



Hochschule für Technik  
und Wirtschaft Berlin

*University of Applied Sciences*

---

# Klassifikation von Maschinengeräuschen mittels

## Spektrogramm und maschinellem Lernen

---

Bachelorarbeit

Name des Studiengangs

Ingenieurinformatik

**Fachbereich 2**

vorgelegt von

Jaspreet Singh Dulai

Datum:

Berlin, 21.09.2023

Erstgutachter: Prof. Dr. Erik Rodner

Zweitgutachter: Marvin Grimm

## Inhaltsverzeichnis

AUFGABENSTELLUNG .....	5
ABBILDUNGSVERZEICHNIS .....	6
TABELLENVERZEICHNIS .....	7
FORMELVERZEICHNIS .....	8
ABKÜRZUNGSVERZEICHNIS.....	9
BEGRIFFS-/DEFINITIONSVERZEICHNIS.....	10
ABSTRACT .....	11
<u>1</u> EINLEITUNG .....	12
1.1 HINTERGRUND .....	12
1.2 ZIEL DER ARBEIT .....	13
1.3 EIGENE LEISTUNG .....	13
<u>2</u> STAND DER TECHNIK/FORSCHUNG .....	14
<u>3</u> THEORETISCHE GRUNDLAGEN .....	16
3.1 GRUNDLAGEN DER AKUSTIK .....	16
3.2 SPEKTROGRAMME .....	17
3.3 GRUNDLAGEN DES MASCHINELLEN LERNENS .....	18
3.3.1 EINFÜHRUNG IN DAS MASCHINELLE LERNEN .....	18
3.3.2 TYPEN DES MASCHINELLEN LERNENS .....	19
3.3.3 GRUNDLEGENDE ALGORITHMEN UND MODELLE .....	20
3.3.4 MODELLAUSWAHL UND -BEWERTUNG .....	21
3.3.5 PRAKTISCHE ASPEKTE .....	25

3.4	TRANSFER LEARNING .....	27
<b>4</b>	<b><u>DATEN.....</u></b>	<b>28</b>
4.1	MIMII-DATENSATZ .....	28
4.1.1	BESCHREIBUNG .....	28
4.1.2	NUTZUNG IM EXPERIMENT .....	28
4.2	ESC-50 DATENSATZ .....	28
4.2.1	BESCHREIBUNG .....	28
4.2.2	NUTZUNG IM EXPERIMENT .....	28
4.3	EIGENER DATENSATZ (PT500).....	29
4.3.1	BESCHREIBUNG .....	29
4.3.2	DATENERSTELLUNG.....	30
4.3.3	NUTZUNG IM EXPERIMENT .....	30
<b>5</b>	<b><u>METHODIK.....</u></b>	<b>32</b>
5.1	VORSTUDIE: MANUELLE KLASSIFIKATION MITTELS SPEKTROGRAMME .....	32
5.1.1	ERSTELLUNG EINFACHER SPEKTROGRAMME .....	32
5.1.2	VISUELLE ANALYSE UND MANUELLE KLASSIFIKATION.....	33
5.2	EXPERIMENT MIT TRANSFER LEARNING: ANWENDUNG DES YAMNET-MODELLS ..... FEHLER! TEXTMARKE NICHT DEFINIERT.	
5.2.1	EINFÜHRUNG IN DAS YAMNET MODELL .....	34
5.2.2	ANWENDUNG DES YAMNET MODELLS MIT ESC50.....	34
5.2.3	CSV-DATENINTEGRATION ZUR ERLEICHTERUNG DES TRANSFER LEARNINGS .....	34
5.2.4	ANWENDUNG DES PUMP-DATENSATZ .....	34
5.2.5	BEWERTUNG DER ERGEBNISSE UND ENTSCHEIDUNG ZUR NEUAUSRICHTUNG.....	34

5.3	ENTWICKLUNG EINES EIGENEN MODELLS .....	34
5.3.1	AUSWAHL UND ERPROBUNG VERSCHIEDENER MODELLE .....	34
5.3.2	IMPLEMENTIERUNG DER KLASSEIFIKATIONSANWENDUNG .....	34
5.3.3	BINÄRE KLASSEIFIKATION MIT DEM PUMP-DATENSATZ .....	34
5.3.4	MULTIKLASSEN-KLASSEIFIKATION MIT DEM PT500-DATENSATZ .....	34
5.3.5	HYPERPARAMETER-OPTIMIERUNG UND CROSS-VALIDATION .....	34
<u>6</u>	<u>ERGEBNISSE UND DISKUSSION.....</u>	<u>36</u>
<u>7</u>	<u>FAZIT UND AUSBLICK .....</u>	<u>37</u>
	<u>LITERATURVERZEICHNIS.....</u>	<u>38</u>
<u>8</u>	<u>ANLAGEN.....</u>	<u>41</u>
<u>9</u>	<u>EIDESSTATTLICHE ERKLÄRUNG.....</u>	<u>42</u>

## Aufgabenstellung

## Abbildungsverzeichnis

Abbildung 1 Typen des maschinellen Lernens .....	19
Abbildung 2 Hyperparametertoptimierung mit Grid Search und Kreuzvalidierung .....	26

## Tabellenverzeichnis

## Formelverzeichnis

Formel 1 Kurzzeit-Fourier-Transformation (STFT) .....	17
Formel 2 Accuracy Binär .....	22
Formel 3 Accuracy Multiklassen .....	22
Formel 4 Loss Binär .....	22
Formel 5 Loss Multiklassen .....	22
Formel 6 Precision Binär .....	23
Formel 7 Precision Multiklassen .....	23
Formel 8 Recall Binär .....	23
Formel 9 Recall Multiklassen .....	23
Formel 10 F-Score Binär .....	24
Formel 11 F-Score Multiklassen .....	24



## Abkürzungsverzeichnis

## Begriffs-/Definitionsverzeichnis

## Abstract

# 1 Einleitung

## 1.1 Hintergrund

In der modernen Industrie sind die Überwachung und Wartung von Maschinen ein kritischer Aspekt, um einen reibungslosen Betrieb und eine lange Lebensdauer der Ausrüstung zu gewährleisten. Eine der Methoden zur Überwachung des Zustands von Maschinen ist die Analyse von Maschinengeräuschen. Diese Geräusche können eine Fülle von Informationen über den Zustand und die Leistung einer Maschine liefern. Die Klassifikation von Maschinengeräuschen kann dabei helfen, Muster und Anomalien zu identifizieren, die auf mögliche Probleme oder Fehlfunktionen hinweisen.

Die Analyse von Maschinengeräuschen kann jedoch eine Herausforderung sein, da sie oft von Natur aus komplex und vielschichtig sind. Hier kommt die Rolle von Techniken wie die Analyse von Spektrogrammen und maschinellem Lernen ins Spiel. Durch die Anwendung von maschinellem Lernen auf Spektrogramme von Maschinengeräuschen können Muster und Charakteristika identifiziert werden, die mit dem menschlichen Ohr möglicherweise nicht wahrnehmbar sind.

In dieser Arbeit wird ein Deep Learning Modell entwickelt, das in der Lage ist, Maschinengeräusche effektiv zu klassifizieren. Dies könnte ein wichtiger Schritt in Richtung automatisierter und präziser Maschinenüberwachung sein, die in der Lage ist, potenzielle Probleme frühzeitig zu erkennen und zu melden, bevor sie zu ernsthaften Störungen führen.

## 1.2 Ziel der Arbeit

Das Hauptziel dieser Bachelorarbeit besteht darin, ein robustes maschinelles Lernmodell zur Klassifikation von Maschinengeräuschen zu entwickeln, wobei Spektrogramme als visuelle Darstellung der Audiodaten verwendet werden. Diese Methode hat das Potenzial, die Zustandsüberwachung von Maschinen in der Industrie erheblich zu verbessern, indem sie eine frühe Erkennung möglicher Mängel oder Störungen ermöglicht.

In der initialen Phase wird eine gründliche Analyse bestehender Methoden zur Geräuscherkennung vorgenommen, um eine fundierte Grundlage für die Entwicklung des eigenen Modells zu schaffen. Besonderes Augenmerk wird auf die Sammlung und Vorbereitung der Daten gelegt, insbesondere durch die Nutzung eines eigens erstellten Datensatzes, der die Geräusche der PT500-Maschine in verschiedenen Zuständen erfasst.

