

Ähnlichkeitssuche

Aktuelle Ansätze aus anderen Fachdomänen

Jonas Hügel^{1,2}, Tamara Martin³, Kais Tahar¹, Holm Graessner³, Dagmar Krefting^{1,2}

1: Institut für Medizinische Informatik, Universitätsmedizin Göttingen, 37075 Göttingen
2: Campus-Institut Data Science, Georg-August-Universität Göttingen, 37075 Göttingen

3: Zentrum für Seltene Erkrankungen und Institut für Medizinische Genetik und Angewandte Genomik, Universität und Universitätsklinikum Tübingen, 72076 Tübingen

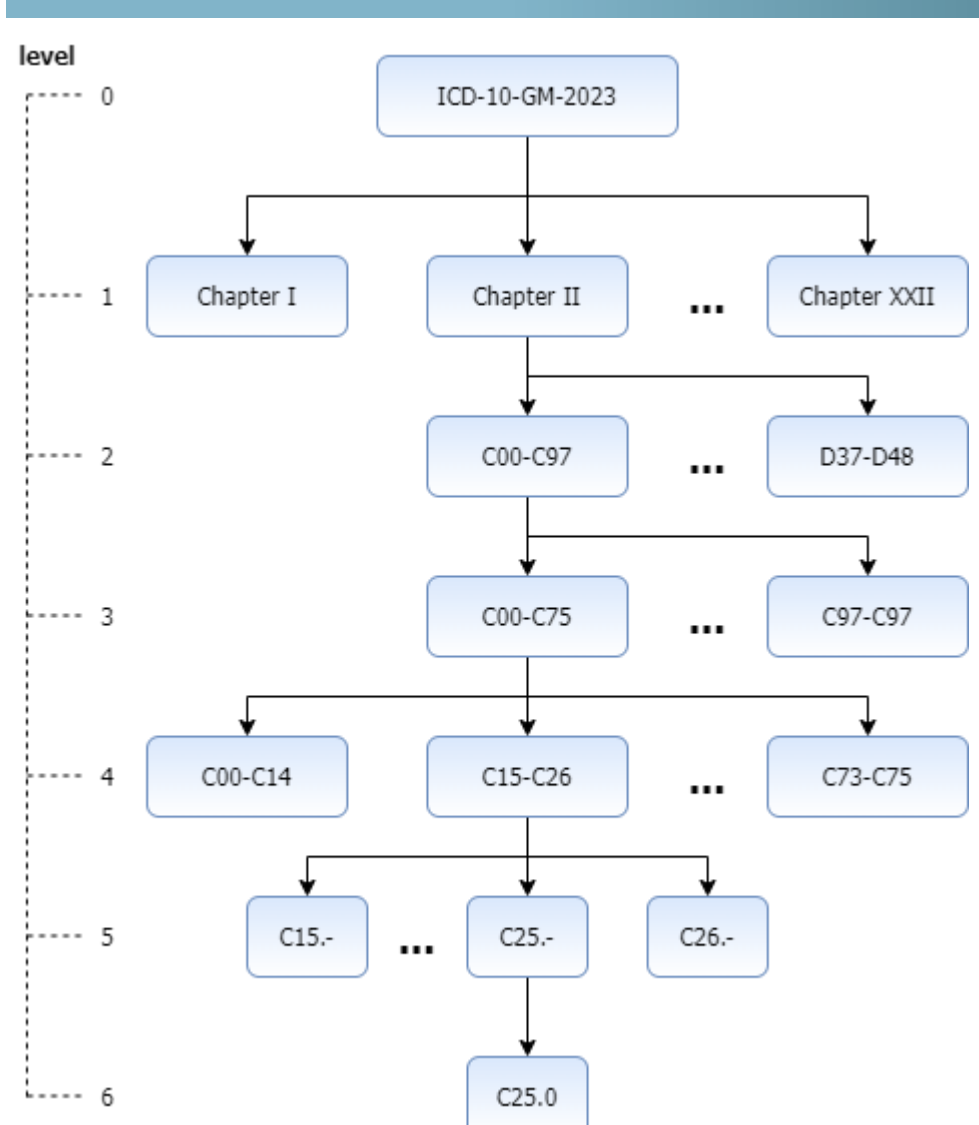
Zusammenfassung:

Durch die Subphenotypisierung häufiger Erkrankungen entstehen häufig auch sehr kleine Subpopulationen, welche spezifische Behandlungsmethoden benötigen. Dadurch verschwimmt die Grenze zwischen Seltenen und Häufigen Erkrankungen. Auf dem Poster diskutieren wir, inwiefern sich Methoden aus den häufigen Erkrankungen für die Suche und Vergleich von ähnlichen Patient:innen auf Seltene Erkrankungen übertragen lassen.

