



Olaf Dössel

Maschinelles Lernen und Künstliche Intelligenz in der Medizin – Eine Einführung und ein Plädoyer

In:

Dössel, Olaf / Schäffter, Tobias / Rutert, Britta (Hrsg.): Künstliche Intelligenz in der Medizin.

ISBN: 978-3-949455-18-6

Berlin: Berlin-Brandenburgische Akademie der Wissenschaften, 2023

S. 16-27

(Denkanstöße aus der Akademie : eine Schriftenreihe der Berlin-Brandenburgischen Akademie der Wissenschaften ; 11)

Persistent Identifier: [urn:nbn:de:kobv:b4-opus4-38039](https://nbn-resolving.org/urn:nbn:de:kobv:b4-opus4-38039)

Die vorliegende Datei wird Ihnen von der Berlin-Brandenburgischen Akademie der Wissenschaften unter einer Creative Commons Namensnennung 4.0 International Lizenz zur Verfügung gestellt.



MASCHINELLES LERNEN UND KÜNSTLICHE INTELLIGENZ IN DER MEDIZIN – EINE EINFÜHRUNG UND EIN PLÄDOYER

Olaf Dössel

Sowohl Patient:innen als auch Ärzt:innen haben sehr unterschiedliche Auffassungen zu den Themen maschinelles Lernen (ML) und künstliche Intelligenz (KI) in der Medizin. Dieser Artikel soll ein Beitrag zur Versachlichung sein. Er beginnt mit einer Einführung in das Thema. Nach einer Abgrenzung werden vier Beispiele vorgestellt, wo ML und KI schon heute in der Medizin eingesetzt werden. Das Thema Apps und Wearables im Zusammenhang mit ML und KI wird kurz beleuchtet. Dann werden Aspekte zu den folgenden Themen beschrieben: mögliche Fehler, Verantwortung, ethische Fragen, Zulassung, Normen und Gesetze. Auch wird die Frage gestellt, wie mehr qualitativ hochwertige und annotierte Daten für die Forschung gewonnen werden können. Der Artikel endet mit Thesen und einem Plädoyer für einen klugen Umgang mit maschinellem Lernen in der Medizin.

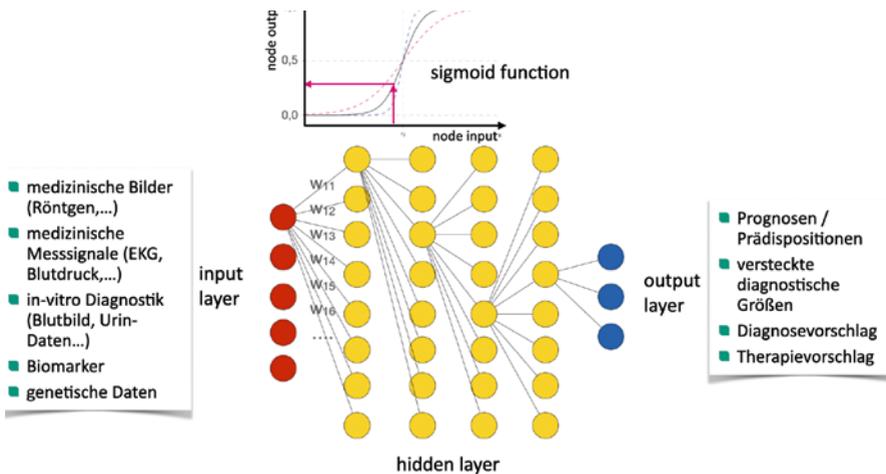


Abb. 1: Grundschemata von maschinellem Lernen in der Medizin

Der „input layer“ ist eine Folge von Zahlen, z. B. das Elektrokardiogramm. Der „output layer“ ist ebenfalls eine Reihe von Zahlen, z. B. steht 1 für „Vorhofflimmern“ und 0 für „kein Vorhofflimmern“. Die gelben Kreise sind sogenannte „Knoten“ (nodes). Der Eingang eines Knotens ist eine gewichtete Summe der Zahlen in der Schicht davor, die w_{ij} sind dabei die Gewichte. Der Ausgang des Knotens wird über eine Sigmoid-Funktion berechnet, bei der eine Steilheit und ein Schwellwert festgelegt werden. Beim Trainieren werden die Gewichtungsfaktoren so bestimmt, dass möglichst viele Eingangsdaten auf den richtigen Ausgang abgebildet werden.

