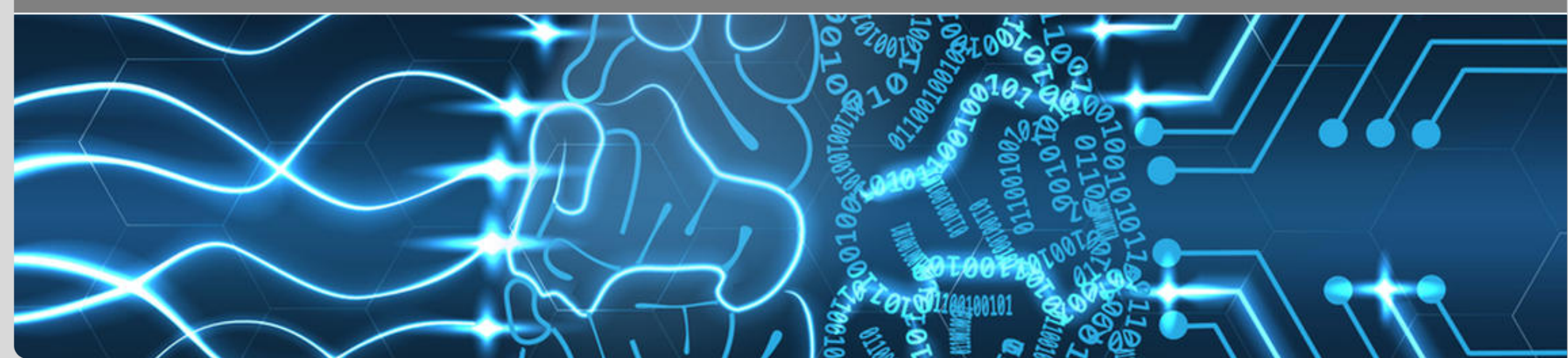


Predictive Analytics by Inferring Structure from Electronic Health Records

Masterarbeit in Kooperation mit INRIA, Grenoble
Luisa Sophie Werner, 06.05.2020

TECO - Technology for Pervasive Computing



Gliederung

- Motivation
- Methode
- Umsetzung
- Ergebnisse
- Erweiterungen
- Zusammenfassung

MOTIVATION

Motivation von Deep Learning im Gesundheitswesen

■ Steigende Popularität von Deep Learning

- Einsatz von GPUs
- Erhöhte Datenverfügbarkeit
- Ende-zu-Ende Lernen

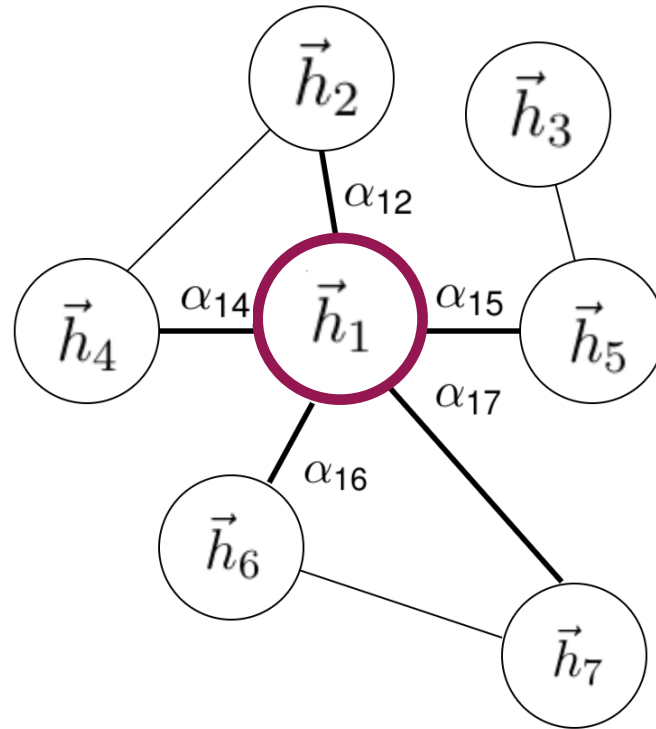
■ Einführung von Electronic Health Record (EHR) Systemen

- Grundlage für Data Science im Gesundheitswesen
- Potential für neue Erkenntnisse, Verbesserungen, Effizienzsteigerung

