



Wie kluge Algorithmen Kinderärzte unterstützen können Diagnostik am Puls der Zeit

Lorenz Grigull

Im täglichen Praxisbetrieb werden seltene Erkrankungen mitunter schnell übersehen. KI-basierte Systeme können hier weiterhelfen, indem sie etwa bei entsprechender Symptomkonstellation einen Warnhinweis geben. Beispiele für die erfolgreiche Zusammenarbeit zwischen Mensch und Maschine gibt es bereits viele.

Die Ausbildung an den medizinischen Fakultäten in Deutschland und Europa war und ist stark wissensorientiert. Studierenden werden viele Fakten vermittelt – zu Symptomen, Inzidenzwerten, Nebenwirkungen – in der Annahme, dass einmal gespeichertes Wissen später automatisch zu korrekten Diagnosen und guten Behandlungsergebnissen führt. Als geradezu logische Konsequenz wurden in den 1980er-Jahren die ersten computerbasierten Expertensysteme entwickelt, die vorhandenes Wissen in wissensbasierte Entscheidungsbäume integrieren, um Systeme zu konstruieren, die bei der Diagnose und klinischen Entscheidung unterstützen („clinical decision support system“, CDSS).

Trotz erheblichen Aufwands waren die Erfolge minimal – kein System erreichte den Standard der ärztlichen

Routine. Medizinische Fragestellungen und Entscheidungsprozesse waren offenbar nicht ausreichend binär zu strukturieren, um ein solides, alltagstaugliches Werkzeug zu etablieren. Zudem war der Programmieraufwand beträchtlich, regelmäßige Anpassungen waren erforderlich und die Leistungsfähigkeit der Computer reichte seinerzeit nicht aus.

Zwei Entwicklungen begünstigten die Renaissance der CDSS: Preiswerte und leistungsfähige Computer wurden flächendeckend verfügbar und dank innovativer mathematischer Klassifikationsverfahren wurde es möglich, Zusammenhänge in großen Datenmengen zu erkennen. Entsprechend hielten computerunterstützte Systeme auch in Teilbereichen der Medizin Einzug und wir finden Schlagwörter wie „Big Data“, „Data Mining“ und „künstliche Intelligenz“

(KI) zunehmend häufiger in ärztlichen Fachjournalen.

Bei seltenen Erkrankungen kann die KI überlegen sein

Ärzt*innen setzen in der Anamnese und Untersuchung tagtäglich „Informations-Puzzlesteine“ zusammen und klassifizieren diese im Rahmen des differenzialdiagnostischen Abklärungsprozesses. Das kann unbewusst passieren, aber auch bewusst durchdacht sein, etwa um bei einer Fallkonferenz oder klinischen Visite die persönliche Einschätzung begründen zu können.

Psychologische Faktoren führen in diesem Denkprozess häufig dazu, dass seltene Erkrankungen (SE) differenzialdiagnostisch übersehen werden. Die „Alarmfunktion“ einer KI, die aufgrund einer bestimmten Symptomkonstellation aktiviert wird, kann dieses Problem überwinden. Denn genauso wie das aktive Nachfragen von Kolleg*innen in der Fallkonferenz ausreichen kann, um eine neue Differenzialdiagnose in Betracht zu ziehen, hilft ein CDSS, weitere Diagnosen zu bedenken und zu prüfen. Von vielen SE hört man vielleicht einmal – etwa am Rande von Prüfungen – oder

entdeckt sie zufällig im „Kleingedruckten“ der Lehrbücher. Im klinischen Alltag – angefüllt mit Otitiden, Vorsorgeuntersuchungen und Impfungen – gerät jedoch das Kleingedruckte automatisch in Vergessenheit. Schon ein kleiner Hinweis einer KI (z. B. „Warnung: Antwortmuster zeigt 80 % Übereinstimmung mit Morbus Pompe“) kann hier dazu führen, eine SE doch in Betracht zu ziehen. Das kann auch ein Impulsgeber sein, um Patient*innen für eine zweite Meinung in eine Spezialsprechstunde oder an ein Zentrum für seltene Erkrankungen (ZSE) zu überweisen, um die Möglichkeit einer SE zu prüfen.

KI lernen Diagnoseprozesse ähnlich wie Studierende

In den letzten zehn Jahren hat die Zahl der Publikationen zum Thema „KI und Diagnostik“ in der Medizin kontinuierlich zugenommen. Die medizinische Datenbank PubMed listete allein im letzten Jahr unter den Stichwörtern „artificial intelligence“ und „diagnosis“ 5.899 Treffer (Stand: 29. April 2021)! Hier findet sich der Einsatz von KI für die Erkennung pathologischer (Mikroskopie-)Befunde, für die automatisierte Analyse von Bilddateien in der Radiologie, für Diagnosevorschläge aus Fragebögen oder auch für eine computerisierte Diagnose bei dermatologischen Fragestellungen (Hautkrebserkennung).