Abbildung 1 zeigt plakativ, wovon dieser Artikel handelt: Wir füttern einen Algorithmus mit medizinischen Daten wie z.B. Bildern, physiologischen Messsignalen, In-Vitro-Messwerten und genetischen Daten, aus denen Prognosen, Prädispositionen, versteckte diagnostische Größen, Diagnose- und Therapieempfehlungen generiert werden können. Der Artikel konzentriert sich also auf maschinelles Lernen und insbesondere auf das „supervised learning“, bei dem ein Trainingsdatensatz mit gut bekannten Paaren aus Eingangs- und Ausgangsdaten zur Verfügung steht. Abbildung 1 visualisiert ein Deep Neural Network. Aber es gibt noch mehr Methoden des maschinellen Lernens: Polynomial-Regression, Support Vector Machines, Decision Trees und Random Forests, k-nearest neighbor und andere mehr.

Die meisten Publikationen zu diesem Thema sind heute im Bereich Bildverarbeitung angesiedelt. Ein wichtiges Beispiel ist die Befundung von Mammographie-Aufnahmen. 2017 nahmen in Deutschland 2,8 Millionen Frauen am Mammographie-Screening teil. Vorgeschrieben ist heute noch ein „Double Reading“, d. h. zwei Expert:innen müssen sich jedes Bild ansehen und entscheiden, ob es Hinweise auf einen bösartigen Brusttumor gibt. Die Frage ist: Kann man dieses Verfahren durch „Single Reading“ plus „Computer Assisted Diagnosis (CAD)“ ersetzen oder vielleicht sogar verbessern? Dazu gibt es eine ganz neue Publikation vom September 2021 mit dem klaren Ergebnis: Single Reading plus CAD ist auf jeden Fall nicht schlechter als Double Reading.¹

Ein ganz anderes Beispiel ist die Apple-Watch. Dabei geht es um die Frage, ob bei einem Anwender der Armbanduhr ein Verdacht auf Vorhofflimmern festgestellt werden kann. Dazu findet man eine Publikation im New England Journal of Medicine mit dem Titel „Large Scale Assessment of a Smart Watch to Identify Atrial Fibrillation“.² 419.000 Menschen haben an der Studie teilgenommen. Die Apple-Watch konnte sehr erfolgreich Vorhofflimmern bei Patient:innen erkennen. Damit konnte gezeigt werden, was maschinelles Lernen in der medizinischen Diagnose ermöglichen kann.

- 1 Graewingholt A & Duffy S. 2021. Retrospective comparison between single reading plus an artificial intelligence algorithm and two-view digital tomosynthesis with double reading in breast screening, *J Med Screen* 28(3): 365–368. doi: 10.1177/0969141320984198
- 2 Perez MV, Kenneth W, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, Balasubramanian V, Russo AM, Rajmane A, Cheung L, Hung G, Lee J. 2019. Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation, *N Engl J Med*; 381: 1909–1917, DOI: 10.1056/NEJMoa1901183

Das dritte Beispiel ist eine App, die verspricht, dass sie zuverlässig bösartige Melanome von gutartigen braunen Flecken auf der Haut unterscheiden kann.³ Diese App hat dank maschinellen Lernens das Potential, Menschenleben zu retten. Ein zu spät erkannter Hautkrebs kann oft nicht mehr geheilt werden. Gleichzeitig kann Leben gefährdet werden, wenn der Benutzer sich auf einen falschen negativen Befund verlässt.

