersuchten Themengebiets.

Literaturverzeichnis

- [AYK23] Wanuji Abewickrema, Mehmet Yildirimoglu und Jiwon Kim. „Multivariate time-varying Kalman filter approach for cycle-based maximum queue length estimation“. In: *Transportation Research Part C: Emerging Technologies* 154 (2023). ISSN: 0968-090X. DOI: 10.1016/j.trc.2023.104238.
- [Agg15] Charu C. Aggarwal. *Data Mining: The Textbook*. Springer International Publishing, 2015. ISBN: 9783319141428. DOI: 10.1007/978-3-319-14142-8.
- [AS94] Rakesh Agrawal und Ramakrishnan Srikant. „Fast Algorithms for Mining Association Rules“. In: *Proceedings of 20th International Conference on Very Large Data Bases*. Hrsg. von Jorge B. Bocca, Matthias Jarke und Carlo Zaniolo. Santiago de Chile, Chile: Morgan Kaufmann, 1994, S. 487–499.
- [Ala+21] Mahbulul Alam, Laleh Jalali, Ahmed Farahat und Chetan Gupta. „Remaining Useful Life Estimation from Event Data“. In: *Annual Conference of the PHM Society* 13.1 (2021). ISSN: 2325-0178. DOI: 10.36001/phmconf.2021.v13i1.3055.
- [Alp14] Ethem Alpaydın. *Introduction to machine learning*. 3. Auflage. Adaptive computation and machine learning. Cambridge, Massachusetts und London, England: MIT Press, 2014. ISBN: 9780262325745.
- [Ape18] Harald. Apel, Hrsg. *Instandhaltungs- und Servicemanagement: Systeme mit Industrie 4.0*. München: Carl Hanser Verlag, 2018. ISBN: 9783446456877. DOI: 10.3139/9783446456877.
- [Are+18] Oluseun Omotola Aremu, Adrià Salvador Palau, Ajith Kumar Parlikad, David Hyland-Wood und Peter Ross McAree. „Structuring Data for Intelligent Predictive Maintenance in Asset Management“. In: *IFAC-PapersOnLine* 51.11 (2018), S. 514–519. ISSN: 2405-8963. DOI: 10.1016/j.ifacol.2018.08.370.

-
- [BKS22] Azzeddine Bakdi, Nicolay Bjorlo Kristensen und Morten Stakkeland. „Multiple Instance Learning With Random Forest for Event Logs Analysis and Predictive Maintenance in Ship Electric Propulsion System“. In: *IEEE Transactions on Industrial Informatics* 18.11 (2022), S. 7718–7728. ISSN: 1551-3203. DOI: 10.1109/TII.2022.3144177.
- [BNS20] Ebru Turanoglu Bekar, Per Nyqvist und Anders Skoogh. „An intelligent approach for data pre-processing and analysis in predictive maintenance with an industrial case study“. In: *Advances in Mechanical Engineering* 12.5 (2020). ISSN: 1687-8140. DOI: 10.1177/1687814020919207.
- [Bis09] Christopher M. Bishop. *Pattern recognition and machine learning*. 8. Aufl. Information Science and Statistics. New York, NY: Springer, 2009. ISBN: 0387310738.
- [Bon+] W. Glenn Bond et al. „A Hybrid Learning Approach to Prognostics and Health Management Applied to Military Ground Vehicles Using Time-Series and Maintenance Event Data“. In: *Annual Conference of the PHM Society*. Bd. 12, S. 10. DOI: 10.36001/phmconf.2020.v12i1.1146.
- [BCG20] Stéphane Bonneval, Jairo Cugliari und Victoria Granger. „Predictive Maintenance from Event Logs Using Wavelet-Based Features: An Industrial Application“. In: *14th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2019)*. Hrsg. von Francisco Martínez Álvarez. Bd. 950. *Advances in Intelligent Systems and Computing Ser.* Cham: Springer International Publishing AG, 2020, S. 132–141. ISBN: 978-3-030-20054-1. DOI: 10.1007/978-3-030-20055-8.
- [Boy10] Stuart A. Boyer. *SCADA: Supervisory control and data acquisition*. 4. Auflage. Research Triangle Park, NC: Instrument Society of America, 2010. ISBN: 1556176600.
- [BMA00] CAREY BUNKS, Dan. McCarthy und TARIK AL-ANI. „Condition-based Maintenance of Machines using Hidden Markov Models“. In: *Mechanical Systems and Signal Processing* 14.4 (2000), S. 597–612. ISSN: 08883270. DOI: 10.1006/mssp.2000.1309.
- [Cai+14] Baoping Cai et al. „Multi-source information fusion based fault diagnosis of ground-source heat pump using Bayesian network“. In: *Applied Energy* 114 (2014), S. 1–9. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2013.09.043.

-
- [CLS06] R. Callan, B. Larder und J. Sandiford. „An Integrated Approach to the Development of an Intelligent Prognostic Health Management System“. In: *2006 IEEE Aerospace Conference*. IEEE / Institute of Electrical and Electronics Engineers Incorporated, 2006, S. 1–12. ISBN: 0-7803-9545-X. DOI: 10.1109/AERO.2006.1656070.
- [Car+19] Thyago P. Carvalho et al. „A systematic literature review of machine learning methods applied to predictive maintenance“. In: *Computers & Industrial Engineering* 137 (2019). ISSN: 0360-8352. DOI: 10.1016/j.cie.2019.106024.
- [Cer19] Carlos Cernuda. „On the Relevance of Preprocessing in Predictive Maintenance for Dynamic Systems“. In: *Predictive Maintenance in Dynamic Systems*. Springer International Publishing, 2019, S. 53–93. ISBN: 978-3-030-05645-2. DOI: 10.1007/978-3-030-05645-2.
- [CBK09] Varun Chandola, Arindam Banerjee und Vipin Kumar. „Anomaly detection“. In: *ACM Computing Surveys* 41.3 (2009), S. 1–58. ISSN: 03600300. DOI: 10.1145/1541880.1541882.
- [CS18] Dan Chen und Jianzhong Sun. „Fuel and emission reduction assessment for civil aircraft engine fleet on-wing washing“. In: *Transportation Research Part D: Transport and Environment* 65 (2018), S. 324–331. ISSN: 1361-9209. DOI: 10.1016/j.trd.2018.05.013.
- [Dei+23] Álvaro Deibe Díaz, José A. Antón Nacimiento, Jesús Cardenal und Fernando López Peña. „A time-varying Kalman filter for low-acceleration attitude estimation“. In: *Measurement* 213 (2023). ISSN: 0263-2241. DOI: 10.1016/j.measurement.2023.112729.
- [DLL97] Thomas G. Dietterich, Richard H. Lathrop und Tomás Lozano-Pérez. „Solving the multiple instance problem with axis-parallel rectangles“. In: *Artificial Intelligence* 89.1-2 (1997), S. 31–71. ISSN: 0004-3702. DOI: 10.1016/S0004-3702(96)00034-3.
- [DIN18] DIN Deutsches Institut für Normung e. V. *DIN EN 13306:2018-02: Instandhaltung - Begriffe der Instandhaltung*. Berlin, 2018. DOI: 10.31030/2641990.
- [DIN19] DIN Deutsches Institut für Normung e. V. *DIN 31051:2019-06: Grundlagen der Instandhaltung*. Berlin, 2019. DOI: 10.31030/3048531.

