

Original-Titel:

Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: A SYSTEMATIC REVIEW AND METAANALYSIS.

Autoren:

Thomas KL. Lui, Chuan-Guo Guo, Wai K. Leung, DOI: <https://doi.org/10.1016/j.gie.2020.02.033>

Kommentar:

Dr. Rainer Bastian, PD. Dr. Brigitte Schumacher, Essen, 24.04.2020

Das kolorektale Karzinom (KRK) ist mit über 73.000 Neuerkrankungen und ca. 27.000 Todesfällen pro Jahr in Deutschland einer der häufigsten malignen Tumoren. Das durchschnittliche Lebenszeitrisko an sporadisch auftretendem Darmkrebs zu erkranken beträgt 6%. Die deutsche Krebsgesellschaft geht davon aus, dass bei rechtzeitiger Diagnose 90 Prozent aller Darmkrebspatienten hätten geheilt werden können. Die hier vorgestellte Metaanalyse prüft die Frage, ob der Einsatz von sogenannter künstlicher Intelligenz (artificial intelligence (AI)) eine höhere diagnostische Genauigkeit, Sensitivität und Spezifität auf die verlässliche Erkennung von kolorektalen Polypen und deren histologische Voreinschätzung liefert. Die Polypen wurden eingeteilt in adenomatöse Polypen und nicht-adenomatöse Polypen, zu denen die hyperplastischen Polypen und die Gruppe der serratierten Läsionen zugeordnet wurde.

Computergestützte, lernfähige Systeme, die derzeit noch nicht im klinischen Einsatz sind, übermitteln während der Darmspiegelung auf einen Polypen hin verdächtige Läsionen in 500-facher Vergrößerung an einen Computer. Eine Software überprüft durch Analyse von etwa 300 Einzelmerkmalen, ob es sich dabei möglicherweise um adenomatöse Polypen oder um nicht-adenomatöse Polypen handelt. Bei dieser Auswertung wird in weniger als einer Sekunde durch einen Ton oder einen Hinweis auf dem Bildschirm der Untersucher bei der Einschätzung der Läsion - mit durch künstliche Intelligenz (AI) generierte Zusatzinformation - unterstützt. Der Arzt entscheidet abschließend darüber, ob er die suspekte Läsion entfernen muss.

Methodisch erfolgte die Studienselektion für diese Metaanalyse entsprechend der PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) und MOOSE (Metaanalysis of Observational Studies in Epidemiology) Leitlinien. Es erfolgte eine systematische Recherche in Embase, PubMed, Medline, Web of Science and Cochrane library, um entsprechende Studien zu identifizieren (Genauigkeit von künstlicher Intelligenz bei der Vorhersage der Histologie und der Erkennung von kolorektalen Polypen). Es wurden nur englischsprachige Publikationen bis September 2019 einbezogen.

Die Datenextraktion und Bewertung der Studienqualität erfolgte hinsichtlich Anzahl der Polypen, richtig positive Erkennung von adenomatösen Polypen durch AI, richtig negative Polypen (nicht-adenomatöse Polypen, eingeschätzt als nicht-adenomatös), sowie entsprechend falsch positiv und falsch negativ. Alle Polypen wurden reseziert und histologisch diagnostiziert, dies war der Referenz- oder Goldstandard. Entsprechende Daten von kleinen < 5mm Polypen wurden – wenn verfügbar – ebenfalls generiert.

Die Prinzipien der evidenzbasierten Medizin (EbM) werden in dieser Metaanalyse zur Bewertung der Diagnostik mittels Einsatzes künstlicher Intelligenz in der Gastroenterologie eingesetzt. Durch spezielle metaanalytische Verfahren lassen sich die Ergebnisse (diagnostische Genauigkeit der AI) der ausgewählten 6.763 Einzelstudien zu einem bestimmten diagnostischen Test statistisch zusammenfassen und somit der Test auf objektivere Art und Weise hinsichtlich seiner klinischen Wertigkeit einschätzen.

Die Qualität des Designs und der Methodik der durch die Autoren eingebrachten diagnostischen Studien ist dabei von großer Bedeutung hinsichtlich der statistischen abschließenden Bewertung. Ergebnisse einer Einzelstudie können durch Verzerrungen (Bias)

beeinflusst sein, die sich wiederum beim metaanalytischen Zusammenfassen (Poolen) verfälschend niederschlagen könnten. Weiterhin stellen Unterschiede in der Qualität einen Hauptfaktor für die statistische Heterogenität zwischen den Primärstudien dar. Hierdurch kann die Analyse in ihrer Aussagekraft erheblich eingeschränkt oder sogar unmöglich werden. Die Beurteilung der Qualität jeder Einzelstudie ist somit ein wesentlicher Bestandteil dieser Metaanalyse.

Um diese Qualitätsbeurteilung zu erleichtern, wurde das 2003 von Whiting und Kollegen publizierte QUADAS-2 Werkzeug (Quality Assessment of Diagnostic Accuracy Studies included in Systematic Reviews) angewendet. Es besteht aus einem Katalog von 14 Fragen, anhand derer die drei Autoren wichtige qualitative Aspekte der einzelnen Studien bewerten konnten. Die meisten der 14 Fragen beziehen sich auf Studienmerkmale mit möglicher systematischer Verzerrung (Bias) der Studienergebnisse und somit auf die methodische Qualität. Diese Merkmale bestimmen die interne Validität (Glaubwürdigkeit) der Studienergebnisse, die umso höher ist, je weniger Bias vorhanden ist.

