

KI soll Diagnostik auf Basis von Messdaten aus Blutproben verbessern

Verbundprojekt der FAU und des Biotech-Unternehmens BioVariance

Mit Hilfe moderner Untersuchungs-Methoden lassen sich heute aus Blutproben riesige Mengen von Informationen gewinnen. Entsprechend komplex ist es, diesen Datenschatz auszuwerten und aus ihm treffsichere Schlüsse zu ziehen, etwa für die Diagnose von Krankheiten. Forschende der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) wollen zu diesem Zweck zusammen mit dem Biotech-Unternehmen BioVariance neue Verfahren aus der künstlichen Intelligenz (KI) entwickeln. Diese sollen sowohl mit tatsächlichen Messdaten als auch mit künstlich generierten synthetischen Datensätzen trainiert werden, so dass sie danach Auffälligkeiten finden, die bei bestimmten Erkrankungen gehäuft vorkommen. BioSamp wird vom Freistaat Bayern mit rund einer Million Euro gefördert; ein Drittel davon geht an die FAU.

Bislang können sich Ärztinnen und Ärzte bei ihren Diagnosen oft gerade einmal auf ein paar Dutzend Kriterien stützen. Die sogenannten Omics-Analysen haben das Zeug dazu, das zu ändern. Aus weniger als einem Tropfen Blut lassen sich mit ihrer Hilfe Zehntausende von Messdaten gewinnen: Welche Proteine enthält die Probe in welchen Mengen? Welche fettähnlichen Verbindungen und Stoffwechsel-Produkte? Welche Gene werden in der Person, aus der das Blut stammt, gerade abgelesen?

„Im Prinzip wird dabei alles gemessen, was im Blut so vorkommt“, erklärt Prof. Dr. Daniel Tenbrinck, Professor für Data Science an der FAU. „Diese riesige Datenmenge hat das Potenzial, uns eine ganze Menge über den Gesundheitszustand von Patientinnen und Patienten zu verraten – nicht nur, an welcher Krankheit sie leiden, sondern möglicherweise sogar, von welcher Variante sie betroffen sind. Oder ob sie zwar ein erhöhtes Risiko für einen Herzinfarkt oder Diabetes haben, aber noch völlig gesund sind, so dass sich die Störung durch prophylaktische Maßnahmen verhindern lässt.“

Suche nach der Nadel im Heuhaufen

Forschende rund um den Globus fahnden deshalb in Omics-Daten nach Auffälligkeiten, die mit bestimmten Krankheiten in Verbindung stehen. Aufgrund der Datenfülle ähnelt diese Aufgabe der sprichwörtlichen Suche nach einer Nadel im Heuhaufen. Daher kommen immer öfter Machine-Learning-Verfahren zum Einsatz, die dabei helfen sollen. „Die künstliche Intelligenz wird mit einer großen Anzahl von Omics-Daten aus Patientinnen und Patienten sowie den bei ihnen diagnostizierten Erkrankungen trainiert“, erklärt Tenbrinck. „Dadurch lernt der Algorithmus, verräterische Spuren in neuen Messwerten zu erkennen und entsprechend zu interpretieren.“

Für das Training der KI werden eigentlich Omics-Daten aus Tausenden von Betroffenen benötigt. Diese zu gewinnen, ist aber ebenso zeitaufwändig wie kostspielig. Tenbrinck möchte daher zusammen mit dem Unternehmen BioVariance eine weitere Strategie nutzen. In der Fachwelt ist sie unter dem Namen „Synthetic Data Generation“, also Erzeugung künstlicher Daten, bekannt. „Wir analysieren dabei mit statistischen Methoden lediglich bis zu 100 Omics-Datensätze und schauen darin nach Mustern und Regelmäßigkeiten“, sagt er. „Diese nutzen wir dann, um neue Datensätze

zu produzieren, die sich statistisch nicht von den Daten tatsächlicher Blutanalysen unterscheiden lassen.“

Mit diesen synthetisch erzeugten Informationen kann man dann die KI trainieren. Was nach einem Taschenspielertrick klingt, hat sich in der Praxis tatsächlich schon vielfach bewährt. „Synthetic Data Generation ist daher in unserem Feld momentan ein sehr aktives Forschungsgebiet“, sagt Tenbrinck. So wird etwa Software für Gesichtserkennung heute oft mit Portraits gefüttert, die zuvor geometrisch etwas verzerrt oder mit Bildrauschen versehen wurden. Der Algorithmus wird auf diese Weise deutlich robuster – er lässt sich nicht mehr so leicht von einem ungünstigen Winkel, unter dem eine Person aufgenommen wurde, oder schlechten Lichtverhältnissen täuschen.

Selbst mit völlig neuen, künstlich erzeugten Bildern lassen sich die Verfahren trainieren. „Dazu muss man aber sicherstellen, dass die synthetischen Gesichter realistisch aussehen“, sagt Tenbrinck. Denn wenn sie etwa alle nur ein Auge haben, wird sich die Erkennungsleistung der damit trainierten Software vermutlich sogar verschlechtern. „Wir untersuchen, wie wir synthetische Omics-Daten erzeugen können, die so realistisch sind, dass sie die Diagnosen der KI tatsächlich robuster und genauer machen“, betont der Wissenschaftler. „Ein wichtiger Punkt dabei ist, dass sich medizinische Expertinnen und Experten die künstlichen Datensätze ansehen und abschätzen, wie plausibel diese sind.“ Bildlich gesprochen, würden die einäugigen Gesichter so direkt aussortiert.

Im Fokus: Long Covid und Depression

Die Partner im BioSamp-Projekt wollen auf diese Weise zunächst die Diagnostik zweier Erkrankungen voranbringen – der schweren Depression und des chronischen Fatigue-Syndroms, einer häufigen Symptomatik bei Long Covid. „Beides sind Störungen, die einen großen Leidensdruck verursachen“, betont Tenbrinck. „Zu Depressionen laufen bei der BioVariance zudem bereits Untersuchungen, auf die wir aufsetzen können.“ Ziel ist es einerseits, diese Störungen sicherer zu identifizieren und möglicherweise in unterschiedliche Varianten einzuteilen. So sprechen zum Beispiel bei einer Depression manche Betroffene auf bestimmte Behandlungsstrategien und Medikamente besser an als andere.

„Wir wollen aber auch dazu beitragen zu erkennen, was bei diesen Krankheiten im Körper genau falsch läuft, wodurch sie also verursacht werden“, erklärt Tenbrinck. Beispielsweise könnte die KI in den Omics-Daten auf ein bestimmtes Gen stoßen, das bei Menschen mit einer Depression besonders aktiv ist. „Dann kann man nachschauen, was zur Funktion dieses Gens in der Forschung bekannt ist, und daraus Schlüsse zur Krankheits-Entstehung ziehen“, sagt der Wissenschaftler. „Unsere Erkenntnisse können also eventuell dazu beitragen, nicht nur die Diagnose von Krankheiten zu verbessern, sondern auch ihre Therapie und Prävention. Das ist es, was ich an dieser Thematik so faszinierend finde.“