

# Medical Image Analysis: Multi Instance Learning for Whole-Slide Image Classification in Pathology

## Hintergrund:

Multi Instance Learning (MIL) ist eine fortgeschrittene Methode im maschinellen Lernen, die es ermöglicht, aus Gruppen von Instanzen zu lernen, anstatt nur aus einzelnen Datenpunkten. Im Kontext der medizinischen Bildanalyse und Bildklassifizierung ermöglicht MIL die effektive Verarbeitung von Whole-Slide Images (WSI), die große Bereiche von Gewebeproben abdecken.

## Anwendungsfall:

Ovarialkarzinom ist der tödlichste Krebs des weiblichen Fortpflanzungssystems. Es gibt fünf häufige Subtypen von Eierstockkrebs: hochgradiges seröses Karzinom, klarzelliges Eierstockkarzinom, endometrioides, niedriggradiges seröses und muzinöses Karzinom. Zusätzlich gibt es mehrere seltene Subtypen (“Outliers”). Diese werden durch unterschiedliche zelluläre Morphologien, Ätiologien, molekulare und genetische Profile sowie klinische Merkmale charakterisiert. Die Identifizierung der Subtypen ist entscheidend für eine gezielte Therapie, was jedoch eine Verbesserung durch Data Science erfahren könnte.

Derzeit erfolgt die Diagnose von Eierstockkrebs durch Pathologen, um die Subtypen zu bewerten. Dies birgt jedoch mehrere Herausforderungen, darunter Meinungsverschiedenheiten zwischen Beobachtern und die Reproduzierbarkeit von Diagnosen. Darüber hinaus fehlt es in unterversorgten Gemeinschaften oft an Zugang zu spezialisierten Pathologen, und selbst gut entwickelte Gemeinschaften haben einen Mangel an Pathologen mit Expertise in gynäkologischen Malignitäten.

Deep-Learning-Modelle haben eine bemerkenswerte Effizienz bei der Analyse von Histopathologiebildern gezeigt. Dennoch gibt es Herausforderungen, wie die Notwendigkeit einer erheblichen Menge an Trainingsdaten, idealerweise aus einer einzigen Quelle. Technische, ethische und finanzielle Einschränkungen sowie Vertraulichkeitsbedenken erschweren das Training. In dieser Abschlussarbeit wird der UBC-OCEAN Datensatz hergenommen, welcher zu dem umfangreichsten und vielfältigsten Datensätzen von Histopathologiebildern des Eierstockkrebses zählt und mehr als 20 Zentren auf vier Kontinenten beigetragen haben.

## Aufgabenstellung:

Im Rahmen dieser Masterarbeit liegt der Fokus auf der Entwicklung einer dynamischen Multi-Instance-Learning-Pipeline für die Klassifizierung von Whole-Slide Images im Kontext des UBC-OCEAN-Anwendungsfalls. Die Hauptaufgaben umfassen die Einarbeitung in MIL für die medizinische Bildanalyse, das Studium der relevanten Frameworks, die Entwicklung einer MIL-Pipeline für

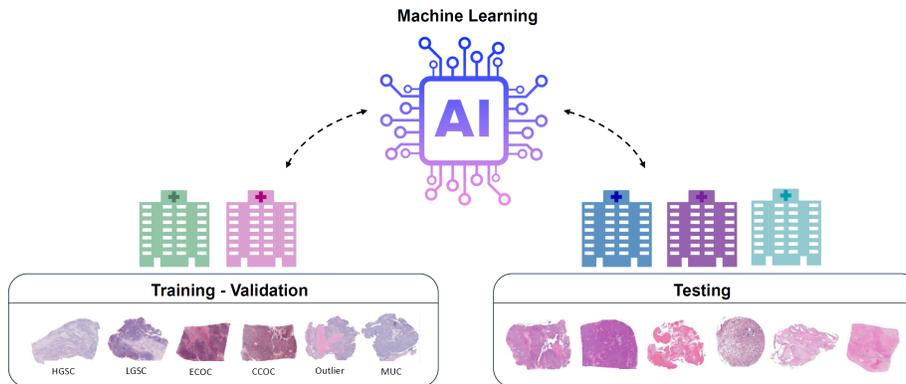


Figure 1: UBC-OCEAN

WSI, die Anwendung, Benchmarking und Evaluierung der Leistung der entwickelten Pipeline.

#### Pipeline Struktur und Methodik:

- Whole-Slide Image (WSI) wird in Patches von 224x224 zerlegt.
- Anwendung von TransPath / Phikon / HistoCartography zur Extrahierung von Pathology Features der jeweiligen Patches.
- Entwicklung eines tiefen Multi-Instance Learning Modells zur Klassifizierung für eine komplette WSI
- Zusammenführung der einzelnen Methoden zu einer dynamischen Pipeline (Git, Dokumentation, ggf. Continuous Integration -> Ziel ist hierbei die Bereitstellung eines open-source Tools für die Community)
- Anwendung, Benchmarking und Evaluierung der Leistung der entwickelten Pipeline

#### Anforderungen:

- Studierende/r im Bereich Informatik, Medizininformatik, Bioinformatik oder einem verwandten Studiengang.
- Programmierkenntnisse in Python sind erforderlich.
- Kenntnisse in Bildverarbeitung sind von Vorteil, jedoch nicht zwingend notwendig.
- Interesse an medizinischen Anwendungen und der Analyse von medizinischen Bildern ist erwünscht.

#### Dauer und Betreuung:

Die Arbeit wird auf eine Dauer von 3 (Bachelorarbeit) bis 6 (Masterarbeit) Monaten ausgelegt. Während der gesamten Arbeit steht Ihnen ein Betreuer zur Seite.

Falls Sie Interesse an dieser Abschlussarbeit haben, freuen wir uns über Ihre Bewerbung. Bitte senden Sie Ihren Lebenslauf und ein kurzes Motivationsschreiben an: dominik.mueller@uni-a.de

Für Rückfragen stehen wir Ihnen gerne zur Verfügung.

Hinweis: Der genaue Inhalt und die Aufgabenstellung können in Absprache mit den Betreuern noch weiter verfeinert werden, um die individuellen Interessen und Fähigkeiten der Studierenden zu berücksichtigen.

**Literatur:**

- TransPath Framework
- HistoSSLscaling Framework
- HistoCartography Framework
- AttentionDeepMIL Framework
- Keras AttentionDeepMIL Beispiel
- UBC-OCEAN Kaggle-Wettbewerb
- UBC-OCEAN Kaggle-Wettbewerb Gewinner