

Deep Learning erkennt molekulare Muster von Krebs

Eine Plattform für künstliche Intelligenz, die am MDC entwickelt wurde, kann genomische Daten extrem schnell analysieren. Sie filtert wesentliche Muster heraus, um Darmkrebs zu klassifizieren und die Entwicklung von Wirkstoffen zu verbessern. Einige Darmkrebs-Arten müssen demnach neu geordnet werden.

Ein neuer Deep-Learning-Algorithmus kann verschiedene Arten von genomischen Daten, die aus kolorektalen Karzinomen gewonnen wurden, schnell und präzise analysieren und so genauer klassifizieren. Dies könnte die Diagnose und damit verbundene Behandlungsoptionen verbessern, berichten die Forscher*innen im Fachjournal Life Science Alliance.

Kolorektale Tumore entwickeln sich äußerst unterschiedlich, sie reagieren auf unterschiedliche Medikamente und auch die Überlebensraten sind sehr verschieden. Meist werden sie auf der Basis von Genexpressionsdaten in Subtypen klassifiziert.

„Die meisten Krankheiten sind deutlich komplexer als ein einzelnes Gen“, sagt Dr. Altuna Akalin, der Leiter der Forschungsgruppe Bioinformatik am Berliner Institut für Medizinische Systembiologie (BIMSB) des Max-Delbrück-Centrums für Molekulare Medizin (MDC). „Um diese Komplexität zu erfassen, brauchen wir irgendeine Art von maschinellern Lernen, die wirklich sämtliche Daten verarbeiten kann.“

Um die zahlreichen im genetischen Material vorhandenen Merkmale, einschließlich der Genexpression, Punktmutationen und strukturelle Veränderungen, bei denen ein DNA-Abschnitt mehrfach erzeugt wird (CNV, copy number variants), analysieren zu können, haben Akalin und sein Doktorand Jonathan Ronen die „Multi-omics Autoencoder Integration“-Plattform, kurz „maui“, entworfen.

Das Prinzip

Für überwachtes maschinelles Lernen braucht man normalerweise menschliche Expert*innen, die die Daten kennzeichnen und den Algorithmus dann so trainieren, dass er diese Kennzeichen vorhersagen kann. Um zum Beispiel die Augenfarbe anhand von Bildern von Augen vorherzusagen, müssen Forscher*innen den Algorithmus zunächst mit Bildern füttern, in denen die Augenfarbe gekennzeichnet ist. Der Algorithmus lernt daraufhin, verschiedene Augenfarben zu erkennen, und kann dann neue Daten selbstständig analysieren.

Beim unüberwachten maschinellen Lernen ist hingegen kein Training notwendig. Ein Deep-Learning-Algorithmus wird mit Daten ohne Kennzeichnungen gefüttert und sucht diese, um gemeinsame Muster oder typische Eigenschaften – in der Fachsprache heißen sie latente Faktoren – zu finden. Diese Art von Algorithmus kann zum Beispiel Bilder von Gesichtern verarbeiten, die in keiner Weise gekennzeichnet sind, und dabei Schlüsselmerkmale wie Augenfarben, Augenbrauen- und Nasenformen oder Lächeln erkennen.

Als Deep-Learning-Plattform ist maui in der Lage, mehrere Omics-Datensätze zu analysieren und die wichtigsten Muster oder Merkmale, in diesem Fall Gensätze oder Indikatoren für Darmkrebs, zu erkennen.

Neuklassifikation der Subtypen?

Die maui-Plattform erkannte in den Daten Muster, die mit den vier bekannten Subtypen kolorektaler Karzinome übereinstimmen, und ordnete Tumore diesen Subtypen mit hoher Präzision zu. Sie hat noch eine interessante Entdeckung gemacht. Sie fand ein Muster, das nahelegt, dass ein Subtyp (CMS2) gegebenenfalls in zwei verschiedene Gruppen unterteilt werden muss. Die Tumore weisen verschiedene Mechanismen und Überlebensraten auf. Das Team schlägt weitere Untersuchungen vor, um festzustellen, ob der Subtyp einzigartig ist oder generell charakteristisch für eine Tumorausbreitung. In jedem Fall zeigt das Ergebnis, wozu die Plattform fähig ist: Sie kann nicht nur die bekannten und bereits mit der Krankheit in Verbindung gebrachten Gene, sondern auch alle anderen Daten berücksichtigen, und ermöglicht damit tiefere Einblicke.

