

KI in der Metabolomics-Forschung

Implementierung einer automatisierten Bildauswertung von GCxGC-qMS-Daten

- In der Metabolomics-Forschung wird eine Vielzahl von Proben analysiert, um Erkenntnisse über den physischen Zustand von Menschen und Pflanzen zu erhalten
- Die implementierten KI-Verfahren helfen bei der automatisierten Auswertung über mehrere Proben hinweg durch eine Harmonisierung der Bildhintergründe, Identifikation von einzelnen Signalen in dichten, überlappenden Peak-Regionen und das Erkennen gleicher Signale in unterschiedlichen Proben
- Dadurch wird eine effiziente und reproduzierbare Auswertung der Bilder ermöglicht

Hintergrund und Fragestellung

Metabolite sind **niedermolekulare Stoffwechselprodukte** aus dem Abbau von Kohlenhydraten, Lipiden und Proteinen. Die Metabolomik ermöglicht die **Analyse** vieler **bekannter und unbekannter Stoffe**, die Aussagen über den Zustand von Menschen und Pflanzen geben können. Damit lassen sich sowohl relative Konzentrationsverhältnisse als auch potenzielle neue Marker für Regulationsmechanismen, Stressantworten und Biomarker identifizieren.

Die **manuelle Auswertung** der entstehenden GCxGC-qMS-Datensätze ist jedoch **zeitintensiv** und limitiert die Skalierbarkeit. Ziel dieses Projekts ist daher der **Einsatz KI-basierter Bildauswertung**, um relevante Signalregionen in zweidimensionalen GC-Darstellungen automatisiert zu erfassen und für weiterführende Analysen nutzbar zu machen.

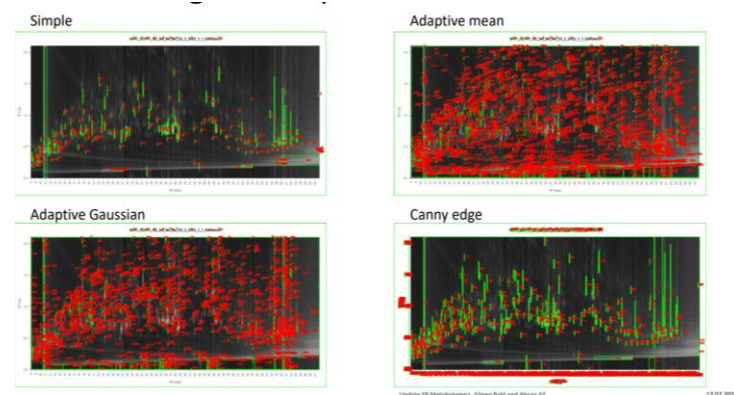
Vorgehensweise

Zur automatisierten Analyse zweidimensionaler GC-Bilddaten wird in diesem Projekt ein mehrstufiges, KI-gestütztes **Bildverarbeitungs- und Clusteringverfahren** entwickelt, um eine robuste Detektion, Segmentierung und konsistente Identifikation metabolischer Signale über mehrere Messungen hinweg zu erreichen.

Zunächst erfolgt eine **globale Hintergrundanpassung**, um systematische Intensitätsunterschiede zwischen einzelnen GC-Aufnahmen auszugleichen. Ergänzend wird eine **lokale Hintergrundkorrektur** angewendet, die langsame Intensitätsdrifts und lokale Artefakte innerhalb einzelner Bilder entfernt. Dadurch werden metabolische Peaks gegenüber dem Hintergrund deutlich hervorgehoben.

Auf Basis der bereinigten Signalintensitäten werden **relevante Peak-Regionen detektiert** und mithilfe von **Watershed-Segmentierung** voneinander getrennt. Dies erlaubt die Auflösung überlappender oder dicht benachbarter Signale.

Zur einheitlichen Identifikation von Metaboliten über mehrere Bilder wurde ein **k-Means-Clustering** verwendet. So erhalten korrespondierende Signale unterschiedlicher Messungen konsistente IDs. Die resultierenden Peak-Informationen stehen für anschließende **statistische Auswertungen** sowie **biologische Interpretationen** zur Verfügung.



Ergebnisse und Schlussfolgerungen

Die Kombination aus Bildverarbeitung und maschinellem Lernen ermöglicht eine **effiziente und reproduzierbare Auswertung** ungerichteter GC-Bilddaten. Dadurch können sowohl bekannte als auch bisher unbekannte Metaboliten systematisch erfasst werden.

Die automatisierte Analyse **verbessert** nicht nur die **Auswertungsgeschwindigkeit** und **Trefferquote**, sondern schafft auch die Grundlage zur Identifikation neuer Biomarker.

Kontakt	Informationen
KI-Beratung (KIDA): kida@bmleh.bund.de	KIDA-Bearbeitende: Aileen Bahl (KIDA, BfR), Ahsan Ali (KIDA, BfR) Leitung KI-Beratung (KIDA): Micha Schneider (KIDA, Thünen)
Björn Egert (MRI): bjoern.egert@mri.bund.de	Anfragender: Björn Egert (MRI)