

## FORSCHUNGSDATENMANAGEMENTPLAN (FDMP)

Bereich	administrative Informationen
Thema des Vorhabens	Medical Information Mart for Intensive Care (MIMIC)-IV
Kurzfassung (max. 1000 Zeichen)	Hier stellen wir den Medical Information Mart for Intensive Care (MIMIC)-IV vor, einen großen deidentifizierten Datensatz von Patienten, die in die Notaufnahme oder eine Intensivstation des Beth Israel Deaconess Medical Center in Boston, MA, eingeliefert wurden. MIMIC-IV enthält Daten von über 65.000 Patienten, die auf einer Intensivstation aufgenommen wurden, und von über 200.000 Patienten, die in der Notaufnahme aufgenommen wurden. MIMIC-IV enthält aktuelle Daten und verfolgt einen modularen Ansatz für die Datenorganisation, der die Herkunft der Daten hervorhebt und sowohl die individuelle als auch die kombinierte Nutzung unterschiedlicher Datenquellen erleichtert.
Zuwendungsempfänger und ausführende Stelle	Lorem ipsum dolor sit amet
Förderkennzeichen	XXXXXXXXXX
Förderprogramm	National Institutes of Health (grant NIH-R01-EB017205)
Projektleitung (Name, ORCID, E-Mail)	Alistair Johnson, <a href="https://orcid.org/0000-0002-8735-3014">0000-0002-8735-3014</a> , <a href="mailto:aewj@mit.edu">aewj@mit.edu</a>
Datum der Erstellung	16.03.2021 (Version v1.0)

### Erhebung der Forschungsdaten, Daten- und Metadatenstandards

#### 1. Beschreibung der Forschungsdaten, Art, Menge, Standards

MIMIC-IV enthält Daten aus den Jahren 2008-2019. Die Daten wurden von bettseitigen Metavision-Monitoren erfasst. Die Daten wurden von zwei verschiedenen klinischen Informationssystemen gesammelt: CareVue und MetaVision.

MIMIC-IV ist in Module unterteilt, die die Herkunft der Daten widerspiegeln. Derzeit gibt es fünf Module:

- hosp - Daten auf Krankenhausebene für Patienten: Labore, Mikro und elektronische Medikamentenverwaltung
- icu - Daten auf ICU-Ebene. Dies sind die Ereignistabellen, deren Struktur mit der von MIMIC-III identisch ist (Chartevents, etc.)
- ed - Daten aus der Notaufnahme
- cxr - Nachschlagetabellen und Metadaten aus MIMIC-CXR, die eine Verknüpfung mit MIMIC-IV ermöglichen
- note- deidentifizierte klinische Freitextnotizen

In MIMIC-IV v3.0 sind insgesamt 364.627 Einzelpersonen enthalten, die jeweils durch eine eindeutige `subject_id` repräsentiert werden. Diese Personen hatten 546.028 Krankenhausaufenthalte und 94.458 eindeutige Aufenthalte auf der Intensivstation.

Die Datenbank wurde in ein zeichenbasiertes, kommagetrenntes Format exportiert. Die Datendateien werden im CSV-Format (Comma Separated Value) gemäß dem RFC 4180-Standard bereitgestellt. Der Datensatz ist auch über Google BigQuery verfügbar.

Anweisungen für den Zugriff auf den Datensatz in BigQuery finden Sie in der Online-Dokumentation zu MIMIC-IV auf der Cloud-Seite [1].

## **2. Dokumentation und Speicherung, Metadaten**

Die aktuelle Dokumentation zu MIMIC-IV ist auf der MIMIC-IV-Website [2] verfügbar. Wir haben ein Open-Source-Repository für den Austausch von Code und die Diskussion über die Datenbank eingerichtet, das so genannte MIMIC Code Repository [3]. Das Code Repository bietet einen Mechanismus für die gemeinsame Diskussion und Analyse aller Versionen von MIMIC, einschließlich MIMIC-IV.

Die Speicherung erfolgt über PhysioNet. Die Module für Krankenhäuser und Intensivstationen in MIMIC-IV sind im MIMIC-IV-Projekt auf PhysioNet verfügbar.

## **Verfügbarmachung der Forschungsdaten**

### **1. Datenselektion**

Die Auswahl der Daten für die langfristige Aufbewahrung und Nutzbarmachung erfolgt nach folgenden Kriterien:

### **2. Datenarchivierung**

Nur die nach den oben erläuterten Selektionsregeln ausgewählten Daten werden archiviert.

