

Prof. Dr. med. Jens Meier

*Klinik für Anästhesiologie und operative Intensivmedizin, Kepler Universitätsklinikum,
Linz Johannes Kepler Universität, Linz*

Anwendung der KI im Krankenhaus

1. Einleitung	46
2. Methoden der Künstlichen Intelligenz	47
3. Probleme der Anwendung des Maschinellen Lernens im klinischen Alltag	55
4. Ausblick	59
Literatur	60

1. Einleitung

Heutzutage gibt es nur wenige Bereiche des täglichen Lebens, die nicht von Maschinellern oder Künstlicher Intelligenz berührt werden.¹ Diese Methoden der modernen Informatik haben mittlerweile Einzug in sämtliche Bereiche unseres Daseins erhalten. Viele der Möglichkeiten, die durch die Anwendung von Maschinellern und Künstlicher Intelligenz entstehen, werden von den allermeisten Menschen in Abhängigkeit von der jeweiligen Persönlichkeit als vorteilhaft und nützlich wahrgenommen.² Die wenigsten von uns haben Probleme damit, sich im täglichen Leben bei der Erstellung von E-Mails, bei der Navigation mit Hilfe von Navigationssystemen oder bei der automatisierten Erstellung von Einkaufslisten helfen zu lassen. In der öffentlichen Diskussion wird zwar vielfach darüber diskutiert, dass die Künstliche Intelligenz auch durchaus eine Gefahr im täglichen Leben darstellen kann, da ihre Eigenschaft, von großen Datenmengen abhängig zu sein, dem generellen Bedürfnis nach Privatsphäre widerspricht, im täglichen Leben aber werden meist die Vorteile größer als die Nachteile eingeschätzt. Man ist häufig überrascht, wie bereitwillig der eine oder andere private Daten zur Nutzung freigibt, wenn man im Gegenzug dafür eine Erleichterung bei der Durchführung unliebsamer Aufgaben erhält.

Ganz anders ist die Situation im Krankenhaus: Hier sind sich die Menschen sehr wohl bewusst, dass Gesundheitsdaten schützenswert sind.³ Aus diesem Grund lässt sich auch eine gewisse Technikskepsis im Bereich der Medizin nicht verleugnen. Zwar wird eine technisierte Medizin im Krankenhaus eher akzeptiert als z.B. in der Hausarztpraxis, allerdings muss man feststellen, die Möglichkeiten des maschinellen Lernens und der Künstlichen Intelligenz bis heute im Spital noch nicht im selben Maße ausgenutzt werden, wie dies in anderen Bereichen außerhalb der medizinischen Domäne der Fall ist.⁴

Das Ziel des vorliegenden Artikels ist es, die Möglichkeiten der Künstlichen Intelligenz im Krankenhaus darzustellen und die zugrunde liegenden mathematischen Mechanismen verständlich zu machen, um dann die technischen Möglichkeiten aufzuzeigen und anschließend die menschlichen Bedenken und Probleme mit dieser Technologie zu diskutieren.

1 Pfannstiel 2022, Furnham 2010

2 Gessl et al. 2019

3 Danaher 2015, Fritsch 2022

4 He 2019

2. Methoden der Künstlichen Intelligenz

Für viele Menschen stellen die Methoden des Maschinellen Lernens und der Künstlichen Intelligenz gewissermaßen eine „Blackbox“ dar. Nur wenige, technik-begeisterte Individuen sind dazu in der Lage, die Hintergründe einzelner Algorithmen zu verstehen, und nur den wenigsten dürfte bewusst sein, dass sämtliche Algorithmen der Künstlichen Intelligenz im Wesentlichen auf drei unterschiedlichen Klassen von Algorithmen basieren. Diese drei Klassen werden als „supervised“, „unsupervised“ und „reinforcement learning“ bezeichnet.

