

Algorithmen lernen in der Biobank alles über Krebs

Grazer Pathologen digitalisieren gemeinsam mit IT-Experten in großem Stil Tumorproben. So können sie Algorithmen trainieren, Krebs zu erkennen und Prognosen zu stellen.

Veronika Szentpétery-Kessler

Wer schon einmal seine alte Diafilm- oder Fotosammlung digitalisiert hat, weiß, wie aufwendig das ist. Für ein gutes Ergebnis braucht es gute Scanner und leistungsfähige Bildsoftware. Die brauchen auch Wissenschaftler, wenn sie Gewebeschnitte aus Tumorproben digitalisieren. Das Diagnostik- und Forschungszentrum für Molekulare Biomedizin an der Medizinischen Universität Graz unter der Leitung von Kurt Zatloukal setzte dabei schon früh auf die neuesten hochauflösenden und mit Hochdurchsatz arbeitenden Scanner-Technologien.

„Wir haben bereits 2017 in einen ersten Prototyp investiert und mit dem Hersteller an der Entwicklung mitgearbeitet, weil wir gesehen haben, dass der Bedarf an Hochdurchsatzdigitalisierung massiv steigen wird“, erzählt Zatloukal, der auf molekulare Pathologie spezialisiert ist. Der Bedarf bestand nicht allein in der Archivierung der Präparate. Durch die Verbreitung von Maschinelernen und künstlicher Intelligenz in der Medizin bekamen große Biobanken wie die seit 1993 in Graz bestehende „eine komplett neue Relevanz“. Denn mit den riesigen Mengen an histologischen Schnitten etwa aus Dickdarm-, Prostata-, Brust- und Lungentumoren lassen sich Algorithmen für das Erkennen wichtiger Veränderungen trainieren. Das Ziel: durch bessere Krebsdiagnosen die Therapie gezielter einsetzen zu können.

Seit drei Jahren arbeiten Pathologen wie Zatloukal am Zentrum für Wissens- und Technologietransfer (ZWT) mit IT-Experten zusammen, um die Grazer Biobank mit inzwischen elf Hochleistungsscannern zu digitalisieren. Pro Jahr können circa 800.000 Präparate gescannt werden. Gut trainierte Algorithmen zeigten eine mindestens ebenso gute Leistung wie ein erfahrener Facharzt – wenn nicht sogar eine bessere, sagt Zatloukal. „Das konnten wir bei Dickdarm- und Prostatakarzinomen sehr schön zeigen. Die