Die Archivierung und Publikation der Daten erfolgt durch PhysioNet mit entsprechender Versionierung.

Die Daten werden für mindestens 10 Jahre zur Verfügung stehen.

### **3. Zugriff, Nutzbarmachung**

Der Zugang zu MIMIC-IV erfolgt über PhysioNet. Der Zugriff ist mit beschränktem Zugang. Um auf die Dateien zugreifen zu können, müssen die folgenden Voraussetzungen erfüllt sein: die erforderliche Ausbildung absolvieren (CITI Data or Specimens Only Research) und Unterzeichnung der Datennutzungsvereinbarung für das Projekt (DUA). Die DUA verlangt von den Nutzern, den Datensatz angemessen zu schützen, nicht zu versuchen, Personen neu zu identifizieren, die Daten nicht weiterzugeben und Probleme im Zusammenhang mit der Deidentifizierung zu melden. Danach kann der Download der Daten begonnen werden.

Es steht eine Demo zu Verfügung ohne Zugangsberechtigungen. Alle Tabellen aus MIMIC-IV wurden in den Demo-Datensatz aufgenommen. Jeder Person in MIMIC-IV wird eine eindeutige subject\_id zugewiesen. Da die subject\_id nach dem Zufallsprinzip generiert wird, führt die Reihenfolge nach subject\_id zu einer zufälligen Teilmenge von Personen. Es wurden nur Personen mit einem anchor\_year\_group-Wert von 2011 - 2013 oder 2014 - 2016 berücksichtigt, um eine Überschneidung mit MIMIC-CXR v2.0.0 zu gewährleisten. Die ersten 100 subject\_id, die die anchor\_year\_group-Kriterien erfüllten, wurden für den Demo-Datensatz ausgewählt.

### **4. Datensicherheit, Datenschutz, rechtliche Aspekte**

Die vom HIPAA vorgeschriebenen Patientenidentifikatoren wurden entfernt. Die Patientenkennungen wurden durch eine zufällige Chiffre ersetzt, was zu deidentifizierten ganzzahligen Kennungen für Patienten, Krankenhausaufenthalte und Aufenthalte auf der Intensivstation führte. Alle Daten wurden de-identifiziert, indem die Daten in einen zukünftigen Zeitraum zwischen 2100 und 2200 verschoben wurden. Strukturierte Daten wurden mithilfe von Nachschlagetabellen und Erlaubnislisten gefiltert. Falls erforderlich, wurde ein Algorithmus zur Deidentifizierung von Freitext angewendet, um PHI aus Freitext zu entfernen. Schließlich wurden Datum und Uhrzeit nach dem Zufallsprinzip mit einem in Tagen gemessenen Offset in die Zukunft verschoben. Jeder subject\_id wurde eine einzige Datumsverschiebung zugewiesen. Dadurch sind die Daten für einen einzelnen Patienten in sich konsistent. Wenn z. B. die Zeit zwischen zwei Messungen in der Datenbank in den

Rohdaten 4 Stunden betrug, dann beträgt die berechnete Zeitdifferenz in MIMIC-IV ebenfalls 4 Stunden. Umgekehrt sind unterschiedliche Patienten zeitlich nicht vergleichbar. Das heißt, dass zwei Patienten, die im Jahr 2130 aufgenommen wurden, nicht unbedingt im selben Jahr aufgenommen wurden.

Dieses Projekt wurde von den Institutional Review Boards des Beth Israel Deaconess Medical Center (Boston, MA) und des Massachusetts Institute of Technology (Cambridge, MA) genehmigt. Da das Projekt keine Auswirkungen auf die klinische Versorgung hatte und alle geschützten Gesundheitsdaten de-identifiziert wurden, wurde auf die erforderliche individuelle Einwilligung der Patienten verzichtet.

## Relevante Richtlinien

- PhysioNet Credentialed Health Data Use Agreement 1.5.0 (<https://physionet.org/content/mimiciv/view-dua/3.0/>)
- RFC 4180-Standard: <http://format.gbv.de/csv/rfc4180>

## Quellen

- [1] Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L.H., Feng, M., Ghassemi, M., ... & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific data, 3(1), 1-9.
- [2] <https://mimic.mit.edu/>
- [3] <https://github.com/MIT-LCP/mimic-code/>