



# Einfluss verschiedener Merkmalsextraktionen auf die KI-basierte prädiktive Instandhaltung

Lisa Schneeweiß, Constantin Pohl, Martin Golz

Hochschule Schmalkalden, Fakultät Informatik, Blechhammer 4, 98573 Schmalkalden

## Problem

- Ziel der Merkmalsextraktion
  - möglichst wenige, charakteristische Merkmale aus den Signalen gewinnen
- Bewertung: keine analytische Methode bekannt
- Empirische Bewertung
  - Basierend auf Stichprobe
  - Basierend auf heuristisch gewählten Methoden
- Merkmalsextraktion hat signifikanten Einfluss auf die mittlere Klassifikationsgenauigkeit
- Data-Leakage-Problem beeinflusst die Wahl der Kreuzvalidierungsmethode
  - Falsche Methode führt zu Verzerrung der Klassifikationsergebnisse

## Ziel

- Auswahl geeigneter Merkmale
- Kriterien
  - Erreichbare Klassifikationsgenauigkeiten
  - Echtzeitfähigkeit
- Kombination aus Merkmalsextraktion und die zum Problem passende Kreuzvalidierungsmethode mit maximaler Klassifikationsgenauigkeit finden

## Datenerfassung

- Signale von zwei vibroakustischen Sensoren an einer Reinstwasseranlage
- Prädiktive Instandhaltung durch frühzeitige Erkennung der Degeneration eines Filters in der Anlage

## Daten

- 2 Klassen Problem
  - 48 Betriebsstunden vor einem Filterwechsel
  - 48 Betriebsstunden nach einem Filterwechsel
- Segmentierung
  - Segmentlänge: 60s
  - Anzahl Segmente: 2.880 ( $2.880 \cdot 60 \text{ s} = 48 \text{ h}$ )
- Insgesamt 5 Filterwechsel
  - 240 Stunden Material eines neuen Filters
  - 240 Stunden Material eines degenerierten Filters

## Angewandte Merkmalsextraktionsmethoden

- Entropie und fraktale Dimensionen
- LogPSD
- Momentanfrequenz mit Binning
- Perzentile
- Wavelet

## Angewandte Skalierungen

- Rohsignal ohne Skalierung
- Min-Max Skalierung
- z-Skalierung
- Quantiltransformation

## Angewandte Kreuzvalidierungen

- Repeated random sub-sampling
  - 5x 80% Trainingsmenge / 20% Testmenge
- Leave-one-subject-out