-
- [DBT19] Duarte Dinis, Ana Barbosa-Póvoa und Ângelo Palos Teixeira. „Valuing data in aircraft maintenance through big data analytics: A probabilistic approach for capacity planning using Bayesian networks“. In: *Computers & Industrial Engineering* 128 (2019), S. 920–936. ISSN: 0360-8352. DOI: 10.1016/j.cie.2018.10.015.
- [DH07] Ming Dong und David He. „Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis“. In: *European Journal of Operational Research* 178.3 (2007), S. 858–878. ISSN: 0377-2217. DOI: 10.1016/j.ejor.2006.01.041.
- [Eck19] Peter P. Eckstein. *Statistik für Wirtschaftswissenschaftler: Eine realdatenbasierte Einführung mit SPSS*. 6. Auflage. Wiesbaden: Springer Fachmedien Wiesbaden, 2019. ISBN: 9783658247980. DOI: 10.1007/978-3-658-24798-0.
- [Eic+13] Steffen C. Eickemeyer, Tim Borchherding, Sebastian Schäfer und Peter Nyhuis. „Validation of data fusion as a method for forecasting the regeneration workload for complex capital goods“. In: *Production Engineering* 7.2-3 (2013), S. 131–139. ISSN: 0944-6524. DOI: 10.1007/s11740-013-0444-8.
- [Ene18] Energias de Portugal. *Energias de Portugal Open Data*. 2018. URL: <https://www.edp.com/en/innovation/open-data> (besucht am 14.07.2023).
- [Esm+13] Bilal Esmael, Arghad Arnaout, Rudolf K. Fruhwirth und Gerhard Thonhauser. „Improving time series classification using Hidden Markov Models“. In: *12th International Conference on Hybrid Intelligent Systems (HIS)*. IEEE, 2013, S. 502–507. ISBN: 978-1-4673-5116-4. DOI: 10.1109/HIS.2012.6421385.
- [FH17] Alireza Fatehi und Biao Huang. „Kalman filtering approach to multi-rate information fusion in the presence of irregular sampling rate and variable measurement delay“. In: *Journal of Process Control* 53 (2017), S. 15–25. ISSN: 0959-1524. DOI: 10.1016/j.jprocont.2017.02.010.
- [FGT12] Philippe Fournier-Viger, Ted Gueniche und Vincent S. Tseng. „Using Partially-Ordered Sequential Rules to Generate More Accurate Sequence Prediction“. In: *Advanced Data Mining and Applications*. Lecture Notes in Computer Science. Alemania: Springer Healthcare Ltd, 2012, S. 431–442. ISBN: 978-3-642-35527-1. DOI: 10.1007/978-3-642-35527-1.

-
- [Fou+16] Philippe Fournier-Viger et al. „The SPMF Open-Source Data Mining Library Version 2“. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2016, S. 36–40. DOI: 10.1007/978-3-319-46131-1{\textunderscore}8.
- [Fri01] Jerome H. Friedman. „Greedy Function Approximation: A Gradient Boosting Machine“. In: *The Annals of Statistics* 29.5 (2001), S. 1189–1232. ISSN: 00905364.
- [Fri02] Jerome H. Friedman. „Stochastic gradient boosting“. In: *Computational Statistics & Data Analysis* 38.4 (2002), S. 367–378. ISSN: 0167-9473. DOI: 10.1016/S0167-9473(01)00065-2.
- [Fro+13] Ilenia Fronza, Alberto Sillitti, Giancarlo Succi, Mikko Terho und Jelena Vlasenko. „Failure prediction based on log files using Random Indexing and Support Vector Machines“. In: *Journal of Systems and Software* 86.1 (2013), S. 2–11. ISSN: 0164-1212. DOI: 10.1016/j.jss.2012.06.025.
- [GVW20] Antoine Guillaume, Christel Vrain und Elloumi Wael. *Predictive maintenance on event logs: Application on an ATM fleet*. Hrsg. von arXiv e-print. 22. Nov. 2020. DOI: 10.48550/arXiv.2011.10996.
- [Gup+14] Manish Gupta, Jing Gao, Charu C. Aggarwal und Jiawei Han. „Outlier Detection for Temporal Data: A Survey“. In: *IEEE Transactions on Knowledge and Data Engineering* 26.9 (2014), S. 2250–2267. ISSN: 1041-4347. DOI: 10.1109/TKDE.2013.184.
- [Gut+19] Clemens Gutschi, Nikolaus Furian, Josef Suschnigg, Dietmar Neubacher und Siegfried Voessner. „Log-based predictive maintenance in discrete parts manufacturing“. In: *Procedia CIRP* 79 (2019), S. 528–533. ISSN: 22128271. DOI: 10.1016/j.procir.2019.02.098.
- [Guy08] Isabelle Guyon. *Feature extraction: Foundations and applications*. Bd. Vol. 207. Studies in fuzziness and soft computing. Berlin und Heidelberg: Springer, 2008. ISBN: 9783540354888.
- [HPK12] Jiawei Han, Jian PEI und Micheline KAMBER. *Data mining: Concepts and techniques*. 3. Auflage. The Morgan Kaufmann series in data management systems. Amsterdam u. a.: Morgan Kaufmann und Elsevier, 2012. ISBN: 978-0-12-381479-1.