Wesentliches Studienergebnis dieser Metaanalyse stellt die Bewertung der diagnostischen Genauigkeit der AI dar, um kolorektale Polypen verlässlich zu erkennen und diese verlässlich histologisch bewerten zu können. Zweitbedeutendes Ergebnis war der Vergleich der AI mit der diagnostischen Qualität erfahrener Untersucher mit mehr als fünf jähriger Praxis in der Koloskopie.

Zur statistischen Datenanalyse wurden statistische Methoden wie das bivariate Modell, um die Sensitivität und Spezifität zu bewerten und das HSROC-Modell (hierarchical summary receiver operating characteristic) eingesetzt. Alle statistischen Analysen wurden mit derselben Software durchgeführt (R Software, Version 3.5.1).

Die diagnostische Leistungsfähigkeit der AI wurde bewertet durch die Analyse von Bildern von 7.680 Darmpolypen in 18 Studien. Die gepoolte Sensitivität hinsichtlich der richtigen histologischen Einschätzung der Läsionen betrug 92% (95% CI, 88,8% - 94,9%) und deren Spezifität betrug 89,8% (95% CI, 0,95 - 0,98).

Bei nicht-adenomatösen Polypen betrug die Sensitivität bezüglich Unterscheidung zwischen hyperplastischen Polypen und SSA 95,2% (95% CI, 83,4% - 98,8%) und die Spezifität 95,9% (95% CI, 0,97 - 0,98).

Fünf Studien prüften den Einfluss von deep learning als Strategie, um mittels AI die Polypdetektion zu verbessern. Die Sensitivität betrug in dieser Gruppe 94,7 % (95% CI, 92,2% - 96,4%) und die Spezifität 91,7% (95% CI, 82,1% - 96,4%).

Auch bezüglich des Nachweises kleiner < 5 mm Polypen konnte durch Analyse von 9 Studien unter Verwendung künstlicher Intelligenz eine Sensitivität (93%) und Spezifität (90,8%) über 90% erzielt werden.

Eine sehr interessante Bewertung erfolgte durch den Vergleich von Endoskopikern mit durch AI gestützte Verfahren bezüglich der Vorhersage der Histologie des Polypen. In 12 Studien betrug die gepoolte Sensitivität und Spezifität aller Endoskopiker bezüglich histologischer Einschätzung 90,5% bzw. 83,9%. Somit ist der erfahrene Koloskopiker in der Erkennung von relevanten Läsionen der mittels AI erkannten Läsionen überlegen. Es zeigte sich jedoch kein wesentlicher Unterschied hinsichtlich der Abschätzung der Histologie.

In weiteren 6 Studien wurden weniger erfahrene Endoskopiker (weniger als 5 Jahre Praxis) einbezogen und in diesem Vergleich zeigte sich, dass die Bewertung mittels künstlicher Intelligenz hinsichtlich Sensitivität signifikant höher war (93,3 % gegenüber 85,8 %), bzgl. der Spezifität ergab sich aber kein signifikanter Unterschied.

Folgende neue und relevante Erkenntnisse resultieren aus der vorgestellten Studie

- Dies ist die erste bisher noch als Pre-proof Version vorliegende Metaanalyse, die die aktuelle Leistungsfähigkeit von künstlicher Intelligenz in der Endoskopie bewertet.

- Obwohl unterschiedliche AI-Modelle in den Studien zum Einsatz kamen, verschiedene Algorithmen, Bildmodalitäten (mit spezieller optischer Technologie wie Narrow Band Imaging) und Patientenselektionen erfolgten, erscheint AI aktuell verlässlich zu sein und beide Ziele zu erreichen:
 - a. eine verlässliche Polypen-Erkennung ist mittels AI-Einsatz möglich.
 - b. ebenso eine verlässliche Voreinschätzung der Polypen hinsichtlich deren Histologie.
- Die Leistungsfähigkeit von AI zur Erkennung von Polypen verbessert sich durch den zusätzlichen Einsatz von deep learning sowie durch den Einsatz der optischen Technologie Narrow Band Imaging (NBI).

Chancen für einen zeitnahen Einsatz von AI in der Koloskopie

- 1) Unterstützung von Ärzten, die die Koloskopie erlernen.
- 2) Unterstützung von Ärzten durch Einholen einer zweiten Meinung hinsichtlich der Bewertung einer Läsion, z.B. in abgelegenen Regionen vorteilhaft.
- 3) Aktuell kein Routineeinsatz von AI gestützten Verfahren, weil noch zu wenig belastbare Daten durch hochwertige Studien vorliegen.
- 4) Künftig wahrscheinlich eine bedeutende Unterstützung des Endoskopikers zu erwarten.
- 5) Autonome Verfahren unter Verwendung von AI in der Darmkrebsvorsorge sind aktuell nicht realisiert.