

Ehrapy: Ein neues Open-Source-Tool für die Analyse komplexer Gesundheitsdaten

New Research Findings, Computational Health, ICB12. September 2024

Unter Federführung von Helmholtz Munich haben Wissenschaftler:innen eine frei zugängliche Softwarelösung entwickelt – speziell für die Analyse komplexer medizinischer Gesundheitsdaten. Das Open-Source-Programm namens „ehrapy“ ermöglicht es Forschenden, große, heterogene Datensätze zu strukturieren und systematisch zu untersuchen. Die Software steht der wissenschaftlichen Gemeinschaft weltweit zur Verfügung, um sie zu nutzen und weiterzuentwickeln.

Ehrapy soll eine zentrale Lücke in der Analyse von Gesundheitsdaten schließen, sagt Lukas Heumos, einer der Hauptentwickler und Wissenschaftler am Institute of Computational Biology bei Helmholtz Munich und der Technischen Universität München (TUM): „Es gab bisher keine standardisierten Tools, um vielfältige und komplexe medizinische Daten systematisch und effizient auszuwerten. Das haben wir mit ehrapy geändert.“ Das Team hinter ehrapy kommt ursprünglich aus der biomedizinischen Forschung und hat große Erfahrung bei der Auswertung komplexer wissenschaftlicher Datensätze. „Der Gesundheitssektor steht vor ähnlichen Herausforderungen bei der Datenauswertung wie Menschen, die im Labor forschen“, war Heumos´ Feststellung zu Beginn des ehrapy-Projekts.

Explorativer Ansatz - hypothesenfreie Analyse

Zusammen mit vielen weiteren Beteiligten hat Heumos deshalb seine Expertise bei der Entwicklung wissenschaftlicher Software genutzt, um eine Lösung zur Auswertung von Patientendaten zu schaffen: „Ehrapy kann auch neue Muster aufdecken und Erkenntnisse möglich machen, ohne dass man die Daten mit einer bestimmten Vorannahme oder Hypothese untersucht.“ Dieser als explorativ bezeichnete Ansatz sei, so Heumos, ein Alleinstellungsmerkmal von ehrapy.

Ehrapy ermöglicht also Forschenden, komplexe, heterogene und sehr große Datenmengen zu sortieren, zu gruppieren und hypothesefrei zu analysieren. Das eröffnet neue Einsichten, die anschließend tiefer erforscht werden können. Heumos: „Der explorative Ansatz bringt frischen Wind in die Analyse von Gesundheitsdaten. Sie werden wegen ihrer Komplexität und Heterogenität oft gar nicht so ausgewertet, wie es eigentlich sinnvoll wäre.“ Ehrapy kann damit neue Wege öffnen, um Gesundheitsdaten besser für die medizinische Forschung und Praxis nutzbar zu machen.

Das langfristige Ziel: routinemäßiger Einsatz in der klinischen Praxis

Ehrapy ist von Anfang an als Open-Source-Software konzipiert worden. „Es war uns wichtig, die Software von Tag eins an der wissenschaftlichen Gemeinschaft zugänglich zu machen“, betont Heumos. Die Software steht als so genanntes Python-Package auf GitHub, einem Online-Dienst zur Softwareentwicklung, zur Verfügung und kann von Forschern weltweit genutzt und weiterentwickelt werden.

Derzeit liegt der Fokus von ehrapy darauf, Forschungsdatensätze, wie sie etwa in großen Gesundheitsforschungszentren gespeichert werden, effizienter und schneller analysieren zu können. „Ein routinemäßiger Einsatz in der klinischen Praxis ist ein langfristiges Ziel, aber zunächst

konzentrieren wir uns darauf, der Forschungsgemeinschaft ein leistungsfähiges Werkzeug zur Verfügung zu stellen“, sagt Heumos.

Zukünftig plant das Team, standardisierte Datenbanken für elektronische Gesundheitsdaten (EHRs) bereitzustellen. Diese Datenbanken sollen es ermöglichen, große Mengen medizinischer Daten besser zu integrieren und zu analysieren. Des Weiteren ermöglicht dies die Entwicklung von EHR-Atlanten, die als Referenzdatensätze zur Kontextualisierung und Annotation neuer Datensätze dienen können.

Ein langer Weg

„Ehrapy erlaubt umfassende Datenanalyse über Einzelsysteme hinweg – das kann einen wichtigen Schritt für zukünftige KI-Systeme in der Medizin liefern, daher erhoffe ich eine relativ zügige Etablierung in verschiedenen Standorten“, sagt der Direktor des Institute of Computational Biology bei Helmholtz Munich und TUM-Professor, Prof. Fabian Theis. „Die Etablierung solcher Technologien in der Medizin ist ein langwieriger Prozess, der Jahrzehnte dauern kann. Unser Ziel ist es, die Brücke zwischen biomedizinischer Forschung und der praktischen Anwendung in der Medizin zu schlagen.“ Dabei, so Theis weiter, konzentrierte sich das Entwicklerteam auf Methoden zur explorativen Analyse der Daten in ganzheitlicher Form, um versteckte Zusammenhänge so leichter offenzulegen. „Wir versuchen damit auch, akademische und kommerzielle Akteure im Gesundheitssektor zu unterstützen.“

Ehrapy auf GitHub: <https://github.com/theislab/ehrapy>

Originalpublikation

Heumos et al. (2024): Exploratory electronic health record analysis with ehrapy. Nature Medicine. DOI: [10.1038/s41591-024-03214-0](https://doi.org/10.1038/s41591-024-03214-0)