Hinter diesen drei Begriffen verbergen sich verschiedene Strategien zur Lösung von unterschiedlichen Problemarten. Im Bereich des „supervised learnings“ werden vor allem Klassifikations- und Regressionsprobleme gelöst. Das bekannteste Beispiel außerhalb der Klinik ist hierfür die Bilderkennung. Mit Hilfe von „supervised learning“ ist es möglich, unterschiedliche Bilder mit vorher gelernten Klassen zu bezeichnen. Derselbe Algorithmus kann im Bereich der klinischen Medizin z.B. für die Unterteilung von Patientinnen und Patienten mit unterschiedlichen Erkrankungen verwendet werden. Weiterhin können die Algorithmen des „supervised learnings“ auch dazu benutzt werden, aus unterschiedlichen Eigenschaften eines Patienten bestimmte kontinuierliche Variablen vorherzusagen.⁵ Im Zentrum des „unsupervised learnings“ steht das clustering, also die Bildung von Gruppen aus einzelnen Datensätzen. „Unsupervised learning“ kann also dazu verwendet werden, eine vorher definierte Menge von Patienten in unterschiedliche Gruppen einzuteilen, die dann im besten Fall individuell unterschiedlich behandelt werden können.⁶ Die dritte Gruppe der typischerweise im klinischen Alltag verwendeten Algorithmen, das sogenannte „reinforcement learning“ beschäftigt sich vor allem mit dem Lernen von Strategien, um eine vorher definierte Aufgabe möglichst erfolgreich lösen zu können. Das bekannteste Beispiel außerhalb der Klinik ist hierbei das Erlernen von Strategien für unterschiedliche Brettspiele wie z.B. Schach oder Go. In diesem Bereich sind die Methoden des „reinforcement learnings“ mittlerweile so erfolgreich, dass die erreichten Ergebnisse nahezu immer besser sind als die der besten Menschen.⁷

5 Hastie (a) 2009

6 Hastie (b) 2009

7 Sutton 2018

2.1. „Supervised learning“ im Krankenhaus

Wie bereits erwähnt, können die Methoden des „supervised learnings“ im klinischen Alltag für eine Vielzahl von Klassifikations- und Regressionsproblemen verwendet werden. Unter einem Klassifikationsproblem versteht man die Bestimmung einer sogenannten „target variable“, also z.B. das Vorliegen einer Erkrankung oder das antizipierte Ergebnis einer Behandlung (z.B. krank vs. nicht krank, infiziert vs. nicht infiziert, verstirbt, verstirbt nicht etc. ...) aus der Kombination einer unterschiedlichen Anzahl von Merkmalen, den sogenannten „features“. „Features“ können hierbei Zahlen (z.B. Mess- oder Laborwerte), aber auch kategorielle Beschreibungen (z.B. männlich/weiblich, behandelt/nicht behandelt) oder jede Kombination der beiden sein. Auch Bilder und Texte werden häufig als „feature“ benutzt, nachdem sie adäquat vorverarbeitet wurden. Das eigentliche Prinzip des „supervised learnings“ ist hierbei unabhängig vom verwendeten Algorithmus und der zugrunde liegenden Fragestellung immer das gleiche: Ausgehend von einem historischen Datensatz, der sowohl die „features“ als auch die „target variable“ erhält, wird ein Modell trainiert. In Abhängigkeit vom verwendeten Algorithmus werden hierbei mit unterschiedlichen mathematischen Verfahren die Zusammenhänge zwischen „features“ und „target variable“ beschrieben. Die wahrscheinlich bekannteste Algorithmenklasse sind hierbei die sogenannten „neuronalen Netze“. Sie ermöglichen es, eine nichtlineare, multidimensionale Modellierung dieser Zusammenhänge im Rahmen mehrerer Optimierungsläufe zu berechnen. Andere bekannte Verfahren sind „random forests“, „support vector machines“, „Hopfield networks“ und das „Xgradient boosting“. Je nach Beschaffenheit und Größe des Datensatzes liefern diese Methoden unterschiedlich gute Ergebnisse, wobei sich aber die inhaltliche Qualität nicht vorhersagen lässt, da sie einerseits von der Beschaffenheit der Daten und andererseits von der Größe des Datensatzes abhängt.

Nach dem Training des Modells wird mit einem weiteren Datensatz die Vorhersagequalität des Modells überprüft. Wichtig ist hierbei, dass für die Überprüfung ein unabhängiger Datensatz verwendet wird, um so feststellen zu können, wie das errechnete Modell mit neuen, bisher unbekannten Daten umgehen kann.

