

Jonas Scherer
Dr. sc. hum.

Decentralized Infrastructure for Medical Image Analysis - The Development and Establishment of Kaapana as an Open Framework for Imaging Platforms in Clinical Environments

Fach/Einrichtung: Deutsches Krebsforschungszentrum (DKFZ)
Doktorvater: Prof. Dr. rer. nat. Klaus H. Maier-Hein

Das Aufkommen neuer daten- und algorithmengestützter Analysemethoden revolutioniert derzeit viele Forschungsbereiche und ermöglicht Lösungen für Probleme, die früher als unlösbar galten. Aber gilt dies auch für die Medizin?

Die Verbesserung der Diagnose, die Feincharakterisierung von Patienten für die personalisierte Medizin, die Überwachung des Krankheitsverlaufs und seiner Prognose oder die Vorhersage der besten Therapie sind nur einige Beispiele, die potenziell von solchen neuen Analysetechniken profitieren könnten. Algorithmen ermöglichen es Maschinen nun, visuelle Bilddaten besser zu verstehen und zu interpretieren, was einerseits vielversprechende Perspektiven für die medizinische Bildverarbeitung bietet.

Andererseits aber auch Herausforderungen mit sich bringt, da große Mengen an annotierten Daten benötigt werden, um ein Overfitting zu verhindern und somit robuste, generalisierende und zuverlässige Modelle zu ermöglichen. Multizentrische Bildgebungsstudien könnten die Verfügbarkeit solcher Daten erheblich verbessern und auch die Heterogenität durch die Zugänglichkeit von Trainingsdaten aus verschiedenen Standorten erhöhen. Die gemeinsame Nutzung von Daten über mehrere Standorte hinweg hat sich jedoch aufgrund des hohen Datenschutzniveaus, das mit medizinischen Daten einher geht und technischer Herausforderungen, wie der Interoperabilität, als schwierig erwiesen. In dieser Dissertation wird daher versucht, die Notwendigkeit eines Datenaustauschs zu vermeiden, indem ein anderer Ansatz verfolgt wird: **"Lasst uns die Algorithmen teilen, nicht die Daten!"**

Ziel und zentrale Forschungsfrage ist dabei, ob die Evaluation und das Training moderner Bildanalyseverfahren in die Kliniken verlagert werden kann. Dieser Ansatz kann zwar den Datenexport und die damit verbundenen datenschutzrechtlichen Herausforderungen umgehen, allerdings müssen die beteiligten Partner auch technisch und organisatorisch dazu in der Lage sein, diese Aufgaben zu erfüllen. Zu diesem Zweck wurden in dieser Dissertation Konzepte, sowie die Entwicklung und Etablierung einer dezentralen Infrastruktur für die klinische medizinische Bildanalyse untersucht, die einen standardisierten Datenzugriff und eine einheitliche Umgebung für die Ausführung von Analysemethoden im Rahmen von multizentrischen Bildgebungsstudien ermöglichen.

Die technische Realisierung dieser Infrastruktur wurde durch die Entwicklung eines Software-Frameworks namens Kaapana erreicht, welches den Aufbau von technischen Plattformen für die Bildanalyse ermöglicht. Die resultierende Software kann auf dedizierten

Servern innerhalb der klinischen IT-Umgebung betrieben werden, um isoliert von jeglicher externer Konnektivität betrieben und mit anderen lokalen klinischen Systemen verbunden zu werden. Durch die Nutzung moderner Cloud-Technologien wie Containern und Kubernetes bietet die lokale Bereitstellung eine private Cloud, auf die von den klinischen Arbeitsstationen über den Webbrowser zugegriffen werden kann. Die standardisierte Anbindung an das klinische PACS über DICOM und die Integration eines Forschungs-PACS ermöglichen eine redundante Datenhaltung und einen konsistenten Datenzugriff für die Durchführung von Analyseverfahren. Dashboards ermöglichen eine effiziente Datenexploration und -filterung, indem sie die DICOM-Metadaten visuell darstellen und über Suchanfragen selektierbar machen. Eine einheitliche Ausführungsumgebung für die Datenverarbeitung ermöglicht es, Algorithmen anzuwenden und sie einheitlich zu verpacken und mit Partnern zu teilen. Leistungsstarke Server-Hardware, einschließlich Grafikkarten, ermöglicht eine Vielzahl von Analyseverfahren wie Deep Learning-basierte Modellinferenz sowie das Training neuer Modelle. Innerhalb der Infrastruktur ist DICOM das grundlegende Datenformat, so dass auch viele Analyseergebnisse standardkonform bereitgestellt werden können. Durch die Verwendung von Formaten wie DICOM SEG, SR oder Encapsulated PDF werden auch Annotationen mit klinischen Arbeitsabläufen und IT-Systemen kompatibel gemacht.

Mit Unterstützung des Radiological Cooperative Network (RACOON) und des Deutschen Konsortiums für Translationale Krebsforschung (DKTK), konnten zwei große deutsche Forschungskonsortien für die Evaluierung gewonnen werden. Dabei haben inzwischen alle 36 Universitätskliniken in Deutschland einen eigenen Server in Betrieb genommen, auf dem die Plattform installiert und getestet wurde. Hierbei wurden die Inbetriebnahme, die Integration, der Betrieb und die Wartung eines solchen dezentralen Netzwerks sowie verschiedene Anwendungsfälle evaluiert, die die typischen Aufgaben eines solchen Systems abdecken. Im Ergebnis konnten diese Anwendungsfälle und die damit verbundenen unterschiedlichen Anforderungen mit dem entwickelten Konzept gut realisiert werden. Ein flexibler Erweiterungsmechanismus erlaubt zudem die Integration von zusätzlichen Diensten wie der MITK-Workbench oder Algorithmen wie dem nnUNet. Darüber hinaus konnten auch bereits erste von externen Partnern entwickelte Analysepipelines integriert und ausgeliefert werden.

Im Rahmen dieser Arbeit wurden die Kliniken durch die Entwicklung, Etablierung und Unterstützung der entwickelten Infrastruktur in die Lage versetzt, aktuelle Forschungsmethoden auf ihre eigenen Datenbestände anzuwenden. Da die Forschungskonsortien, an denen bereits alle deutschen Universitätskliniken beteiligt sind, ihre Tätigkeit gerade erst aufgenommen haben, ergeben sich vielfältige Möglichkeiten, die hochwertigen Daten der Partnerstandorte für die Forschung in Zukunft zugänglich und nutzbar zu machen. Aufgrund des offenen Quellcodes von Kaapana, seiner auf gängigen Industriestandards basierenden Architektur und der bereits weiten Verbreitung könnte dieses Framework auch als Basis für die Datenverarbeitung in anderen Forschungsbereichen Anwendung finden. Somit könnte auch eine verbesserte Datenintegration für klinische Datenverarbeitung im Allgemeinen unterstützt werden.