-
- [HTF09] Trevor Hastie, Robert Tibshirani und Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2. Auflage. Springer Series in Statistics. New York, NY: Springer-Verlag New York, 2009. ISBN: 9780387848587.
- [HC94] L. E. Holloway und S. Chand. „Time templates for discrete event fault monitoring in manufacturing systems“. In: *American Control Conference, 1994*. IEEE / Institute of Electrical and Electronics Engineers Incorporated, 1994, S. 701–706. ISBN: 0-7803-1783-1. DOI: 10.1109/ACC.1994.751830.
- [Hub+19] Steffen Huber, Hajo Wiemer, Dorothea Schneider und Steffen Ihlenfeldt. „DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model“. In: *Procedia CIRP* 79 (2019), S. 403–408. ISSN: 22128271. DOI: 10.1016/j.procir.2019.02.106.
- [Int03] International Standard Organisation. *ISO 13374-1: Condition monitoring and diagnostics of machines - Data processing, communication and presentation -: Part 1: General guidelines*. 15.03.2003.
- [Ise84] Rolf Isermann. „Process fault detection based on modeling and estimation methods—A survey“. In: *Automatica* 20.4 (1984), S. 387–404. ISSN: 0005-1098. DOI: 10.1016/0005-1098(84)90098-0.
- [Ise06] Rolf Isermann. *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Berlin, Heidelberg: Springer, 2006. ISBN: 9783540303688. DOI: 10.1007/3-540-30368-5.
- [JS16] Jannis Weinert und Simon Watson. „Wind turbine fault detection by normal behaviour modelling [Poster]“. In: (2016).
- [JAM87] A. K. S. Jardine, P. M. Anderson und D. S. Mann. „Application of the weibull proportional hazards model to aircraft and marine engine failure data“. In: *Quality and Reliability Engineering International* 3.2 (1987), S. 77–82. ISSN: 07488017. DOI: 10.1002/qre.4680030204.
- [JLB06] Andrew K.S. Jardine, Daming Lin und Dragan Banjevic. „A review on machinery diagnostics and prognostics implementing condition-based maintenance“. In: *Mechanical Systems and Signal Processing* 20.7 (2006), S. 1483–1510. ISSN: 08883270. DOI: 10.1016/j.ymsp.2005.09.012.

-
- [Jia+13] Jianfeng Si et al. „Exploiting Topic based Twitter Sentiment for Stock Prediction“. In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (2013), S. 24–29.
- [Joa98] Thorsten Joachims. „Text categorization with Support Vector Machines: Learning with many relevant features“. In: *Machine Learning: ECML-98*. Hrsg. von Claire Nédellec und Céline Rouveirol. Bd. 1398. Lecture Notes in Computer Science, Lecture Notes in Artificial Intelligence. Berlin und Heidelberg: Springer, 1998, S. 137–142. ISBN: 978-3-540-69781-7. DOI: 10.1007/BFb0026683.
- [Ke+17] Guolin Ke et al. „LightGBM: A Highly Efficient Gradient Boosting Decision Tree“. In: *Advances in Neural Information Processing Systems* 30 (2017). DOI: 10.5555/3294996.3295074.
- [KLF05] E. Keogh, J. Lin und A. Fu. „HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence“. In: *Fifth IEEE International Conference on Data Mining*. IEEE / Institute of Electrical and Electronics Engineers Incorporated, 2005, S. 226–233. ISBN: 0-7695-2278-5. DOI: 10.1109/ICDM.2005.79.
- [KAC17] Nam-Ho Kim, Dawn An und Joo-Ho Choi. *Prognostics and Health Management of Engineering Systems: An Introduction*. Cham und s.l.: Springer International Publishing, 2017. ISBN: 9783319447421. DOI: 10.1007/978-3-319-44742-1.
- [KS22] Uwe Klingauf und Henrik Simon. *Intelligente Datennutzung in der Instandhaltung (INDI: Schlussbericht der TU Darmstadt: Berichtszeitraum: 01.01.2018 bis 31.12.2021)*. 2022. DOI: 10.2314/KXP:1853252077. URL: <https://www.tib.eu/de/suchen/id/TIBKAT:1853252077>.
- [KJ97] Ron Kohavi und George H. John. „Wrappers for feature subset selection“. In: *Artificial Intelligence* 97.1-2 (1997), S. 273–324. ISSN: 0004-3702. DOI: 10.1016/S0004-3702(97)00043-X.
- [KBV18] Panagiotis Korvesis, Stephane Besseau und Michalis Vazirgiannis. „Predictive Maintenance in Aviation: Failure Prediction from Post-Flight Reports“. In: *IEEE 34th International Conference on Data Engineering*. Piscataway, NJ: IEEE, 2018, S. 1414–1422. ISBN: 978-1-5386-5520-7. DOI: 10.1109/ICDE.2018.00160.
- [Lal+06] Thomas Navin Lal, Olivier Chapelle, Jason Weston und André Elisseeff. „Embedded Methods“. In: *Feature Extraction*. Springer, Berlin, Heidelberg, 2006, S. 137–165. DOI: 10.1007/978-3-540-35488-8.

-
- [Lan01] J. E. Land. „HUMS-the benefits-past, present and future“. In: *Aerospace Conference, 2001, IEEE Proceedings*. IEEE, 2001, S. 3083–3094. ISBN: 0-7803-6599-2. DOI: 10.1109/AERO.2001.931326.
- [LM14] Quoc Le V und Tomas Mikolov. *Distributed Representations of Sentences and Documents*. Hrsg. von arXiv e-print. 16. Mai 2014. DOI: 10.48550/arXiv.1405.4053.
- [Lee+18] Gil-Yong Lee et al. „Machine health management in smart factory: A review“. In: *Journal of Mechanical Science and Technology* 32.3 (2018), S. 987–1009. ISSN: 1976-3824. DOI: 10.1007/s12206-018-0201-1.
- [Lee+14] Jay Lee et al. „Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications“. In: *Mechanical Systems and Signal Processing* 42.1-2 (2014), S. 314–334. ISSN: 08883270. DOI: 10.1016/j.ymssp.2013.06.004.
- [Lei+20] Yaguo Lei et al. „Applications of machine learning to machine fault diagnosis: A review and roadmap“. In: *Mechanical Systems and Signal Processing* 138 (2020), S. 106587. ISSN: 08883270. DOI: 10.1016/j.ymssp.2019.106587.
- [LH03] Lei Yu und Huan Liu. „Feature selection for high-dimensional data: A fast correlation-based filter solution“. In: *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*. 2003.
- [Lei14] Bernhard Leidinger. *Wertorientierte Instandhaltung*. Wiesbaden: Springer Fachmedien Wiesbaden, 2014. ISBN: 978-3-658-04400-8. DOI: 10.1007/978-3-658-04401-5.
- [Li+17] Chenzhao Li, Sankaran Mahadevan, You Ling, Sergio Choze und Liping Wang. „Dynamic Bayesian Network for Aircraft Wing Health Monitoring Digital Twin“. In: *AIJA Journal* 55.3 (2017), S. 930–941. ISSN: 0001-1452. DOI: 10.2514/1.J055201.
- [Li+07] Zhiguo Li, Shiyu Zhou, Suresh Choubey und Crispian Sievenpiper. „Failure event prediction using the Cox proportional hazard model driven by frequent failure signatures“. In: *IIE Transactions* 39.3 (2007), S. 303–315. DOI: 10.1080/07408170600847168.
- [Liu+18] Ruonan Liu, Boyuan Yang, Enrico Zio und Xuefeng Chen. „Artificial intelligence for fault diagnosis of rotating machinery: A review“. In: *Mechanical Systems and Signal Processing* 108 (2018), S. 33–47. ISSN: 08883270. DOI: 10.1016/j.ymssp.2018.02.016.