Ein letztes Beispiel: In Deutschland versterben jedes Jahr über 100.000 Menschen an Kammerflimmern. Der automatische Defibrillator kann das Leben dieser Menschen retten. Allerdings muss das Gerät zuverlässig erkennen, ob es sich um ein schockpflichtiges EKG handelt – schnell, zuverlässig und ohne einen Arzt⁴. In der nächsten Generation wird das noch besser gelingen – mit maschinellem Lernen.⁵

Man erkennt das extrem weite Spektrum an Anwendungen von ML in der Medizin. Die folgende Übersicht ist sicherlich nicht vollständig:

- **Diagnostische Bildgebung:** Trennen von Bildern mit und ohne Befund, Markieren auffälliger Bereiche, Registrieren, Segmentieren
- **Onkologie:** Identifikation der spezifischen Tumorklasse, genetische Eingruppierung
- **Strahlentherapie:** Vorschlag eines Bestrahlungsplans
- **Anästhesie:** Vorhersage kritischer Kreislaufzustände
- **Intensiv-Medizin:** Erkennung lebensbedrohlicher Zustände, Alarme
- **Ophthalmologie:** Befundung von OCT-Bildern

3 Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM & Thrun S. 2017. Dermatologist-level classification of skin cancer with deep neural networks, Nature, volume 542: 115–118.

4 Zur besseren Lesbarkeit werden in diesem Artikel abwechselnd die männliche und weibliche Form verwendet.

5 Shen C, Khedraki R, Cohoon T, Barakat A, Elashery AR, Shah K, Freed B, Walters D, Gibson DN, Olson N, Ng FS, Perry P, Rogers J and Bhavnani S. 2020. Development of a convolutional neural network for shockable arrhythmia classification within a next generation automated external defibrillator, Journal of the American College of Cardiology. Volume 75, Issue 11 Supplement 1, DOI: 10.1016/S0735-1097(20)34095-X

- **Kardiologie:** Auswertung von Langzeit-EKGs, z.B. Identifikation von Extrasystolen, Erkennen von Phasen mit Vorhofflimmern
- **Endokrinologie:** Ratgeber für die Insulingabe bei Patienten mit Diabetes mellitus
- **Dermatologie:** Erkennung maligner Melanome
- **Neurologie:** Vorhersage epileptischer Anfälle, Anfallstagebuch
- **Pathologie:** Erkennung suspekter Gebiete oder Befundung histologischer Schnitte von Gewebe
- **Notfall-Medizin:** Triage (schnelle Einschätzung, welche Patientin sofort ärztliche Hilfe benötigt und welcher Patient noch etwas warten kann), automatischer Defibrillator

In einem Positionspapier der acatech⁶ wird folgendes Ordnungsschema vorgeschlagen:

- **Gruppe 1:** ML-Applikationen für den Laien ohne Einbeziehung eines Arztes. Beispiele: Smart Watch, Wearables, Apps für Laien.⁷
- **Gruppe 2:** Einteilung von Patienten in Gruppen unterschiedlicher Priorität. Beispiel: Mammographie Screening.
- **Gruppe 3:** Unterstützung der Ärztin bei der Diagnose und Therapieauswahl. Beispiel: Anästhesie-Assistent, Computer Assisted Radiology, hier entscheidet am Ende immer noch eine Ärztin.
- **Gruppe 4:** medizinische Anwendungen, ohne dass ein Arzt dabei ist. Beispiel: automatischer Defibrillator.

6 Acatech (Hrsg.). 2020. Machine Learning in der Medizintechnik. Analyse und Handlungsempfehlungen: 10–13, <https://www.acatech.de/publikation/machine-learning-in-der-medizintechnik/>

7 Siehe auch Dössel O, Schäffter T, Kutyniok G, Rutert B (Hrsg.). 2021. Apps und Wearables für die Gesundheit: Chancen und Herausforderungen“, Denkanstoß 7, Berlin, <https://edoc.bbaw.de/frontdoor/index/index/docId/3696>

Quer dazu verläuft die Einteilung in zwei Fälle: (a) das Lernen ist beim Ausliefern des Medizinproduktes abgeschlossen (möglicherweise erfolgt nach einer gewissen Zeit ein Update) oder (b) das System passt sich auch nach der Auslieferung kontinuierlich an den Patienten an (selbstlernendes System). So könnte sich ein Assistent für den Diabetiker bei der Beratung für die Insulinabgabe kontinuierlich an den Patienten anpassen. Dem Autor sind keine zugelassenen Medizinsysteme vom Typ (b) bekannt. Trotzdem spielen sie in der öffentlichen Diskussion eine dominante Rolle.

