

KI identifiziert Krebszellen

Wie unterscheiden sich krebskranke von gesunden Zellen? Ein neuer Machine-Learning-Algorithmus namens „ikarus“ kennt die Antwort, berichtet ein Team um den Bioinformatiker Altuna Akalin vom MDC nun im Fachjournal „Genome Biology“. Das Programm hat eine charakteristische Gensignatur gefunden.

Wenn es darum geht, in Datenbergen Muster zu identifizieren, ist ein Mensch einer künstlichen Intelligenz (KI) chancenlos unterlegen. Besonders das maschinelle Lernen, ein Teilbereich der KI, wird oft eingesetzt, um Gesetzmäßigkeiten in Datensätzen zu finden – sei es zur Aktienmarktanalyse, Bild- und Spracherkennung oder der Klassifizierung von Zellen. Um Krebszellen zuverlässig von gesunden Zellen zu unterscheiden, hat ein Team um Dr. Altuna Akalin, Leiter der Technologieplattform „Bioinformatik und Omics-Datenwissenschaft“ am Max-Delbrück-Centrum für Molekulare Medizin in der Helmholtz-Gemeinschaft (MDC), nun ein Machine-Learning-Programm namens „ikarus“ entwickelt. In den Tumorzellen fand das Programm ein krebsübergreifendes Muster, bestehend aus einer charakteristischen Kombination an Genen. Der Algorithmus entdeckte in dem Muster außerdem Arten von Genen, die man bislang nicht eindeutig mit Krebs in Verbindung gebracht hatte, schreibt die Forschungsgruppe im Fachjournal „Genome Biology“.

Maschinelles Lernen bedeutet im Grunde, dass ein Algorithmus anhand von Trainingsdaten selbstständig lernt, bestimmte Fragestellungen zu beantworten. Seine Strategie ist dabei, nach Mustern in den Daten zu suchen, die ihm bei der Problemlösung helfen. Nach der Trainingsphase kann das System das Gelernte verallgemeinern und somit unbekannte Daten beurteilen. „Eine große Herausforderung war, geeignete Lerndatensätze zu bekommen, bei denen Fachleute bereits eine präzise Einteilung der Zellen in ‚gesund‘ und ‚krebskrank‘ vorgenommen hatten“, erzählt Jan Dohmen, der Erstautor der Studie.

Eine überraschend gute Trefferquote

Obendrein sind Datensätze aus Einzelzell-Sequenzierungen häufig verrauscht. Das bedeutet: Die Informationen über die molekularen Eigenschaften der einzelnen Zellen sind nicht ganz genau – weil zum Beispiel in jeder Zelle eine unterschiedliche Anzahl Gene erkannt wird oder die Proben nicht immer gleich verarbeitet werden. Sie hätten unzählige Publikationen durchforstet und etliche Forschungsgruppen kontaktiert, um ausreichend gute Datensätze zu bekommen, berichten Dohmen und sein Kollege Dr. Vedran Franke, der Ko-Leiter der Studie. Mit Daten von Lungen- und Darmkrebszellen trainierten das Team den Algorithmus schließlich, bevor sie ihn auf Datensätze von weiteren Tumorarten anwendeten.

In der Trainingsphase musste ikarus eine Liste charakteristischer Gene finden, anhand derer das Programm die Zellen einteilen konnte: „Wir haben verschiedene Ansätze ausprobiert und verfeinert“, sagt Dohmen. Eine zeitintensive Arbeit, wie sich alle drei Forscher erzählen. „Ausschlaggebend war, dass ikarus letztlich zwei Listen nutzte: eine für Krebsgene und eine für Gene anderer Zellen“, erklärt Franke. Nach der Lernperiode konnte der Algorithmus auch bei anderen Krebsarten zuverlässig zwischen gesunden und krebskranken Zellen unterscheiden, etwa in Gewebeproben von Leberkrebs oder Neuroblastomen. Seine Trefferquote lag meist nur wenige Prozent daneben. Das hat auch die Forschungsgruppe überrascht: „Wir haben nicht erwartet, dass eine gemeinsame Signatur existiert, die Tumorzellen von verschiedenen Krebsarten so genau

definiert“, sagt Akalin. „Noch können wir allerdings nicht sagen, dass die Methode für alle Krebsarten funktioniert“, fügt Dohmen hinzu. Damit iKarus zuverlässig bei der Krebsdiagnose helfen kann, wollen die Forschenden ihn noch an weiteren Tumorarten testen.

KI als vollautomatische Diagnose-Hilfe

Die Klassifizierung „gesund“ versus „krebskrank“ ist dabei längst nicht das Ende des Projekts. In ersten Tests konnte iKarus bereits zeigen, dass sich die Methode auch andere Zelltypen oder bestimmte Subtypen von Tumorzellen unterscheiden kann. „Wir wollen den Ansatz verallgemeinern“, sagt Akalin, „also ihn derart weiterentwickeln, dass er alle möglichen Zelltypen in einer Biopsie unterscheiden kann“.

In der Klinik schauen sich Pathologen Gewebeproben von Tumoren meist nur unter dem Mikroskop an und identifizieren so die unterschiedlichen Zelltypen. Da ist mühsam und kostet viel Zeit. Mit iKarus könnte dieser Schritt irgendwann vollautomatisch ablaufen. Außerdem könne man aus den Daten zusätzlich etwas über die unmittelbare Umgebung des Tumors ableiten, sagt Akalin. Das wiederum könnte den Ärztinnen und Ärzten helfen, eine optimale Therapie auszuwählen. Denn oftmals deute die Zusammensetzung des Krebsgewebes und der Mikroumgebung darauf hin, ob eine bestimmte Behandlung oder ein Medikament anschlagen wird oder nicht. Darüber hinaus hilft die KI möglicherweise, neue Medikamente zu entwickeln: „Wir können mit iKarus Gene identifizieren, die potenzielle Treiber der Krebserkrankung sind“, sagt Akalin. Neuartige Wirkstoffe könnten dann an diesen molekularen Zielstrukturen ansetzen.

Zusammenarbeit im Home-Office

Bemerkenswert an der Publikation sei, dass die notwendigen Arbeiten vollständig während der Coronapandemie durchgeführt wurden. Alle Beteiligten waren zu der Zeit nicht an ihren normalen Arbeitsplätzen im Berliner Institut für Medizinische Systembiologie (BIMSB), das zum MDC gehört. Sie hielten im Home-Office nur über digitale Kanäle Kontakt. „Das Projekt beweist, dass man eine digitale Struktur schaffen kann, die wissenschaftliche Arbeiten unter diesen Bedingungen ermöglicht“, findet Franke daher.

Berliner Institut für Medizinische Systembiologie (BIMSB)

[Weitere Informationen](#)

Literatur

Jan Dohmen et al. (2022): „Identifying tumor cells at the single-cell level using machine learning“. Genome Biology, DOI: [10.1186/s13059-022-02683-1](https://doi.org/10.1186/s13059-022-02683-1)