

Trainingsdaten für medizinische KI auf Herz und Nieren prüfen

Systematischer Prüfkatalog aus der PTB - für verlässlichen KI-Einsatz

Künstliche Intelligenz (KI) in der Medizin kann helfen, Krankheiten früher zu erkennen, Menschen besser zu versorgen und die Gesundheitsausgaben zu senken. Die breite Verwendung hängt aber stark davon ab, ob Vertrauen in die Algorithmen aufgebaut werden kann. Eine entscheidende Frage lautet: Sind die Algorithmen mithilfe qualitativ hochwertiger Daten trainiert worden? In der Physikalisch-Technischen Bundesanstalt (PTB) ist ein neuer Prüfkatalog für solche Daten zur Entwicklung von KI-Medizinprodukten erstellt worden. Er bildet auch die Basis für einen agilen Zulassungsprozess solcher Softwareprodukte. Das Projektteam stellt dieses sogenannte METRIC-Framework in der aktuellen Ausgabe der (zur Nature-Gruppe gehörenden) Fachzeitschrift *npj Digital Medicine* vor.

Künstliche Intelligenz bietet ein hohes Potenzial: So können KI-Algorithmen beispielsweise helfen, schnelle MRT-Aufnahmen des bewegten Herzens gestochen scharf zu machen. (MRT: Magnetresonanztomografie) Oder sie können die Diagnosearbeit drastisch beschleunigen, indem sie große Mengen etwa von Vitaldaten (z. B. EKG), Laborwerten (z. B. Blutbild) und medizinischen Bildern schneller auswerten, als ein Mensch das kann. „Die Einsatzmöglichkeiten sind vielfältig und faszinierend. Aber alles steht und fällt damit, ob es gelingt, bei Ärzt*innen und Patient*innen großes Vertrauen in die Sicherheit dieser Verfahren aufzubauen“, sagt Daniel Schwabe, Mathematiker bei der PTB. Hier setzte das Teilprojekt im Rahmen des europäischen TEF-Health-Projektes an, das Schwabe seit Anfang 2023 koordiniert und dessen erstes Ergebnis er jetzt mit seinem Team vorstellt.

KI-Algorithmen müssen trainiert werden, indem ihnen beispielsweise Hunderte und Tausende medizinischer Daten zusammen mit der entsprechenden Diagnose vorgelegt werden. Daraus lernen sie, welche Daten und welche Diagnose zusammenpassen, und können schließlich selber Diagnosen stellen. „Es ist allgemein klar, dass die Qualität der Trainingsdaten ein entscheidender Punkt ist. Aber was steckt eigentlich genau hinter dem Begriff „Datenqualität“ im Kontext medizinischer Anwendungen? Das hat unser Team sich sehr gründlich angesehen“, erklärt Schwabe. Während einer akribischen Recherche, in die insgesamt 5408 Publikationen eingeflossen sind, haben die Forschenden den Begriff in einzelne Eigenschaften aufgespalten. Diese Systematik haben sie in einer Grafik aufgearbeitet, die sie das „Rad der Datenqualität“ nennen. Es legt den Grundstein für die systematische Analyse der Qualität medizinischer Trainingsdaten. Dies dient vor allem Entwicklerinnen und Entwicklern von KI-Medizinprodukten, um die Nutzbarkeit und Eignung ihrer Datensätze besser einzuschätzen. Wenn sie mehr über die Trainingsdaten wissen, verringert sich die Gefahr von Fehleinschätzungen. Das Endprodukt, also der KI-Algorithmus, wird robuster und besser interpretierbar. „Dies leistet einen großen Beitrag zum Ziel der Vertrauenswürdigkeit von KI-Medizinprodukten, wie es vom neuen EU-Gesetz (EU AI Act) gefordert wird“, sagt Schwabe.

Das Rad der Datenqualität kann auch zum Testen von KI-Algorithmen durch Verwendung von Referenzdatensätzen genutzt werden. Das wird die PTB auch in ihrem eigenen [Projekt TraCIM](#) nutzen, das im Frühjahr als Demonstrator auf der Hannover Messe vorgestellt wurde. Damit kann sie KI-Algorithmen für die Medizin bewerten. In Zukunft soll TraCIM es Herstellerfirmen ermöglichen, ihre KI-Algorithmen von der PTB als neutraler Stelle überprüfen zu lassen.

Als nächsten Schritt befasst sich das Projektteam um Daniel Schwabe mit der Frage, wie sich die gefundenen Eigenschaften der Datenqualität automatisiert messen lassen. Außerdem untersucht das Projektteam wie stark der Einfluss von einzelnen Faktoren der Datenqualität auf die Eigenschaften (z. B. Vorhersage-Qualität, Robustheit, Unsicherheit) eines KI-Medizinprodukts ist.

Das Projekt TEF-Health

Künstliche Intelligenz (KI) treibt aktuell eine Vielzahl von Innovationen in der Medizin an. Gerade in diesem Gebiet stellt sich die kritische Frage: Wann ist ein neues KI-System verlässlich und vertrauenswürdig? Um innovative Ansätze in KI und Robotik im Gesundheitswesen hinsichtlich des neuen EU AI Acts zu prüfen und schneller zur Marktreife zu bringen, hat die EU im Januar 2023 ein Projekt im Rahmen des Programms TEF - Testing and Experimentation Facilities - im Bereich Gesundheit gestartet. Das Projekt mit Namen [TEF-Health](#) ist Teil des Programms „Digital Europe“ und wird mit 60 Millionen Euro gefördert. Die PTB ist maßgeblich an einer Projektgruppe beteiligt, die agile Zulassungsprozesse für vertrauenswürdige KI etablieren wird. Innerhalb der PTB ist das Projekt Teil des Innovationsclusters Systemische Metrologie, um Qualitätsstandards für vernetzte Daten und KI auf messbare Größen zurückzuführen.

es/ptb

Wissenschaftliche Veröffentlichung

D. Schwabe, K. Becker, M. Seyferth, A. Klaß, T. Schäffter: [The METRIC-framework for assessing data quality for trustworthy AI in medicine: a systematic review](#). npj Digital Medicine 7, 203 (2024)