Es folgt ein ganz kurzer Blick auf die Apps, die Smart Watches und die Wearables, also Anwendungen der Gruppe 1. Die gute Nachricht ist, dass erfolgreiche Apps dem Hersteller Millionen von Daten liefern, mit denen man hervorragend ML-Algorithmen trainieren kann. Damit sind auch gute Ergebnisse zu erwarten. Die schlechte Nachricht ist demgegenüber, dass die Erfassung der Daten oft unzuverlässig ist und dass in den Daten viele Fehler auftreten können. Zudem erfährt der Hersteller sehr persönliche Gesundheitsdaten vom Patienten, die er wiederum geschäftlich nutzen kann.

Im folgenden Abschnitt geht es um Fehler, die ein ML-Algorithmus machen kann. Es gibt zwei ganz unterschiedliche Arten von Fehlern, und es ist wichtig, dass diese Fehler in der Diskussion immer deutlich auseinandergehalten werden.

Es gibt Fehler bei der Implementation: Der Trainingsdatensatz enthält nicht ausgewogen alle Menschen, auf die der Algorithmus später angewendet werden soll (Bias), er enthält zu viele falsche Klassifizierungen, die Zahl der Trainingsdaten ist zu klein, der Algorithmus ist zu sehr an den Trainingsdatensatz angepasst und nicht generalisierbar (Overfitting und Leakage).

Auf der anderen Seite gibt es Fehler aufgrund einer unsicheren Datenlage: die „Wahrheit“⁸ ist zum Zeitpunkt, zu dem eine Entscheidung getroffen werden muss, nicht genau bekannt. Viele Entscheidungen in der Medizin basieren auf einer unsicheren Datenlage. Medizinische Studien zeigen immer eine statistische Schwankung. Es gibt fast keine Aussagen in der Medizin, die sich im Nachhinein immer als richtig erweisen.

8 Wahrheit ist hier gemeint im Sinne von Ground Truth, also im Sinne einer „bestmöglichen Entscheidung“.

Zum Fehler der ersten Art ist zu sagen: Die Qualität, die der Hersteller verspricht, muss er auch im klinischen Alltag halten können. Sonst haftet der Hersteller. Zum Fehler der zweiten Art ist zu sagen: Solche Fehler werden passieren. Das ist heute nicht anders. Diese Fehler treten bei allen Entscheidungen in der Medizin – mit oder ohne ML – auf. Aber wenn ein ML-Algorithmus nachweisen kann, dass er statistisch bessere Entscheidungen fällt als jede einzelne Ärztin, ist die Zahl der Fehler – gemittelt über alle Patienten – mit dem ML-Algorithmus kleiner als ohne ihn.

Kommen wir zum Thema Verantwortung als einem im Zusammenhang mit maschinellem Lernen besonders häufig genannten Begriff. Wer trägt bei der Verwendung von Systemen der Medizintechnik welche Verantwortung? Wir haben bereits ein sehr differenziertes System von Verantwortung und Haftung im Bereich der Medizin. Es ist fast immer eine Dreiecksbeziehung zwischen dem Patienten, der Ärztin, dem Hersteller. Jeder trägt für seinen Teil die Verantwortung. Die Patientin muss zu ihrer Behandlung eine „informierte Zustimmung“ geben. Darüber hinaus gelten die Arzthaftung und die Produkthaftung. Der Hersteller muss so weit wie möglich sicherstellen, dass sein Produkt die Normen erfüllt und „hält, was es verspricht“. Der Hersteller und sein Produkt werden von einer sogenannten benannten Stelle überwacht. Die benannten Stellen wiederum müssen sich gegenseitig überwachen. Täglich kommt es zu Streitfällen zwischen Patient:innen und Ärzt:innen. Der Patient meint: Meine falsche Behandlung hätte eine gute Ärztin besser gemacht. Die Ärztin entgegnet: Die „richtige“ Behandlung konnte man zu dem Zeitpunkt gar nicht kennen. Meistens endet das mit dem Satz: Der unglückliche Ausgang war „schicksalhaft“, ein in diesem Zusammenhang häufig verwendeter Begriff. Dann übernimmt tatsächlich niemand die Verantwortung für den schlechten Ausgang.

