KI-Modell prognostiziert Krankheitsrisiken Jahrzehnte im Voraus

Wissenschaftlerinnen und Wissenschaftler vom European Molecular Biology Laboratory (EMBL) und vom Deutschen Krebsforschungszentrum (DKFZ) haben ein KI-Modell entwickelt, das das langfristige individuelle Risiko für mehr als 1.000 Erkrankungen einschätzt. Das Modell, das auf anonymisierten medizinischen Daten aus Großbritannien und Dänemark trainiert und getestet wurde, kann Gesundheitsereignisse für eine Zeitspanne von über einem Jahrzehnt prognostizieren. Das in der Fachzeitschrift Nature vorgestellte Modell ist noch nicht für den klinischen Einsatz bereit, eröffnet aber schon jetzt neue Möglichkeiten, um Gesundheitsstrategien zu entwickeln.

Lässt sich anhand Ihrer persönlichen Krankengeschichte vorhersagen, mit welchen Gesundheitsproblemen Sie in den nächsten zwei Jahrzehnten konfrontiert sein könnten? Dass dies möglich ist, zeigen nun Forschende vom EMBL, vom DKFZ und der Universität Kopenhagen. Sie haben ein generatives KI-Modell entwickelt, das auf der Basis umfangreicher Gesundheitsdaten abschätzt, mit welchen gesundheitlichen Beeinträchtigungen der oder die Einzelne im Laufe der Zeit rechnen muss. Es kann das Risiko und den Zeitpunkt von über 1.000 Krankheiten prognostizieren und Gesundheitsentwicklungen über einen Zeitraum von zehn Jahren vorhersagen.

Die Algorithmen, auf deren Basis das neue generative KI-Modell entwickelt wurde, ähneln denen, die in großen Sprachmodellen (LLMs) verwendet werden. Das Modell wurde zunächst an anonymisierten Patientendaten von 400.000 Teilnehmern aus der UK Biobank trainiert. Anschließend prüften die Forscher es erfolgreich mit Daten von 1,9 Millionen Personen aus dem dänischen nationalen Patientenregister. Das Modell ist die bislang umfassendste Demonstration dafür, wie generative KI den Verlauf menschlicher Krankheiten in großem Maßstab modellieren kann, und wurde anhand von Daten aus zwei völlig getrennten Gesundheitssystemen geprüft.

„Unser KI-Modell ist ein Machbarkeitsnachweis, der zeigt, dass es möglich ist, viele langfristige Gesundheitsmuster zu erkennen und diese Informationen zu nutzen, um aussagekräftige Vorhersagen zu generieren", sagt Ewan Birney vom EMBL. „Indem wir modellieren, wie sich Krankheiten im Laufe der Zeit entwickeln, können wir untersuchen, wann bestimmte Risiken auftreten und wie frühzeitige Interventionen am besten geplant werden können. Das ist ein großer Schritt in Richtung personalisierter und präventiverer Ansätze in der Gesundheitsversorgung."

Die „Grammatik" der Gesundheitsdaten

„So wie große Sprachmodelle aus der Abfolge von Wörtern in Texten die Grammatik unserer Sprache lernen können, lernt dieses KI-Modell die Logik der zeitlichen Abfolge von Ereignissen in Gesundheitsdaten, um ganze Krankengeschichten zu modellieren", erklärt Moritz Gerstung vom DKFZ. Zu diesen Ereignissen gehören medizinische Diagnosen oder auch Lebensstilfaktoren wie Rauchen. An der Reihenfolge, in der die Ereignisse eintreten, und der Zeit, die zwischen diesen Ereignissen vergeht, lernt das Modell, das Krankheitsrisiko vorherzusagen.

„Medizinische Ereignisse folgen oft vorhersehbaren Mustern", sagt Tom Fitzgerald vom Europäischen Bioinformatik-Institut des EMBL (EMBL-EBI). „Unser KI-Modell lernt diese Muster und kann zukünftige Gesundheitsergebnisse prognostizieren. Es gibt uns die Möglichkeit, auf der Grundlage der Krankengeschichte einer Person und anderer wichtiger Faktoren zu untersuchen, was passieren könnte. Entscheidend ist, dass es sich dabei nicht um eine Gewissheit handelt, sondern um eine Einschätzung der potenziellen Risiken."

Das Modell eignet sich für diverse Erkrankungen, insbesondere solche mit klaren und konsistenten Verlaufsmustern, wie Diabetes, Herzinfarkte oder auch Sepsis. Bei Diagnosen wie Infektionskrankheiten, die von unvorhersehbaren Lebensereignissen abhängen, oder sehr seltenen Erkrankungen ist es jedoch weniger zuverlässig.

