



Thema Projektstudie

Vorhersage von Überlebenszeiten aufgrund von multivariaten prognostischen Faktoren mit Maschinellern Lernen

Univ.-Prof. Dr. Georg Heinze¹, Assoz.-Prof. Dr. Daniela Dunkler¹ und Ao Univ.-Prof. Dr. Georg Dorffner²

¹ Institut für Klinische Biometrie

² Institut für Artificial Intelligence

Das Ziel dieser Projektstudie ist, mit neuen Verfahren des Maschinellen Lernens, den sogenannten Transformer-Modellen, aus multivariaten longitudinalen Prozessen individuelle Überlebenszeiten über lange Sicht, etwa 30 Jahre, vorherzusagen. Dazu sollen individuelle mittelfristige Beobachtungen von prognostischen Faktoren, etwa über 5-10 Jahre, benutzt werden. Wir nehmen an, dass solche Daten in einer Kohorte von Proband:innen unterschiedlichen Lebensalters vorliegen, gemeinsam mit dem Überlebensstatus am Ende der individuellen Beobachtungsperiode. Mit diesen Daten soll nun ein sogenanntes TransformerJM-Deep-Learning Modell bzw. ein TransformerLSR-Modell zur Langzeitprognose trainiert werden (Lin & Luo, 2022; Zhang et al, 2025). Diese Methoden wurden kürzlich zum gemeinsamen Modellieren der multivariaten prognostischen Faktoren sowie des Überlebens der Proband:innen zum Zwecke der Langzeitprognose vorgeschlagen.

In der Projektarbeit verwenden wir zwei synthetische Datensätze mit Kurzzeit-Nachbeobachtung; RIPOSTE-V4 und RIPOSTE-V5. Datensatz V4 simuliert das unregelmäßige und zwischen den Proband:innen sehr unterschiedliche Nachbeobachtungsmuster einer realen retrospektiven Kohortenstudie. V5 wiederum gleicht einer typischen prospektiven Kohortenstudie mit regelmäßigen Nachbeobachtungen. Für V4 und V5 gibt es jeweils zwei passende Testdatensätze, die zunächst unter Verschluss gehalten werden, aber im späteren Verlauf des Projekts zur Validierung genutzt werden sollen. Im ersten Testdatensatz können die im Trainingsdatensatz prognostizierten Überlebensprozesse mit den an den gleichen synthetischen Proband:innen über lange Zeit beobachteten Überlebensprozessen verglichen werden und somit die Güte der Langzeitprognose validiert werden (interne Validierung). Im zweiten Testdatensatz liegen über lange Zeit beobachtete Überlebensprozesse von einer weiteren Kohorte unabhängiger synthetischer Proband:innen derselben Population vor, mit denen die Validität der Transformer-Modelle außerhalb der Trainingskohorte geprüft werden sollen (externe Validierung). Abb. 1 fasst den Ablauf der Projektstudie zusammen.

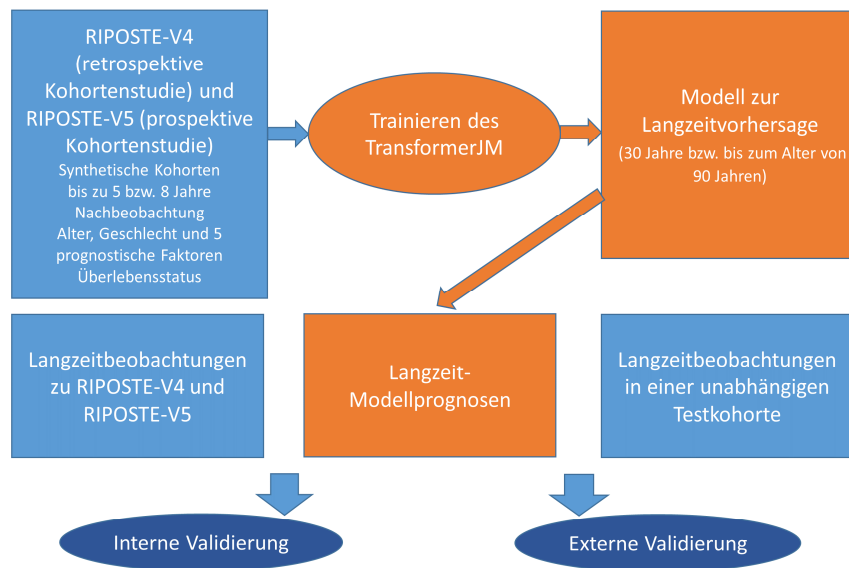


Abb. 1: Schematischer Verlauf der Projektstudie.

Die Berechnungen sollen mittels den von den Autor:innen bereitgestellten Pythoncodes erfolgen. Alle Arbeitsschritte sind sorgfältig zu planen und zu dokumentieren. Die trainierten Modelle sind geeignet zu beschreiben, und die Resultate der Validierungen sollen visuell und numerisch dargestellt werden.

Referenzen

Lin, J., Luo, S., 2022. Deep learning for the dynamic prediction of multivariate longitudinal and survival data. *Statistics in Medicine* 41, 2894–2907. <https://doi.org/10.1002/sim.9392>

Zhang, Z., Zhao, Y., Xu, Y., 2025. TransformerLSR: Attentive joint model of longitudinal data, survival, and recurrent events with concurrent latent structure. *Artificial Intelligence in Medicine* 160, 103056. <https://doi.org/10.1016/j.artmed.2024.103056>

Anforderungen:

- Interesse an prognostischen Fragestellungen und am wissenschaftlichen Arbeiten
- Kenntnis von Python und R, sowie Github
- Fähigkeit, Arbeitsschritte sorgfältig zu planen, zu dokumentieren und adäquat zu berichten

Aufgaben:

- Existierenden Code für neue Probleme adaptieren
- Trainieren der Transformer-Modelle mit Hilfe der bereitgestellten Datensätze
- Berechnung von Langzeitprognosen aus den Modellen
- Vergleich der Langzeitprognosen mit beobachteten Überlebenszeiten mittels geeigneter statistischer Maßzahlen und Grafiken
- Darstellung der Ergebnisse in übersichtlicher Form