Aufgabenstellung

- Vorhersage der **Mortalität** bei Patienten auf Basis von verabreichten **Medikamenten** und **Diagnosen** am ersten **Tag** des Krankenhausaufenthaltes
- **Graph Neural Networks**
 - Erlernen und Ausnutzen der inhärenten Graphstruktur von EHR
 - Finden von Vektor Repräsentationen der Medikamente und Diagnosen, die die Graphstruktur nutzen

Motivation: Graph Neural Networks [1]



$$\vec{h}_i' = \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} \mathbf{W} \vec{h}_j \right)$$

Aktivierungsfunktion

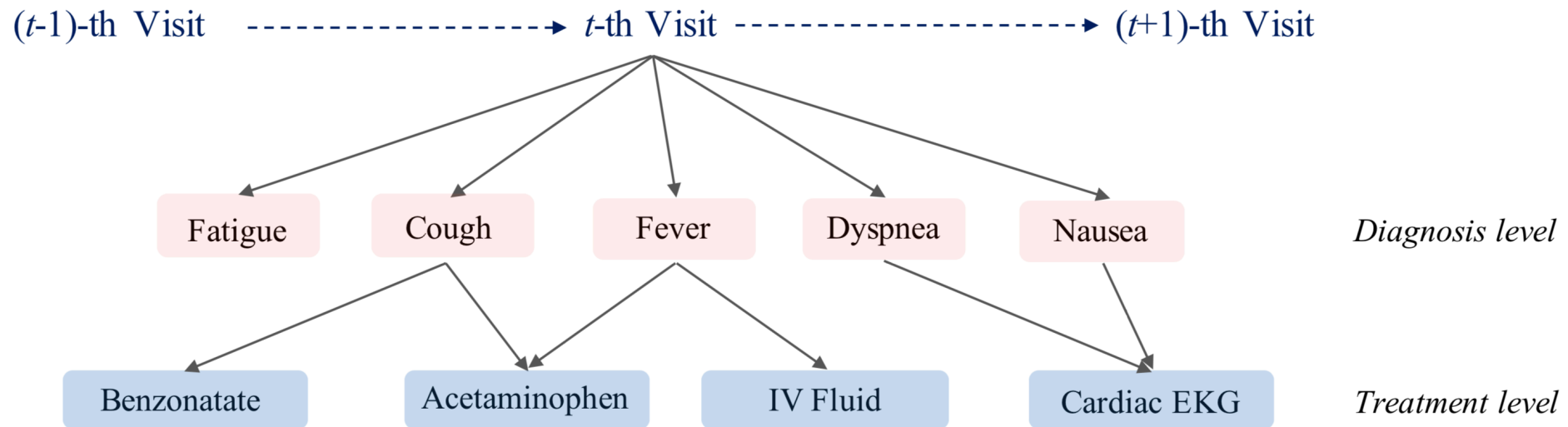
Lernbare Parameter

Nachbarschaft eines Knotens i

MiME: Multilevel Medical Embedding of Electronic Health Records for Predictive Healthcare [2]

- Vorhersage von Herzversagen
- Mehrstufiges Lernen von Features auf Basis der Graphstruktur
- Deutliche Verbesserung der Vorhersagegüte
- Privater Datensatz mit inhärenter Graphstruktur

MiME: Multilevel Medical Embedding of Electronic Health Records for Predictive Healthcare



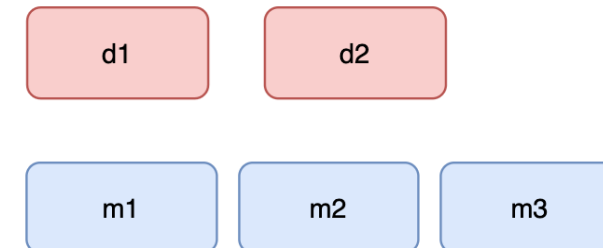
METHODE

Methode: Graph Convolutional Transformer [3]

- Graphstruktur ist in den meisten EHR Datensätzen unbekannt
- Annahme, dass Graphstruktur trotzdem existiert
- **Ziel: Erlernen der Graphstruktur**
- Anwendung von GNNs oder MiME auf Basis der erlernten Graphstruktur