-
- [Liu+17] Yu-Ying Liu et al. „Learning Continuous-Time Hidden Markov Models for Event Data“. In: *Mobile Health*. Springer International Publishing, 2017, S. 361–387. ISBN: 978-3-319-51394-2. DOI: 10.1007/978-3-319-51394-2.
- [MTI97] Heikki Mannila, Hannu Toivonen und A. Inkeri Verkamo. „Discovery of Frequent Episodes in Event Sequences“. In: *Data Mining and Knowledge Discovery* 1.3 (1997), S. 259–289. ISSN: 1573-756X. DOI: 10.1023/A:1009748302351.
- [Mey14] Martin Meyer. *Signalverarbeitung: Analoge und digitale Signale, Systeme und Filter*. 7. Auflage. SpringerLink Bücher. Wiesbaden: Springer Fachmedien Wiesbaden, 2014. ISBN: 9783658026127. DOI: 10.1007/978-3-658-02612-7.
- [Mik+13] Tomas Mikolov, Kai Chen, Greg Corrado und Jeffrey Dean. *Efficient Estimation of Word Representations in Vector Space*. Hrsg. von arXiv. 16. Jan. 2013. DOI: 10.48550/arXiv.1301.3781.
- [Nat] National Institute of Standards and Technology. *Guide to Computer Security Log Management*. Gaithersburg.
- [NWC21] Huihui Ni, Shuting Wang und Peng Cheng. „A hybrid approach for stock trend prediction based on tweets embedding and historical prices“. In: *World Wide Web* 24.3 (2021), S. 849–868. DOI: 10.1007/s11280-021-00880-9.
- [Oss14] Daniel Ossmann. *Fehlerdetektion, -Isolation und -Identifikation in elektrohydraulischen Aktuatoren systemen moderner, ziviler Flugzeuge: Zugl.: München, Techn. Univ., Diss., 2014*. 1. Aufl. Luftfahrt. München: Verl. Dr. Hut, 2014. ISBN: 3843918562.
- [PEI+04] Jian PEI et al. „Mining sequential patterns by pattern-growth: the Prefix-Span approach“. In: *IEEE Transactions on Knowledge and Data Engineering* 16.11 (2004), S. 1424–1440. ISSN: 1041-4347. DOI: 10.1109/TKDE.2004.77.
- [PLD05] Hanchuan Peng, Fuhui Long und Chris Ding. „Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy“. In: *IEEE transactions on pattern analysis and machine intelligence* 27.8 (2005), S. 1226–1238. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2005.159.

-
- [Phi+17] Philippe Fournier Viger, Jerry Chun-Wei Lin, Ragu Uday Kiran, Yun Sing Koh und Rincy Thomas. „A Survey of Sequential Pattern Mining“. In: *Data Science and Pattern Recognition* 1.1 (2017), S. 54–77.
- [RYM18] Maik Reeder, Nurseda Y. Yürüşen und Julio J. Melero. „Data-driven learning framework for associating weather conditions and wind turbine failures“. In: *Reliability Engineering & System Safety* 169 (2018), S. 554–569. ISSN: 09518320. DOI: 10.1016/j.ress.2017.10.004.
- [Rom+16] A. Romero, Y. Lage, S. Souza, B. Wang und T.-H. Gan. „Vestas V90-3MW Wind Turbine Gearbox Health Assessment Using a Vibration-Based Condition Monitoring System“. In: *Shock and Vibration* 2016 (2016), S. 1–18. ISSN: 1070-9622. DOI: 10.1155/2016/6423587.
- [RCB00] Evan L. Russell, Leo H. Chiang und Richard D. Braatz. *Data-driven Methods for Fault Detection and Diagnosis in Chemical Processes*. Advances in Industrial Control. London: Springer London, 2000. ISBN: 9781447104094. DOI: 10.1007/978-1-4471-0409-4.
- [Sal06] Felix Salfner. *Modeling Event-driven Time Series with Generalized Hidden Semi-Markov Models*. 2006. DOI: 10.18452/9265.
- [Sax+08a] Abhinav Saxena, Kai Goebel, Don Simon und Neil Eklund. „Damage propagation modeling for aircraft engine run-to-failure simulation“. In: *International Conference on Prognostics and Health Management, 2008*. Piscataway, NJ: IEEE, 2008. ISBN: 978-1-4244-1935-7. DOI: 10.1109/PHM.2008.4711414.
- [Sax+08b] Abhinav Saxena et al. „Metrics for evaluating performance of prognostic techniques“. In: *International Conference on Prognostics and Health Management, 2008*. Piscataway, NJ: IEEE, 2008, S. 1–17. ISBN: 978-1-4244-1935-7. DOI: 10.1109/PHM.2008.4711436.
- [Sch10] Michael Schenk, Hrsg. *Instandhaltung technischer Systeme: Methoden und Werkzeuge zur Gewährleistung eines sicheren und wirtschaftlichen Anlagenbetriebs*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, 2010. ISBN: 9783642039492.
- [SKS22] Meik Schlechtingen, Martin Kato und Henrik Simon. *Austausch Promotionsvorhaben mit EnBW: Datenverfügbarkeit Windkraftanlagen für Forschungszwecke*. online, 21.03.2022.