Die Datenmenge ist dabei nicht automatisch ein Garant für die Qualität der KI-gestützten Diagnostik und nicht jede Information ist für ein entsprechendes Computersystem gleichermaßen nutzbar. Grundsätzlich gilt, dass lernende Systeme (Menschen und Maschinen) während des Lernprozesses Informationen sortieren und ordnen. So lernt auch ein Kind anhand der Informationen „kuscheliges Fell“ und „bellen“, einen Hund zu erkennen; Studierende der Medizin lernen analog anhand der Informationen „Bilytopenie“ und „Blasten im Blutausschrieb“ sowie „Patient ist drei Jahre alt“ zur Verdachtsdiagnose „Leukämie“ zu gelangen.

Für den Trainingsprozess von Mensch und Maschine gilt vergleichbar, dass unter anderem die Qualität der Trainingsdaten entscheidend für den Erfolg des Lernprozesses ist. Vereinfacht kann man

sich das so vorstellen: Studierende lernen schrittweise, das Krankheitsbild „akute lymphoblastische Leukämie (ALL)“ unter dem Mikroskop zu erkennen: Zunächst lernen sie die Einstellungen am Mikroskop, dann die verschiedenen Zelltypen (ggf. deren Färbeverhalten) und schließlich deren korrekte Zuordnung. Unreife Zellen bereiten anfangs möglicherweise besondere Schwierigkeiten, aber nach dem zehnten „typischen Blutausschrieb bei einer ALL“ werden die ALL-Blasten wahrscheinlich recht zuverlässig erkannt. Im Weiteren lernen Studierende, atypische Blasten zu erkennen und werden schließlich mit Ausstrichpräparaten konfrontiert, ohne vorab die zugehörigen Diagnosen zu kennen. Bei einer optimalen Lernsituation sollten sie dann typische Blasten mit einer Zuverlässigkeit von etwa 80 % erkennen und dem Krankheitsbild ALL zuordnen können.

Im Prinzip vergleichbar trainieren KI-basierte Systeme. Hierfür müssen jeweils geeignete Voraussetzungen geschaffen werden, was aber technisch gut lösbar ist. Und analog kann auch eine Maschine „Muster“ in anderen (Bild-) Datenbanken erkennen und so „lernen“, einem Datensatz eine entsprechende (Verdachts-)Diagnose zuzuordnen (bzw. einen Hund auf einem Bild zu „erkennen“). Hierfür werden bei Maschinen mathematische Standardverfahren für die Klassifikation von Datensätzen – zum Beispiel künstliche neuronale Netze, nearest neighbor, random forest oder support vector machine – eingesetzt und dann programmiert. Für die jeweilige Anforderung an die KI erfolgt anschließend ein zeitaufwendiger Lernprozess, an dessen Abschluss die zuverlässige Ergebnisreproduktion steht, also die sekundenschnelle Zuordnung eines unbekanntes Datensatzes zu einer zuvor gelernten Kategorie.

Wo die KI schon heute in der Medizin eingesetzt wird

In Zusammenarbeit mit dem KI-Experten Dr. Werner Lechner (KImedi GmbH) und dem Mathematiker und Statistiker Professor Frank Klawonn (Ostfalia Hochschule/Helmholtz Institut für Infektionsforschung) wird am Zentrum für seltene Erkrankungen in Bonn (ZSEB) ein Konzept umgesetzt,

das mittels einer Kombination aus Fragebogen und KI den Weg zur Diagnose unterstützt. Der hier verfolgte Ansatz hat sich bislang in unterschiedlichen klinischen Szenarien bewährt und wird kontinuierlich überprüft und erweitert. Ein Element dieses Diagnose-unterstützenden Verfahrens ist in Form einer Diagnose-App bereits ein zertifiziertes Medizinprodukt.