Im Mittelpunkt der Arbeit steht die Entwicklung eines CNN-Modells (Convolutional Neural Network), das in der Lage ist, feine Unterschiede in den Geräuschmustern zu erkennen und zu klassifizieren. Dabei wird eine systematische Hyperparameteroptimierung durchgeführt, um die beste Modellleistung zu erreichen, während gleichzeitig Maßnahmen gegen Überanpassung ergriffen werden. Schließlich zielt die Arbeit darauf ab, das entwickelte Modell kritisch zu bewerten, wobei verschiedene Metriken wie Genauigkeit, Verlust, Präzision und Recall zur Bewertung der Modellleistung herangezogen werden. Das Endziel ist es, ein effektives und zuverlässiges Werkzeug zur Überwachung von Maschinengeräuschen zu präsentieren, das in industriellen Anwendungen implementiert werden kann.

## 1.3 Eigene Leistung

Im Rahmen dieser Arbeit wurden mehrere Schritte unternommen, um ein zuverlässiges Modell für die Klassifikation von Maschinengeräuschen zu entwickeln. Dazu gehören:

- Die Erstellung und Vorbereitung eines Datensatzes der PT500
- Die Entwicklung CNN-Modells für die Klassifikation von Maschinengeräuschen, basierend auf Spektrogrammen.
- Die Durchführung einer Hyperparameteroptimierung zur Verbesserung der Modellleistung und zur Vermeidung von Überanpassung.
- Die Bewertung des Modells.

## 2 Stand der Technik/Forschung

In der aktuellen Forschungslandschaft zur Klassifikation von Maschinengeräuschen durch maschinelles Lernen und Spektrogramme gibt es unterschiedliche Ansätze. Die vorliegende Arbeit nutzt diese Vielfalt an Ansätzen zur Entwicklung eines Modells zur Klassifikation von Maschinengeräuschen. Im Folgenden werden einige bisherige Arbeiten in diesem Bereich vorgestellt und mit dem aktuellen Forschungsprojekt verglichen.

Carlos Alfredo Catalán und sein Team haben eine innovative Methode zur Überwachung des Verschleißzustandes von Werkzeugen während des Fräsprozesses entwickelt. Dabei nutzten sie Spektrogrammbilder zur Schulung des Klassifikationsmodells, ähnlich dem Ansatz dieser Bachelorarbeit, bei der Spektrogramme als visuelle Darstellung der Audiodaten verwendet werden. (Cruz-González et al., 2023)

Ohini Kafui Toffa und M. Mignotte untersuchten in Parallelität einen neuartigen Ansatz zur Klassifizierung von Umgebungsgeräuschen, welcher die Zusammenarbeit zwischen lokalen binären Mustern (LBPs) und Audiomeerkmalen fördert. Diese Methode könnte potenziell neue Wege für die Merkmalsextraktion und Zusammenarbeit eröffnen. Eine Entwicklung, die die Genauigkeit und Zuverlässigkeit von Geräuschklassifizierungsmodellen erheblich verbessern könnte. Zudem könnte sie als Inspirationsquelle für zukünftige Erweiterungen der vorliegenden Arbeit dienen. (Toffa & Mignotte, 2021)

Auf einer ähnlichen Note fokussierte Nikhil Bhatia auf die Anwendung von Transfer Learning und Spektrogramm-Audioklassifikationsmethoden zur Vermittlung grundlegender Konzepte des maschinellen Lernens. Die mögliche Anwendung dieser Techniken in der aktuellen Arbeit wird in Bezug auf die Erweiterung der Modellkapazitäten durch Transfer Learning untersucht (Bhatia, 2020).

Ebenso wurde von Xunsheng Ji und Mitarbeitenden ein verbessertes Verfahren zur Merkmalsauswahl für die Klassifikation von Vogelgesängen vorgestellt, das auf dem Local Binary Pattern (LBP) basiert. Eine Anwendung dieser Forschung auf die Klassifikation von Maschinengeräuschen könnte die Leistung des hier entwickelten Modells weiter verbessern (Ji et al., 2021).

Salamon und Bello lieferten einen umfassenden Überblick über aktuelle Trends und Techniken zur Klassifikation von Umgebungsgeräuschen, welcher als fundierte Basis für die aktuellen Forschungsanstrengungen dienen kann (Salamon & Bello, 2017).

Darüber hinaus gab Piczak wertvolle Einblicke in die Anwendung von Convolutional Neural

Networks (CNNs) zur Klassifikation von Umgebungsgeräuschen, eine Technik, die auch in dieser Bachelorarbeit eingehend untersucht und angewendet wird (Piczak, 2015).

Insgesamt liefern diese Arbeiten eine Vielzahl von Perspektiven und Methoden zur Klassifizierung von Geräuschen, die als inspirierende Quelle und fundierte Grundlage für die Entwicklung des Modells zur Klassifizierung von Maschinengeräuschen in der vorliegenden Arbeit dienen können. Durch einen präzisen Vergleich dieser Arbeiten lässt sich ein tiefergehendes Verständnis für die Vor- und Nachteile der unterschiedlichen Ansätze erlangen. Dies könnte dazu beitragen, die Methodik der vorliegenden Arbeit effektiv anzupassen und zu optimieren.

## 3 Theoretische Grundlagen

### 3.1 Grundlagen der Akustik

Die Akustik, ein Kerngebiet der Physik, beschäftigt sich intensiv mit der Untersuchung mechanischer Schallwellen in verschiedenen Medien wie Luft, Wasser oder Festkörpern. Im Kontext der vorliegenden Arbeit sind einige spezifische Aspekte der Akustik von besonderer Bedeutung.

Schallwellen werden durch ihre physikalischen Eigenschaften wie Frequenz, Wellenlänge und Amplitude charakterisiert. Die Frequenz, ausgedrückt in Hertz (Hz), bezieht sich auf die Anzahl der Schwingungen pro Sekunde, während die Wellenlänge den Abstand zwischen zwei aufeinander folgenden Wellenbergen beschreibt. Die Amplitude hingegen gibt die Größe der Schwingung an, die oft mit der Lautstärke des Schalls in Verbindung gebracht wird. (Kinsler et al., 1963)

Die Schallausbreitung ist ein komplexer Prozess, der von verschiedenen Faktoren beeinflusst wird, darunter die Eigenschaften des Mediums (wie Dichte und Temperatur) und die Eigenschaften der Schallquelle selbst. Diese Faktoren können die Geschwindigkeit und die Art der Schallausbreitung beeinflussen, was sich wiederum auf die Qualität der erfassten akustischen Daten auswirkt. Ein umfassendes Verständnis der Schallausbreitung ist daher für eine effektive Analyse und Klassifizierung von Maschinengeräuschen unerlässlich.

In der Praxis der Maschinengeräuschklassifikation können diese Grundlagen der Akustik als solides Fundament dienen, um effektive und präzise maschinelle Lernmodelle zu entwickeln, die in der Lage sind, komplexe akustische Muster zu erkennen und zu klassifizieren.