-
- [SSA13] Meik Schlechtingen, Ilmar Ferreira Santos und Sofiane Achiche. „Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: System description“. In: *Applied Soft Computing* 13.1 (2013), S. 259–270. ISSN: 1568-4946. DOI: 10.1016/j.asoc.2012.08.033.
- [Sip+14] Ruben Sipos, Dmitriy Fradkin, Fabian Moerchen und Zhuang Wang. „Log-based predictive maintenance“. In: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM Digital Library. New York, NY: ACM, 2014, S. 1867–1876. ISBN: 9781450329569. DOI: 10.1145/2623330.2623340.
- [SJS06] Marina Sokolova, Nathalie Japkowicz und Stan Szpakowicz. „Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation“. In: *Advances in artificial intelligence: proceedings*. Bd. 4304. Lecture Notes in Computer Science. Berlin: Springer, 2006, S. 1015–1021. ISBN: 978-3-540-49787-5. DOI: 10.1007/11941439\{ \text{extunderscore} \} 114.
- [SA96] Ramakrishnan Srikant und Rakesh Agrawal. „Mining sequential patterns: Generalizations and performance improvements“. In: *Advances in database technology*. Lecture Notes in Computer Science. Berlin: Springer-Verlag, 1996, S. 1–17. ISBN: 978-3-540-49943-5. DOI: 10.1007/BFb0014140.
- [Stö23] Störungsanalysen DB Fernverkehr. *Potenziale von Schadcodes im Fernverkehr: Meeting*. Dezember 2023.
- [Str12] Matthias Strunz. *Instandhaltung*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. ISBN: 978-3-642-27389-6. DOI: 10.1007/978-3-642-27390-2.
- [Sus+15] Gian Antonio Susto, Andrea Schirru, Simone Pampuri, Sean McLoone und Alessandro Beghi. „Machine Learning for Predictive Maintenance: A Multiple Classifier Approach“. In: *IEEE Transactions on Industrial Informatics* 11.3 (2015), S. 812–820. ISSN: 1551-3203. DOI: 10.1109/TII.2014.2349359.
- [TL] Tobias Wensky und Jens Friedrichs Lutz Winkler. „Environmental Influences on Engine Performance Degradation“. In: *ASME Turbo Expo 2010: Power for Land, Sea and Air (Hg.)*

-
- [V F+09] V. Fathabadi, M. Shahbazian, K. Salahshour und Lotfollah Jargani. „Comparison of Adaptive Kalman Filter Methods in State Estimation of a Non-linear System Using Asynchronous Measurements“. In: *Proceedings of the World Congress on Engineering and Computer Science* (2009).
- [Ver22a] Verein Deutscher Ingenieure. *Intelligente Zustandsprognose und vorausschauende Instandhaltung: Prognostics and health management. VDI-Statusreport Juli 2022*. Hrsg. von Verein Deutscher Ingenieure. Düsseldorf, 2022.
- [Ver22b] Verein Deutscher Ingenieure. *Künstliche Intelligenz im Ingenieuralltag: Erwartungen und Realität. VDI-Statusreport Juli 2022*. Hrsg. von Verein Deutscher Ingenieure. Düsseldorf, 2022.
- [Ves07] Vestas Central Europe, Hrsg. *Ergänzungsheft zur Betriebsanleitung: V90-2.0 MW VMP 5000-Steuerung*. 15. Feb. 2007. URL: <https://de.scribd.com/document/39801351/Fehlerliste-V90-2-944665-R11> (besucht am 05. 10. 2023).
- [VWZ04] Pieter-Jan Vlok, Maciej Wnek und Maciej Zygmunt. „Utilising statistical residual life estimates of bearings to quantify the influence of preventive maintenance actions“. In: *Mechanical Systems and Signal Processing* 18.4 (2004), S. 833–847. ISSN: 08883270. DOI: 10.1016/j.ymssp.2003.09.003.
- [VWD14] Gregory W. Vogl, Brian A. Weiss und M. Alkan Donmez. *Standards Related to Prognostics and Health Management (PHM) for Manufacturing*. National Institute of Standards and Technology, 2014. DOI: 10.6028/nist.ir.8012.
- [Wan+17] J. Wang, C. Li, S. Han, S. Sarkar und X. Zhou. „Predictive maintenance based on event-log analysis: A case study“. In: *IBM Journal of Research and Development* 61.1 (2017), 11:121–11:132. ISSN: 0018-8646. DOI: 10.1147/JRD.2017.2648298.
- [WAH22] Jingyi Wang, Yousef Alipouri und Biao Huang. „Multirate Sensor Fusion in the Presence of Irregular Measurements and Time-Varying Time Delays Using Synchronized, Neural, Extended Kalman Filters“. In: *IEEE Transactions on Instrumentation and Measurement* 71 (2022), S. 1–9. ISSN: 0018-9456. DOI: 10.1109/TIM.2021.3135537.
- [Web+12] P. Weber, G. Medina-Oliva, C. Simon und B. Iung. „Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas“. In: *Engineering Applications of Artificial Intelligence* 25.4 (2012),

S. 671–682. ISSN: 09521976. DOI: 10.1016/j.engappai.2010.06.002.

- [WHN08] Weizhong Yan, Hai Qiu und Naresh Iyer. *Feature Extraction for Bearing Prognostics and Health Management (PHM) - A Survey (Preprint)*. Hrsg. von United States Air Force. 2008.
- [Zak+06] James J. Zakrajsek et al. *Rotorcraft Health Management Issues and Challenges*. Hrsg. von NASA STI Program Office. 2006.
- [Zhe+09] Ziming Zheng, Zhiling Lan, Byung H. Park und Al Geist. „System log pre-processing to improve failure prediction“. In: *Networks, International Conference on Dependable Systems and Networks*. IEEE, 2009, S. 572–577. ISBN: 978-1-4244-4422-9. DOI: 10.1109/DSN.2009.5270289.