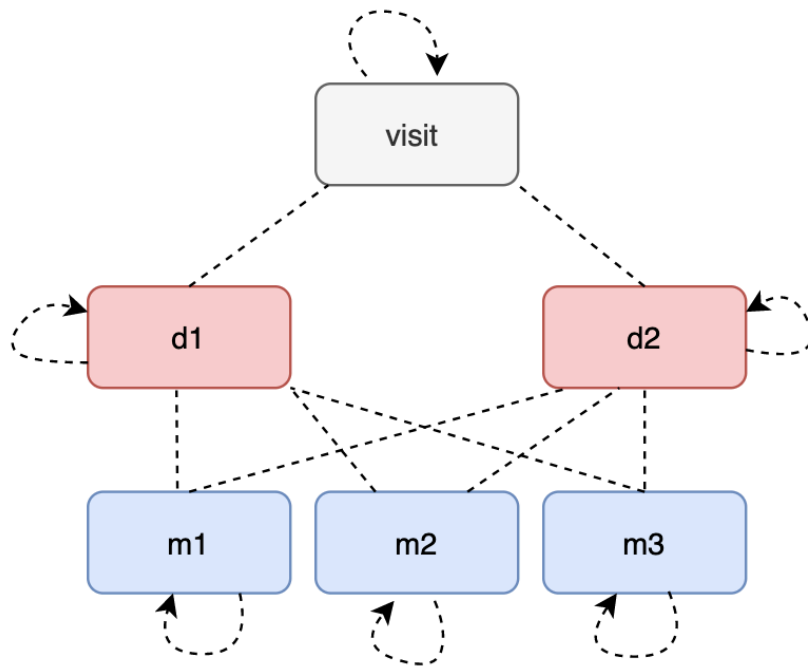
Methode: Graph Convolutional Transformer

- Jeder Krankenhausaufenthalt wird als hierarchischer Graph G modelliert
- Adjazenzmatrix A beschreibt Kanten
- Beispiel:
 - Diagnosen: $\{d1, d2\}$
 - Verabreichte Medikamente: $\{m1, m2, m3\}$



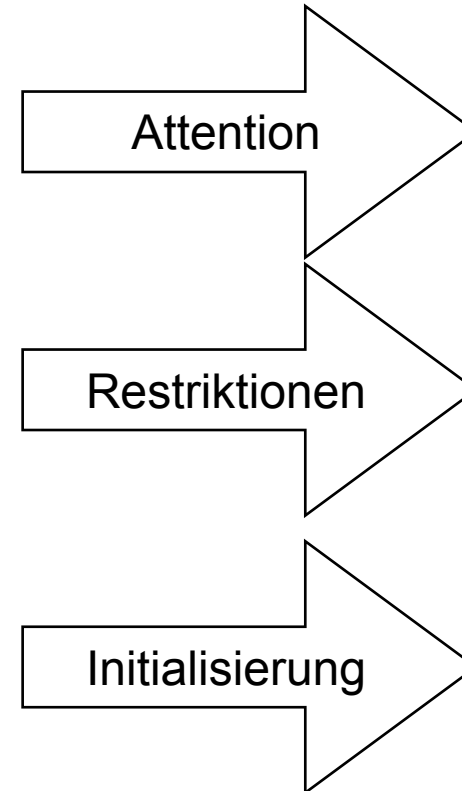
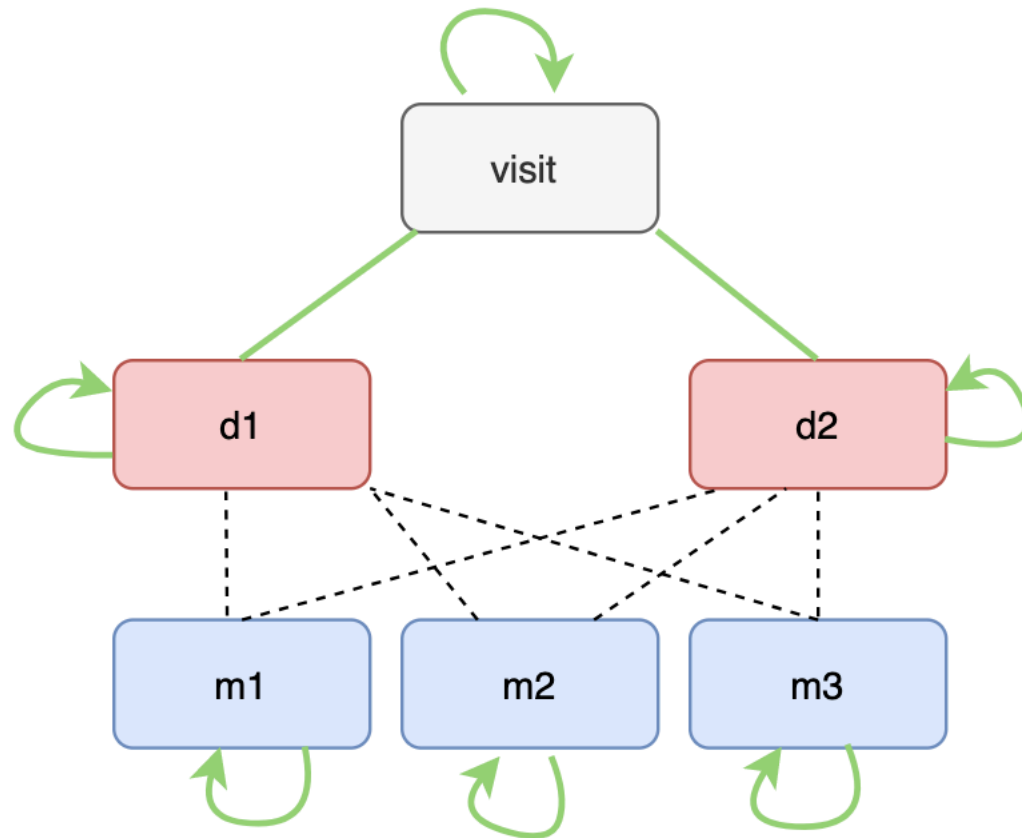
Methode: Graph Convolutional Transformer