## 3.2 Spektrogramme

Ein Spektrogramm, wie von Fulop und Fitz definiert, ist "eine zweidimensionale Darstellung eines Signals, bei der die Zeit auf der x-Achse, die Frequenz auf der y-Achse und die Amplitude in Farbe dargestellt wird". Diese visuelle Darstellung ist ein zentrales Werkzeug in der Analyse von Maschinengeräuschen, da sie detaillierte Einblicke in die Frequenzkomponenten eines Signals über die Zeit bietet, wodurch einzigartige Muster und Charakteristiken identifiziert werden können, die für die Klassifikation von Maschinengeräuschen essentiell sind. (Fulop & Fitz, 2006)

Die moderne Forschung hat die Anwendung von Spektrogrammen in der Klassifikation von Maschinengeräuschen erheblich vorangetrieben. Beispielsweise werden sie in Fahrzeug-zu-Fahrzeug-Netzwerken zur Verbesserung der Straßensicherheit genutzt, indem sie die Identifikation von Fahrzeugdynamiken während Manövern relativ zueinander ermöglichen, was eine zuverlässige Bewusstseinsbildung mit hoher Genauigkeit und wenigen Fehlklassifikationen sowohl in Sichtlinien- als auch in Nicht-Sichtlinien-Bedingungen. (Kihei et al., 2017)

Ein weiterer Fortschritt ist die Entwicklung der zeitkartierten Spektrogrammanalyse, die eine kontinuierliche Kartierung des sich ändernden Spektrums eines Eingangssignals entlang der Zeitdomäne ermöglicht, wodurch neue Möglichkeiten zur Entspannung der Spezifikationen des Detektionsgeräts eröffnet werden. (Konatham et al., 2020)

Die Integration von Spektrogrammen mit maschinellem Lernen eröffnet neue Perspektiven in der Diagnose und Überwachung von Maschinen. Durch die Analyse komplexer akustischer Szenarien können Maschinengeräusche effektiv klassifiziert werden, was zu einer verbesserten Überwachung und Wartung von Maschinen führt.

Ein wichtiger Faktor bei der Erstellung eines Spektrogramms ist die Nutzung der Kurzzeit-Fourier-Transformation (STFT), welches ein Verfahren ist, um Frequenzbestandteile eines Signals über die Zeit zu extrahieren. Die mathematische Definition der STFT lautet:

$$STFT(x(t)) = X(\tau, \omega) = \int_{-\infty}^{\infty} x(t)w(t - \tau)e^{-j\omega t}dt$$

*Formel 1 Kurzzeit-Fourier-Transformation (STFT)*

wo  $x(t)$  das Signal,  $w(t)$  das Fenster und  $\tau$  die Zeitverschiebung darstellen. Diese Transformation analysiert ein Signal durch seine Komponenten unterschiedlicher Frequenzen und ermöglicht somit eine tiefere Untersuchung der enthaltenen Informationen. Besonders in der vorliegenden Arbeit, welche die Klassifikation von Maschinengeräuschen behandelt, ist die STFT eine grundlegende Methode, um präzise Einblicke in die Zeit-Frequenz-Charakteristika der Geräusche zu gewinnen.

### 3.3 Grundlagen des maschinellen Lernens

#### 3.3.1 Einführung in das maschinelle Lernen

Das maschinelle Lernen, ein Teilgebiet der künstlichen Intelligenz, hat sich als revolutionäre Technologie in verschiedenen Bereichen, einschließlich der Analyse und Klassifizierung von Maschinengeräuschen, etabliert. Im Wesentlichen geht es darum, Computern die Fähigkeit zu verleihen, aus Daten zu lernen und Muster zu erkennen, ohne dass sie explizit programmiert werden müssen – die Maschine lernt.

Grundlage des maschinellen Lernens ist die Entwicklung von Algorithmen, die auf der Analyse großer Datenmengen basieren. Diese Algorithmen ermöglichen es Maschinen, komplexe Muster in den Daten zu erkennen und Vorhersagen oder Entscheidungen zu treffen, die auf neuen, noch nie zuvor gesehenen Daten basieren.

(Bild: Hier das Bild aus der Quelle einfügen später... nicht vergessen1!1!).

Die Anwendungsbereiche von ML sind vielfältig und reichen von der Bild- und Spracherkennung bis hin zur Vorhersage von Markttrends und der Verbesserung von Geschäftsstrategien. In der Industrie wird es zunehmend für die Zustandsüberwachung von Maschinen eingesetzt, um potenzielle Ausfälle frühzeitig zu erkennen und vorbeugende Maßnahmen zu ergreifen.

Die Entwicklung eines solchen Modells umfasst mehrere Schritte, darunter die Datensammlung, die Datenvorverarbeitung, die Modellentwicklung und die Evaluierung. Diese Schritte sind entscheidend, um ein Modell zu entwickeln, welches genaue und zuverlässige Ergebnisse liefert.

In den folgenden Abschnitten werden die verschiedenen Arten des maschinellen Lernens, die grundlegenden Modelle und Algorithmen sowie die Auswahl und Bewertung von Modellen im Detail besprochen. Darüber hinaus werden praktische Aspekte wie Datenvorverarbeitung und Feature-Engineering behandelt, die für die Entwicklung eines erfolgreichen maschinellen Lernmodells von entscheidender Bedeutung sind.

### 3.3.2 Typen des maschinellen Lernens

Maschinelles Lernen kann in verschiedene Typen unterteilt werden, basierend auf der Art und Weise, wie Modelle aus Daten lernen. Die drei Haupttypen sind überwachtes Lernen, unüberwachtes Lernen und bestärkendes Lernen. (Khandelwal, 2022)

Beim **überwachten Lernen** werden Modelle mit einem Datensatz trainiert, der sowohl Eingabevariablen als auch die entsprechenden korrekten Ausgaben enthält. Das Ziel ist es, eine Funktion zu lernen, die eine Eingabe (Features) auf eine Ausgabe (Label) abbildet. Dies ermöglicht dem Modell, Vorhersagen oder Entscheidungen zu treffen, wenn es mit neuen Daten konfrontiert wird. Typische Anwendungen sind die Klassifikation und die Regression. Überwachte Lernmodelle sind besonders nützlich in Anwendungen, bei denen die Vorhersage von zukünftigen Ereignissen auf der Grundlage von historischen Daten erforderlich ist.

Im Gegensatz dazu arbeitet das **unüberwachte Lernen** mit Datensätzen ohne beschriftete Antworten. Das Ziel hier ist es, die zugrunde liegenden Strukturen in den Daten zu entdecken. Häufige Techniken sind Clustering und Assoziationsregel-Lernen. Diese Art des Lernens ist besonders nützlich, wenn die Struktur der Daten unbekannt ist und wir Modelle erstellen möchten, die helfen, komplexe Muster in den Daten zu identifizieren.

**Bestärkendes Lernen** ist ein Typ des maschinellen Lernens, bei dem ein Agent lernt, wie er in einer Umgebung handeln soll, um einige langfristige Belohnungen zu maximieren. Es ist besonders nützlich in Situationen, in denen ein Agent automatisch die optimale Strategie lernen soll, um ein bestimmtes Ziel zu erreichen

In den folgenden Abschnitten werden wir tiefer in die verschiedenen Modelle und Algorithmen des maschinellen Lernens eintauchen, sowie die praktischen Aspekte der Modellauswahl und -bewertung und die notwendigen Schritte zur Optimierung von Modellen diskutieren.

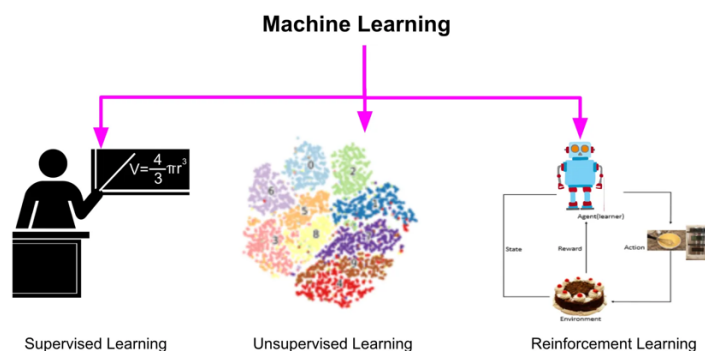


Abbildung 1 Typen des maschinellen Lernens (Khandelwal, 2022)

### 3.3.3 Grundlegende Algorithmen und Modelle

In diesem Abschnitt untersuchen wir einige der grundlegenden Modelle und Algorithmen des maschinellen Lernens, die als Bausteine für komplexere Systeme dienen können.