Anhang

A. Anhang Kapitel 4

Tabelle A.1.: Verwendete Python Distributionen und Bibliotheken

Package/Bibliothek	Version
Python	3.9.7
Matplotlib	3.5.1
Pandas	1.4.2
Scikit-Learn	1.3.0
Scipy	1.8.0
SPMF	1.4

A.1. Ergänzende Informationen zum Datensatz

Tabelle A.2.: Technische Informationen zu den Anlage des EDP
Datensatzes

Power	
Rated power (kW)	2000
Cut-in wind speed (m/s)	4
Rated wind speed (m/s)	12
Cut-out wind speed (m/s)	25
Rotor	
Diameter	90
Number of blades	3
Max Rotor speed	14,9
Power density (W/m ²)	314,4
Gearbox	
Type	Planetengetriebe
Stages	3
Generator	
Type	Asynchron
Max Speed (U/min)	2016
Grid frequency (Hz)	50
Tower	
Hub height (m)	80
Type	Stahl

Tabelle A.3.: Liste der Ausfälle und Fehler nach WKA

WKA	Komponente	Zeitstempel	Eintrag
T01	Getriebe	2016-07-18T02:10	Gearbox pump damaged
T06	Generator	2016-07-11T19:48	Generator replaced
T06	Generator	2016-07-24T17:01	Generator temperature sensor failure
T06	Generator	2016-09-04T08:08	High temperature generator error
T06	Generator	2016-10-27T16:26	Generator replaced
T06	Generator	2016-10-02T17:08	Refrigeration system and temperature sensors in generator replaced
T06	Hydraulik	2016-04-04T18:53	Error in pitch regulation
T07	Generatorlager	2016-04-30T12:40	High temperature in generator bearing (replaced sensor)
T07	Transformator	2016-07-10T03:46	High temperature transformer
T07	Transformator	2016-08-23T02:21	High temperature transformer. Transformer refrigeration repaired
T09	Getriebe	2016-10-11T08:06	Gearbox repaired
T09	Generatorlager	2016-06-07T16:59	High temperature generator bearing
T09	Generatorlager	2016-08-22T18:25	High temperature generator bearing
T09	Generatorlager	2016-10-17T09:19	Generator bearings replaced
T11	Generator	2016-03-03T19:00	Electric circuit error in generator
T11	Hydraulik	2016-10-17T17:44	Hydraulic group error in the brake circuit
T01	Transformator	2017-08-11T13:14	Transformer fan damaged
T06	Getriebe	2017-10-17T08:38	Gearbox bearings damaged
T06	Hydraulik	2017-08-19T09:47	Oil leakage in Hub
T07	Generatorlager	2017-08-20T06:08	Generator bearings damaged
T07	Generator	2017-08-21T14:47	Generator damaged
T07	Hydraulik	2017-06-17T11:35	Oil leakage in Hub
T07	Hydraulik	2017-10-19T10:11	Oil leakage in Hub
T09	Getriebe	2017-10-18T08:32	Gearbox noise
T09	Generatorlager	2017-01-25T12:55	Generator bearings replaced
T09	Hydraulik	2017-09-16T15:46	Pitch position error related GH
T11	Hydraulik	2017-04-26T18:06	Hydraulic group error in the brake circuit
T11	Hydraulik	2017-09-12T15:30	Hydraulic group error in the brake circuit

A.2. Ergänzende Informationen zur statistischen Analyse

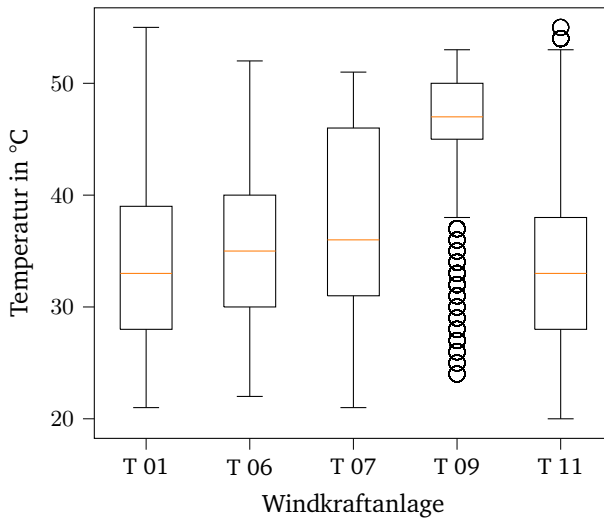


Abbildung A.1.: Verteilung der Hydrauliköltemperatur

Tabelle A.4.: Abschaltzeiten der Anlagen aufgrund von
Instandhaltungsereignissen und Einspeisemanagement

Anlage Instandhaltungsereignisse	relative Zeit	absolute Zeit
T01	0.595	4 days 08:26:30
T06	1.235	9 days 00:36:06
T07	1.7	12 days 10:18:06
T09	1.595	11 days 15:48:35
T11	0.428	3 days 03:07:50
mittlere	1,11	8 days 02:51:25
EINSMAN		
T01	0.447	3 days 06:29:18
T06	0.414	3 days 00:34:34
T07	0.394	2 days 21:09:23
T09	0.403	2 days 22:45:55
T11	0.471	3 days 10:39:26
mittlere	0,426	3 days 02:43:43

A.3. Ergänzende Informationen zur Hyperparameteroptimierung

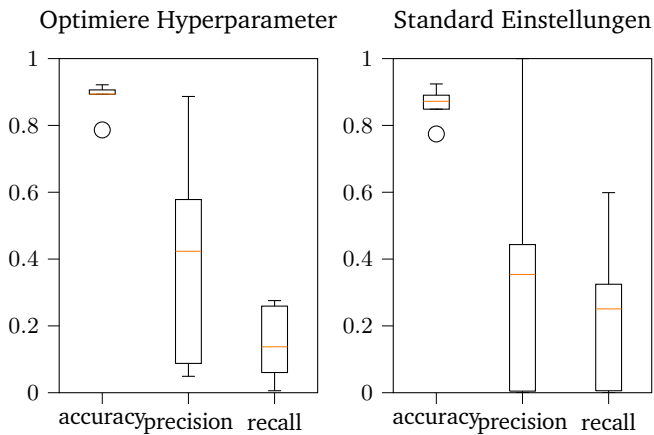


Abbildung A.2.: Vergleich des Gradient Boosting mit Standardeinstellungen und optimierten Hyperparametern an Bsp. von WKA T01 und auf Basis von allen Ereignismerkmalen