Es stellt sich nun die Frage, was genau bei Medizinprodukten, die ML-Algorithmen enthalten, eigentlich anders ist? Die meisten der oben genannten Aspekte gelten unverändert. Ein besonders wichtiger Aspekt ist aus meiner Sicht, dass die benannten Stellen prüfen, ob bei einem Medizinprodukt alle Normen eingehalten werden. Hier sehe ich Herausforderungen. Wir müssen möglicherweise die Normen anpassen, so dass die benannten Stellen klare Richtlinien bekommen, was sie bei einem ML-Medizinsystem genau prüfen müssen. Hierfür benötigen wir:

- Begriffsbestimmungen und genaue Definitionen,
- Anforderungen an die Qualität und die Größe der Datenbasis,
- Anforderungen an die Trennung von Trainings-, Validierungs- und Testdaten,
- Anforderungen an den Ausschluss/die Vermeidung von Bias,

- Anforderungen an die Angaben zur Treffsicherheit, zu Ausreißern und zur Robustheit,
- Anforderungen an die Transparenz,
- Anforderungen an die Interaktion mit der Ärztin (user interface).

Auf das Thema „Vermeidung von Bias“, also der unbewussten Bevorzugung bzw. Benachteiligung von Bevölkerungsgruppen, wird später noch genauer eingegangen.

Der Begriff „Treffericherheit“ erfordert eine genauere Definition. Bei einer Klassifikationsaufgabe können generell die Sensitivität und die Spezifität⁹ bestimmt werden. Weiter kann man im Einzelfall fragen, mit welcher Wahrscheinlichkeit eine andere Klassifizierung herausgekommen wäre (Klasse A: 51 % und Klasse B: 49 % oder Klasse A: 10 % und Klasse B: 90 %). Oder man kann fragen, ob eine leicht veränderte Eingangsgröße (z. B. durch einen Messfehler) zu dem gleichen oder zu einem ganz anderen Ergebnis geführt hätte. Diese letzten beiden Angaben sind für einen Arzt wichtige Informationen. Sie sollten immer angegeben werden, auch weil sie dem Arzt deutlich vor Augen führen, dass keine mit ML erzeugte Klassifikation oder Vorhersage mit 100 % Sicherheit richtig ist (Anforderung an das User Interface).

Auch der Begriff „Transparenz“ muss genauer betrachtet werden. Man könnte Angaben darüber fordern, welches Merkmal in den Eingangsdaten zu der Entscheidung geführt hat. Einige der ML-Algorithmen (z. B. Entscheidungsbäume (Decision Trees)) können schon heute diese Information liefern. Bei den neuronalen Netzen ist das schwieriger. Daher spricht man dort oft von einer Black Box. Mathematiker haben aber Methoden erarbeitet, mit denen auch bei neuronalen Netzen aus der Black Box eine Gray Box wird, d. h. die Entscheidung des Algorithmus wird teilweise nachvollziehbar gemacht.¹⁰ Diese Transparenz ist natürlich sehr wünschenswert. Sie sollte, wo immer möglich, dargestellt werden. Sie darf aber keine zwingende Voraussetzung sein, da diese Forderung nicht immer erfüllt werden kann. Übrigens basieren auch heute viele Leitlinien der Medizin nicht auf genau bekannten Ursache-Wirkung-Prinzipien, sondern auf statistischem Wissen aus klinischen Studien (Grundprinzip der evidenzbasierten Medizin).

