

# Masterarbeit: Identifizierung kleinster Peaks und Einzelevents im Massenspektrum mittels Machine Learning

## Hintergrund / Motivation

Die Hochleistungsflüssigchromatographie (HPLC) gekoppelt mit der Massenspektrometrie (MS) ist eine essenzielle Technik zur Analyse komplexer Proben. Ein zentraler Bestandteil der Auswertung ist die Interpretation der Peaks in den Spektren, die je nach Zielsetzung auf der Identifikation bekannter Signale an bestimmten Positionen basiert. In vielen Fällen erfordert dies jedoch die manuelle Analyse großer Spektren durch erfahrene Experten – eine zeitaufwendige und anspruchsvolle Aufgabe, insbesondere bei uneindeutigen oder kleinsten Peaks.

Um diesen Prozess effizienter und skalierbarer zu gestalten, bietet sich eine Automatisierung mittels maschinellen Lernens an. Gut annotierte Trainingsdaten mit einer klaren Trennung zwischen Eingabe (Spektrum) und erwarteter Ausgabe (Interpretation) sowie die Ähnlichkeit zur Mustererkennung legen den Einsatz von Klassifikationsmodellen nahe.

Im Rahmen dieser Masterarbeit soll erforscht werden, wie maschinelles Lernen zur Automatisierung der Spektrenanalyse beitragen kann. Dabei stehen sowohl die Modellwahl als auch die Optimierung der Datenaufbereitung und Trainingsstrategie im Fokus.

## Aufgabenpakete

1. **Literaturrecherche:**
  - Einarbeitung in die theoretischen Grundlagen der HPLC-MS-Technik.
  - Überblick über aktuelle Methoden zur Peak-Erkennung und -Quantifizierung im Massenspektrum
  - Untersuchung bestehender Ansätze im Bereich maschinelles Lernen und künstliche Intelligenz für die Auswertung von Massenspektren
2. **Datenanalyse:**
  - Analyse und Aufbereitung der vorhandenen Massenspektren
  - Identifikation relevanter Merkmale zur Peak-Erkennung
3. **Modellentwicklung:**
  - Auswahl geeigneter maschineller Lernverfahren zur Peak-Erkennung
  - Training und Validierung des Modells mit den vorhandenen Daten
  - Optimierung der Modellparameter zur Verbesserung der Erkennungsrate
4. **Implementierung und Test:**
  - Implementierung des Algorithmus in einer geeigneten Programmiersprache (z.B. Python, PyTorch)
  - Test und Validierung des Algorithmus anhand neuer, unabhängiger Datensätze
  - Vergleich der Ergebnisse mit manuell ausgewerteten Daten inklusive Unterscheidung von false positives im Untergrund
5. **Dokumentation und Präsentation:**
  - Ausführliche Dokumentation der durchgeführten Arbeiten und Ergebnisse

- Präsentation der Ergebnisse

### **Anforderungen an den/die Bewerber/in:**

- Abgeschlossenes Bachelorstudium in Chemie, Bioinformatik, Informatik oder einem verwandten Fachgebiet.
- Grundkenntnisse in HPLC und Massenspektrometrie.
- Erfahrung in der Programmierung, vorzugsweise in Python.
- Grundkenntnisse in maschinellem Lernen und Datenanalyse.
- Selbstständige und strukturierte Arbeitsweise.

### **Betreuung**

- Invite GmbH
- Merck
- Universität

### **Zeitrahmen**

- ASAP
- Für 6 Monate

### **Ansprechpartner**

Maja Diebig-Lorenz

[Info@invite-research.com](mailto:Info@invite-research.com)