2.1.1. Anwendung in der Klinik

In der klinischen Forschung sind die Algorithmen des „supervised learnings“ die mit Abstand am häufigsten verwendeten Methoden überhaupt. Die klassische Domäne dieser Methoden ist hierbei die Bilderkennung.⁸ Sei es, die Diagnose von unterschiedlichen Karzinomformen in der Dermatologie, die Klassifizierung von histologischen Befunden in der Pathologie, die Bildsegmentierung oder die Dia-

⁸ Aljuaid 2022

gnosestellung in der Radiologie oder die Analyse von EKG-Daten im Hinblick auf Arrhythmie- oder Ischämiedetektion, immer kommen Verfahren aus dem Bereich des „supervised learnings“ zur Anwendung.

Die hierbei erreichten Ergebnisse sind beeindruckend. Es existiert eine Vielzahl von Anwendungen, mit denen genau umschriebene Probleme entweder auf dem gleichen oder einem besseren Niveau als von Menschen gelöst werden können. Hierzu zählen vor allem Analysen aus dem Bereich der Bilderkennung, zum Beispiel im Bereich der Dermatologie, der Radiologie oder der Histopathologie.⁹ Darüber hinaus haben diese Modelle mehrere entscheidende Vorteile gegenüber menschlichen Ärztinnen und Ärzten: Sie arbeiten mit der gleichen Qualität zu jeder Tages- und Nachtzeit, sind beliebig skalierbar, die Fehlerquote ist gering und vorhersehbar, und sie besitzen eine hohe Ausfallsicherheit. Aus diesem Grund mag es nicht verwundern, dass vielfach kolportiert wird, dass die Methoden des „supervised learnings“ innerhalb kurzer Zeit die ärztliche Expertise ersetzen könnten. Insbesondere dann, wenn die bisher eingesetzten Modelle noch genauer und aussagekräftiger werden sollten. Dem gegenüber steht aber eine prinzipielle Einschränkung dieser Methoden, nämlich die Beschränkung der Aussagekraft des Modells auf eine zu Beginn definierte Domäne.

2.1.2. Limitierungen des „supervised learnings“

Im Bereich des „supervised learnings“ wird jedes Modell auf einen bestimmten Datensatz trainiert. In den allermeisten Fällen handelt es sich hierbei um historische Daten, die aus den Aufzeichnungen bisheriger Behandlungen gewonnen werden. Um ein aussagefähiges Modell zu trainieren, ist hierbei eine Vielzahl von Daten notwendig. Datensätze mit wenigen hundert Einzelfällen sind für das Training nicht geeignet. Erst ab vielen tausend bis hunderttausenden Datensätzen können Modelle trainiert werden, die verlässlich im klinischen Alltag einsetzbar sind. Nur wenige Institutionen besitzen Datensätze für ein bestimmtes Modell in der ausreichenden Anzahl. Werden Modelle auf kleineren Datensätzen trainiert, besteht die Gefahr des sogenannten „overfittings“. Das Modell lernt hierbei die Spezifika eines speziellen Datensatzes, ohne ausreichend zu generalisieren, um dadurch auch mit anderen, neuartigen Daten umgehen zu können. Zum Beispiel kann es möglich sein, dass die Vorhersagequalität eines Modells im Bereich des Spitals, aus dem die Trainingsdaten kommen, extrem hoch ist, dass dies aber an anderen Spitälern mit andersartiger Patientenselektion nicht der Fall ist (z.B. periphere Klinik vs. Maximalversorger, unterschiedliche Altersstruktur der PatientInnen etc.). Aus diesem Grund ist es sehr schwierig, allgemeingültige Modelle an unterschiedlichen Kliniken oder in unterschiedlichen Situationen zu verwenden.

9 Hogarty 2020, Chang 2019, Hosny 2018

Streng genommen kann nur garantiert werden, dass ein Modell nur in Situationen verwendet werden kann, wo die features aus derselben Grundgesamtheit kommen, auf die das Modell auch trainiert worden ist. Bereits beim Wechsel von einem zum anderen Spital ist diese Voraussetzung nicht mehr vollständig gegeben, was die Vorhersagequalität zwangsläufig reduzieren muss. Aus diesem Grund existieren auch fast keine allgemeingültigen Modelle, die universell eingesetzt werden können. Andererseits haben einzelne Kliniken nur in Einzelfällen Datensätze, die es erlauben würden, individuelle Modelle zu trainieren und anschließend zu validieren.