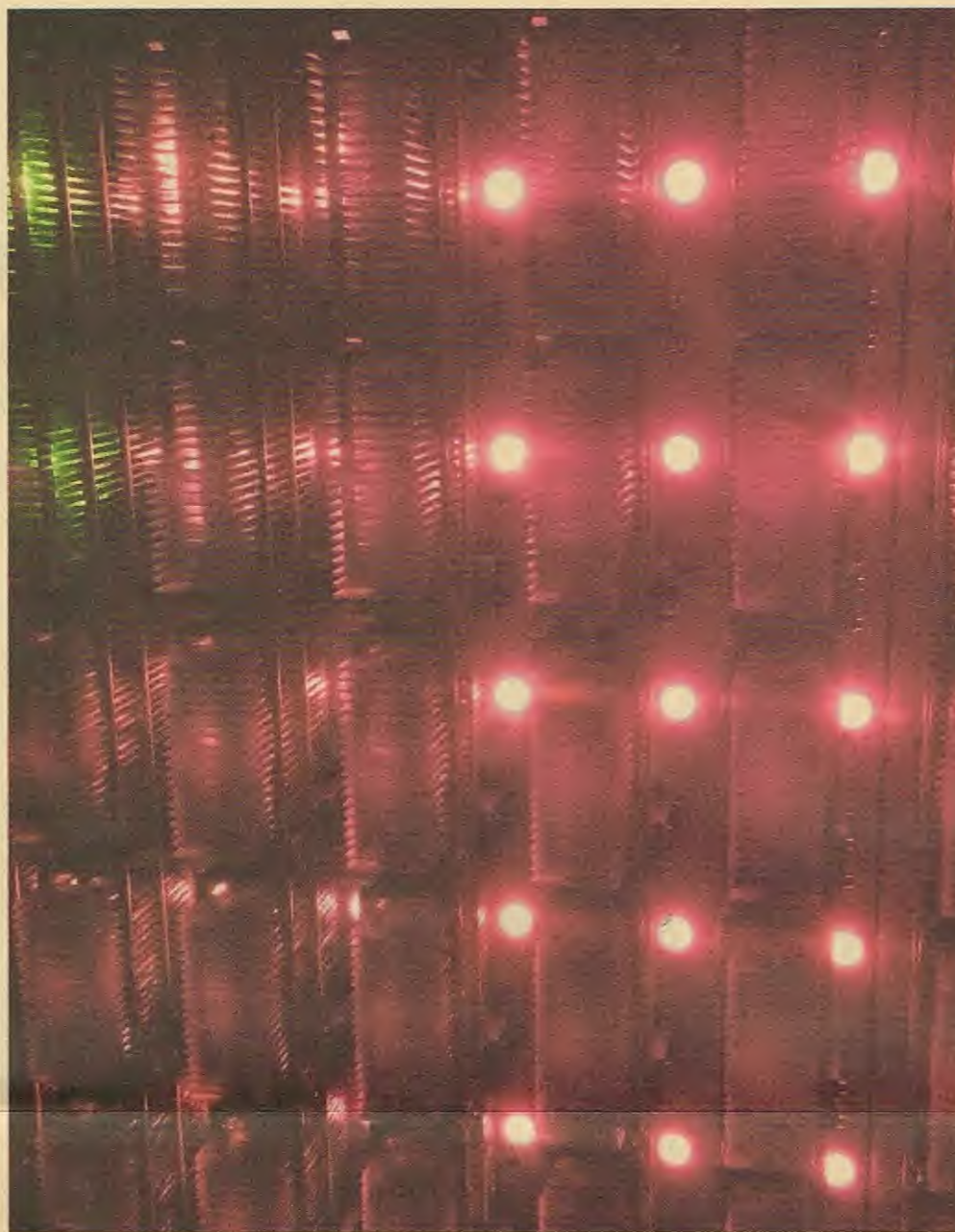


Foto: Med-Uni Graz

Mit speziellen Hochleistungsscannern werden im Zentrum für Wissens- und Technologietransfer in Graz jährlich bis zu 800.000 Biobankproben digitalisiert.

Ergebnisse sind gerade im Publikationsprozess.“ Denn Algorithmen ermüden nicht und geraten nicht in Stress, wenn ein Telefon läutet oder drei Leute zugleich etwas wollen.

Standardisierte Diagnosen

Bisher war die Gewebeauswertung eine Domäne der Pathologen. Ihr Urteil über angefärbte Biomarkerveränderungen und andere Muster ist zum Beispiel für die personalisierte Krebsmedizin, also die Auswahl der individuell am besten geeigneten Mittel, wichtig. Die Mediziner verlassen sich dabei auf ihre Erfahrung, doch die kann variieren. Nicht jeder liest also dasselbe aus einem Schnitt heraus. Algorithmen würden dagegen immer gleich hohe Kriterien anlegen und standardi-

sierte und damit besser vergleichbare Diagnosen liefern, sagt Zatloukal.

„Wir Menschen erkennen komplexe Strukturen optisch sehr gut, aber wir quantifizieren sehr schlecht. Ich sage immer: Einen Baum erkennen wir sofort. Aber wir können nicht sagen, wie viele Blätter er hat. Das ist die Stärke von Algorithmen.“ Insgesamt sollen aber „die kognitiven Stärken des Menschen durch die analytischen Kapazitäten eines Algorithmus ergänzt werden“, ohne jemanden zu ersetzen.

Wie aber lernen die Algorithmen? Der klassische Weg ist Zatloukal zufolge, dass der Experte festlegt, welche Veränderungen wie klassifiziert werden. Nach vielen hundert bis tausend Wiederholungen klassifiziert der Algorithmus dann selbst-

ständig. „Wir haben einen neuen Weg versucht und den Algorithmus gleich selbst klassifizieren lassen, basierend auf der Information, wie lange ein Mensch mit einem bestimmten Tumor überlebt hat“, schildert der Pathologe. Die Algorithmen fanden selbst Merkmale, die mit einer Krankheitsprognose in Zusammenhang stehen.“ Künftig, so die Hoffnung der Forscher, könne man auf diesem Weg auch neue morphologische Informationen nutzen, die wegen zu großer Komplexität bisher diagnostisch nicht nutzbar waren.

Dabei dürfe der Algorithmus „keine Blackbox“ sein. „Fachleute müssen nachvollziehen können, was er in seine Berechnung einbezogen hat. Dann kann der Mensch

überlegen, ob das plausibel ist“, beschreibt Zatloukal. „Denn ein Arzt kann nur Verantwortung für etwas übernehmen, das er nachvollziehen kann.“ Sein Kollege Andreas Holzinger arbeitet intensiv an dieser Plausibilitätsprüfung.

Diagnostische Algorithmen seien genauso Diagnostika wie Labortests und müssen für die Zulassung ihre analytische Leistung zeigen. Deshalb entwickelt Zatloukals Team einen eigenen ISO-Standard, der definiert, wie die Daten erhoben werden, wie sich ein Algorithmus trainieren lässt und innerhalb welcher Grenzen er ein qualitativ gutes Ergebnis liefern kann.

Sichere Datenberge

Zatloukal hat sich auch über die sichere Speicherung der riesigen Datenmengen Gedanken gemacht, die beim Scannen produziert werden. Die Digitalisierung eines einzelnen Tumorschnitts erzeugt etwa zehn Gigabyte an Daten. Für das Algorithmentraining brauche man aber hunderttausende Schnitte. Ein Trainingsprojekt für das Klassifizieren von Dickdarmkarzinomen produzierte gar zwei Petabyte an Daten. Solche Größenordnungen waren vor wenigen Jahren nur aus der Astronomie – etwa in Verbindung mit Radioteleskopen – bekannt.

Um solche Datenmengen sicher zu speichern, setzen die Forscher auf eine neue Cloud-Technologie des Austrian Institute of Technology (AIT) und von fragmentIX Storage Solutions. Ihr Secret-Sharing-Verfahren unterteilt die Daten zunächst in Fragmente und speichert sie dann verschlüsselt auf verschiedenen Servern und bei verschiedenen Cloud-Providern. Selbst wenn ein Server gehackt und ein Fragment ausgelesen wird, lässt sich daraus keine Information rekonstruieren. Das gleichzeitige Hacken aller beteiligten Server wäre nur schwer möglich. Umgekehrt könnten die Forscher ein verlorengegangenes Fragment wieder gut rekonstruieren.