Wahrscheinlichkeiten, keine Gewissheiten

Wie Wettervorhersagen liefert auch das neue KI-Modell Wahrscheinlichkeiten und keine Gewissheiten. Es kann das Schicksal einer bestimmten Person nicht genau vorhersagen, sondern bietet gut kalibrierte Schätzungen darüber, wie wahrscheinlich

bestimmte Erkrankungen in einem bestimmten Zeitraum auftreten werden. Zum Beispiel die Wahrscheinlichkeit, innerhalb des nächsten Jahres eine Herzerkrankung zu entwickeln. Diese Risiken werden als Zeitraten ausgedrückt, ähnlich wie bei der Vorhersage einer 70-prozentigen Regenwahrscheinlichkeit für morgen. Hier sind Vorhersagen über einen kürzeren Zeitraum für gewöhnlich präziser als langfristige Prognosen.

Die Forschenden konnten zeigen, dass die vom Modell berechneten Wahrscheinlichkeiten tatsächlich mit der erwarteten Häufigkeit eintraten. Wie jedes KI-Modell hat es jedoch auch seine Grenzen. Da die Trainingsdaten aus der UK Biobank beispielsweise hauptsächlich von Personen im Alter von 40 bis 60 Jahren stammen, sind Gesundheitsereignisse im Kindes- und Jugendalter unterrepräsentiert, das gilt auch für bestimmte ethnische Gruppen.

Beispiel Herzinfarkt

Das vom KI-Modell berechnete Risiko eines Herzinfarkts bei Männern im Alter zwischen 60 und 65 variiert zwischen einer Wahrscheinlichkeit von 4 pro 10.000/Jahr und etwa 100 pro 10.000/Jahr, abhängig von früheren Diagnosen und dem Lebensstil der Männer. Frauen haben im Durchschnitt ein geringeres Herzinfarktrisiko, aber eine ähnlich breite Streuung.

Darüber hinaus steigt das Herzinfarkt-Risiko bei Männern und Frauen mit zunehmendem Alter. Eine systematische Bewertung dieser berechneten Risiken in verschiedenen Alters- und Geschlechtsgruppen zeigt, dass sie gut mit der Anzahl von Fällen übereinstimmen, die in einem Teil der UK Biobank Kohorte, die nicht für das Training des Modells benutzt wurden, beobachtet wurden.

Das Modell ist noch nicht für den klinischen Einsatz bereit, könnte aber bereits jetzt Forschern helfen …

- zu verstehen, wie Krankheiten sich im Laufe der Zeit entwickeln und fortschreiten.
- zu untersuchen, wie sich Lebensstil und frühere Erkrankungen auf das langfristige Krankheitsrisiko auswirken.
- Gesundheitsergebnisse anhand künstlicher Patientendaten zu simulieren, wenn reale Daten schwer zu beschaffen oder zugänglich sind.

In Zukunft könnten KI-Tools, die auf repräsentativeren Datensätzen trainiert wurden, Ärzten dabei helfen, Hochrisikopatienten frühzeitig zu identifizieren. Angesichts der alternden Bevölkerung und der steigenden Rate chronischer Erkrankungen könnte die Fähigkeit, zukünftige Gesundheitsbedürfnisse vorherzusagen, den Gesundheitssystemen helfen, besser zu planen und Ressourcen effizienter zuzuweisen. Bevor KI-Modelle jedoch in einer klinischen Umgebung eingesetzt werden können, sind noch viele weitere Tests sowie robuste regulatorische Rahmenbedingungen erforderlich.

„Das ist der Beginn einer neuen Art, die menschliche Gesundheit und den Verlauf von Krankheiten zu verstehen", prognostiziert Moritz Gerstung. „Solche generativen Modelle könnten eines Tages dazu beitragen, die Versorgung zu personalisieren und den Bedarf an medizinischer Versorgung in großem Maßstab zu antizipieren. Durch das Lernen aus großen Populationen bieten diese Modelle einen aussagekräftigen Einblick in den Verlauf von Krankheiten und könnten letztendlich frühzeitigere, maßgeschneiderte Interventionen unterstützen."

Das KI-Modell wurde unter strengen ethischen Regeln mit anonymisierten Gesundheitsdaten trainiert. Die Teilnehmer der UK Biobank gaben ihre Einwilligung, und auf die dänischen Register wurde gemäß den nationalen Vorschriften zugegriffen, die vorschreiben, dass die Daten innerhalb Dänemarks bleiben müssen. Die Forscher verwendeten sichere, virtuelle Systeme, um die Daten zu analysieren, ohne sie über Grenzen hinweg zu übertragen. Diese Sicherheitsvorkehrungen tragen dazu bei, dass KI-Modelle unter Wahrung der Privatsphäre und unter Einhaltung ethischer Standards entwickelt und eingesetzt werden.

Die Arbeit wurde durch Beiträge der EMBL-Mitgliedstaaten, Mittel des DKFZ und einen Zuschuss der Novo Nordisk Foundation finanziert.

Publikation:
Artem Shmatko, Alexander Wolfgang Jung, Kumar Gaurav, Søren Brunak, Laust Mortensen, Ewan Birney, Tom Fitzgerald & Moritz Gerstung:
Learning the natural history of human disease with generative transformers, Nature 2025
DOI: 10.1038/s41586-025-09529-3

Pressemitteilung

17.09.2025
Quelle: Deutsches Krebsforschungszentrum (DKFZ)

Weitere Informationen

- Deutsches Krebsforschungszentrum (DKFZ)