9 https://de.wikipedia.org/wiki/Beurteilung_eines_binären_Klassifikators#Sensitivität_und_Falsch-negativ-Rate

10 Binder A, Montavon G, Bach S, Müller KR, Samek W. 2016. Layer-wise Relevance Propagation for Neural Networks with Local Renormalization Layer, arXiv:1604.00825 or <https://doi.org/10.48550/arXiv.1604.00825>

Die an ML angepassten Normen müssen EU-weit gelten, besser noch weltweit und sie müssen nachprüfbar sein. Dabei stellt sich immer die Frage nach Referenzdatensätzen. Das sind Datensätze, die unabhängig vom Hersteller generiert werden und mit denen so objektiv wie möglich die Treffsicherheit quantitativ bestimmt werden kann. Der Vorschlag der EU zu einer „Verordnung zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz“¹¹ plädiert für eine risiko-basierte Festlegung von Mindeststandards. Auch in Deutschland wird an einer „Normungsroadmap künstliche Intelligenz“¹² gearbeitet, die auch den Bereich „KI in der Medizin“ umfasst.

Noch ein kurzer Blick auf die Gesetzeslage: Es sind in der 19. Legislaturperiode des Bundestages sehr viele Gesetze verabschiedet worden, die das Thema KI in der Medizin berühren. Insbesondere das Patientendatenschutzgesetz¹³ und die digitale Gesundheitsanwendungen Verordnung¹⁴. Im Patientendatenschutzgesetz ist auch die elektronische Patientenakte (ePA) geregelt, und das hat Bezug zu unserem Thema, denn mit Hilfe der ePA können wichtige Daten für die Forschung generiert werden. Das Gesetz sieht eine freiwillige Datenspende für die Forschung ab 2023 vor. Hierauf soll später noch einmal genauer eingegangen werden.

Im nächsten Abschnitt soll kurz auf ethische Aspekte eingegangen werden. Es gibt mittlerweile sehr viele Publikationen zum Thema Ethik und KI in der Medizin.¹⁵ Viele wiederholen nur Aspekte, die schon einmal vorher von anderen beschrieben wurden. Viele beschäftigen sich auch mit Problemen, die wir eigentlich gar nicht haben.

11 https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75e-d71a1.0019.02/DOC_1&format=PDF, 2021.

12 <https://www.din.de/resource/blob/772438/6b5ac6680543eff9fe372603514be3e6/normungsroadmap-ki-data.pdf>, 2020.

13 Patientendatenschutzgesetz PDSG 2020. <https://www.bundesgesundheitsministerium.de/patientendaten-schutz-gesetz.html>

14 Digitale Gesundheitsanwendungen Verordnung 2020. <https://www.bundesgesundheitsministerium.de/service/gesetze-und-verordnungen/guv-19-lp/digav.html>

15 Siehe z. B. AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations, in Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, Luetge C, Madelin R, Pagallo U, Rossi F, Schafer B, Valcke P, Vayena E. 2018. Minds and Machines. 28: 689–707, <https://doi.org/10.1007/s11023-018-9482-5>

Worauf wir bei ML in der Medizin achten müssen, ist völlig unbestritten¹⁶:

- Vorrang menschlichen Handelns und menschliche Aufsicht
- Technische Robustheit und Sicherheit
- Schutz der Privatsphäre und Datenqualitätsmanagement
- Transparenz
- Vielfalt
- Nichtdiskriminierung und Fairness,
- Gesellschaftliches und ökologisches Wohlergehen
- Rechenschaftspflicht

Die Themen „Technische Robustheit“, „Transparenz“ und „Rechenschaftspflicht“ (Verantwortung) wurden schon oben ausführlich betrachtet. Der „Schutz der Privatsphäre“ ist in der Europäischen Datenschutz-Grundverordnung und in der dazu gehörenden deutschen Datenschutz-Grundverordnung (DSGVO)¹⁷ sehr stringent geregelt. Es ist selbstverständlich, dass auch Medizinprodukte, die Algorithmen des maschinellen Lernens enthalten, diese Gesetze einhalten müssen. Auch bei der Entwicklung solcher Produkte müssen diese Gesetze beachtet werden.