Motivation

- Verschwimmende Grenzen zwischen seltenen Erkrankungen (SE) und häufigen Erkrankungen (HE) auf Grund von Subphenotypisierung durch umfassender Diagnostik
- Kaum Evidenz für spezifische Behandlungsmethoden für viele einzelne Subphenotypen HE aufgrund von geringer Fallzahlen
- Abschätzen der Behandlungserfolge anhand ähnlicher Patienten
- Vielversprechende Ansätze in anderen Fachdomänen, z. B. der Onkologie, Alzheimer und COVID Forschung, zur Suche nach Ähnlichen Patienten, die auf SE übertragbar seien sollten

Ontologie/Taxonomie basierte Verfahren



- Ontologien/Taxonomien (z.B. ICD-10) sind aus informationstechnischer Sicht Graphen
- Anwendung bewährter Algorithmen aus der Graphentheorie, um Ähnlichkeit von Knoten (ICD-Codes) zu berechnen
- Patientenhistorie kann als Menge (Set) von ICD Codes betrachtet werden
- Anwendung von Set Similarity Algorithmen möglich

Abb. 1: ICD-10 Taxonomie Struktur [1].

- (Paarweiser) Vergleiche von multiple Patienten miteinander möglich (siehe Abb. 2)
- Gute Ergebnisse in anderen Fachdomänen, z.B. Pankreaskarzinom [1]
- Integration der Alpha-ID-SE möglich
- Bereits gut erprobte Algorithmen vorhanden