Das Prinzip hinter dem Ansatz sind dabei Fragen (an Menschen ohne Diagnose), die ihren Ursprung in Interviews mit Menschen *mit* nachgewiesener SE haben. Hierfür wurden Patient*innen im Kontext wissenschaftlicher Forschungsprojekte gebeten, ihre Erfahrungen aus der Zeit bis zur Diagnose „zu erzählen“. Aus diesem Material entstanden in einem qualitativen Analyseprozess Fragen, die in einem finalen Fragebogen für bestimmte Leitsymptome (z. B. „Muskelschwäche“, „chronischer Husten“ oder „Infektanfälligkeit“) zusammengestellt wurden. Diese Fragebögen wurden – in enger Kooperation mit Patientenorganisationen – an Betroffene mit *bekannt*er Diagnose versendet mit der Bitte, den Fragebogen zu beantworten (und zwar aus Sicht der Zeit *vor* Diagnosestellung). So entstand eine Datenbank (bzw. verschiedene Datenbanken, bezogen auf das jeweilige Leitsymptom), anhand derer ein KI-System trainieren konnte, ein Antwortmuster einer bestimmten Diagnose (z. B. primäre ciliäre Dyskinesie, Muskeldystrophie oder primärer Immundefekt) zuzuordnen. Die Fragen sind laienverständlich und orientieren sich stark am prädiagnostischen Erfahrungswissen Betroffener. Die Sensitivität der KI plus Fragebogen-Tools lag in den Projekten zuverlässig über 80 %.

Aktuell wird der Fragebogen für das Leitsymptom „Muskelschwäche“ im Projekt ARTIS („artificial intelligence solution“) für Erwachsene ohne Diagnose eingesetzt. Hierbei erhalten Menschen, die aufgrund ihrer Beschwerden nach diagnostischen Antworten suchen und bei ihrer individuellen Recherche auf unseren Fragebogen geleitet werden, im Falle eines pathologischen Antwortmusters einen „Voucher-Code“. Dieser dient gegenüber den beteiligten Ärzt*innen im Projekt zur datenschutzkonformen

Zuordnung des Fragebogens. Idealerweise gibt die KI einen wertvollen Hinweis, der für Betroffene den Weg zur Diagnose verkürzt. Die große diagnostische Sicherheit der vorgestellten Werkzeuge erscheint insofern nicht überraschend, als die wichtigste Zutat die strukturierte Anamnese in Form von fundiert ausgewählten Fragen an die Betroffenen ist.

Weitere KI-basierte Systeme

Es gibt viele weitere Ansätze, um mit KI den Weg zur Diagnose zu unterstützen: Sehr populär ist die Diagnose-App Ada. Diese wird auf ein mobiles Endgerät geladen und führt im Dialog – ausgehend von den Symptomen – zu einer Verdachtsdiagnose. Auch diese App lässt sich intuitiv bedienen und hat eine gute Ergebnisqualität, wie auch Ronicke et al. publiziert haben. Für den Bereich SE wird für Ada aktuell ein zusätzliches Modul entwickelt.

Große Aufmerksamkeit hat auch der Ansatz von Face2Gene gefunden. Hierbei wird ein Foto des Betroffenen hochgeladen, anschließend macht das Programm 30 diagnostische Vorschläge, basierend auf dem Vergleich mit Portraitfotos von bekannten Betroffenen. Die hier eingesetzte KI „DeepGestalt“ hat sich in unterschiedlichen klinischen

Kontexten als Erkennungs-Tool für genetische Erkrankungen bewährt. Ada und Face2Gene sind kostenfrei nutzbar.

Schließlich sei noch auf eine Pionierin der Diagnose-Unterstützung verwiesen: Isabel Healthcare. Die Entstehungsgeschichte von Isabel: Eine pädiatrische Patientin verstirbt beinahe an einer seltenen Komplikation einer häufigen Krankheit. Erst Behandelnde einer Spezialklinik erkennen die Komplikation und retten die Patientin. Der Vater der Patientin ist entsetzt, dass es nur einem glücklichen Zufall zu verdanken ist, dass seine Tochter überlebt hat. Er entschließt sich, gemeinsam mit dem behandelnden Pädiater ein Werkzeug zur Diagnosefindung zu entwickeln – und benennt es nach seiner Tochter. In das Programm werden Alter, Herkunft und Symptome des Betroffenen eingetragen und Isabel generiert Diagnosevorschläge. Isabel kann umsonst ausprobiert werden, nach Ablauf einer Testperiode ist die Nutzung kostenpflichtig.