Zum Thema „Vielfalt und Nicht-Diskriminierung“: Jede klinische Studie muss heute „Einschlusskriterien“ und „Ausschlusskriterien“ festlegen. Eine klinische Studie durchzuführen, die alle Menschen einschließt, ist fast unmöglich. Die Ergebnisse der Studie dürfen nur für die Patienten angewendet werden, die die Einschlusskriterien erfüllen. Dabei passieren auch Fehler – das war (leider) schon immer so. Das bedeutet für Medizinprodukte, die ML beinhalten: Eine ML-Anwendung muss genau spezifizieren, wofür und für wen sie gut ist (Medical Device Regulation (MDR)¹⁸). Im Datensatz und bei der klinischen Studie müssen alle Patient:innen, die die Einschlusskriterien erfüllen, gleichmäßig vertreten sein. Auf Gruppen von Patient:innen, die nicht beteiligt waren, darf das ML-Programm nicht angewendet werden. Das bedeutet zwar nicht, dass das Problem mit der Nichtdiskriminierung und Fairness (Bias) abschließend gelöst ist, aber es bedeutet, dass das Problem in der Medizin schon seit langem bekannt ist und es schon viele Richtlinien dazu gibt.

16 High-Level Expert Group on Artificial Intelligence of the European Commission (AI-HLEG), Ethics Guidelines for Trustworthy AI, 08.04.2019, <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

17 Datenschutz-Grundverordnung (DSGVO), 2018, <https://dsgvo-gesetz.de>

18 <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32017R0745&from=DE>

Zum Schluss noch ein Blick auf das Thema „Daten für die Forschung“. Zurzeit werden drei Säulen der Datengewinnung für die medizinische Forschung aufgebaut: NFDI4Health¹⁹, Gaia-X²⁰ und das Forschungsdatenzentrum am BfArM, das hier ausschließlich betrachtet werden soll.

Das Patientendatenschutzgesetz²¹ regelt die Möglichkeit, Daten aus der elektronischen Patientenakte (ePA) für die Forschung freizugeben. Die gematik²² hat den Auftrag, bis 2023 die Schnittstellen von der ePA zum Forschungsdatenzentrum am BfArM zur Verfügung zu stellen, das die Weitergabe der Daten für die Forschung organisiert. Die Daten aus der ePA können dann – nach der Einverständniserklärung der Patienten – schnell für Forschungsprojekte zur Verfügung stehen („Datenspende“). Das ist insbesondere aus dem Blickwinkel der Forschung und Entwicklung für neue medizinische Systeme mit maschinellem Lernen sehr zu begrüßen, da so hochwertige Datensätze mit Millionen von Patientendaten entstehen können. Aber wie kompliziert wird diese Einverständniserklärung („broad consent“) sein? Und wer darf Anträge auf Datenzugriff beim BfArM stellen? Hier gibt es noch offene Fragen. Im Patientendatenschutzgesetz steht folgendes: „Das Forschungsdatenzentrum [...] macht die übermittelten Daten [...] folgenden Nutzungsberechtigten zugänglich: den Hochschulen, den nach landesrechtlichen Vorschriften anerkannten Hochschulkliniken, öffentlich geförderten außeruniversitären Forschungseinrichtungen und sonstigen Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung, sofern die Daten wissenschaftlichen Vorhaben dienen“. Eine offene Frage ist, ob die medizintechnische Industrie und die Pharma-Industrie im Verbund mit einer Universität Zugriff auf die Daten bekommen darf? Insgesamt haben wir in Deutschland einen Flickenteppich von Landesdatenschutzverordnungen und Krankenhausgesetzen, die in unterschiedlicher Weise mal das eine erlauben und das andere verbieten. Und die Patienteneinwilligung ist offenbar ein sehr kompliziertes Thema. Der sogenannte „broad consent“, der von Expertinnen und Experten der Medizininformatikinitiative erarbeitet und zunächst von allen 16 Datenschutzbeauftragten für gut befunden wurde,²³ ist jetzt wieder umstritten.