Fragestellung: Welches Medikament wurde für welche Diagnose verschrieben ?



	visit	d1	d2	m1	m2	m3
visit	?	?	?	?	?	?
d1	?	?	?	?	?	?
d2	?	?	?	?	?	?
m1	?	?	?	?	?	?
m2	?	?	?	?	?	?
m3	?	?	?	?	?	?

Methode: Graph Convolutional Transformer



	visit	d1	d2	m1	m2	m3
visit						
d1				?	?	?
d2				?	?	?
m1		?	?			
m2		?	?			
m3		?	?			

Methode: Graph Convolutional Transformer

- Erlernen von Kantengewichten mit Attention
- Restriktionen
- Initialisierung mit bedingten Wahrscheinlichkeiten
- Finden von Vektorrepräsentationen
- Vorhersagen

UMSETZUNG

Premier Healthcare Database [5]

- Premier Healthcare Database: Eine der umfangreichsten EHR Datenbanken aus den USA
- Datensatz aus dem Jahr 2006
- Mehr als 1 Mio. Krankenhausbesuche

Premier Healthcare Database

■ Anwendung verschiedener Filter

- Reduktion des Datensatzes auf knapp 900.000 Patienten

■ Unbalancierter Datensatz

- ~ 3.3% der Patienten sind verstorben
- Zufälliges Undersampling der Klasse der nicht verstorbenen Patienten
- Reduktion des Datensatz auf ~ 50.000 Patienten

Baseline Modelle: Logistische Regression

- Keine Graphstruktur
- Keine automatisches Erlernen von Repräsentationen

Ist ein neuronales Netz hilfreich bei der der Aufgabenstellung?

Ist die Graphstruktur hilfreich bei der der Aufgabenstellung?

Baseline Modelle: GCT ohne Initialisierung mit bedingten Wahrscheinlichkeiten

- Erlernen der Graphstruktur
- Gleichverteilte Initialisierung der Adjazenzmatrix

Welchen Nutzen hat die Initialisierung mit bedingten Wahrscheinlichkeiten für GCT?

Spielt die Initialisierung überhaupt eine Rolle?

