



Zentralinstitut
kassenärztliche
Versorgung

Identifikation vulnerabler Patientengruppen mit Pat2Vec & GERBEHRT

Routinedaten als Vulnerabilitätsschicht für die Krisenvorsorge

Dr. Edgar Steiger, Zi
Berlin, 16. September 2026

● Krisen treffen nicht alle gleich

- Strom- oder Wasserausfall:
z.B. Dialyse, Heimbeatmung und kühlpflichtige Arzneimittel
- Hitze:
Hochaltrige, Menschen mit u.a. Demenz, Herz- und Nierenerkrankte
- Unterbrochene Versorgung:
z.B. Dauermedikation, chronisch Erkrankte, Pflegebedürftige, Immobilität
- Leitfrage:
Wer benötigt bei welcher Krise welche Unterstützung und wo?

● Morbidität ist nur ein Teil der Krisenvulnerabilität

- Risiko = Gefährdung × Exposition × Vulnerabilität × Kapazität
- Routinedaten beschreiben vor allem die medizinische Ausgangslage (Morbidität und bestehende Versorgung).
Frühwarnsysteme beschreiben die aktuelle Gefährdung.
- Vulnerabilität: Versorgungsabhängigkeit, Erreichbarkeit, besondere Bedürfnisse im Krisenfall.
- Datenbasierte Analyse von Krisenresilienz braucht eigene Outcomes und handlungsbezogene Schwellenwerte.

● Vollständige Routinedaten als Ausgangspunkt

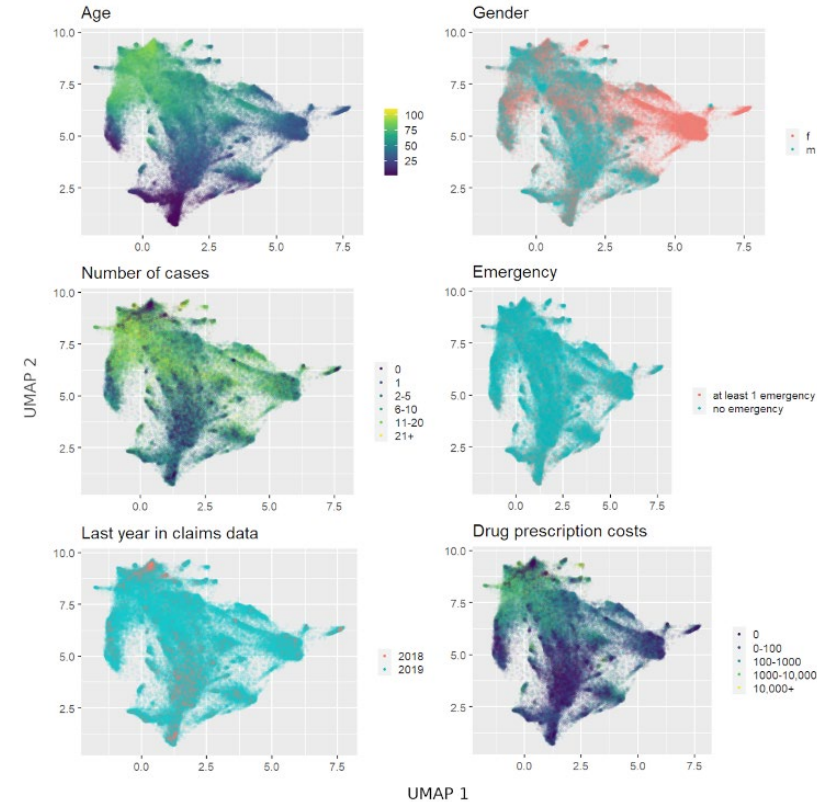
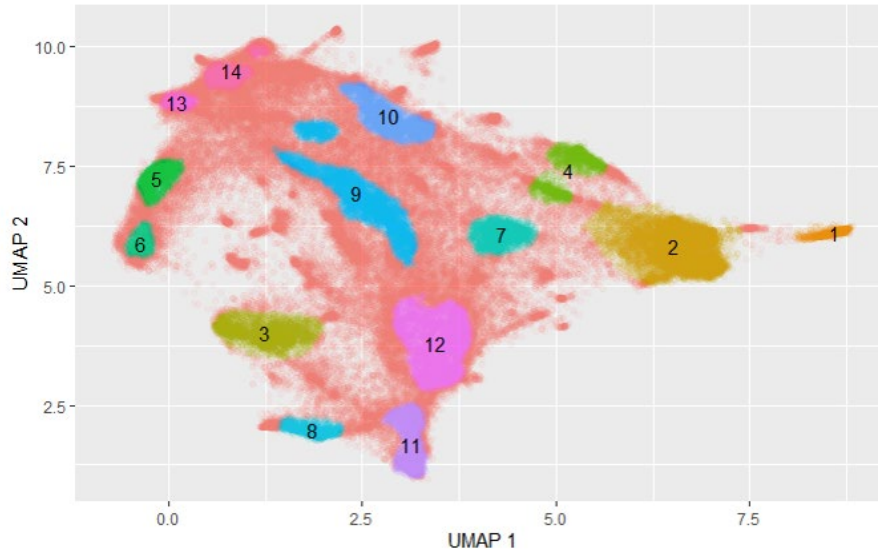
- Bundesweite vertragsärztliche Abrechnungsdaten (Diagnosen, Arzneiverordnungen, Leistungen) bilden longitudinale Behandlungsverläufe ab
- Vollständige Verläufe enthalten mehr Information als eine Auswahl etablierter Risikofaktoren
- Für Krisenfragen müssen Gefährdungs-, Register-, Erreichbarkeits- und Kapazitätsdaten hinzukommen
- Grenzen bleiben Abrechnungsverzug, ambulante GKV-Perspektive und fehlende krisenspezifische Outcomes