Potenzielle Risiken beim Einsatz von KI

Die Frage nach der Fehldiagnose

In vielen Diskussionen zum Einsatz von Diagnose-unterstützenden Verfahren wurde mir wiederholt die Frage nach den Fehldiagnosen der Maschine ge-

stellt. Nachdem ich zum Beispiel auf einer Fachtagung eines unserer Diagnosewerkzeuge vorgestellt hatte, das mit einer Zuverlässigkeit von 85 % erkennen konnte, ob ein Antwortmuster zu einem Kind mit primärem Immundefekt gehörte, fragte mich ein ärztlicher Kollege: „Herr Grigull, 85 % sind ja schön und gut, aber wie stehen Sie zu der Rate von 15 % Fehldiagnosen ihrer KI?“ Meine Antwort – damals wie heute – lautete, dass jede Fehldiagnose für Betroffene schlimm und unbedingt zu vermeiden ist. Für die mir bekannten Systeme gilt, dass die Fehlerquote durch Verbesserung der Datensätze sowie kontinuierliches Training der Algorithmen gesenkt werden kann.

Dem kritischen Kollegen antwortete ich zudem mit einem Augenzwinkern, dass die Ärzteschaft der diagnostischen Güte einer KI offenbar kritischer als der eigenen Diagnosequalität gegenüber steht. Schließlich ist mir – leider! – in 25 Jahren Tätigkeit noch kein*e Kolleg*in begegnet, die mir Auskunft oder gar eine Statistik zur persönlichen diagnostischen Zuverlässigkeit geben konnte. Ebenso wenig kenne ich Statistiken aus deutschen Krankenhäusern oder von ärztlich Tätigen in der Niederlassung, bei denen die Anzahl richtiger und falscher Diagnosen erfasst wird. Aus meiner Sicht wäre es ein großer Fortschritt, wenn hier eine transparente Dokumentation stattfände.

Intransparenz bei der Entscheidungsfindung

Ein weiterer häufig genannter Kritikpunkt an KI-gestützten Verfahren ist die Intransparenz des Prozesses. Diese Kritik ist verständlich, denn bei den bekannten Systemen zur diagnostischen Unterstützung können die Anwendenden nicht erkennen, warum die KI eine bestimmte Differenzialdiagnose favorisiert.

Diese fehlende Nachvollziehbarkeit ist systemimmanent, sollte aber nicht als grundsätzliches K.O.-Kriterium gegenüber einer KI oder einer computergestützten Entscheidung dienen. Denn wie oft verlässt man sich bei der Diagnostik im Alltag auf die Ergebnisse eines CT oder MRT, ohne die computergestützten Algorithmen hinter der Bildkonstruktion zu verstehen? Wie oft muss das



© sturti / Getty Images / iStock (Symbolbild mit Fotomodell)

Je öfter sie unter das Mikroskop blicken, desto zuverlässiger erkennen Studierende pathologische Befunde – ähnlich lernt auch die KI.

Hier steht eine Anzeige.



Weiterführende Links

- Aktion ARTIS; Fragebogen zum Leitsymptom Muskelschwäche: www.itsinyourmuscles.eu/de/ihr-symptome-uberprufen/
- Fragebogen des ZSE Bonn zu neuromuskulären Erkrankungen: www.patientconcept.de/portal/public/fragebogen/neuromuscular-diseases
- Diagnose-App Ada: ada.com/de/app/
- Face2Gene: www.face2gene.com
- Isabel Healthcare: www.isabelhealthcare.com
- PatientConcept-App: www.patientconcept.de
- Society to Improve Diagnosis in Medicine: www.improvediagnosis.org
- Aktionsbündnis Patientensicherheit: www.aps-ev.de
- SE-Atlas: www.se-atlas.de
- Datenbank Orphanet: www.orpha.net

„Bauchgefühl“ erhalten, wenn es um die Begründung einer klinischen Entscheidung geht, ohne dass wir deswegen grundsätzlich der Empfehlung oder Entscheidung misstrauen? Schließlich ist es ja gerade das „Gesamtmuster“ eines Betroffenen, das bei Mensch (und Maschine) einen Impuls auslöst.

Datenschutz umsetzen

Zwei weitere Risiken müssen im Kontext KI-basierter Systeme beachtet und ausgeräumt werden: Zum einen müssen die Systeme datenschutzkonform sein. Hier kann es Konflikte geben – je nachdem, welche Informationen verwendet werden; gerade bei den sehr seltenen Er-

Fazit für die Praxis

- KI-basierte Systeme können den Behandelnden helfen, sich seltene Erkrankungen als Differenzialdiagnose in Erinnerung zu rufen.
- Es gibt bereits zahlreiche Apps und Systeme, die zur Diagnosefindung eingesetzt werden.
- Stetes Lernen führt dazu, dass die Vorhersagen einer KI immer zuverlässiger werden.
- Viele Systeme erreichen eine Trefferquote von 80–85% – und sind damit eine wertvolle Hilfe für Ärzt*innen.

krankungen können schon wenige Hinweise Rückschlüsse auf die Identität des Betroffenen möglich machen. Gute Erfahrungen haben wir damit gemacht, dass das KI-System in ein bestehendes datenschutzkonformes Konzept (z. B. die App PatientConcept) integriert wird. So lassen sich zwei Ziele erreichen: Datenschutz und gute Verfügbarkeit Diagnose-unterstützender Systeme.