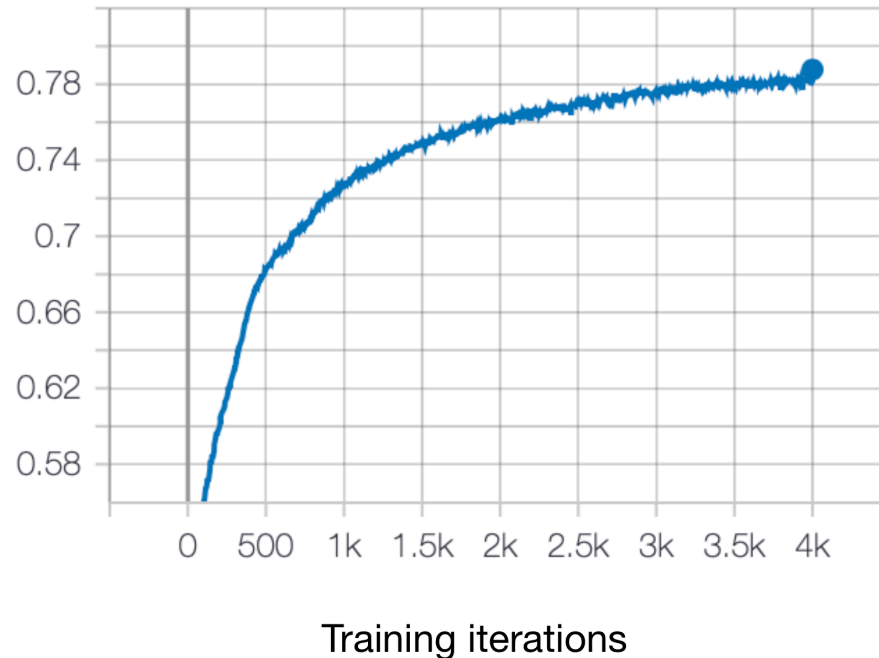
Training und Evaluierung

- Unterteilung des Datensatzes in Train, Validation und Test set (8:1:1)
- Training/Evaluierung von GCT und den Baselines auf Train/Validation set
- Prediction auf dem Test set
- Vergleich der AUC-ROC der Modelle

ERGEBNISSE

Ergebnisse: GCT mit Initialisierung

AUC-ROC



AUC-ROC auf dem Validation Set: **78.2 %**
AUC-ROC auf dem Test Set: **78.8 %**

GCT: AUC-ROC auf dem Validation Set während des Trainings

Motivation ➤ Methode ➤ Umsetzung ➤ **Ergebnisse** ➤ Erweiterungen ➤ Zusammenfassung

Interpretation der Ergebnisse

Modell	AUC-ROC (Test Set)
Logistische Regression	80.0 %
GCT ohne Initialisierung	78.2 %
GCT mit Initialisierung	78.8 %

■ Logistische Regression liefert das beste Ergebnis

■ GCT mit und ohne Initialisierung liefern ähnliche Ergebnisse

Interpretation der Ergebnisse

- Logistische Regression ist fuer die Aufgabenstellung besser geeignet als ein neuronales Netz
- Die Graphstruktur hat keinen deutlichen Nutzen für die Vorhersage der Mortalität
- Weitere Untersuchungen sind notwendig, um GCT als Methode zu beurteilen

Erweiterungsmöglichkeiten von GCT

- Training und Validierung auf weiteren Train/Test Divisionen
- Umfangreiche Hyperparameter Optimierung
- Umgang mit unbalancierten Klassen
- Erweiterung von GCT auf andere Fragestellungen
- Weitere Informationen in die Graphstruktur einbetten

Zusammenfassung

- GCT als Vorhersagemodell für Mortalität
- Keine deutlichen Verbesserungen
- ABER: Potential für weitere Anwendungen
- Weitere Untersuchungen notwendig
- Zahlreiche Erweiterungen denkbar

Literatur

- [1] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Li`o, and Yoshua Bengio. Graph attention networks. <https://arxiv.org/abs/1710.10903>, 2017.
- [2] Edward Choi, Cao Xiao, Walter F. Stewart, and Jimeng Sun. Mime: Multilevel medical embedding of electronic health records for predictive healthcare. <https://arxiv.org/abs/1810.09593v1>, October 2018.
- [3] Edward Choi, Zhen Xu, Yujia Li, Michael W Dusenberry, Gerardo Flores, Yuan Xue, and Andrew M Dai. Graph convolutional transformer: Learning the graph- ical structure of electronic health records. arXiv preprint arXiv:1906.04716, 2019.
- [4] Jay Alammar. The illustrated transformer [blog post on jalammar]. <http://jalammar.github.io/illustrated-transformer/>, 2018.
- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Premier healthcare database: Data that informs and performs. <https://www.premierinc.com>, 2015.
- [6] Nitish Shirish Keskar and Richard Socher. Improving generalization performance by switching from adam to sgd, 2017.

FRAGEN ?