2.2. „Unsupervised learning“ im Krankenhaus

Das Grundprinzip des „unsupervised learnings“ ist es, Ordnung in einem vorher ungeordneten Datensatz zu schaffen. Ein typisches Beispiel ist eine riesige Sammlung von Büchern ohne Beschriftungen oder Kategorien, wobei diese auf der Grundlage von inhärenten Mustern oder Ähnlichkeiten geordnet werden sollen. „Unsupervised learning“ ist mit diesem Szenario vergleichbar, bei dem der Algorithmus Muster in den Daten erforscht und erkennt, ohne dass er explizite Anleitungen oder gekennzeichnete Beispiele erhält.

Im Gegensatz zum „supervised learning“, bei dem der Algorithmus auf markierten Daten mit vordefinierten Ergebnissen trainiert wird, arbeitet das „unsupervised learning“ auf eine eher explorative Weise. Das Hauptziel ist es, verborgene Strukturen, Beziehungen oder Cluster in den Daten aufzudecken, was es zu einem unschätzbaren Werkzeug für die Entdeckung von Mustern macht, die sich der menschlichen Beobachtung entziehen könnten. Die Methoden des „unsupervised learnings“ eignen sich für das sogenannte „clustering“ und die Dimensionalitätsreduktion.

2.2.1. Clustering

Clustering ist ein grundlegendes Konzept des unüberwachten Lernens, bei dem ähnliche Datenpunkte zu Gruppen zusammengefasst werden.

In der medizinischen Forschung kann das unüberwachte Clustering beispielsweise auf Patientendaten angewendet werden, um verschiedene Untergruppen mit ähnlichen Merkmalen zu identifizieren. Dies kann bei personalisierten Behandlungsplänen und dem Verständnis von Krankheiten helfen.¹⁰

¹⁰ Jiang 2023

2.2.2. Dimensionalitätsreduktion

Unüberwachte Lerntechniken eignen sich auch hervorragend zur Dimensionalitätsreduzierung, d.h. zur Vereinfachung komplexer Datensätze durch Beibehaltung wesentlicher Merkmale.

Im medizinischen Bereich kann dies bei der Analyse hochdimensionaler Daten wie Genomik- oder Bildgebungsdaten von Nutzen sein und eine effizientere Analyse und Interpretation ermöglichen.

2.2.3. Anwendung im Krankenhaus

Unüberwachtes Lernen wurde bereits mehrfach eingesetzt, um auf der Grundlage von Patientendaten Subtypen von Krankheiten zu identifizieren. In der Krebsforschung kann es zum Beispiel helfen, zwischen verschiedenen molekularen Subtypen zu unterscheiden und so maßgeschneiderte Behandlungsstrategien zu entwickeln.¹¹ Darüber hinaus ist es auch möglich, durch die Analyse molekularer und chemischer Daten bei der Identifizierung potenzieller Arzneimittelkandidaten und dem Verständnis ihrer Wechselwirkungen mit biologischen Systemen Fortschritte zu erzielen. Dies hat das Potenzial, die Prozesse der Arzneimittelentdeckung zu revolutionieren.¹² Unüberwachtes Lernen könnte aber auch einen wichtigen Beitrag zur Patientenstratifizierung liefern, bei der Personen mit ähnlichen klinischen Profilen in Gruppen zusammengefasst werden. Dies ermöglicht gezieltere und personalisierte Untersuchungen und therapeutische Ansätze.¹³ Das Erkennen von Anomalien in medizinischen Daten, z.B. die Identifizierung von Unregelmäßigkeiten in Patientenakten oder Anomalien in medizinischen Bildern, ist eine weitere wichtige Anwendung des „unsupervised learnings“.

