

Hightech im Darm: Wo KI den Menschen (noch) (nicht) übertrifft

Kann ein KI-gestütztes System zur Polypenerkennung die Darmkrebsvorsorge verbessern?

Prof. Dr. Alexander Hann vom Uniklinikum Würzburg (UKW) hat mit EndoMind ein KI-gestütztes System (Computer-Aided Detection, CADE) entwickelt, welches die Detektionsrate von Adenomen (ADR) erhöhen soll. Die KI EndoMind markiert bei einer Darmspiegelung erkannte Adenome mit einem blauen Rechteck. EndoMind wurde nun in gastroenterologischen Schwerpunktpraxen getestet. Die Ergebnisse hat das Nature-Journal [NPJ Digital Health](#) veröffentlicht. Sie zeigen: Die KI hilft erfahrenen Gastroenterologinnen und Gastroenterologen kaum dabei, die Detektionsrate von Polypen bei einer Darmspiegelung (Koloskopie) zu verbessern. Dennoch bietet KI viel Potenzial.

Ein Interview mit Alexander Hann, Professor für Digitale Transformation in der Gastroenterologie am UKW und Leiter der Arbeitsgruppe InExEn - Interventional and Experimental Endoscopy.

Herr Hann, die Ergebnisse der aktuellen Studie sind ernüchternd. Doch Sie wirken gar nicht enttäuscht.

Zunächst einmal freut es mich, dass wir mit dieser Arbeit zeigen konnten, dass anhand ambulanter Daten aus Deutschland in einem universitären Umfeld eine KI für die Darmkrebsvorsorge entwickelt werden konnte, die mit KI-Systemen millionenschwerer Unternehmen mithalten kann, und die dort getestet wurde, wo Vorsorge jeden Tag durchgeführt wird: in gastroenterologischen Schwerpunktpraxen. Diese betreiben in der Regel keine Forschung. Dort geht es vornehmlich darum, die Patientinnen und Patienten bestmöglich zu versorgen. Eine Studie mitsamt Aufklärung etc. bedeutet zusätzlichen Aufwand. Wir haben es geschafft, diese Schwelle so niedrig wie möglich zu gestalten und zum Beispiel die Verschlüsselung der Daten so sicher, aber auch so einfach wie möglich gemacht. Darauf bin ich schon etwas stolz.

Ich freue mich, dass dieses Projekt, das das längste wissenschaftliche Projekt meiner Karriere war, nun endlich veröffentlicht wurde. Dass wir die Studie in einem so hochrangigen Journal publizieren konnten, macht zudem Mut. Das zeigt, dass wir an einer Uniklinik alles haben, um etwas zu entwickeln und zu validieren, was sonst nur große Unternehmen können.

Aktuelle Publikation: Lux, T.J., Saßmannshausen, Z., Kafetzis, I. *et al.* Artificial intelligence assisted colorectal lesion detection in private practices a randomized controlled study. *npj Digit. Med.* 9, 284 (2026). <https://doi.org/10.1038/s41746-026-02576-8>

Wie lange dauerte denn das Forschungsprojekt?

Die Ausschreibung zur Förderung von neuen Technologien in der Medizin erfolgte im Dezember 2018, zu dieser Zeit war ich noch an der Uniklinik Ulm tätig. Mein ehemaliger Chef aus Stuttgart, Wolfram Zoller, machte mich darauf aufmerksam. Ich war gerade im Urlaub und hatte nur wenig Zeit, mich zusammen mit ihm zu bewerben, freute mich aber über die Gelegenheit. Ich notierte meine Ideen und Projektskizzen auf Servietten. Das Konzept reichten wir dann im Januar 2019 ein. Anfang 2020 wurden wir aufgefordert, den Vollantrag zu schreiben. Zu diesem Zeitpunkt war ich bereits in Würzburg. Der Antrag wurde genehmigt, sodass wir im Juli 2020 starten konnten.

Innerhalb eines knappen Jahres haben wir gemeinsam mit dem Lehrstuhl für Künstliche Intelligenz unter der Leitung von Professor Puppe (jetzt am CAIDAS) die KI erstellt. Neben den Informatikern Adrian Krenzer und Michael Bank war auch der Facharzt Daniel Fitting maßgeblich beteiligt. Die KI haben wir in einer Pilotstudie an verschiedenen Zentren getestet und 2022 publiziert.

Publikation: Lux, T. J., Banck, M., Saßmannshausen, Z. et al.: Pilot Study of a New Freely Available Computer-Aided Polyp Detection System in Clinical Practice. Int J Colorectal Dis 37, 1349-1354 (2022). <https://doi.org/10.1007/s00384-022-04178-8>).