Daneben wird zunehmend von dem Phänomen einer „Cyberchondrie“ berichtet, dass also Menschen ihre Symptome im Internet recherchieren und aufgrund der gefundenen „Diagnosen“ in manchen Fällen verunsichert sind. Entsprechend sollten alle Systeme so ausgelegt sein, dass Verdachtsdiagnosen exklusiv von Behandelnden übermittelt und entsprechend klinisch eingeordnet werden. Gleichzeitig sollten sie den Rechercheergebnissen Betroffener immer auf Augenhöhe begegnen und in den eigenen Diagnosefindungsprozess integrieren.

Schlusswort und Ausblick

Das deutsche Gesundheitswesen war in den letzten 20 Jahren wenig fortschrittsfreundlich. Der lange Kampf für die Befundübermittlung per Fax, handschriftliche Notizen, wie sie etwa im gelben Kinderuntersuchungsheft seit 50 Jahren gepflegt werden, sowie die ungenügende Verfügbarkeit einheitlicher Datensätze kennzeichnen eine sehr konservative und fortschrittsskeptische Grundhaltung. Entsprechend fehlen den Behandelnden regelmäßig Informationen – mit der Folge überflüssiger Doppel-Untersuchungen oder Fehlentscheidungen aufgrund von Informationslücken. Fortschritte wurden systematisch verhindert, Insel-Lösungen gefördert und privatwirtschaftlichen Alternativen Vorschub geleistet, bei denen ärztliche Mitsprache zu fehlen droht. Die COVID-19-Pandemie zeigt diese Defizite gnadenlos auf. Daher sind kluge, gemeinsame Initiativen zu fördern und zu unterstützen. Diese sollten neben den Bedürfnissen der Ärzt*innen aus den verschiedenen Sektoren auch die Sicht der KI-Fachleute, besonders aber auch die der Betroffenen berücksichtigen.

So könnte am Ende ein System stehen, in dem eine gut geführte elektronische

Patientenakte hilft, Fehldiagnosen zu reduzieren, Warnhinweise bei Arzneimittelwechselwirkungen erfolgen, patientennahe Forschung ermöglicht wird und Befunde rasch verfügbar sind. Über die portionierte und gerichtete Verwendung von Daten (Stichwort: „Datenspende“) kann dann ebenfalls diskutiert werden.

Ich möchte alle Leser*innen ermutigen, sich aktiv an dieser Diskussion zu beteiligen, damit am Ende ein besseres Gesundheitssystem steht, das Ressourcen sinnvoll nutzt und in dem Menschen entscheiden, an welcher Stelle im System Computer und KI uns helfen könnten, unsere Patient*innen besser zu behandeln.

Weiterführende Literatur

1. Pantel JT et al. Efficiency of computer-aided facial phenotyping (DeepGestalt) in individuals with and without a genetic syndrome: diagnostic accuracy study. *J Med Internet Res* 2020;22:e19263
2. Ronick S et al. Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study. *Orphanet J Rare Dis* 2019;14:69
3. Cheraghi-Sohi S et al. Assessing the utility of a differential diagnostic generator in UK general practice: a feasibility study. *Diagnosis (Berl)* 2020;8:91-9
4. Grigull L et al. Common pre-diagnostic features in individuals with different rare diseases represent a key for diagnostic support with computerized pattern recognition? *PLoS One* 2019;14:e0222637
5. Grigull L et al. Diagnostic support for selected neuromuscular diseases using answer-pattern recognition and data mining techniques: a proof of concept multicenter prospective trial. *BMC Med Inform Decis Mak* 2016;16:31



Prof. Dr. Lorenz Grigull, MBA, MME
 Leiter Sektion Kinder Zentrum für Seltene Erkrankungen
 Universitätsklinikum Bonn
 Venusberg-Campus 1
 53127 Bonn
 E-Mail: Lorenz.grigull@ukbonn.de

Interessenkonflikt

Lorenz Grigull ist Mitgründer der Klmedi GmbH, Donauwörth. Er erhielt Vortragshonorare von Sanofi und Shire. Das Projekt ARTIS wird gefördert von Sanofi.

Hier steht eine Anzeige.