Die wichtigsten Methoden des „unsupervised learnings“ sind das „K-Means clustering“, das „hierarchische Clustering“ und die „principal components analysis“ (PCA). „K-Means“ ist ein beliebter Clustering-Algorithmus, der Daten auf der Grundlage von Ähnlichkeit in „K“-Cluster unterteilt. In der medizinischen Forschung kann er verwendet werden, um Patientenkohorten mit ähnlichen klinischen Merkmalen zu identifizieren. Die Anzahl der Cluster muss allerdings vom Untersucher vorgegeben werden, was zwangsläufig die erhaltenen Ergebnisse beeinflusst. Bei „hierarchischem Clustering“ wird ein Baum von Clustern erstellt, der hierarchische Beziehungen zwischen Datenpunkten aufzeigt. In der Genetik kann hierarchisches Clustering eingesetzt werden, um die Beziehungen zwischen verschiedenen Genausdrücken zu verstehen. Die PCA ist ein Verfahren zur Dimen-

¹¹ Wang 2023

¹² Stevenson 2023

¹³ Maslove 2022

sionalitätsreduzierung, das hochdimensionale Daten in eine niedrigdimensionale Darstellung umwandelt. In der medizinischen Bildgebung kann die PCA die Interpretation komplexer Bilder vereinfachen, wobei die wesentlichen Informationen erhalten bleiben.

2.2.4. Einschränkungen

Die Modelle des „unsupervised learnings“ können manchmal schwierig zu interpretieren sein, insbesondere wenn es sich um hochdimensionale Daten handelt. Es ist wichtig, die Implikationen der entdeckten Muster zu verstehen, was insbesondere im medizinischen Kontext nicht immer ganz einfach ist. Darüber hinaus hängt die Wirksamkeit des „unsupervised learnings“ stark von der Qualität und Repräsentativität der Eingabedaten ab. Verrauschte oder verzerrte Daten können zu ungenauen Ergebnissen führen. Die Wahl des richtigen unüberwachten Lernalgorithmus hängt von der Art der Daten und den spezifischen Zielen der Analyse ab. Eine sorgfältige Abwägung der verschiedenen Algorithmen ist unerlässlich.

Aber selbst nach optimaler Auswahl des zu verwendeten Verfahrens sind die Ergebnisse vielfach schwierig zu interpretieren. Aus diesem Grund werden diese Modelle im klinischen Alltag bisher kaum verwendet.

2.3. „reinforcement learning“ im Krankenhaus

In der dynamischen Landschaft der künstlichen Intelligenz erweist sich das „reinforcement learning“ als ein faszinierendes Paradigma, das die Art und Weise widerspiegelt, wie Menschen und Tiere durch Interaktion mit ihrer Umwelt lernen. Die zugrundeliegenden Prinzipien des Lernens werden hierbei verwendet, um Computer auf eine ähnliche Art und Weise Strategien lernen zu lassen. Stellen Sie sich vor, Sie bringen einem Hund einen neuen Trick bei. Sie programmieren nicht explizit jede Bewegung, sondern geben Feedback – Belohnungen für korrekte Handlungen und sanfte Korrekturen für Fehler. Das „reinforcement learning“ funktioniert nach einem ähnlichen Prinzip: Ein Agent lernt, Entscheidungen zu treffen, indem er mit der Umgebung interagiert und Rückmeldungen in Form von Belohnungen oder Bestrafungen erhält.

Beim „reinforcement learning“ ist ein „Agent“ der Lernende oder Entscheidungsträger, der mit einer „Umgebung“ interagiert. Der Agent führt „Aktionen“ durch, die die Umgebung beeinflussen. Im medizinischen Kontext könnte ein Agent ein System für Behandlungsempfehlungen sein, das mit Patientendaten interagiert, um personalisierte Interventionen vorzuschlagen. Im Mittelpunkt des „reinforcement learnings“ steht das Konzept der Belohnungen und Bestrafungen. Der Agent

zielt darauf ab, die kumulativen Belohnungen im Laufe der Zeit zu maximieren, indem er lernt, welche Aktionen zu positiven Ergebnissen führen. Im Gesundheitswesen könnte dies zur Optimierung von Behandlungsplänen führen, um die Genesung des Patienten zu maximieren und gleichzeitig die negativen Auswirkungen zu minimieren. Hierbei steht der Agent allerdings vor einem Dilemma zwischen der Erkundung neuer Aktionen, um bessere Strategien zu entdecken, und der Anwendung bereits bekannter, bewährter Aktionen, um den unmittelbaren Nutzen zu maximieren. Das Gleichgewicht zwischen Exploration und Anwendung bereits bewährter Aktionen ist eine der zentralen Herausforderungen des „reinforcement learnings“.

