



 AUS DEM NETZWERK EVIDENZBASIERTE MEDIZIN

Fluch oder Segen?

Symptom Checker und Diagnostik-Apps

VON PROF. DR. ANDREAS SÖNNICHSEN IM AUFTRAG
DES NETZWERKS EVIDENZBASIERTE MEDIZIN E. V. (WWW.EBM-NETZWERK.DE)



Im November 2018 teilte die Techniker Krankenkasse in einer Pressemitteilung mit: „Ein auf künstlicher Intelligenz (KI) basierender Symptomcheck erweitert seit heute das digitale Angebot der Techniker Krankenkasse (TK). Nutzer der App "Ada", die TK-versichert sind, können dort ihre Beschwerden eingeben, erhalten eine persönliche Analyse und werden auf Wunsch über passende digitale Versorgungsangebote der TK informiert.“ Und weiter: „Aus Sicht der TK können Angebote wie der Symptomcheck mit Künstlicher Intelligenz kranken Menschen den Alltag spürbar erleichtern und sie dabei unterstützen, beim Thema Gesundheit kompetenter zu werden. Solche digitalen Angebote sind deshalb auch ein wichtiger Beitrag, um die Patientensouveränität zu fördern.“⁽¹⁾

Aus Sicht der evidenzbasierten Medizin gehen wir der Frage nach, ob die hoffnungsvollen Aussagen einer der führenden Krankenversicherungen in Deutschland einer Überprüfung standhalten.

Die Idee, medizinische Diagnostik durch elektronische Algorithmen zu unterstützen, ist nicht neu. Als einer der ersten „Symptom Checker“ wurde „Isabel“ bereits im Jahr 1999 gegründet – als Reaktion auf eine schwerwiegende Fehldiagnose bei einem dreijährigen Mädchen namens Isabel. Inzwischen sind mehrere Dutzende solcher Diagnose-Tools auf dem Markt. Doch halten sie wirklich, was sie versprechen?

STUDIENEVIDENZ ZUM EINSATZ VON SYMPTOM CHECKERN

Im Jahr 2015 wurde im British Medical Journal eine an der Harvard University durchgeführte Evaluationsstudie (Audit Study) über 23 Symptom Checker publiziert, die sehr enttäuschend ausfiel (2). Die Autoren testeten 45 Fallvignetten, 15 Notfälle, 15 Fälle, die eine hausärztliche Behandlung erforderten, und 15 Fälle, bei denen ein Arztbesuch überflüssig war. Nur in 34 % der Fälle nannten die Apps die korrekte Diagnose an erster Stelle. Die Schwankungsbreite lag zwischen 5 % und 50 %, das heißt, selbst die beste App lag nur in der Hälfte der Fälle richtig. Die Unterscheidung zwischen häufigen und seltenen Krankheiten zeigte, dass häufige mit größerer Treffsicherheit, nämlich in 38 % der Fälle, seltene aber nur in 28 % der Fälle korrekt erkannt wurden. Diagnosen, die keinen Arztbesuch erforderten, wurden häufiger korrekt klassifiziert (40 %) als Diagnosen, die eine notfallmäßige Versorgung erforderten (24 %). Auch die Triage-Korrektheit wurde untersucht: Immerhin wurden 80 % der Notfälle von den Apps als solche identifiziert. Erschreckend ist allerdings, dass bei Beschwerden, die keinen Arztbesuch erforderten, in fast 70 % der Fälle