● Pat2Vec: unsupervised Mustererkennung

- Machine Learning-Ansatz „Doc2Vec“ lernt aus kompletten Diagnoseverläufen kompakte Patient*innenprofile ohne vorgegebenes Ziel-Outcome
- Ziel 1: bisher unbekannte Morbiditätsmuster und Patient*innengruppen explorieren
- Ziel 2: ein Embedding als Grundlage für präzisere nachgelagerte Prädiktionsmodelle bereitstellen
- Trade-off: hohe Informationsverdichtung, aber schwer interpretierbare Dimensionen
- keine Krisenvalidierung

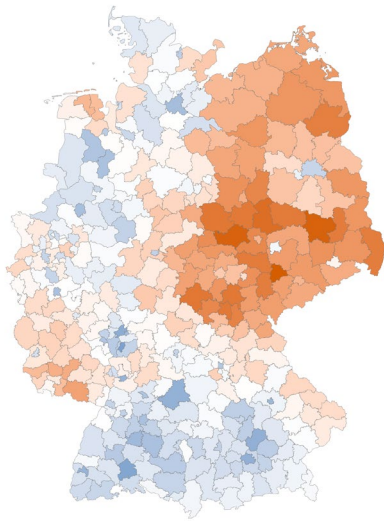
● Pat2Vec: Exploration

- „Morbiditätskontinuum“
- Ohne geeignetes Outcome identifiziert Pat2Vec keine validierten vulnerablen Gruppen.



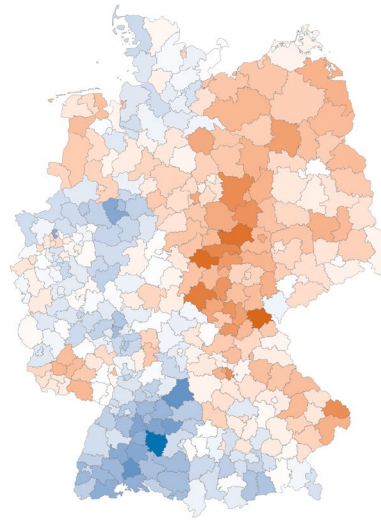
- **Höhere Modellgüte garantiert keine Krisenresilienz**

Anwendung Pat2Vec: Simulation ambulante Bedarfsplanung
R2 von 35 auf 50 %, mehr hausärztliche Kassensitze für deprivierte Regionen
Generelle Morbiditätsbelastung $\neq$ Allgemeines Gefährdungsrisiko



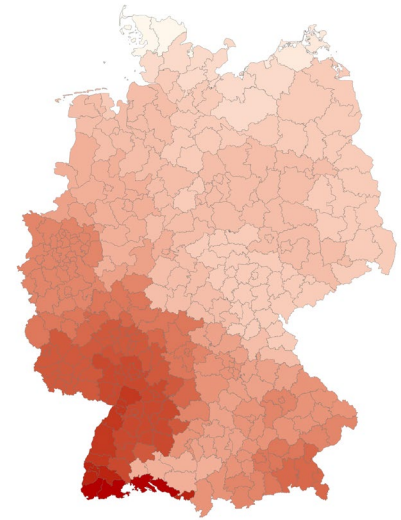
% Sitze vs. nicht adjustierte Allokation

-10 0 10 20



Pat2Vec+AGG: Prozentpunkte vs. 16-Zellen-Modell

-5 0 5



DWD-Hitzewarntage 2025 (bevoelkerungsgew. Mittel)

