

Studienprotokoll

-Nur für Lehrzwecke -

Erstellt im Rahmen des MI-Lab für den MIMIC IV Datensatz als synthetisches Studiendokument

1. Hintergrund und Zielsetzung der Studie

In den letzten Jahren gab es eine konzertierte Aktion zur Einführung digitaler Patientendatensysteme in Krankenhäusern. In den USA verfügten im Jahr 2015 fast 96 % der Krankenhäuser über ein digitales elektronisches Patientendatensystem (EHR) [1]. Retrospektiv erhobene medizinische Daten werden zunehmend für die Epidemiologie und prädiktive Modellierung genutzt. Letzteres ist zum Teil auf die Wirksamkeit von Modellierungsansätzen bei großen Datensätzen zurückzuführen [2]. Trotz dieser Fortschritte bleibt der Zugang zu medizinischen Daten zur Verbesserung der Patientenversorgung eine große Herausforderung. Die Gründe für die eingeschränkte gemeinsame Nutzung medizinischer Daten sind zwar vielfältig, doch die Sorge um den Schutz der Privatsphäre der Patienten wird als eines der wichtigsten Probleme hervorgehoben. Obwohl Patientenstudien gezeigt haben, dass fast alle der Meinung sind, dass de-identifizierte medizinische Daten zur Verbesserung der medizinischen Praxis genutzt werden sollten, diskutieren Fachleute weiterhin über die optimalen Mechanismen, um dies zu erreichen. Die MIMIC-III-Datenbank hat als einzige ein permissives Zugriffssystem eingeführt, das eine breite Wiederverwendung der Daten ermöglicht [3]. Dieser Mechanismus hat sich bei der breiten Nutzung von MIMIC-III in einer Vielzahl von Studien bewährt, die von der Bewertung der Behandlungswirksamkeit in genau definierten Kohorten bis zur Vorhersage wichtiger Patientenergebnisse wie der Sterblichkeit reichen. MIMIC-IV soll an den Erfolg von MIMIC-III anknüpfen, wobei eine Reihe von Änderungen vorgenommen wurde, um die Nutzbarkeit der Daten zu verbessern und mehr Forschungsanwendungen zu ermöglichen.

2. Studienteilnehmer

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.

MIMIC-IV enthält Patienten, die in der Notaufnahme und auf der Intensivstation aufgenommen wurden. Während Patienten, die in die Intensivstation eingewiesen werden, einen entsprechenden Krankenhausaufenthalt haben müssen, können Patienten in die Notaufnahme eingewiesen werden, ohne anschließend ins Krankenhaus eingewiesen zu werden. Daher ist die Zahl der Patienten in MIMIC-IV viel höher als die Zahl der einzelnen Patienten mit Krankenhausaufenthalten. In MIMIC-IV v3.0 gibt es 364.627 eindeutige Patienten, von denen 223.452 mindestens einen Krankenhausaufenthalt hatten (d. h. mindestens einen Eintrag in der Aufnahmetabelle). Die übrigen 141.175 Patienten wurden nur in der Notaufnahme behandelt, was anhand der Tabelle der Verlegungen überprüft werden kann.

3. Datenerhebung

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. Einem krankenhausweiten EHR und einem ICU-spezifischen klinischen Informationssystem. Die Erstellung von MIMIC-IV wurde in drei Schritten durchgeführt:

1. Datenerfassung:

Die Daten von Patienten, die in die Notaufnahme des BIDMC oder in eine der Intensivstationen eingeliefert wurden, wurden aus den jeweiligen Krankenhausdatenbanken extrahiert. Es wurde eine Master-Patientenliste erstellt, die alle Krankenaktennummern von Patienten enthielt, die zwischen 2008 und 2022 auf einer Intensivstation oder in der Notaufnahme aufgenommen wurden. Alle Quelltabellen wurden so gefiltert, dass sie nur Zeilen enthielten, die sich auf Patienten in der Hauptpatientenliste bezogen

2. Vorbereitung:

Die Daten wurden umstrukturiert, um die retrospektive Datenanalyse zu erleichtern. Dazu gehörten die Entnormalisierung von Tabellen, die Entfernung von Prüfpfaden und die Umstrukturierung in weniger Tabellen. Das Ziel dieses Prozesses ist es, die retrospektive Analyse der Datenbank zu vereinfachen. Wichtig ist, dass keine Datenbereinigungsschritte durchgeführt wurden, um sicherzustellen, dass die Daten einen realen klinischen Datensatz widerspiegeln.

3. Deidentifizieren:

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. 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

4. Datenarchivierung

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.