In einer weiteren Studie haben wir unsere KI in einem Kopf-an-Kopf-Rennen mit kommerziell erhältlichen KIs verglichen.

Publikation: Troya, J. et al.: Direct comparison of multiple computer-aided polyp detection systems. Endoscopy 2024; 56(01): 63-69, DOI: [10.1055/a-2147-0571](https://doi.org/10.1055/a-2147-0571).

Dabei zeigte sich, dass unsere KI nicht schlechter als die anderen ist. Die randomisierte Studie dazu lief schließlich von November 2021 bis November 2022 und wurde nun endlich publiziert.

Wenn die KI nicht besser ist als der Mensch und erfahrene Untersuchende somit immer noch den Goldstandard darstellen, sollte sie dennoch zum Einsatz kommen?

Das ist eine gute Frage. Sie wird aber nicht durch eine einzige Studie beantwortet, sondern durch viele Studien. Wissenschaft ist schließlich auch dazu da, Dinge zu hinterfragen. Inzwischen gibt es über 45 randomisierte Studien zu diesem Thema. Als wir das Projekt gestartet haben, gab es nur wenige, die sich aber optimistisch zeigten. Wir konnten für unser Kollektiv das realistische Bild aufzeigen, dass die KI in diesem Fall nicht so viel bringt. Sie detektiert vielleicht ein bisschen mehr, aber nicht signifikant mehr.

Es gibt jedoch auch Negativdaten. Ein Stichwort ist hier Deskillung. Führt der Einsatz von KI bei Darmspiegelungen dazu, dass Ärztinnen und Ärzte ihre eigenen diagnostischen Fähigkeiten teilweise verlernen?

Eine polnische Studie zeigte ein widersprüchliches Bild: Während die Ärztinnen und Ärzte mit eingeschalteter KI mehr Polypen fanden, sank ihre Leistung im Vergleich zu früher, an den Tagen als die KI ausgeschaltet war. Dies deutet darauf hin, dass sie sich möglicherweise an die Unterstützung gewöhnt haben und selbst weniger aufmerksam nach Polypen gesucht haben. Diese Interpretation wird durch frühere Eye-Tracking-Daten gestützt, die zeigen, dass sich das Blickverhalten unter KI verändert.

Publikation: Troya J, Fitting D, Brand M, Sudarevic B. et al.: „The influence of computer-aided polyp detection systems on reaction time for polyp detection and eye gaze. Endoscopy. 2022 Oct;54(10):1009-1014. doi: [10.1055/a-1770-7353](https://doi.org/10.1055/a-1770-7353)

Es gibt allerdings auch eine länger angelegte Studie aus Japan, die über mehrere Jahre keinen solchen negativen Effekt feststellen konnte. Insgesamt ist die Studienlage uneinheitlich, sodass noch keine eindeutige Aussage darüber getroffen werden kann, ob und wie stark Deskillung durch KI tatsächlich auftritt.

Eine Metaanalyse hat gezeigt, dass die KI 7,4 Prozent mehr Polypen detektiert.

Genau. Es hat sich herausgestellt, dass es sich um eher kleine Polypen handelt. Daraufhin wurde eine Mikrosimulation durchgeführt. Wenn ich 10.000 Personen mit KI untersuche und zehn Jahre lang begleite, sinkt die Rate neu aufgetretener Kolonkarzinome von 82 auf 71 Prozent. Das heißt, bei

100 Darmspiegelungen wird nur noch bei 71 Personen ein neu aufgetretenes Kolonkarzinom gefunden. Die Rate an kolonkarzinombedingten Todesfällen reduziert sich von 15 auf 13.

Diese Daten wurden drei Fachgesellschaften vorgelegt. Jede von ihnen gab eine andere Empfehlung ab.

Die American Society of Gastroenterology sprach keine Empfehlungen aus und positionierte sich somit weder gegen KI bei der Vorsorge noch dafür. Das British Medical Journal (BMJ) sah nach dieser Mikrosimulation wenig Nutzen und sprach sich gegen den Einsatz von KI aus. Die Europäische Fachgesellschaft für Endoskopie (ESGI) sprach hingegen eine schwache Empfehlung aus.

Wie lautet Ihre Empfehlung?

Das Risiko, dass aus einem kleinen Adenom Krebs entsteht, ist zwar sehr gering. Wenn ich Patient wäre, würde ich mir jedoch wünschen, dass auch die kleinen Adenome entfernt werden. Denn ich möchte nicht zur Risikogruppe gehören. In der Ausbildung sollten wir allerdings auf Bewährtes setzen, bis eindeutig geklärt ist, ob es negative Einflüsse gibt. Es ist jedoch sinnvoll, junge Auszubildende auf einen Polypen hinzuweisen, sei es durch KI oder den Auszubildenden. Dadurch schulen sie ihr Auge.