## Untersuchungsgegenstand

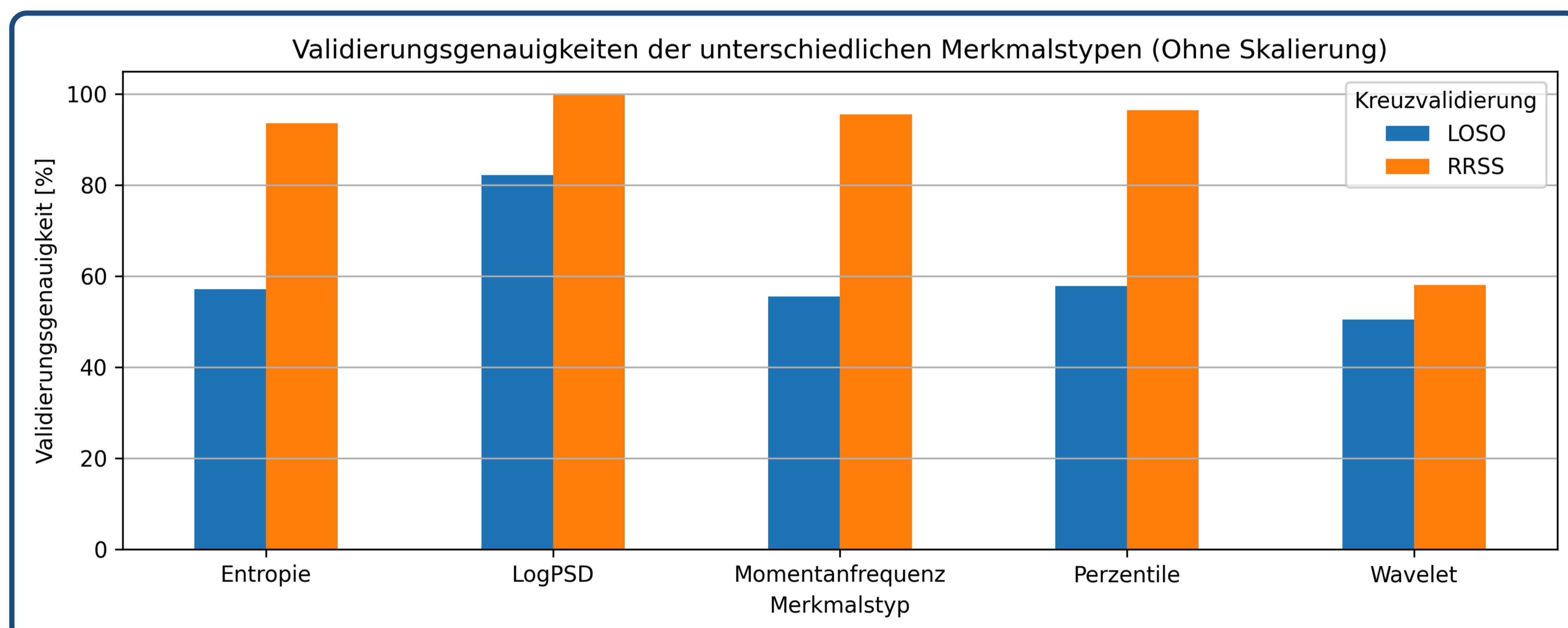
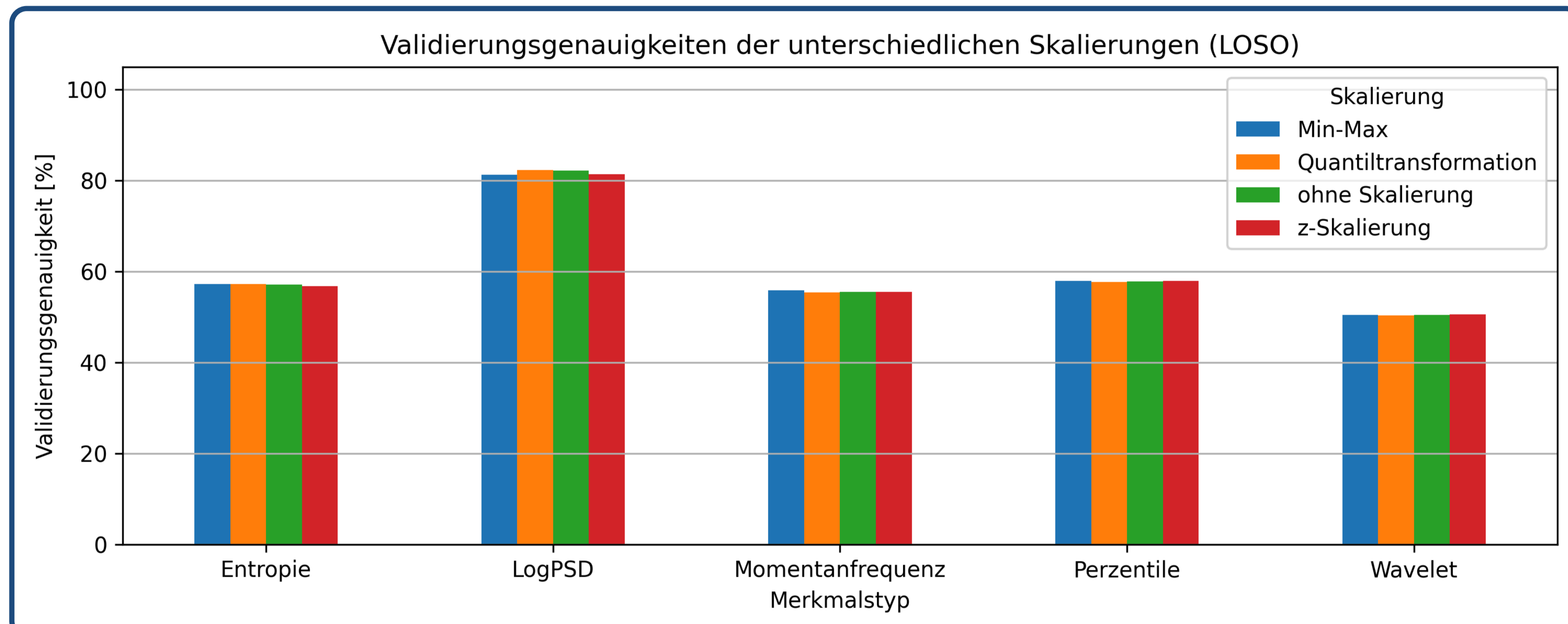
- Mikroelektronik-Industrie: Reinstwasser-Aufbereitung
  - Basierend auf Umkehr-Osmose-Aggregaten
- Mindestdurchsatzmenge muss garantiert sein
- Risiko
  - Vorfilterstufen können degenerieren
  - Zeiträume: schwer vorhersagbar
  - Keine direkte Beobachtbarkeit
- Ausfall der gesamten Anlage droht

## Empirische Bewertung

- Klassifikationsgenauigkeit mit Machine-Learning
  - Gradient Boosting Machine (LightGBM)
- Kriterium
  - Klassifikationsgenauigkeit an Validierungsmengen
- Vergleich verschiedener Merkmalstypen, Skalierungsmethoden sowie Kreuzvalidierungsmethoden

## Merkmalstypen

- Drei Entropie-Maße
  - drei Maße der fraktalen Dimension
  - trendeliminierte Fluktuation
  - Poincaré-Parameter
  - Lyapunov-Exponent
- Logarithmierte spektrale Leistungsdichten
  - in 248 Spektralbändern, direkt geschätzt mittels diskreter Fourier-Transformation
- Hilbert-Transformation
  - Momentan-Frequenz, sowie Mittelung der Transformierten in 300 Kurz-Zeit-Intervallen
- 29 Perzentile der Amplitudenverteilung
  - im Zeitbereich, ohne Berücksichtigung der Zeitabhängigkeit
- Diskrete Wavelet-Transformation
  - Wavelet-Entropie über 13 Koeffizientenfolgen



## Ergebnisse

- Merkmalstypen beeinflussen die Val.-Genauigkeit
- Kreuzvalidierungen beeinflussen die Val.-Genauigkeit
- Skalierung hat bei den vorliegenden Daten minimalen Einfluss auf die Val.-Genauigkeit
- RRSS führt bei den getesteten Merkmalstypen immer zu einer höheren Genauigkeit als LOSO
  - Strikte Trennung der Filterwechsel ist notwendig
- LogPSD führt zu der besten RRSS sowie LOSO Val.-Genauigkeit (82.23%)

## Diskussion

- Skalierung hat wenig Einfluss, weitere Untersuchungen können ohne Skalierung fortgesetzt werden
- Unterschied zwischen RRSS und LOSO zeigt das ein Data-Leakage Problem vorliegt
  - Es muss eine Kreuzvalidierungsmethode mit strikter Trennung der Subjekte für die Validierung genutzt werden
- Optimierung der LogPSD Merkmalsextraktion könnte Val.-Genauigkeit positiv beeinflussen