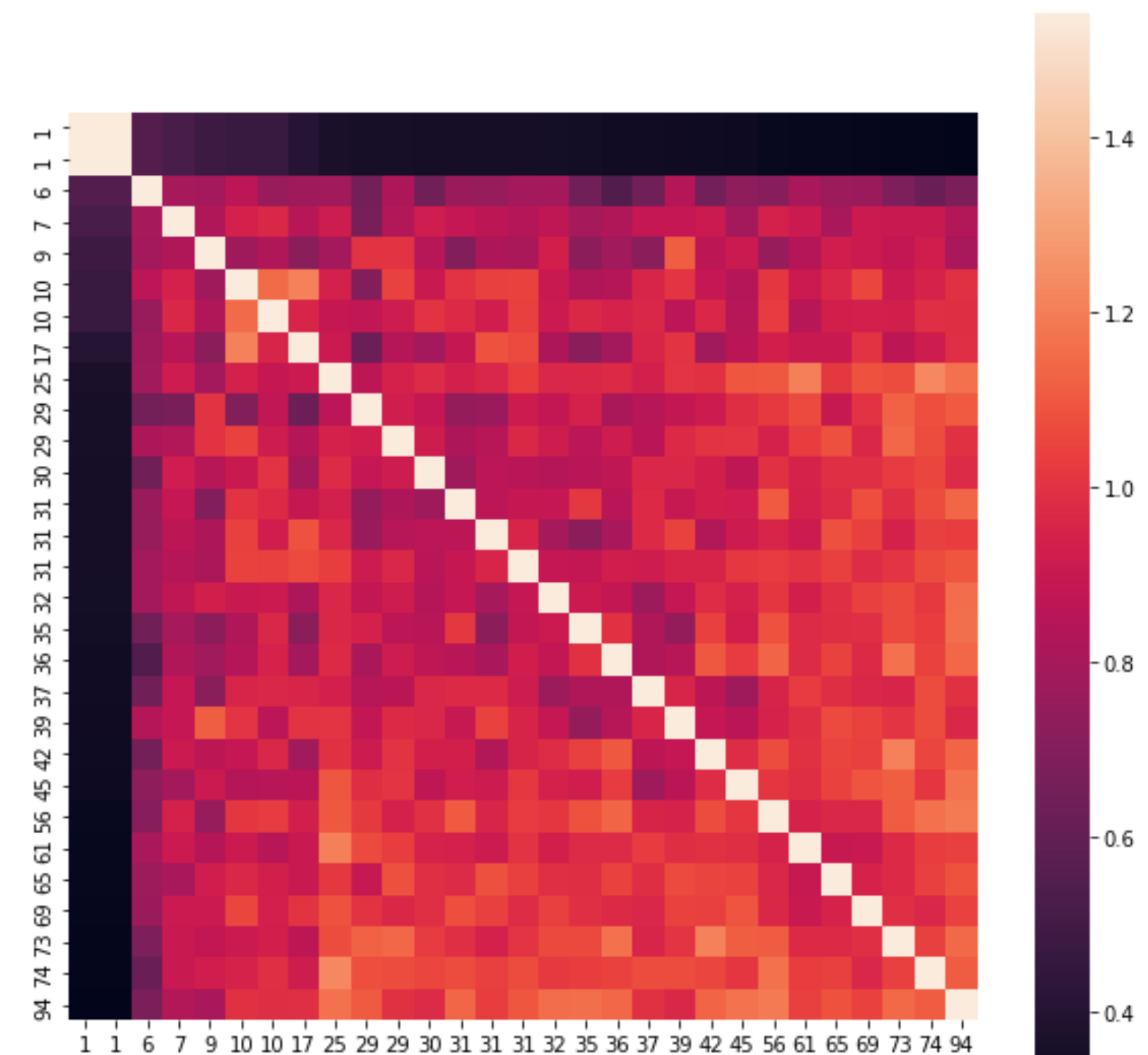


Abb. 2: zeigt die Heatmap für paarweise Ähnlichkeit-Scores von Patienten. Die Achsenbeschriftung gibt die Menge an ICD Codes für den jeweiligen Patienten an [1].

Funding:

This work is partially funded by the VolkswagenStiftung (ZN3424), the German Research Foundation (426671079), the "Collaboration on Rare Diseases" of the Medical Informatics Initiative (CORD-MI) funded by the German Federal Ministry of Education and Research (BMBF), under grant number: FKZ-01ZZ1911R.