**Lineare Modelle**, einschließlich der linearen Regression und der logistischen Regression, gehören zu den einfachsten und am häufigsten verwendeten Algorithmen des maschinellen Lernens. Diese Modelle versuchen, eine lineare Beziehung zwischen den Eingabevariablen und der Zielvariablen herzustellen. Sie eignen sich besonders für Aufgaben mit klar definierten linearen Beziehungen und sind oft ein guter Ausgangspunkt für die Modellentwicklung. (Lu, 2022)

**Entscheidungsbäume** sind ein weiteres beliebtes Modell, das auf der Segmentierung des Merkmalsraums in eine Reihe einfacher Regeln beruht. Sie sind leicht zu verstehen und zu interpretieren, was sie zu einer guten Wahl für explorative Datenanalysen macht. (Quinlan, 2014)

**Neuronale Netze** stehen im Mittelpunkt dieser Arbeit und sind eine Familie von Algorithmen, die sich durch ihre Fähigkeit auszeichnen, komplexe Muster in Daten zu erkennen. Sie sind von der Funktionsweise des menschlichen Gehirns inspiriert und bestehen aus Schichten von Neuronen, die miteinander verbunden sind. In dieser Arbeit wird ein spezielles neuronales Netz, ein Convolutional Neural Network (CNN), entwickelt, um Maschinengeräusche zu klassifizieren. CNNs sind besonders effektiv bei der Analyse von visuellen und akustischen Daten, da sie in der Lage sind, lokale Muster in den Daten zu erkennen und zu lernen. (Goodfellow et al., 2016)

In den folgenden Abschnitten werden Methoden zur Modellauswahl und -bewertung sowie praktische Aspekte der Datenvorverarbeitung und Modelloptimierung diskutiert.

### 3.3.4 Modellauswahl und-bewertung

Die Auswahl und Bewertung des Modells stellen einen entscheidenden Schritt im Prozess des maschinellen Lernens dar, um die Leistungsfähigkeit und Zuverlässigkeit des entwickelten Modells zu gewährleisten. Im Folgenden werden spezifische Aspekte wie Über- und Unteranpassung, Kreuzvalidierung sowie verschiedene Metriken zur Bewertung eines Modells detailliert diskutiert, um ein tiefes Verständnis der Best Practices in diesem Bereich zu fördern.

**Überanpassung (Overfitting)** tritt auf, wenn das Modell die Trainingsdaten zu gut lernt, aber schlecht auf neue, unbekannte Daten generalisiert. Tritt Overfitting auf, dann hat das Modell den Trainingsdatensatz „auswendig“ gelernt. **Unteranpassung (Underfitting)** hingegen liegt vor, wenn das Modell nicht in der Lage ist, die zugrunde liegenden Muster in den Trainingsdaten zu erfassen. Die Balance zwischen diesen beiden Zuständen zu finden, ist entscheidend für die Entwicklung eines erfolgreichen Modells. (Goodfellow et al., 2016)

Die k-fache **Kreuzvalidierung** ist eine wichtige Methode, insbesondere bei der Handhabung der Datensatz-Aufteilung in einen festen Trainings- und Testdatensatz. Dies kann problematisch sein, insbesondere wenn der Testdatensatz klein ist, da eine begrenzte Größe statistische Unsicherheiten bei der Abschätzung des mittleren Testfehlers verursachen kann und somit die Bewertung der Leistung verschiedener Algorithmen erschwert wird.

Bei sehr großen Datensätzen mit Hunderttausenden oder mehr Beispielen ist dies weniger problematisch. Bei kleineren Datensätzen sind jedoch alternative Ansätze erforderlich. Diese ermöglichen die Schätzung des mittleren Testfehlers unter Einbeziehung aller Beispiele, erfordern jedoch einen höheren Rechenaufwand. Solche Methoden wiederholen Trainings- und Testberechnungen auf verschiedenen zufällig ausgewählten Teilmengen oder Segmenten des ursprünglichen Datensatzes. Im Bereich der Datenanalyse gilt die k-fache Kreuzvalidierung als eine etablierte Methode. Hierbei wird der Datensatz in k Teilmengen aufgeteilt, die sich nicht überschneiden. Der Testfehler wird durch die Berechnung des durchschnittlichen Testfehlers über k Iterationen geschätzt. Dabei dient in jeder Iteration eine andere Teilmenge als Testdatensatz, während der Rest als Trainingsdatensatz verwendet wird. (Goodfellow et al., 2016)

Zur Beurteilung der Wirksamkeit und Zuverlässigkeit des entwickelten Modells ist es unerlässlich, verschiedene **Bewertungsmetriken** zu verwenden. In dieser Arbeit werden insbesondere die Metriken Accuracy (Genauigkeit), Loss (Verlust), Precision (Präzision), Recall und der F-Score herangezogen. Diese ausgewählten Metriken bieten einen umfassenden Einblick in die Leistung des Modells aus verschiedenen Perspektiven (Sokolova & Lapalme, 2009).

**Accuracy:** Diese Metrik ermöglicht eine objektive Beurteilung der Effektivität des Modells, indem sie das Verhältnis der korrekt klassifizierten Beispiele zur Gesamtanzahl der Beispiele angibt.

Binäre Klassifikation

$$\frac{tp + tn}{tp + fn + fp + tn}$$

*Formel 2 Accuracy Binär*

Multiklassen Klassifikation

$$\frac{\sum_{i=1}^l \frac{tp + tn}{tp + fn + fp + tn}}{l}$$

*Formel 3 Accuracy Multiklassen*

Der Verlust liefert Informationen über die Diskrepanz zwischen den vom Modell vorhergesagten Labels und den tatsächlichen Labels. Ein niedrigerer Wert zeigt eine bessere Modellleistung an.

Binäre Klassifikation (Binary Crossentropy)

$$-\frac{1}{N} \sum_{i=1}^N (y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i)))$$

*Formel 4 Loss Binär*

Multiklassen Klassifikation (Categorical Crossentropy)

$$-\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{ic} \log(p(y_{ic}))$$

*Formel 5 Loss Multiklassen*

Hierbei ist  $N$  die Anzahl der Beobachtungen,  $C$  die Anzahl der Klassen,  $y_{ic}$  die wahre Klassenzugehörigkeit (0 oder 1) und  $p(y_{ic})$  die vorhergesagte Wahrscheinlichkeit.

Die Präzision gibt Auskunft darüber, wie zuverlässig das Modell positive Vorhersagen trifft. Dabei wird das Verhältnis der korrekt positiv vorhergesagten Beispiele zur Gesamtanzahl der positiv vorhergesagten Beispiele gemessen.

Binäre Klassifikation

$$\frac{tp}{tp + fp}$$

*Formel 6 Precision Binär*

Multiklassen Klassifikation

$$\frac{\sum_{i=1}^l tp_i}{\sum_{i=1}^l (tp_i + fp_i)}$$

*Formel 7 Precision Multiklassen*

Der Recall bewertet die Fähigkeit des Modells, alle tatsächlichen positiven Beispiele zu identifizieren. Hierbei wird das Verhältnis der korrekt positiv vorhergesagten Beispiele zur Gesamtanzahl der tatsächlichen positiven Beispiele berechnet.

Binäre Klassifikation

$$\frac{tp}{tp + fn}$$

*Formel 8 Recall Binär*

Multiklassen Klassifikation

$$\frac{\sum_{i=1}^l tp_i}{\sum_{i=1}^l (tp_i + fn_i)}$$

*Formel 9 Recall Multiklassen*

Der F-Score dient als einheitliches Maß, das sowohl die Präzision als auch den Recall berücksichtigt. Er bietet eine ausgewogene Bewertung der Modellleistung, indem er beide Metriken in einem harmonischen Mittelwert zusammenfasst.

Binäre Klassifikation

$$\frac{(\beta^2 + 1)tp}{(\beta^2 + 1)tp + \beta^2 fn + fp}$$

Formel 10 F-Score Binär

Multiklassen Klassifikation

$$\frac{(\beta^2 + 1)(Precision Recall)}{(\beta^2 Precision) + Recall}$$

Formel 11 F-Score Multiklassen

Hierbei stehen tp (True Positives), tn (True Negatives), fp (False Positives) und fn (False Negatives) für die Anzahl der korrekt positiv, korrekt negativ, falsch positiv und falsch negativ klassifizierten Beispiele, während l die Anzahl der Klassen repräsentiert und  $\beta$  ein Gewichtungsfaktor ist, der das Verhältnis von Präzision zu Recall steuert.

Diese Kennzahlen dienen nicht nur zur objektiven Bewertung der Modellleistung, sondern auch als leitende Instrumente zur Identifizierung von Verbesserungsbereichen und zur gezielten Anpassung zur Optimierung der Modellleistung.