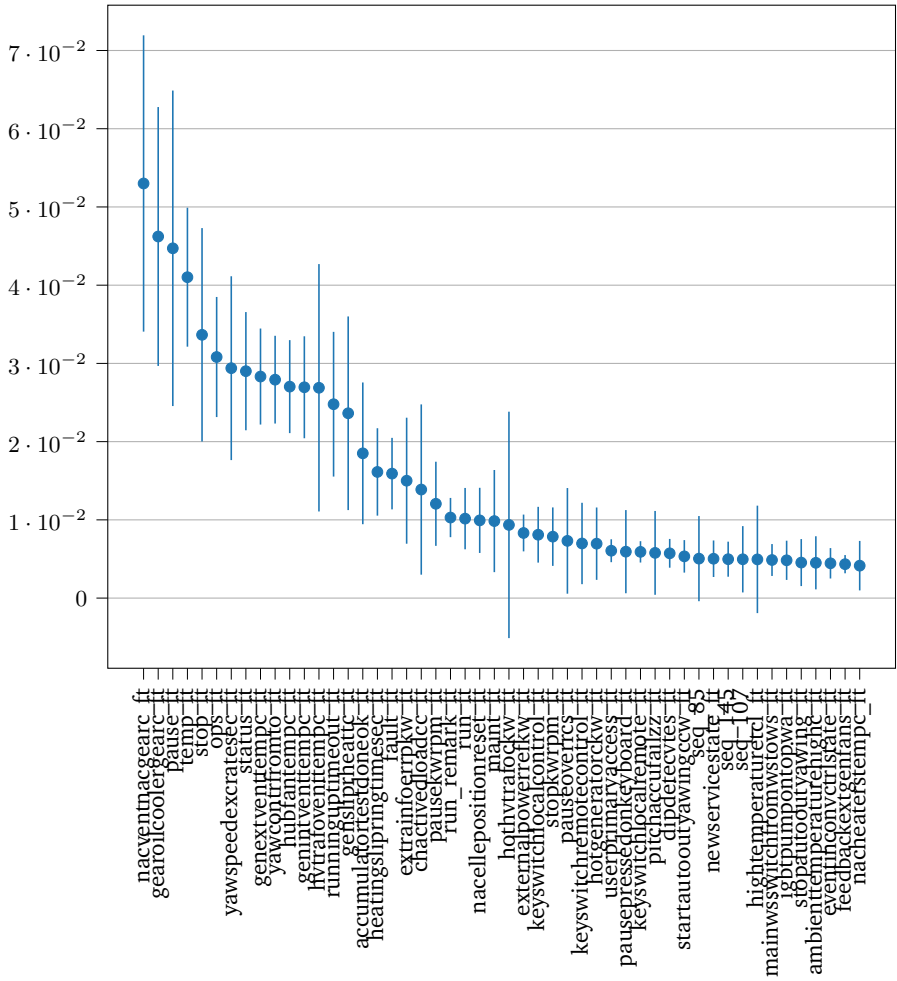


Abbildung A.3.: Ranking der wichtigsten Merkmale aus Ereignisdaten gemittelt über 5 Modelle

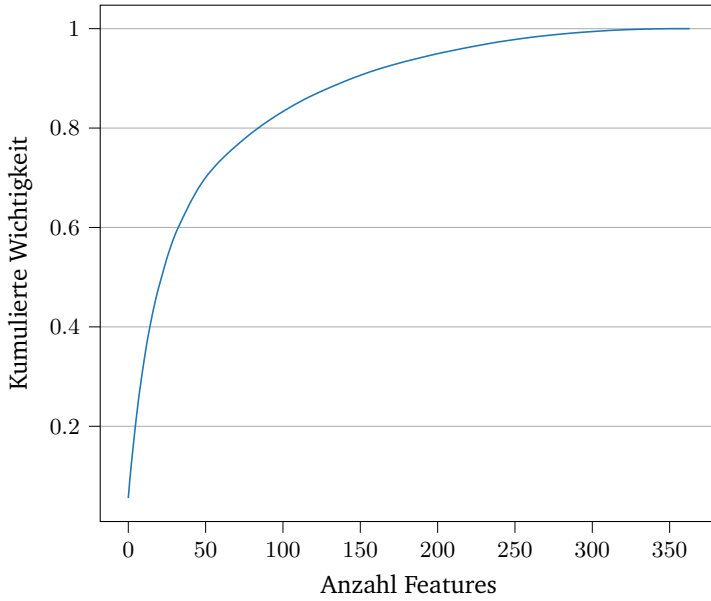


Abbildung A.4.: Kumulierte Wichtigkeit der Merkmale aus Ereignisdaten gemittelt über 5 Modelle

Tabelle A.5.: Für den Vergleichsansatz verwendete Merkmale aus den Zeitreihendaten

Descriptor	Description
Gen_RPM_Max	Maximum generator rpm in latest average period
Gen_RPM_Min	Minimum generator rpm in latest average period
Gen_RPM_Avg	Average generator rpm
Gen_RPM_Std	Std. generator rpm in latest average period
Gen_Phase1_Temp_Avg	Average temperature inside generator in stator windings phase 1
Gen_Phase2_Temp_Avg	Average temperature inside generator in stator windings phase 2
Gen_Phase3_Temp_Avg	Average temperature inside generator in stator windings phase 3
Hyd_Oil_Temp_Avg	Average temperature oil in hydraulic group
Gear_Oil_Temp_Avg	Average temperature oil in gearbox
Gear_Bear_Temp_Avg	Average temperature in gearbox bearing on high speed shaft
Nac_Temp_Avg	Average temperature in nacelle
Rtr_RPM_Max	Maximum rotor rpm in latest average period
Rtr_RPM_Min	Minimum rotor rpm in latest average period
Rtr_RPM_Avg	Average rotor rpm
Amb_WindSpeed_Max	Maximum windspeed within average timebase
Amb_WindSpeed_Min	Minimum windspeed within average timebase
Amb_WindSpeed_Avg	Average windspeed within average timebase
Amb_WindSpeed_Std	Std windspeed within average timebase
Amb_WindDir_Relative_Avg	Average wind relative direction
Amb_WindDir_Abs_Avg	Average wind absolute direction
Amb_Temp_Avg	Average ambient temperature
Prod_LatestAvg_ActPwrGen0	Active power - generator disconnected (yaw motor hydraulic motor etc.)
Prod_LatestAvg_ActPwrGen1	Active power - generator connected in delta
Prod_LatestAvg_ActPwrGen2	Active power - generator connected in star
Prod_LatestAvg_TotActPwr	Total active power
Prod_LatestAvg_ReactPwrGen0	Reactive power - generator disconnected (yaw motor hydraulic motor etc.)
Prod_LatestAvg_ReactPwrGen1	Reactive power - generator connected in delta
Prod_LatestAvg_ReactPwrGen2	Reactive power - generator connected in star
Prod_LatestAvg_TotReactPwr	Total reactive power
HVTrafo_Phase1_Temp_Avg	Average temperature in HV transformer phase L1
HVTrafo_Phase2_Temp_Avg	Average temperature in HV transformer phase L2
HVTrafo_Phase3_Temp_Avg	Average temperature in HV transformer phase L3
Grd_InverterPhase1_Temp_Avg	Average temperature measured by the IGBT-driver on the grid side inverter
Cont_Top_Temp_Avg	Average temperature in the top nacelle controller
Cont_Hub_Temp_Avg	Average temperature in the hub controller
Cont_VCP_Temp_Avg	Average temperature on the VCP-board
Gen_SlipRing_Temp_Avg	Average temperature in the split ring chamber
Spin_Temp_Avg	Average temperature in the nose cone
Blds_PitchAngle_Min	Minimum angle in latest average period
Blds_PitchAngle_Max	Average angle
Blds_PitchAngle_Avg	Std. angle in latest average period