Longitudinale Sequenzen der Patienten:innenhistorie

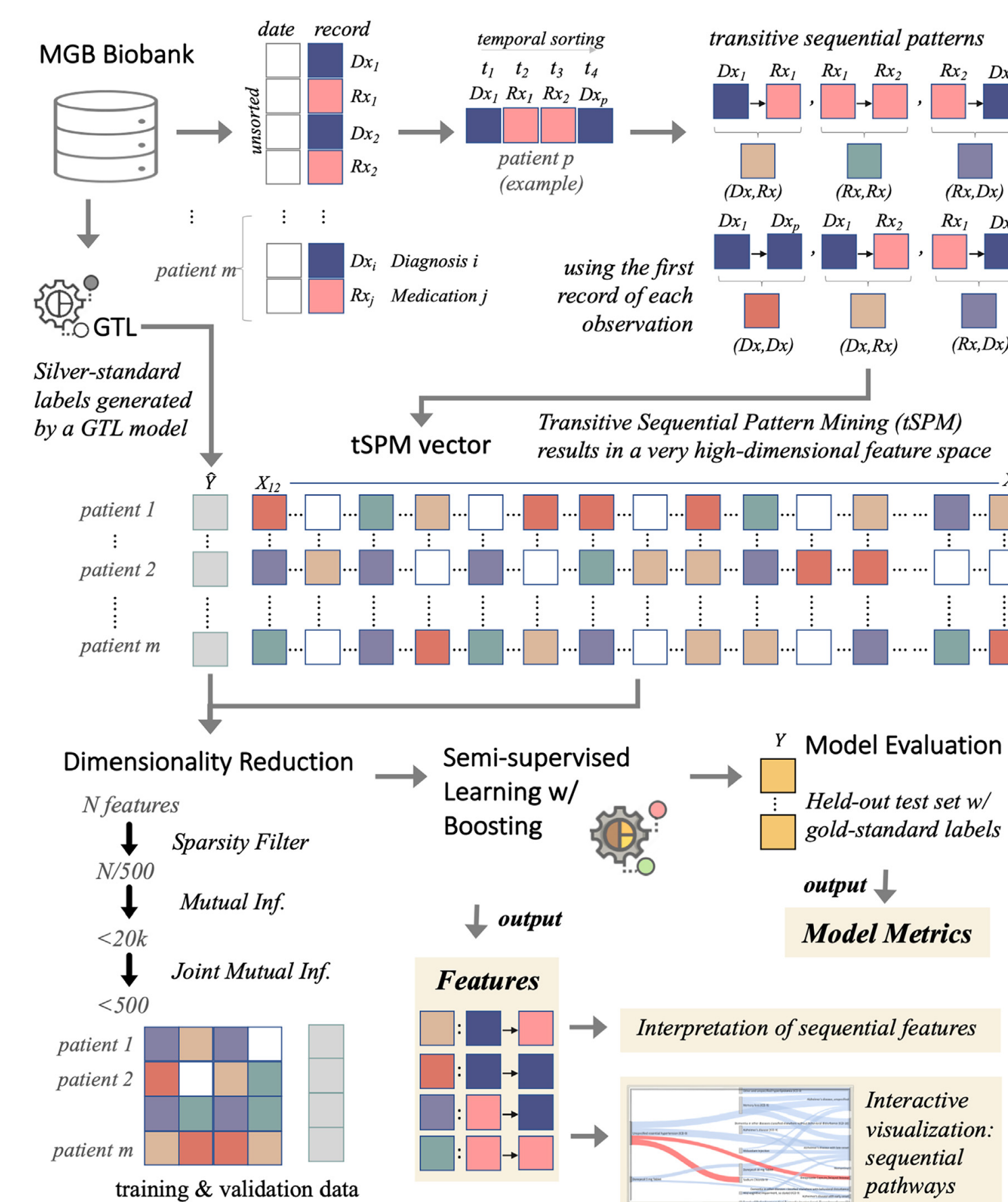


Abb. 2: Beispielhafter Workflow zur Generierung und Integration der Sequenzen zur nachfolgenden Anwendungen im Maschinellen Lernen [4].

- Kombinieren von einzelnen Einträgen aus der Patientenakte zu Sequenzen [2, 3]
- Inkludiert die inhärente temporale Komponente der Patient:innenhistorie
- Erzeugen großer Menge an Sequenzen
- Dimensionalitätsreduktion nach Erstellung der Sequenzen notwendig
- Verbesserte Ergebnisse bei Nutzung der Sequenzen in nachfolgenden Maschinellen Lernen Applikation, z.B. bei Alzheimer [4] und COVID [5]

- Ermöglichen der Implementierung von komplexen Definitionen, wie z.B. der WHO POST COVID-19 Definition [3]
- Gut geeignet für chronische und komplexe Krankheiten
- Integriert lange Behandlungshistorie der Patient:innen von SE
- Gute Verfügbarkeit der Algorithmen [2, 3]