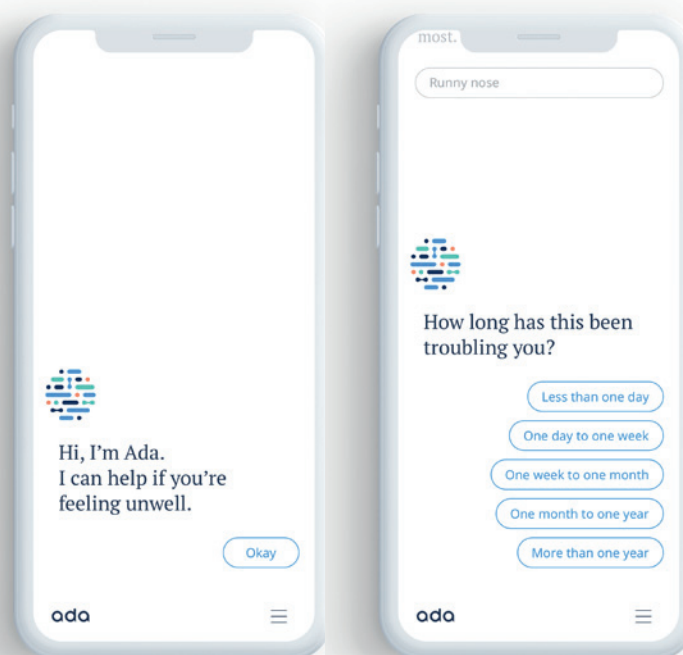
Diagnostik-App "Ada"

ein Arztbesuch empfohlen wurde. Bei seltenen Erkrankungen wurde in nur 62 % der Fälle eine korrekte Triage-Empfehlung abgegeben.

Insgesamt ist die Studienlage zur Evaluation von Diagnostik-Apps spärlich. Zu „Ada“ findet sich bisher keine einzige publizierte Evaluationsstudie, obwohl auf der Ada-Webseite behauptet wird, dass das Tool auf Forschung über acht Jahre basiert (<https://ada.com/de/milestones/>).

Neben der Audit Study gibt es nur wenige in PubMed gelistete Evaluationsstudien zum Stichwort "Symptom Checker", die sich überwiegend mit bestimmten Symptomen oder einer Gruppe definierter Symptome befassen. So wurden beispielsweise ophthalmologische Symptome in einer App nur in 11 von 42 Fällen (26 %) der korrekten Diagnose zugeordnet (3). Eine andere Autorengruppe testete die Fähigkeit von fünf Symptom-Checker-Apps (Mayo, Isabel, WebMD, Symptomate, Symcast), aus 8.363 konsekutiven Patienten einer Notaufnahme 168 Patienten mit HIV, Hepatitis C oder beidem zu identifizieren. Im Durchschnitt wurden nur 20 % der Fälle korrekt erkannt, und in nur 45 % der Fälle wurde eine der Zieldiagnosen als mögliche Differentialdiagnose gelistet.

Eine neuere „Studie“ gibt vor, ein Update der Harvard Audit Study unter Einbeziehung von fünf neueren Symptom Checkern vorzunehmen, unter anderen Ada und Symptoma (4). In dieser „Studie“ wird gezeigt, dass sich die neuen Symptom Checker inklusive Ada in ihrer



Treffsicherheit nicht wesentlich von den in der Audit Study getesteten unterscheiden, mit der Ausnahme von Symptoma, der angeblich in 82,2 % der Fälle die richtige Diagnose an erster Stelle nennt. Es bleibt aber unklar, welche Fälle hier wie analysiert wurden, und die Studienmethodik wird nicht adäquat beschrieben. Im weiteren Verlauf des Artikels wird dann auch sehr schnell klar, dass es sich um eine als Studie getarnte Werbebotschaft

für Symptoma handelt, die von einem der Symptoma-Geschäftsführer als Erstautor verfasst wurde.

In einer weiteren kleineren Studie wurden Patienten mit rheumatoider Arthritis gebeten, den NHS-Symptom Checker und den WebMD-Checker mit ihren Symptomen zu benutzen (5). Elf der 34 Patienten wurden (unnötigerweise, da es sich durchweg um chronische Zustände handelte) aufgefordert, sofort einen Notarzt zu rufen. Nur bei 20 % der Patienten wurde die korrekte Diagnose an erster Stelle genannt.