A.4. Ergänzende Darstellungen zur Sensitivitätsanalyse

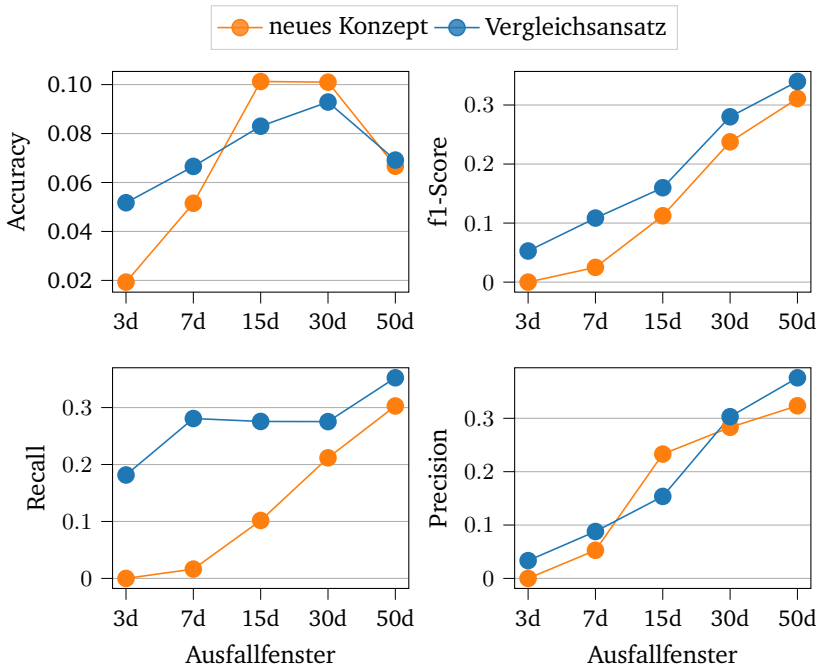


Abbildung A.5.: Standardabweichung der fünf Modelle in Abhängigkeit variierender Fenstergrößen für das Ausfallfenster

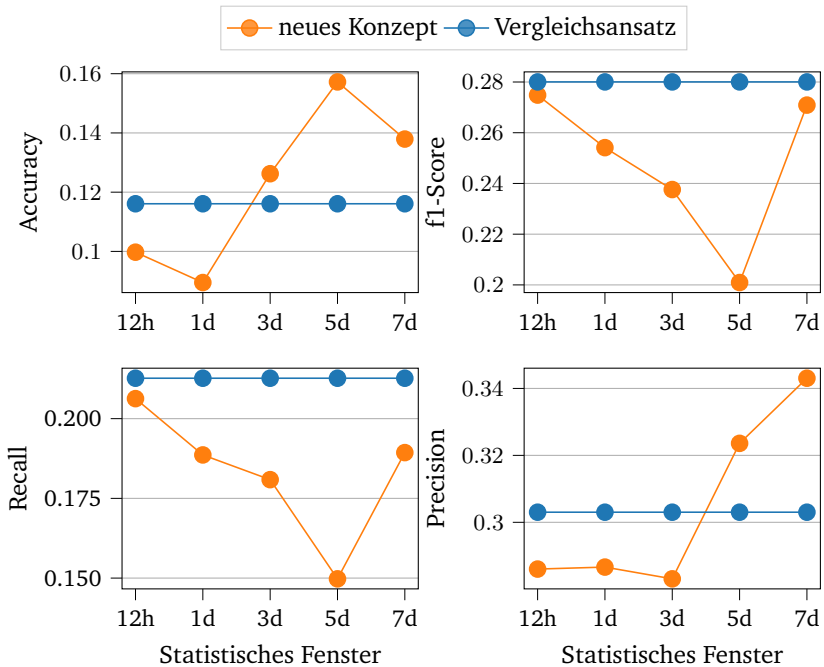


Abbildung A.6.: Standardabweichung der fünf Modelle in Abhängigkeit variierender Fenstergrößen für statistisch basierte Ereignismerkmale

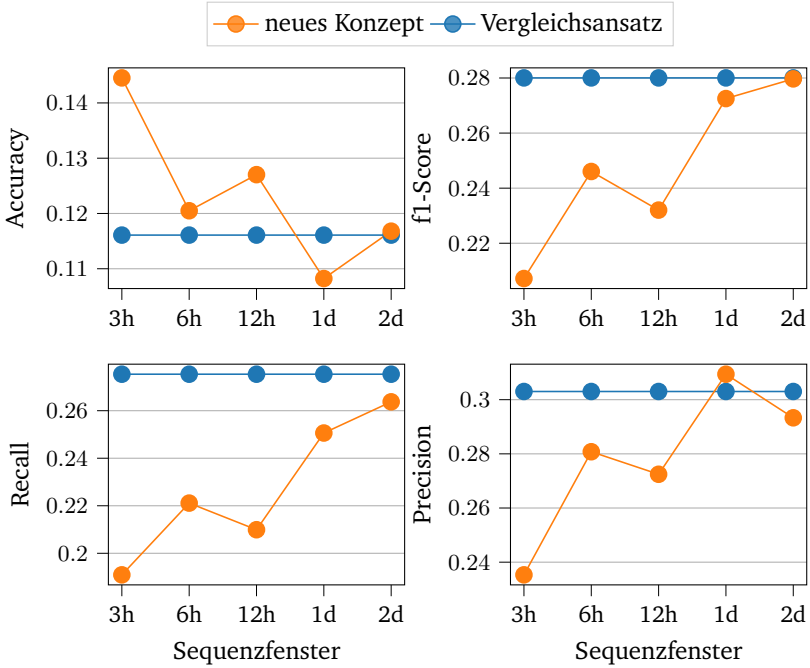


Abbildung A.7.: Standardabweichung der fünf Modelle in Abhängigkeit variierender Fenstergrößen für sequenzbasierte Ereignismerkmale