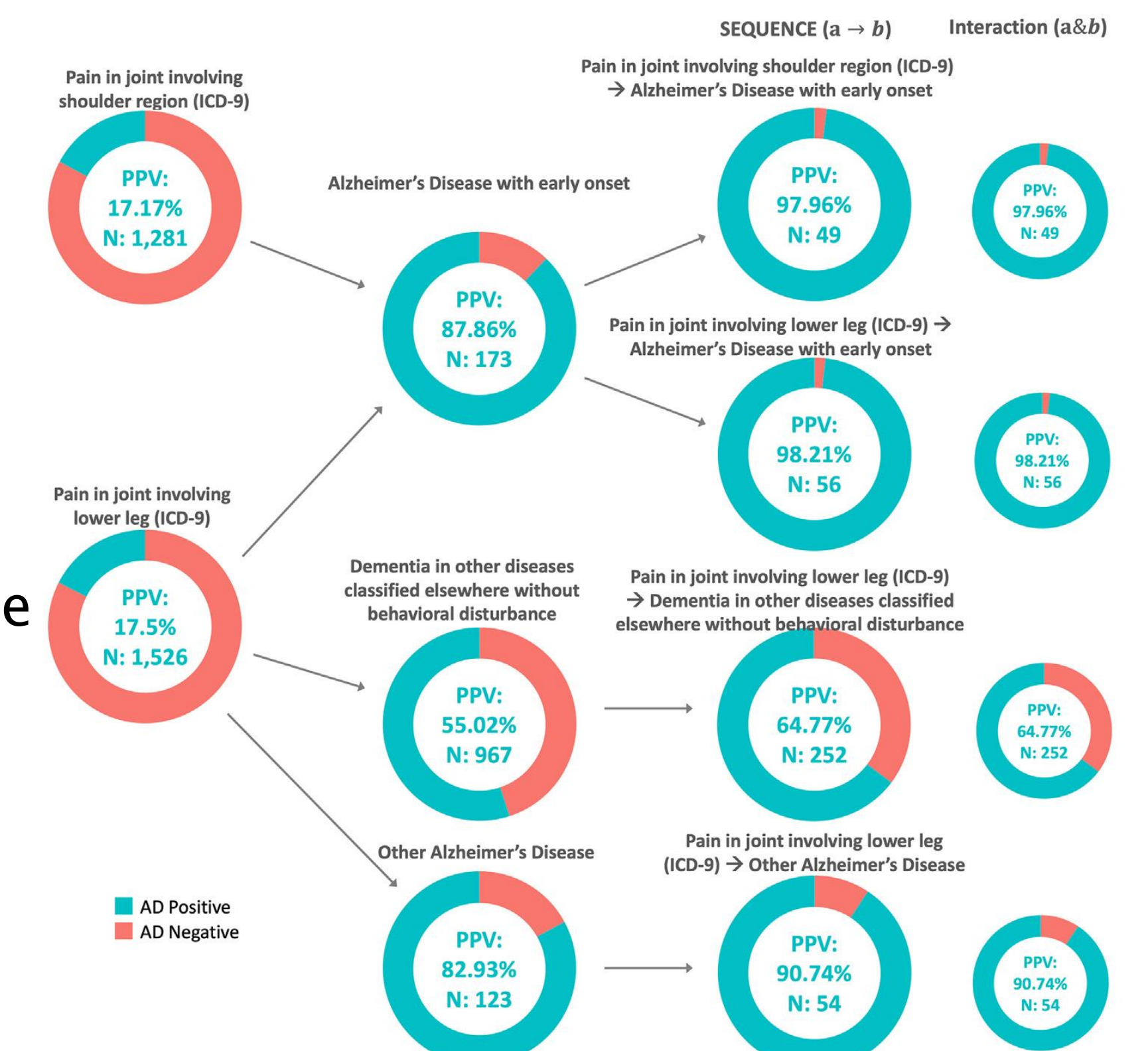


Abb. 4: zeigt wie die Sequenzen aus mehreren Einträgen im Vergleich zu einzelnen Einträgen aus der Patientenakte die Klassifizierung verbessern [4].

Diskussion

- Ansätze bieten ein hohes Potential für SE
- Algorithmen sind öffentlich verfügbar
- Weitere umfassende fachliche Bewertungen und Analysen werden benötigt:
 - In Einzelne erste Projekte für unterschiedliche SE
 - Integration in Konsil-SE könnte Unterstützung für Fallkonferenzen bieten
 - Auswahl der Datenpunkt für die Integration abhängig von Algorithmen und fachlicher Bewertung

Referenzen

- [1] Schneider et al., "Improving ICD-based semantic similarity by accounting for varying degrees of comorbidity", Pre-Print (2023) doi: 10.48550/arXiv.2308.07359.
- [2] Estiri et al., "Transitive Sequencing Medical Records for Mining Predictive and Interpretable Temporal Representations", Patterns (2020).
- [3] Hügel et al., "tSPM+: a high-performance algorithm for mining transitive sequential patterns from clinical data", Pre-print (2023) doi: 10.48550/arXiv.2309.05671.
- [4] Estiri et al., "Temporal Cohort Identification for Alzheimer's Disease with Sequences of Clinical Records", eBioMedicine (2023).
- [5] Estiri et al., "Individualized prediction of COVID-19 adverse outcomes with MLHO", Scientific Reports (2021).

jonas.huegel@med.uni-goettingen.de <https://medizininformatik.umg.eu>