19 <https://www.nfdi4health.de>

20 <https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Artikel/UseCases/framework-of-medical-records-in-europe.html>

21 Patientendatenschutzgesetz PDSG <https://www.bundesgesundheitsministerium.de/patientendaten-schutz-gesetz.html>

22 <https://www.gematik.de/anwendungen/e-patientenakte/>

23 <https://www.medizininformatik-initiative.de/de/mustertext-zur-patienteneinwilligung>, 2020.

Eine Reihe von Umfragen zu diesem Thema zeigt, dass es in der Bevölkerung eine hohe Bereitschaft gibt, Gesundheitsdaten anonymisiert für die wissenschaftliche Forschung freizuschalten.²⁴ Aber meistens können die Daten nur pseudonymisiert werden, da es irgendwo einen Schlüssel gibt, mit dem die Identität des Datenspenders zurückgewonnen werden kann. Ohne diesen Schlüssel können Datenbanken nicht kontinuierlich um neue Daten der Patient:innen erweitert werden. Es gibt Bedenken in der Bevölkerung, Daten aus der klassischen Gesundheitsversorgung für kommerzielle Anbieter freizuschalten. Aber wie sollen die Ergebnisse der Forschung beim Patienten ankommen, wenn wir kommerzielle Anbieter ausschließen? Und bei Gesundheits-Apps geben viele Menschen ihre Daten völlig bedenkenlos kommerziellen Anbietern.

Zum Schluss sollen neun Thesen, die zum Teil aus der acatech-Arbeitsgruppe²⁵ hervorgegangen sind, vorgestellt werden.

- Maschinelles Lernen in der Medizin wird den Arzt nicht ersetzen, sondern ihn unterstützen.
- In wenigen Ausnahmen, zum Beispiel in Notfällen, ist ein sofortiges Handeln – auch ohne Anwesenheit einer Ärztin – durch das System mit maschinellem Lernen erforderlich.
- Medizinsysteme mit maschinellem Lernen sollten – wo immer möglich – erklären können, warum sie zu einer bestimmten Aussage gekommen sind („Transparenz“), und eine „Treffsicherheit“ angeben.
- Die Prinzipien von Arzt- und Produkthaftung müssen in ihrer Anwendung auf maschinelles Lernen in der Medizin genauer untersucht werden. Rechtssicherheit ist eine notwendige Voraussetzung für die Einführung von ML in der Medizin.
- Die Forschungsförderung im Bereich des maschinellen Lernens in der Medizin wird als „gut“ eingeschätzt. Trotzdem sind weitere Anstrengungen nötig, zum Beispiel in den Bereichen Transparenz, Treffsicherheit, Evaluierung und klinische Anwendung.

24 <https://www.pwc.de/de/gesundheitswesen-und-pharma/healthcare-barometer-2020.html>, 2020.

25 <https://www.acatech.de/publikation/machine-learning-in-der-medizintechnik/>, 2020.

- Der Schutz der persönlichen Daten muss – wie durch die Datenschutz-Grundverordnung (EU-DSGVO), das Bundesdatenschutzgesetz (BDSG) und das Patientendaten-Schutz-Gesetz (PDSG) vorgeschrieben – gewährleistet sein.
- Es sind Lösungen notwendig, um große medizinische Datenbanken für die Forschung und Entwicklung zusammenzustellen. Die Generierung von Referenzdatensätzen für die objektive Qualitätsmessung von ML-Algorithmen sollte gefördert werden. Die Einrichtung verschiedener Datenzentren für die Forschung wird begrüßt. Die Datenspende muss als „etwas Gutes“ und für alle Menschen Nützlich angesehen werden.
- Es muss geklärt werden, wie klinische Studien für den Beweis der Wirksamkeit eines ML-Systems aussehen müssen.
- Die regulatorischen Aspekte der Zulassung und Zertifizierung von Medizinprodukten (Normen) müssen an neue Aspekte von ML angepasst werden.

Alle diese Überlegungen führen zu einem Plädoyer für einen klugen und besonnenen Einsatz von maschinellem Lernen in der Medizin. Die großen Vorteile sollten genutzt werden, ohne dabei die berechtigten Bedenken zu vernachlässigen.