

Open-Source-Lösung „ehrapy“ bietet neue Möglichkeiten der Gesundheitsdatenanalyse

Wissenschaftler:innen von Helmholtz Munich und der Technischen Universität München (TUM) haben eine frei zugängliche Softwarelösung entwickelt - speziell für die Analyse komplexer medizinischer Gesundheitsdaten. Das Open-Source-Programm „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.

Mit-Entwickler und DZL-Wissenschaftler am Standort München (CPC-M) ist Fabian Theis. Der Direktor des Institute of Computational Biology hofft auf eine zügige Etablierung des Tools in verschiedenen Standorten: „Ehrapy erlaubt umfassende Datenanalyse über Einzelsysteme hinweg – das kann einen wichtigen Schritt für zukünftige KI-Systeme in der Medizin liefern. Unser Ziel ist es, die Brücke zwischen biomedizinischer Forschung und der praktischen Anwendung in der Medizin zu schlagen. Dabei 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.“

Explorativer Ansatz - hypothesenfreie Analyse

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. 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 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.“

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.

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)