DIESE TABELLE WIRD WAHRSCHEINLICH WIEDER ENTFERNT (Das war der erste Ansatz, jedoch ist es so nicht so schön mit den Beschriftungen der Formeln

	Binäre Klassifikation	Multiklassen Klassifikation
Accuracy (Genauigkeit)	$\frac{tp + tn}{tp + fn + fp + tn}$	$\frac{\sum_{i=1}^l \frac{tp + tn}{tp + fn + fp + tn}}{l}$
Precision (Präzision)	$\frac{tp}{tp + fp}$	$\frac{\sum_{i=1}^l tp_i}{\sum_{i=1}^l (tp_i + fp_i)}$
Recall (Sensitivität)	$\frac{tp}{tp + fn}$	$\frac{\sum_{i=1}^l tp_i}{\sum_{i=1}^l (tp_i + fn_i)}$
F-score (F-Wert)	$\frac{(\beta^2 + 1)tp}{(\beta^2 + 1)tp + \beta^2 fn + fp}$	$\frac{(\beta^2 + 1)(Precision Recall)}{(\beta^2 Precision) + Recall}$



### 3.3.5 Praktische Aspekte

Im Folgenden werden die entscheidenden Aspekte der praktischen Anwendung des maschinellen Lernens erläutert. Ein besonderes Augenmerk wird dabei auf die Datenvorverarbeitungs- und Modelloptimierungsphasen gelegt, um ein umfassendes Verständnis für die Herausforderungen und Lösungsansätze in diesem Bereich zu vermitteln.

Die **Datenvorverarbeitung** ist ein zentraler Schritt in der Datenanalyse, insbesondere im Bereich des maschinellen Lernens. Sie stellt sicher, dass die Rohdaten in ein Format umgewandelt werden, das für die weitere Analyse und Modellierung besser geeignet ist. Diese Phase kann mehrere kritische Schritte umfassen, von denen jeder dazu beiträgt, die Qualität und Effizienz des nachfolgenden Analyseprozesses zu verbessern.

Die erste Phase, die Datenbereinigung, befasst sich mit der Bereinigung der Daten von Inkonsistenzen, Fehlern oder Ausreißern, die die Leistung des Modells negativ beeinflussen könnten. Hier kommen Techniken zur Behandlung fehlender Werte und zur Normalisierung der Daten zum Einsatz, um eine konsistente Grundlage für die Analyse zu schaffen.

Die nächste Phase ist die Datenintegration, in der Daten aus verschiedenen Quellen und Formaten integriert werden, um eine umfassende Datenbasis zu schaffen. Diese Integration erleichtert eine ganzheitliche Analyse und eröffnet neue Perspektiven, indem komplexere Muster und Beziehungen in den Daten erkannt werden können.

Auf die Integration folgt die Datentransformation, bei der die integrierten Daten in eine geeignete Struktur oder ein geeignetes Format umgewandelt werden, das den spezifischen Anforderungen der Analyse oder des Modells besser entspricht. Dies kann durch verschiedene Techniken erreicht werden, die darauf abzielen, die Daten in eine Form zu bringen, die eine effektivere Analyse ermöglicht.

Schließlich umfasst die Phase der Datenreduktion Techniken, um die Dimensionalität der Daten zu reduzieren bzw. zu erhöhen, ohne dabei wichtige Informationen zu verlieren. Dies trägt dazu bei, die Effizienz des Modells zu erhöhen und die Komplexität des Analyseprozesses zu verringern, indem unnötige oder redundante Daten entfernt werden (García et al., 2015).

Insgesamt ist die Datenvorverarbeitung ein entscheidender Schritt bei der Vorbereitung von Daten für effektive Analysen und Modellierungsprozesse im Bereich des maschinellen Lernens. Sie trägt dazu bei, die Qualität und Relevanz der Daten, die für die Analyse verwendet werden, zu verbessern.

Die **Optimierung von Modellen** ist ein zentraler und kontinuierlicher Prozess, der darauf abzielt, die maximale Leistungsfähigkeit eines Modells zu erreichen. Ein kritischer Teil dieses Prozesses ist die Anpassung der Hyperparameter. Dabei wird häufig die Grid-Search-Methode verwendet, eine Technik, bei der verschiedene Kombinationen von Hyperparametern systematisch untersucht werden, um die optimale Konfiguration zu identifizieren. In Kombination mit Kreuzvalidierung ermöglicht Grid Search eine robuste Bewertung der Modellperformance, indem verschiedene Teilmengen der Daten zur Validierung verwendet werden. Dieser Ansatz fördert nicht nur die Verbesserung der Modellleistung, sondern hilft auch, das Risiko des Overfittings zu minimieren, indem sichergestellt wird, dass das Modell gut auf unbekannte Daten generalisiert. Letztendlich zielt diese Methode darauf ab, die Hyperparameter zu finden, die die beste Leistung in Bezug auf die vorgegebenen Kriterien liefern. (Ruder, 2017)

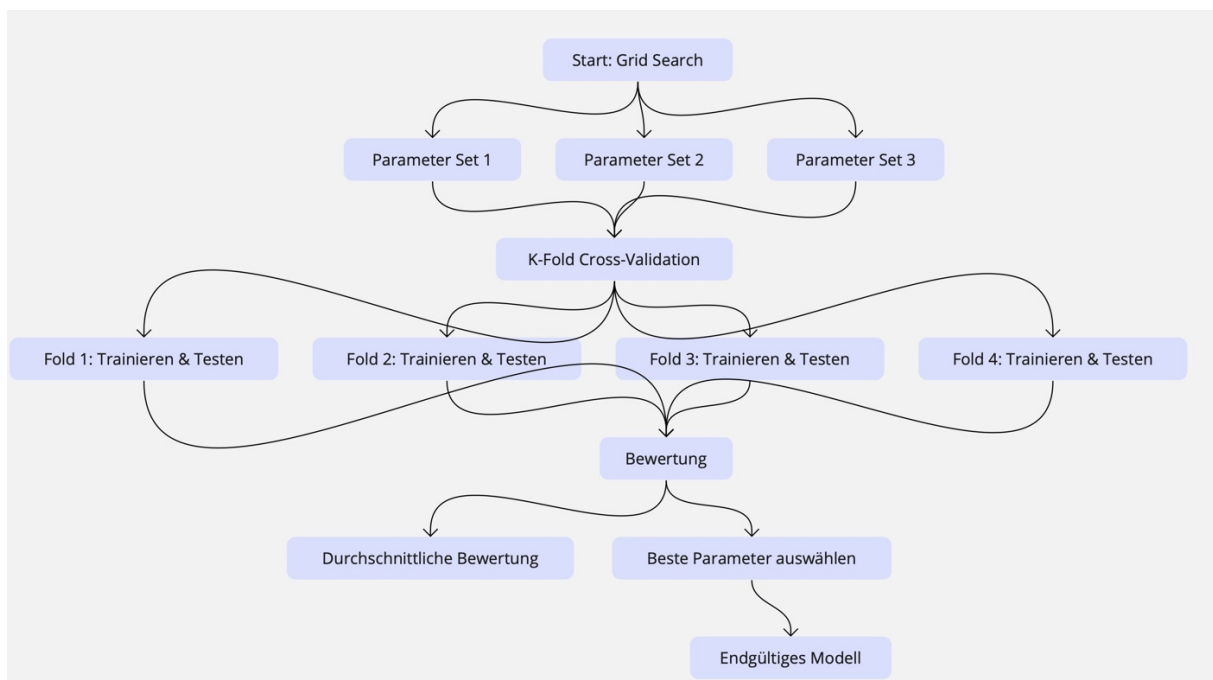


Abbildung 2 Hyperparameteroptimierung mit Grid Search und Kreuzvalidierung

### 3.4 Transfer Learning

Transfer Learning ist eine Methode im Bereich des maschinellen Lernens, bei der ein für eine Aufgabe entwickeltes Modell als Ausgangspunkt für ein Modell für eine zweite, ähnliche Aufgabe verwendet wird. Diese Technik ist besonders nützlich, wenn die Daten für die zweite Aufgabe begrenzt sind, da sie es ermöglicht, von den bereits erlernten Merkmalen der ersten Aufgabe zu profitieren. (Pan & Yang, 2010)

Eine der prominenten Anwendungen des Transfer Learnings ist die Verwendung von vortrainierten Modellen, die auf umfangreichen Datensätzen trainiert wurden. Ein solches Modell ist YAMnet, ein tiefes neuronales Netzwerk, das von TensorFlow entwickelt wurde. YAMnet ist darauf trainiert, eine Vielzahl von Audiomeerkmalen zu erkennen, und kann als Feature-Extractor für neue Aufgaben im Bereich der Audioanalyse verwendet werden. (Tensorflow, 2023)

In der Praxis könnte dies bedeuten, dass YAMnet zur Extraktion von Merkmalen aus Maschinengeräuschen verwendet wird, wobei die extrahierten Merkmale dann als Eingabe für ein neues Modell dienen, das speziell für die Klassifikation von Maschinengeräuschen entwickelt wurde. Dies könnte die Entwicklungszeit verkürzen und die Leistung des Modells verbessern, insbesondere wenn die verfügbaren Daten begrenzt sind.

