

Pharmaziepraktikum/Master-Projekt: Identifizierung kleinster Peaks und Einzelevents in Massenspektren mittels Machine Learning

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 Arbeit soll erforscht werden, wie maschinelles Lernen zur Automatisierung der Spektrenanalyse beitragen kann. Hier stehen sowohl die Modellwahl als auch die Optimierung der Datenaufbereitung und Trainingsstrategie im Fokus.

Ihre Aufgaben (Zeitraumen: 6 Monate):

1. **Literaturrecherche:**
 - Einarbeitung in die theoretischen Grundlagen der HPLC-MS-Technik.
 - Überblick über aktuelle Methoden zur Peak-Analyse 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 (z.B. in Python oder R)
 - Test und Validierung des Algorithmus anhand neuer, unabhängiger Datensätze
 - Vergleich der Ergebnisse mit manuell ausgewerteten Daten
5. **Dokumentation und Präsentation:**
 - Ausführliche Dokumentation und Präsentation der Arbeiten und Ergebnisse

Ihr Profil:

- Studium in Pharmazie, Chemie, Bioinformatik, Informatik oder einem verwandten Fachgebiet.
- Grundkenntnisse in **HPLC und Massenspektrometrie**
- **Interesse und/oder Erfahrung** in der Programmierung (Python/R), **maschinellernem Lernen**, oder der **Datenanalyse**
- Selbstständige und strukturierte Arbeitsweise

Wir bieten:

- Interdisziplinäre Betreuung durch die **Universität, INVITE und Merck**
- Zusammenarbeit mit einem **HPLC-MS Expertenteam bei Merck (Darmstadt)**
- **Attraktive Vergütung** während des Projektes

Interesse?

Ihr Ansprechpartner:

Maja Diebig-Lorenz
diebig-lorenz@invite-research.com