EndoMind steht auf Ihrer Website zum freien Download zur Verfügung. Wie oft und von wem wird die Software heruntergeladen?

Bis April 2026 haben 366 Personen EndoMind heruntergeladen, nach Zugriffsstatistiken der Homepage hauptsächlich aus den USA, Russland, China und vor allem aus Japan. Das liegt wahrscheinlich daran, dass fast alle großen Endoskopiefirmen aus Japan kommen. Sie wollen sich die Software anschauen und mit ihrer eigenen vergleichen. Da EndoMind kein Medizinprodukt ist, darf es außerhalb von Studien nicht getestet werden. Wir liefern jedoch mit EndoMind und anderen KIs Ideen für künftige Medizinprodukte oder Verbesserungen.

Was müsste an den aktuellen KI-Modellen noch verbessert werden?

Ein Problem sind die eigentlich relevanten fortgeschrittenen Adenome. Die KIs werden jedoch hauptsächlich mit kleineren Polypen trainiert. Meines Wissens gibt es noch keine KI, die große Adenome oder solche mit deutlich mehr krankhaften Veränderungen ebenso verlässlich wie das menschliche Auge erkennt - oder gar besser. Das ist leider nicht so trivial. Wir bekommen häufiger Fälle zugewiesen, in denen wir größere Polypen finden, insbesondere flache Polypen, die teilweise übersehen wurden.

Sie arbeiten ja auch an einigen Verbesserungen und Qualitätsindikatoren. Im Projekt „Poseidon“ beschäftigen Sie sich beispielsweise mit einer KI-gestützten Größenbestimmung der Polypen. Dabei hilft ein Wasserstrahl bei der Einordnung der Größe. Ein weiteres Thema ist die Rückzugszeit. Was heißt das?

Um den gesamten Darm zu untersuchen, also vom Beginn des Kolons, dem Zökum, bis zum Ende, dem Rektum, muss man sich mindestens sechs bis neun Minuten Zeit lassen. Studien zeigen, dass Patientinnen und Patienten zwischen zwei Darmspiegelungen seltener an Krebs erkranken oder versterben, wenn die Rückzugszeit mindestens sechs Minuten betrug. Wenn während des Rückzugs jedoch Polypen abgetragen oder Wunden versorgt werden, zählt das nicht zu dieser Rückzugszeit. Wir haben eine KI entwickelt, die zeigt, ob wir uns beim Rückzug genug Zeit gelassen haben, um alles zu untersuchen. Dazu haben wir auch zwei Studien publiziert. In der ersten haben wir anhand

von Videos die Diskrepanz zwischen der Selbsteinschätzung der untersuchenden Person und der mit einer Stoppuhr gemessenen Zeit untersucht ([DOI: 10.1055/a-2122-1671](https://doi.org/10.1055/a-2122-1671)). Zusätzlich wurde eine selbstgebaute KI zur Rückzugszeitbestimmung retrospektiv evaluiert. In der zweiten Studie haben wir die KI zur Bestimmung der Rückzugszeit prospektiv an Patientinnen und Patienten des UKW getestet und gezeigt, dass die KI genauso präzise misst wie eine Stoppuhr ([DOI: 10.1055/a-2721-6798](https://doi.org/10.1055/a-2721-6798)). Die Industrie hat die Ideen hinter unseren KIs bereits aufgegriffen; die ersten entsprechenden Medizinprodukte sind im Begriff auf den Markt zu kommen.

Werden all diese Qualitätsindikatoren am Ende in einer KI zusammengeführt?

Das ist das Ziel. Die KI detektiert, was gesehen wird, und erstellt einen Bericht. Wurde der tiefste Punkt des Kolons erreicht, also der Wurmfortsatz und der Blinddarm? Check. Haben wir Polypen festgestellt? Check. Wie groß sind die Polypen? Das können wir jetzt mithilfe der Wasserstrahlmethode bestimmen. Check. Dann haben wir eine KI entwickelt, die die Instrumente erkennt, mit denen wir Gewebe abtragen. Check. Schließlich kommt die KI zum Einsatz, die Gefäßmissbildungen erkennt. All diese einzelnen Puzzlestücke führen wir am Ende zu einem Bericht zusammen, der ein gewisses Qualitätsniveau erreicht.

Sie möchten noch mehr über die Arbeitsgruppe InExEn erfahren? In [diesem Interview](#) berichten Alexander Hann und Jana Theile, Ärztin und wissenschaftliche Mitarbeiterin der AG InExEn über KI-gestützte Diagnostik, Angiodysplasien und digitale Trainingskonzepte.

Das Interview führte Kirstin Linkamp / Wissenschaftskommunikation