In dieser Arbeit wird ein kleiner Versuch unternommen, das Potenzial des Transfer Learnings mit YAMnet zur Verbesserung der Klassifikation von Maschinengeräuschen zu erkunden.

## 4 Daten

### 4.1 MIMII-Datensatz

#### 4.1.1 Beschreibung

Der MIMII-Datensatz ist eine Sammlung von Geräuschaufnahmen aus der Industrie, die normale und anomale Geräusche von Maschinen wie Ventilen und Pumpen umfasst. Die Aufnahmen wurden mit einer Acht-Kanal-Mikrofonanordnung in realen Fabrikumgebungen erstellt, um eine realistische Darstellung von Industriegeräuschen zu bieten. Dieser Datensatz dient als Benchmark für die akustische Fehlerdiagnose von Maschinen. (Purohit et al., 2019)

#### 4.1.2 Nutzung im Experiment

Zunächst wurden simple Spektrogramme mit einem einfachen Code erstellt. Diese wurden zunächst visuell analysiert, um festzustellen, ob eine manuelle Klassifizierung der Geräusche auf der Grundlage menschlicher Beobachtungen möglich ist. Dies diente als eine Art Vorstudie, um das Potential der Daten für eine maschinelle Lernanalyse zu evaluieren.

Im weiteren Verlauf des Experiments wurde speziell der Pumpendatensatz aus dem MIMII-Kompendium verwendet. Das anfängliche Modell, das entwickelt wurde, zielte darauf ab, zwischen normalen und anomalen Geräuschen zu unterscheiden, die spezifisch für Pumpengeräusche sind. Dieser Schritt war entscheidend, um die Effektivität des Modells bei der Erkennung und Klassifizierung von Maschinengeräuschen zu bewerten und zu optimieren. Die Fokussierung auf Pumpengeräusche lieferte eine solide Grundlage für eine eingehende Analyse, die für die weitere Entwicklung und Verfeinerung des Modells unerlässlich war.

### 4.2 ESC-50 Datensatz

#### 4.2.1 Beschreibung

Der ESC-50 Datensatz ist eine beschriftete Sammlung von 2000 Umgebungsgeräuschaufnahmen, die sich hervorragend für Benchmarking-Methoden zur Klassifikation von Umgebungsgeräuschen eignen. Dieser Datensatz bietet eine vielfältige Palette von Umgebungsgeräuschen, die eine breite Basis für das Training und die Validierung von Modellen zur Geräuscherkennung bieten. (Piczak, 2015/2023)

#### 4.2.2 Nutzung im Experiment

In der Anfangsphase wurde mit dem YAMNet-Modell und dem ESC-50 Datensatz experimentiert, um ein grundlegendes Verständnis für die Funktionsweise des Transfer Learning zu entwickeln. In späteren Phasen des Experiments wurde versucht, das Modell mit dem

Pumpendatensatz aus dem MIMII-Datensatz weiter zu trainieren. Dies diente als praktische Anwendung des Transfer Learning, bei der die Fähigkeiten des Modells zur Klassifikation von Maschinengeräuschen weiter untersucht und optimiert wurden.

### 4.3 Eigener Datensatz (PT500)

#### 4.3.1 Beschreibung

Das PT500-System, entwickelt von der renommierten Firma GUNT, ist ein unverzichtbares Werkzeug für Lehre und Forschung in den Bereichen Maschinenbau, Verfahrenstechnik und Energietechnik. Dieses System wurde speziell entwickelt, um den Zustand von Maschinen oder deren Komponenten durch Schwingungsanalyse zu bewerten. Insbesondere ermöglicht es die Simulation und Untersuchung verschiedener Arten von Maschinenschäden, wobei der Schwerpunkt auf Wälzlagern liegt.

Wesentliche Bestandteile des Grundgerätes sind mechanische Komponenten wie Kupplungen, Lagerböcke, Wellen bzw. Rotoren, ein drehzahlvariabler Antriebsmotor, der über einen Frequenzumrichter gesteuert wird, sowie ein Anzeige- und Bediengerät mit digitalen Anzeigen für Leistung und Drehzahl. Eine großflächige Aufspannplatte aus Aluminium ermöglicht eine einfache, flexible und präzise Montage der verschiedenen Versuchskomponenten. Zur Sicherheit während des Betriebs und zur ungestörten Beobachtung der Versuche ist eine transparente Schutzhaube vorhanden (GUNT, 2023).

Die durchgeführten Untersuchungen erstreckten sich auf insgesamt sechs verschiedene Zustände der Wälzlager, die intern zur Unterscheidung mit den Buchstaben A bis F gekennzeichnet sind. Diese Zustände umfassten:

- A unbeschädigten Wälzlagers
- B Schäden am Außenring
- C Schäden am Innenring
- D Schäden am Wälzkörper
- E Schadenskombination aus B-D
- F starke Abnutzung

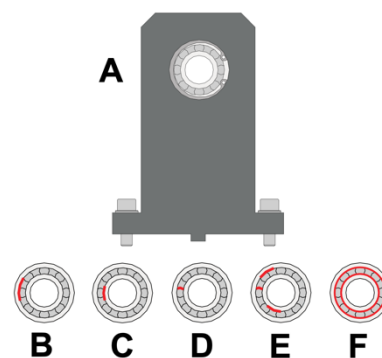


Abbildung 3 Wälzlagerzustände (GUNT, 2014)

Diese Vielfalt an Zuständen ermöglichte eine umfassende Untersuchung und Bewertung der Maschinenschäden im Rahmen dieses Forschungsprojektes.

#### 4.3.2 Datenerstellung

Die Datenerfassung für dieses Projekt wurde mit einem TASCAM DR-05X Mikrofon durchgeführt. Das Mikrofon wurde auf einer weichen Unterlage etwa 30 cm von den kritischen Wälzlager entfernt aufgestellt, um die Geräusche der Maschine aufzunehmen. Die Drehzahlen wurden manuell auf 500, 1000 und 1500 U/min eingestellt, um eine breite Datenerfassung zu gewährleisten, ohne das Risiko einzugehen, nicht getestete höhere Drehzahlen zu verwenden.

Um die Datenqualität zu gewährleisten und Störungen und mögliche Beeinträchtigungen anderer Personen zu minimieren, wurden die Aufnahmen bewusst zu späterer Stunde durchgeführt, wenn weniger Personen im Gebäude anwesend waren. Dieser zeitliche Ansatz wurde gewählt, um sicherzustellen, dass die aufgezeichneten Maschinengeräusche in einer möglichst kontrollierten Umgebung aufgenommen wurden.