„Mittels datenwissenschaftlicher Methoden lassen sich Erkenntnisse auch aus normalerweise schwer interpretierbaren komplexen Daten gewinnen“, sagt Akalin. „Man kann Algorithmen mit allen Daten, die zu Tumoren vorliegen, füttern und sie werden sinnvolle Muster finden.“

Schneller und besser

Das Programm war nicht nur genauer, es arbeitet auch schneller als andere Algorithmen des maschinellen Lernens – nur drei Minuten braucht es, um 100 Muster herauszufiltern. Andere Programme benötigten dafür 20 Minuten oder sogar elf Stunden.

„Das Programm ist in der Lage, in einem Bruchteil der Rechenzeit eine um Größenordnungen höhere Zahl an latenten Faktoren zu lernen“, erläutert Jonathan Ronen, Erstautor der Studie.

Das Team war überrascht davon, wie schnell das System arbeitet, insbesondere weil die Forschenden keine Grafikkarten verwendeten, die die Berechnungen normalerweise beschleunigen. Das zeigt, wie ausgesprochen gut optimiert und effizient der Algorithmus bereits ist, auch wenn das Team weiter daran arbeitet, das System noch zu verbessern.

Wirkstoffentwicklung verbessern

Um die Wirkung potenzieller Medikamente zu untersuchen, passte das Team, dem auch Sikander Hayat von der Bayer AG angehörte, das Programm etwas an: Es kann nun auch Zelllinien analysieren, die Tumoren entnommen bzw. im Labor gezüchtet wurden. Auf molekularer Ebene unterscheiden sich Zelllinien jedoch auf vielerlei Weise von echten Tumoren. Um das Ausmaß der Unterschiede abzuschätzen, verglich das Team mithilfe von maui Zelllinien, an denen derzeit Wirkstoffe gegen Darmkrebs getestet werden, mit Zellen aus echten Tumoren. Knapp die Hälfte der Zelllinien war demnach enger mit anderen Zelllinien verwandt als mit echten Tumoren. Nur eine Handvoll Linien ähneln den verschiedenen Arten kolorektaler Karzinome am meisten.

Die Suche nach neuen Medikamenten verlässt sich zwar längst nicht nur auf Zelllinien, diese Erkenntnis könnte aber dazu beitragen, das volle Potenzial der Zelllinienforschung besser auszuschöpfen. Möglicherweise lässt sie sich auch für andere Arten der Wirkstoffprüfung anpassen, die auf genetischen Informationen basiert.

Google für Tumore

Nachdem die Deep-Learning-Plattform für Darmkrebs umfassend getestet wurde, könnten damit auch Daten neuer Patient*innen analysiert werden. „Man kann es sich wie eine Suchmaschine vorstellen“, sagt Akalin.

Ein Arzt oder eine Ärztin könnte die genetischen Daten eines Erkrankten in maui einspeisen, um die

beste Übereinstimmung zu finden und so den Tumor schnell und genau zu klassifizieren. Die Plattform könnte dann Medikamente empfehlen, die bei ähnlichen Tumoren gut angeschlagen haben. So könnte sie voraussagen helfen, ob eine bestimmte Therapie etwas nützt und wie die Überlebensrate sind.

Derzeit ist dies nur in einem akademischen Umfeld möglich und wenn die Ärzt*innen zuvor alle vorhandenen klinischen Protokolle ausprobiert hatten. Es sei ein langer Weg bis zur Zulassung eines Tests oder Systems für den klinischen Einsatz, sagt Akalin. Das Team wägt mit der Unterstützung des [Digital Health Accelerator Programms des Berlin Institute of Health](#) das Potenzial für die Vermarktung des Systems ab. Darüber hinaus entwickeln sie mau für die Anwendung auf andere Krebsarten weiter.

Text: Laura Petersen

Literatur

Jonathan Ronen et al. (2019): „Evaluation of colorectal cancer subtypes and cell lines using deep learning“, Life Science Alliance, DOI: [10.26508/lsa.201900517](https://doi.org/10.26508/lsa.201900517)