Der vielleicht gravierendste Nachteil aller Bemühungen, Diagnosen mit Hilfe von vermeintlich künstlicher Intelligenz zu stellen, ist das Risiko massiver Überdiagnostik. Wenn man sich die Zahlen der Audit Study vor Augen hält, in der fast 70 % der gesunden Personen mit unspezifischen, selbstlimitierenden Beschwerden, die keinen Arztbesuch erfordern, zur Arztkonsultation aufgefordert wurden, dann führen diese Apps zu einer Flut von Behandlungsfällen, die das Gesundheitssystem



bei breiter Nutzung nicht entlasten, sondern lahmlegen. Sie tragen so ganz offensichtlich auch nicht zur Mündigkeit, Selbstverantwortung und Health Literacy von Patienten bei, wie von der TK fälschlicherweise erhofft wird, sondern führen genau zum Gegenteil: zu Verunsicherung und übermäßiger, unnötiger Inanspruchnahme von Gesundheitsleistungen. Zumindest müsste vor der breiten Einführung solcher Systeme in qualitativ hochwertigen Studien überprüft und belegt werden, dass dieses nicht passiert.

Ebenso gibt es bisher keinen einzigen Nachweis aus qualitativ akzeptablen Studien, dass Symptom Checker in der Entdeckung seltener Erkrankungen einer lege artis durchgeführten ärztlichen Anamnese überlegen sind. Ärztin und Arzt können nämlich zumindest eines, was der Symptom Checker nie leisten kann: Sie können zuhören und offen formulierte Sprache verstehen, inklusive der mitgelieferten nonverbalen Elemente komplexer menschlicher Kommunikation: Intonation, Mimik, Gestik, Körpersprache. Dazu kommt vor allem in der Allgemeinmedizin die erlebte Anamnese, die durch eine langjährige Arzt-Patient-Beziehung gegeben ist

und die in einem Symptom Checker vollständig verlorengeht.

Vielleicht gibt es einen Platz für Symptom Checker als differentialdiagnostisches Tool in ärztlicher Hand, wenn man bei unklaren Beschwerden nicht weiterkommt, hinter denen möglicherweise eine seltene Erkrankung stecken könnte, auf die der Arzt aufgrund von Wissensdefiziten nicht kommt. Aber auch das sollte vor euphorischen Lobliedern des allgemeinen KI-Hypes in prospektiven Studien nachgewiesen werden. Für eine breite Anwendung von Symptom Checkern gibt es jedenfalls derzeit definitiv keinen Anlass. ■



Prof. Dr. Andreas Sönnichsen
Vorsitzender
des Netzwerks Evidenzbasierte
Medizin und Leiter der Abteilung für
Allgemein- und Familienmedizin,
Zentrum für Public Health an der
Medizinischen Universität Wien

Literatur

1. Techniker Krankenkasse. Pressemitteilung: TK setzt auf künstliche Intelligenz für bessere Versorgung 2018 [Available from: <https://www.tk.de/presse/themen/digitale-gesundheit/kuenstliche-intelligenz/start-von-ada-2053188>.
2. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ*. 2015;351:h3480.
3. Shen C, Nguyen M, Gregor A, Isaza G, Beattie A. Accuracy of a Popular Online Symptom Checker for Ophthalmic Diagnoses. *JAMA Ophthalmol*. 2019.
4. Nateqi J, Lin S, Kroboth H, Gruarin S, Lutz T, Dvorak T, et al. [From symptom to diagnosis-symptom checkers re-evaluated : Are symptom checkers finally sufficient and accurate to use? An update from the ENT perspective]. *HNO*. 2019;67(5):334-42.
5. Powley L, McIlroy G, Simons G, Raza K. Are online symptoms checkers useful for patients with inflammatory arthritis? *BMC Musculoskelet Disord*. 2016;17(1):362.