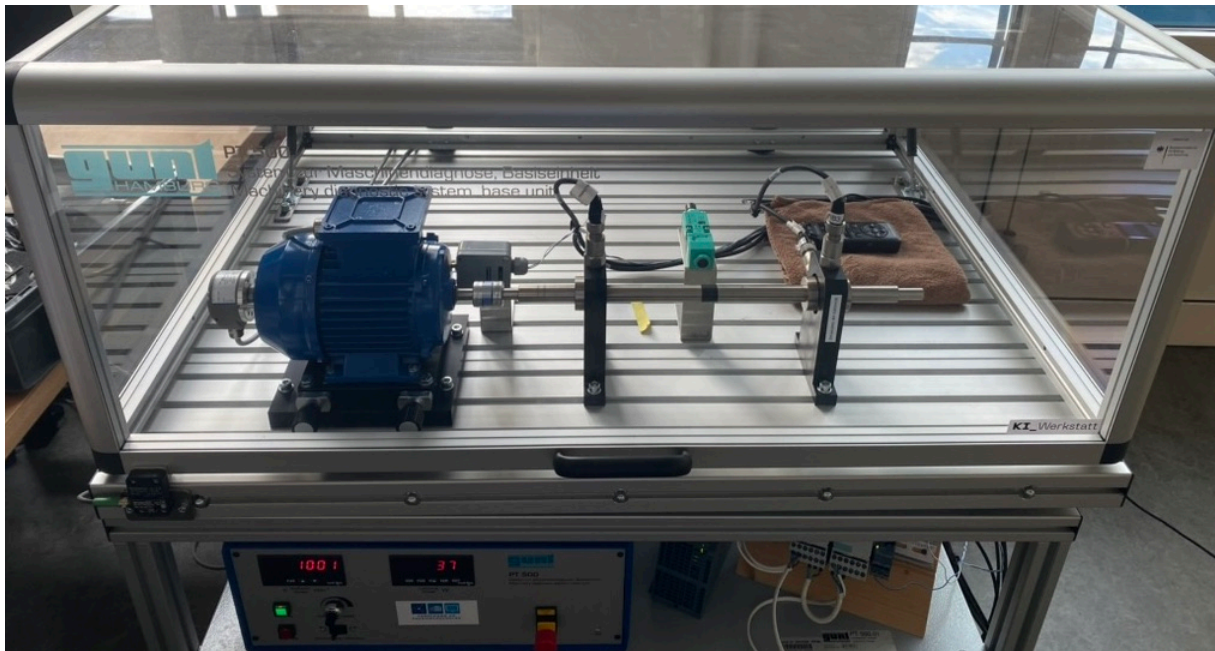


Abbildung 4 Versuchsaufbau für die Aufnahme

#### 4.3.3 Nutzung im Experiment

Im Experiment wurden die aufgezeichneten Audiodaten als primäre Ressource für das Training und die Erstellung des maschinellen Lernmodells verwendet. Die Daten wurden in 10-Sekunden-Segmente segmentiert und dienten als Grundlage für die Analyse und Klassifikation von Maschinengeräuschen, insbesondere von Wälzlagergeräuschen.

Die Segmentierung der Daten in 10-Sekunden-Stücke wurde durch einen einfachen Code, welcher später zum gesamten Projekt hinzugefügt wurde, ermöglicht. Wodurch nicht nur die Effizienz des Trainingsprozesses erhöhte, sondern auch eine präzise Analyse der aufgenommenen Geräusche entstand. Dieser Ansatz erleichterte die Integration der Daten in das Projekt und ermöglichte eine detaillierte Leistungsbewertung des Modells, die in den folgenden Abschnitten der Arbeit näher erläutert wird.

## 5 Methodik

### 5.1 Vorstudie: Manuelle Klassifikation mittels Spektrogramme

#### 5.1.1 Erstellung einfacher Spektrogramme

In der Initialphase des Projektes lag der Schwerpunkt auf der Erstellung einer visuellen Darstellung der Maschinengeräusche, um eine solide Grundlage für die manuelle Analyse und Klassifikation der Daten zu schaffen. Diese visuelle Darstellung erfolgte durch die Erstellung von Spektrogrammen, die die Frequenzspektren der Audiosignale im zeitlichen Verlauf visualisieren.

Zur Erstellung der Spektrogramme wurde ein Python-Skript genutzt, welches auf der weit verbreiteten Python-Bibliothek librosa zur Analyse von Musik und Audio basiert (transcriptaze, 2020/2023). Zusätzlich wurden die Bibliotheken numpy für numerische Operationen und PIL für die Bildverarbeitung implementiert. Der entwickelte Code nutzt die Funktionen dieser Bibliotheken, um Audiodaten zu laden, Spektrogramme zu generieren und diese als Bilder zu speichern. Eine detaillierte Beschreibung des verwendeten Codes kann im Anhang dieser Arbeit gefunden werden

Der angezeigte Codeabschnitt zeigt die wav2stft-Funktion, die für die Generierung von Spektrogrammen aus Audio-Dateien zentral ist. Die Funktion akzeptiert mehrere Parameter, einschließlich des Dateipfads und verschiedener Parameter zur Steuerung der Spektrogrammerstellung. Sie verwendet die librosa-Bibliothek zum Laden der Audiodatei und zur Anwendung der Kurzzeit-Fourier-Transformation (STFT).

```
def wav2stft(fname, sr, fftSize, fftHop, dur) :  
    try:  
        audiodata, samplerate = librosa.load(fname, sr=sr, mono=True, duration=dur)  
    except:  
        print('can not read ' + fname)  
        return  
  
    if sr == None:  
        print('Using native samplerate of ' + str(samplerate))  
    S = np.abs(librosa.stft(audiodata, n_fft=fftSize, hop_length=fftHop, win_length=fftSize, center=True))  
    return S
```

Abbildung 5 Funktion zur Berechnung des Spektrogramms (transcriptaze, 2020/2023)

Nach der Generierung der Spektrogramme wurde eine visuelle Analyse durchgeführt. In diesem Schritt wurden die verschiedenen Geräuschklassen anhand ihrer spektralen Eigenschaften manuell identifiziert. Diese manuelle Klassifikation diente als erste Einschätzung des Potenzials der Daten für die Verwendung in maschinellen Lernmodellen.



### 5.1.2 Visuelle Analyse und manuelle Klassifikation

In dieser Projektphase wurde eine visuelle Analyse der zuvor erstellten Spektrogramme durchgeführt. Ziel war es, die zwei verschiedenen Arten von Maschinengeräuschen, in diesem Fall die Pumpe aus dem MIMII Datensatz, zu identifizieren und charakteristische Merkmale zu erfassen. Diese Analyse wurde unabhängig durchgeführt, um eine erste Einschätzung der Klassifizierbarkeit der Geräusche zu erhalten.

Die visuelle Analyse erfolgte durch genaues Betrachten der erstellten Spektrogramme. Dabei wurden Muster und Merkmale identifiziert, die auf unterschiedliche Geräuschtypen hindeuteten. Diese Analyse basierte auf subjektiven Beobachtungen und meiner eigenen Expertise auf dem Gebiet der Geräuscherkennung.

Die Ergebnisse dieser visuellen Analyse führten zur Identifizierung verschiedener Arten von Geräuschen in den Spektrogrammen. Markante Muster und Charakteristika wurden identifiziert und dokumentiert.

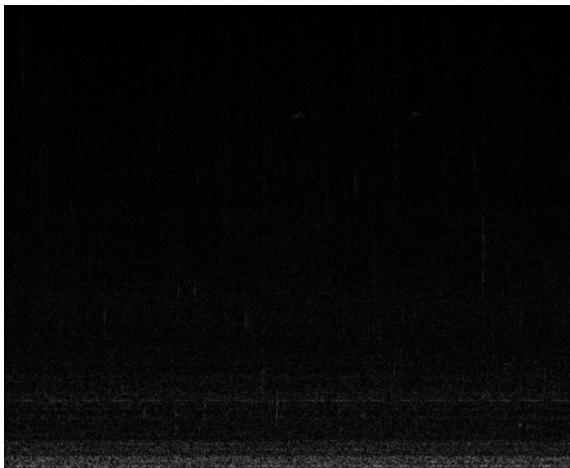


Abbildung 7 Normale Pumpe

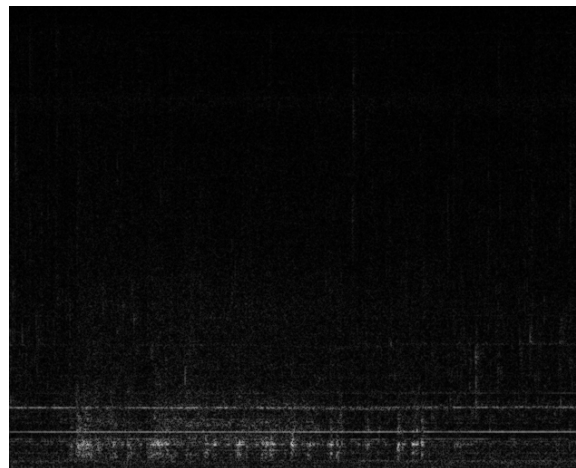


Abbildung 6 Anomale Pumpe

Die beiden Spektrogramme veranschaulichen die Unterschiede zwischen normalen und anomalen Maschinengeräuschen. Das Spektrogramm im normalen Zustand (Abbildung 7) zeigt eine regelmäßige und kontinuierliche Energieverteilung, was auf einen stabilen Betriebszustand hinweist. Im Gegensatz dazu weist das Spektrogramm im anomalen Zustand (Abbildung 6) unterbrochene und variable Frequenzmuster auf, die mögliche Störungen oder Anomalien anzeigen. Diese optische Unterscheidung ist ein grundlegender Schritt für die manuelle Klassifikation der Zustände.

