

# Entwurf eines Frameworks zur Bewertung der Datenqualität in intensivmedizinischen Datenbanken am Beispiel ARDS

(Bachelorarbeit)



CHARLOTTE WOLF

## Motivation

Auf Intensivstationen werden kontinuierlich Patientendaten gesammelt und aufgezeichnet. In der Forschung werden solche Datensätze als Sekundärdaten – nicht primär für diese Zwecke erhobene Daten – genutzt, da häufig keine anderen Daten vorliegen. Die Daten dienen zum Beispiel dem Training von Machine Learning Modellen zur Früherkennung von Krankheiten wie z.B. dem akuten Lungenversagen. Aufgrund technischer und menschlicher Faktoren auf Intensivstationen sowie der Datenmenge kommt es im Prozess der Aufzeichnung aber auch zu Inkonsistenzen. Häufig werden diese nicht korrigiert oder bemerkt, da der Fokus primär auf der Behandlung der Patienten liegt. Die Nutzbarkeit der Forschungsergebnisse ist jedoch stark von der Qualität der genutzten Daten abhängig. Eine

Einschätzung der Datenqualität ist also essentiell. Im Kontext des ASIC-Use Cases des SMITH-Projektes und in Kooperation mit dem Uniklinikum Aachen soll in dieser Bachelorarbeit eine effiziente Einschätzung ermöglicht werden.

## Stand der Technik

Es gibt bereits verschiedene Standardisierungsansätze zur Datenqualitätsbewertung, wie zum Beispiel ISO Standards oder die Technologie- und Methodenplattform für die vernetzte medizinische Forschung (TMF). Die gewählten Kategorien der Datenqualität sind abhängig von der Art der Daten – je nach Krankheitsbild oder Studie werden bestimmte Kategorien priorisiert. Auch existieren bereits Frameworks, die sich mit der Qualität verschiedener Datenbanken beschäftigen. Allerdings gibt es bisher keine eindeutigen Metriken für Kategorien oder die genaue Bedeutung der Terminologien selbst. Ein erster Entwurf zur Plausibilitätsbewertung einer intensivmedizinischen Datenbank wurde bereits an diesem Lehrstuhl im Rahmen einer Bachelorarbeit vorgestellt. Während diese eine gute Basis für einige Kriterien bildet, wurden andere Kriterien kaum oder noch gar nicht beachtet.

## Zielsetzung

In dieser Arbeit soll ein Framework konzipiert werden, mit dem eine Datenbank mit Routinedaten von Intensivstationspatienten bewertet werden kann. Zudem sollen die genutzten Terminologien eindeutig definiert werden. Nach dem Entwurf des Konzepts ist das zweite Ziel eine algorithmische Umsetzung der Qualitätsprüfung der Datenbank und die Implementierung einer Applikation, welche die Anzeige und Überprüfung für heterogene Datensätze und effiziente Auswertung der Daten ermöglicht.

## Geplante Vorgehensweise

Auf Basis der Literaturrecherche werden bisher veröffentlichte Ansätze und Metriken evaluiert. Danach werden Methoden zur Überprüfung der am geeignetsten Kategorien entwickelt und in einem Framework zusammengeführt. Das Framework soll die im SMITH-Projekt zur Verfügung stehenden Forschungsdatenbanken von Intensivstationspatienten verwenden und bewerten. Zuletzt soll die Bewertung zur Datenqualität der Datenbanken ausgegeben und die Applikation evaluiert werden.