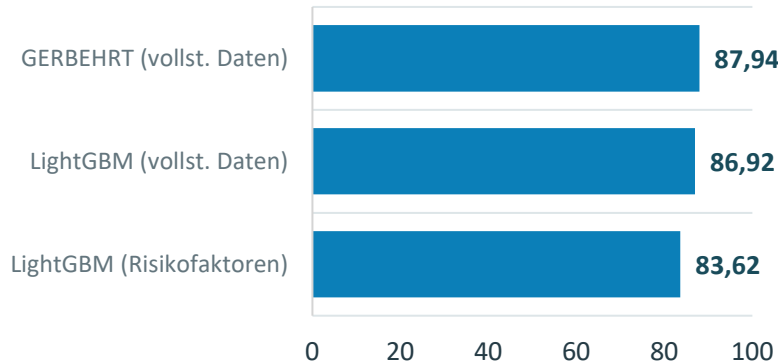
4 8 12 16

● GERBEHRT: semi-supervised Prädiktions-Framework

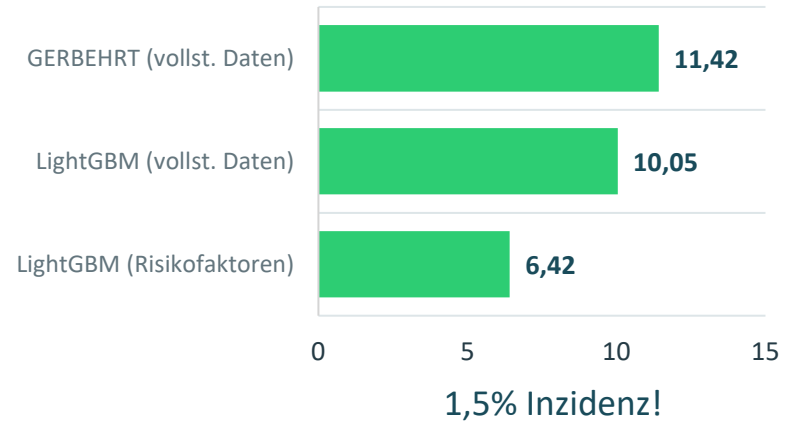
- Vortraining auf vollständigen ambulanten Verläufen von etwa 9 Mio. GKV-Versicherten: u.a. Diagnosen, Verordnungen, Zeitkomponente
- Prospektive Vorhersage des 3-Jahres-Risikos von inzidenter CKD (etwa 1,5%) nach Beobachtungszeitraum und einem Jahr Datenlücke
- Das Framework lässt sich durch erneutes Fine-Tuning an weitere definierte Outcomes anpassen
- Trade-off: höhere Prädiktionsgüte bei geringerer Transparenz
- Abhängigkeit vom gewählten Outcome

- **Vollständige Daten wichtiger als komplexere Modelle**

AUROC (%)



Average Precision (%)



Der große Gewinn entsteht durch vollständige Routinedaten.
GERBEHRT verbessert die Ergebnisse gegenüber LightGBM nur noch begrenzt.

Für Krisenfragen spricht das zunächst für vollständige Daten und eine transparente Baseline.

- CKD ist relevant, aber kein Krisenoutcome

- Dialyse und Nierenerkrankungen sind wegen ihrer Infrastruktur- und Versorgungsabhängigkeit krisenrelevant
- Die CKD-Prädiktion schätzt Erkrankungsrisiko. Sie misst nicht, ob eine laufende Behandlung in der Krise gesichert ist
- Für die Vorsorge wäre eine aktuelle, rechtlich nutzbare Registrierung dialyseabhängiger Personen mit Transport-, Infrastruktur- und Kapazitätsdaten zunächst direkter handlungsrelevant
- GERBEHRT wird interessant, sobald geeignete Krisenoutcomes für prospektive Planung vorliegen

● Pragmatischer Einstieg: vollständig und transparent

- Daten verbreitern: vollständige Routinedaten mit Registern, Gefährdungsinformationen und Versorgungskapazitäten verbinden
- Transparente Baseline: relevante Diagnosen, Arzneimittel und Versorgungsabhängigkeiten zunächst einfach zählen oder regelbasiert gewichten
- Outcome und Handlung definieren: Was soll verhindert werden und welche Maßnahme folgt aus einem erhöhten Risiko?
- ML-Zusatznutzen prüfen: Prädiktionsgewinn gegen Erklärbarkeit/Transparenz, Subgruppenleistung und praktische Umsetzbarkeit abwägen
- Beispiel Impfkampagne COVID-19: Priorisierung nach Alter und Morbidität (ärztliche Einschätzung auf Basis von Richtlinien)

● Fazit: Operationalisierung vor Modellkomplexität

- Pat2Vec entdeckt unsupervised Muster. GERBEHRT lernt supervised definierte Outcomes. Beide sind anpassbare Methoden, aber keine validierten Krisenmodelle.
- Bei vulnerablen Gruppen wiegen Transparenz und Handlungsbezug schwer. Eine einfache, nachvollziehbare Baseline kann zunächst genügen.
- Allgemein höhere Prädiktionsgüte erzeugt keine Krisenresilienz. Dafür braucht es eigene Outcomes, verknüpfte Daten und eine eigene Planungslogik.