## 5.2 Transfer Learning mit dem YAMNet Modell

### 5.2.1 Einführung in das YAMNet Modell

Im zweiten Projektabschnitt lag der Fokus auf der Erkundung des Potenzials des Transfer Learnings im Hinblick auf die Klassifikation von Maschinengeräuschen unter Nutzung des YAMNet-Modells. YAMNet ist ein vortrainiertes Deep-Learning-Modell, das von Google entwickelt wurde und die Abkürzung für "Yet Another MobileNet" steht.

Die Architektur von YAMNet beruht auf der MobileNetV1-Struktur, die für ihre geringen Rechenanforderungen und schnelle Verarbeitungszeiten bekannt ist. Diese Eigenschaften machen es ideal für den Einsatz in Umgebungen mit begrenzten Ressourcen wie beispielsweise auf mobilen Geräten. Das YAMNet-Modell wurde speziell für die Analyse von Audiodaten entwickelt und hat sich als effektiv bei der Klassifikation verschiedener Umgebungsgeräusche erwiesen.

Es wurde mit dem umfangreichen Datensatz AudioSet-YouTube trainiert, der ein breites Spektrum an Umgebungsgeräuschen repräsentiert. Diese gründliche Schulungsphase sorgt dafür, dass YAMNet eine bemerkenswerte Genauigkeit bei der Klassifizierung von Audioereignissen aufweist, was es zu einer vielversprechenden Grundlage für die Klassifizierung von Maschinengeräuschen in diesem Projekt macht.

### 5.2.2 Anwendung des YAMNet Modells

#### 5.2.3 CSV-Datenintegration zur Erleichterung des Transfer Learnings

#### 5.2.4 Anwendung des Pump-Datensatzes

#### 5.2.5 Bewertung der Ergebnisse und Entscheidung zur Neuausrichtung

## 5.3 Entwicklung eines eigenen Modells

### 5.3.1 Auswahl und Erprobung verschiedener Modelle

### 5.3.2 Implementierung der Klassifikationsanwendung

### 5.3.3 Binäre Klassifikation mit dem Pump-Datensatz

### 5.3.4 Multiklassen-Klassifikation mit dem PT500-Datensatz

### 5.3.5 Hyperparameter-Optimierung und Cross-Validation



## 6 Ergebnisse und Diskussion

## 7 Fazit und Ausblick

## Literaturverzeichnis

Bhatia, N. (2020). *Using transfer learning, spectrogram audio classification, and MIT app inventor to facilitate machine learning understanding* [Thesis, Massachusetts Institute of Technology]. <https://dspace.mit.edu/handle/1721.1/127379>

Cruz-González, C. E., Catalán-Catalán, C. A., Torres-Arellano, M., & Canales-Siller, H. (2023). Cutting tool wear and work hardening monitoring through cutting sound classification and machine learning in 304 stainless steel. *MRS Advances*, 8(2), 52–58. <https://doi.org/10.1557/s43580-023-00506-4>

Fulop, S. A., & Fitz, K. (2006). Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications. *The Journal of the Acoustical Society of America*, 119(1), 360–371. <https://doi.org/10.1121/1.2133000>

García, S., Luengo, J., & Herrera, F. (2015). *Data Preprocessing in Data Mining* (Bd. 72). <https://doi.org/10.1007/978-3-319-10247-4>

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>

GUNT. (2014). *PT 500.12 Zubehörsatz Schäden bei Wälzlagern*.

GUNT. (2023). *PT500*. <https://www.gunt.de/de/produkte/system-zur-maschinendiagnose-basiseinheit/052.50000/pt500/glct-1:pa-119:pr-1022>

Ji, X., Jiang, K., & Xie, J. (2021). LBP-based bird sound classification using improved feature selection algorithm. *International Journal of Speech Technology*, 24(4), 1033–1045. <https://doi.org/10.1007/s10772-021-09866-4>

Khandelwal, R. (2022, Juli 20). Supervised, Unsupervised, and Reinforcement Learning. *Medium*. <https://arshren.medium.com/supervised-unsupervised-and-reinforcement-learning-245b59709f68>

Kihei, B., Copeland, J. A., & Chang, Y. (2017). Automotive Doppler sensing: The Doppler profile with machine learning in vehicle-to-vehicle networks for road safety. *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 1–5. <https://doi.org/10.1109/SPAWC.2017.8227637>

- Kinsler, L. E., Frey, A. R., & Mayer, W. G. (1963). Fundamentals of Acoustics. *Physics Today*, 16(8), 56–57. <https://doi.org/10.1063/1.3051072>
- Konatham, S. R., Chatellus, H., & Azana, J. (2020). Photonics-Based Real-Time Spectrogram Analysis of Broadband Waveforms. *Journal of Lightwave Technology*, PP, 1–1. <https://doi.org/10.1109/JLT.2020.3010516>
- Lu, J. (2022). *A rigorous introduction to linear models* (arXiv:2105.04240). arXiv. <https://doi.org/10.48550/arXiv.2105.04240>
- Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
- Piczak, K. J. (2015). Environmental sound classification with convolutional neural networks. *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*, 1–6. <https://doi.org/10.1109/MLSP.2015.7324337>
- Piczak, K. J. (2023). *Karolpiczak/ESC-50* [Python]. <https://github.com/karolpiczak/ESC-50> (Original work published 2015)
- Purohit, H., Tanabe, R., Ichige, K., Endo, T., Nikaido, Y., Suefusa, K., & Kawaguchi, Y. (2019). *MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection* (public 1.0) [dataset]. Zenodo. <https://doi.org/10.5281/zenodo.3384388>
- Quinlan, J. R. (2014). *C4.5: Programs for Machine Learning*. Elsevier.
- Ruder, S. (2017). *An overview of gradient descent optimization algorithms* (arXiv:1609.04747). arXiv. <https://doi.org/10.48550/arXiv.1609.04747>
- Salamon, J., & Bello, J. P. (2017). Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification. *IEEE Signal Processing Letters*, 24(3), 279–283. <https://doi.org/10.1109/LSP.2017.2657381>
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45, 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Tensorflow. (2023). *Sound classification with YAMNet | TensorFlow Hub*. TensorFlow. <https://www.tensorflow.org/hub/tutorials/yamnet>

Toffa, O. K., & Mignotte, M. (2021). Environmental Sound Classification Using Local Binary Pattern and Audio Features Collaboration. *IEEE Transactions on Multimedia*, 23, 3978–3985. <https://doi.org/10.1109/TMM.2020.3035275>

transcriptaze. (2023). *Wav2png* [Go]. <https://github.com/transcriptaze/wav2png> (Original work published 2020)



## 8 Anlagen

## 9 Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass

- ich die vorliegende wissenschaftliche Arbeit selbstständig und ohne unerlaubte Hilfe angefertigt habe,
- ich andere als die angegebenen Quellen und Hilfsmittel nicht benutzt habe,
- ich die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe,
- die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfbehörde vorgelegen hat.

Berlin, 05.09.2023

Unterschrift: \_\_\_\_\_