In klinischen Studien kann das „reinforcement learning“ die Auswahl von Behandlungsoptionen optimieren, indem es ein Gleichgewicht zwischen der Erkundung neuartiger Therapien und der Anwendung bewährter wirksamer Maßnahmen herstellt.

2.3.1. Anwendungen in der Medizin

Verstärkungslernen wird zur Optimierung von Behandlungsplänen für chronische Erkrankungen eingesetzt, um personalisierte und adaptive Interventionen im Laufe der Zeit sicherzustellen.¹⁴ Ein zentrales Problem im klinischen Alltag ist hierbei die Tatsache, dass die eigentliche Aktion nicht einfach durchgeführt werden kann. Es muss im klinischen Setting deshalb mit Datensätzen trainiert werden, die möglichst viele unterschiedliche Therapieoptionen umfassen. Eine weitere Anwendungsmöglichkeit ist das Design klinischer Studien, das komplexe Entscheidungsprozesse beinhaltet. „Reinforcement learning“ kann aber auch zur Optimierung von Arzneimitteldosierungen eingesetzt werden, da es dazu in der Lage ist, die Medikamentendosierung auf der Grundlage der Patientenreaktionen anzupassen und so die therapeutischen Ergebnisse zu optimieren.¹⁵ Eine weitere Domäne des „reinforcement learnings“ ist die Optimierung der Ressourcenzuweisung im klinischen Alltag. Krankenhäuser und Gesundheitsdienstleister stehen vor der Herausforderung, Ressourcen effizient zuzuweisen. „Reinforcement learning“ kann dabei helfen, die Ressourcenzuweisung zu optimieren und eine effektive Patientenversorgung sicherzustellen.¹⁶

¹⁴ Komorowski 2018

¹⁵ Zeng 2022

¹⁶ Xu 2023

2.3.2. Einschränkungen des „reinforcement learnings“ für die klinische Anwendbarkeit

„Reinforcement Learning“ im Gesundheitswesen wirft ethische Überlegungen auf, wie z.B. die Sicherstellung von Fairness bei Behandlungsempfehlungen und der Umgang mit Verzerrungen in den Trainingsdaten. Einige Algorithmen des Verstärkungslernens können darüber hinaus eine große Anzahl von Interaktionen mit der Umgebung erfordern, um effektiv zu lernen, was in Szenarien, in denen die Datenerfassung ressourcenintensiv ist, eine Herausforderung darstellt. Und letztlich ist das Verständnis und die Interpretation der von Modellen des verstärkten Lernens getroffenen Entscheidungen eine ständige Herausforderung, insbesondere in kritischen Bereichen wie dem Gesundheitswesen, da insbesondere in diesem Bereich vielfach der Interpretierbarkeit der erhaltenen Ergebnisse eine große Bedeutung beigemessen wird.

Nichts destoweniger verspricht „reinforcement learning“, das sich an der Art und Weise orientiert, wie lebende Organismen von ihrer Umgebung lernen, eine Revolutionierung der Entscheidungsfindung im Gesundheitswesen und darüber hinaus. Seine Anwendungen bei der Optimierung von Behandlungen, der Planung klinischer Studien und der Ressourcenzuteilung unterstreichen sein Potenzial, die Ergebnisse für die PatientInnen zu verbessern und die Prozesse im Gesundheitswesen zu rationalisieren. Bei der Erforschung des Bereichs des „reinforcement learnings“ durch die Wissenschaft ist die Berücksichtigung ethischer Implikationen und fortlaufender Fortschritte bei den algorithmischen Techniken von entscheidender Bedeutung, um die transformative Kraft dieser Technologie verantwortungsvoll zu nutzen.

3. Probleme der Anwendung des Maschinellen Lernens im klinischen Alltag

Auf den ersten Blick mag es verwundern, dass die oben genannten Methoden des maschinellen Lernens bis heute kaum Einzug in den klinischen Alltag gefunden haben. Hierfür sind mehrere unterschiedliche Ursachen verantwortlich, die einen breiten Einsatz bisher effektiv verhindert haben.

Wie bereits weiter oben ausgeführt, sind sämtliche Methoden des maschinellen Lernens von der Verfügbarkeit von Trainingsdaten abhängig, die im besten Fall ein genaues Abbild der Situation darstellen, in der die entwickelten Algorithmen später verwendet werden sollen. Für die meisten Kliniken stehen diese Daten nicht in der notwendigen Anzahl oder Genauigkeit zur Verfügung. Dies führt dazu, dass viele Modelle nur auf Daten anderer Kliniken trainiert werden können. In der Folge muss also immer davon ausgegangen werden, dass diese Daten unter Umständen nicht die PatientInnen abbilden, auf die die trainierten Modelle angewendet werden sollen. Dies schränkt aber die Vorhersagequalität stark ein, da die Übertragbarkeit der errechneten Modelle nicht garantiert werden kann. Eine denkbare Lösung wäre hier das „pooling“ von mehreren unterschiedlichen Datenquellen, was aber aufgrund der Datenschutzgrundverordnung (DSGVO) nur mit großem Aufwand möglich ist. Gänzlich rechtlich ungeklärt bis zum heutigen Zeitpunkt ist darüber hinaus auch die Frage, ob historische Daten von PatientInnen ohne deren Zustimmung für solche kombinierten Modelle verwendet werden dürfen. Ohne Einwilligung ist dies nur sehr schwer vorstellbar, was die Entwicklung solcher Modelle nahezu unmöglich macht. Bekanntestes Beispiel in diesem Zusammenhang ist „Google Health“. Diesem Zweig des Alpha-Konzerns wurde mehrfach vorgeworfen, Daten von PatientInnen verwendet zu haben, ohne diese ausreichend zu informieren oder sie an der Verwertung ausreichend zu beteiligen.¹⁷

¹⁷ Mearian 2019, 2020

3.2. Interpretierbarkeit der Ergebnisse von Machine-learning-Modellen

Viele ÄrztInnen und Ärzte haben bis heute große Bedenken hinsichtlich einer Verwendung von „machine learning“ und Künstlicher Intelligenz im klinischen Alltag, da die Ergebnisse für sie zwar leicht interpretierbar, aber kaum herleitbar sind. Die zugrundeliegenden Algorithmen werden im Allgemeinen als „black box“ verstanden, was häufig zu Unbehagen und einem gewissen Vertrauensverlust führt. Überraschend ist hierbei, dass wir im täglichen Alltag oft keine Probleme damit haben, Ergebnisse von Technologien zu akzeptieren, die nur von einer kleinen Teilgruppe der Bevölkerung verstanden werden, solange sie nicht mit Künstlicher Intelligenz in Verbindung gebracht werden. Das klassische Beispiel hierfür ist z.B. die weit verbreitete Anwendung von Navigationssystemen. Nur die wenigsten Menschen „verstehen“ die zugrundeliegenden technischen Details, es existiert aber kaum ein Neuwagen, der ohne ein funktionierendes Navigationssystem verkauft wird, und kaum ein Autofahrer bzw. eine Autofahrerin verlässt sich heute im Alltag rein auf gedrucktes Kartenmaterial. Umso mehr überrascht, dass in anderen Bereichen die Methoden der Künstlichen Intelligenz auf breiter Front abgelehnt werden. Dies mag zu einem großen Teil an der negativen Berichterstattung in den Medien liegen, ist aber trotzdem nicht immer ganz verständlich. Überraschenderweise steigt die Akzeptanz aber wieder, wenn die Ergebnisse einer Künstlichen Intelligenz so beeindruckend sind, dass sie im Alltag faszinierend wirken. Bestes Beispiel sind hier aktuell die sogenannten „large language models“ (LLM), deren Ergebnisse eine so hohe Qualität besitzen, dass sie zwar vielfach angewendet werden, die zugrundeliegenden Modelle aber nur von den wenigsten Menschen verstanden werden. Ob eine bessere Aufklärung über die Funktionsweise von Machine-learning-Modellen hier in Zukunft zu einer Verbesserung dieser Situation führen wird, muss sich erst noch zeigen.

3.3. Rechtliche Bedenken

In Österreich existiert der sogenannte Arztvorbehalt, der zum Inhalt hat, dass bestimmte Tätigkeiten und Maßnahmen aufgrund einer gesetzlichen Festlegung nur von einem/r ordnungsgemäß ausgebildeten und approbierten ÄrztIn ausgeübt beziehungsweise durchgeführt werden dürfen. Mit dieser Festlegung ist aber auch die Verantwortlichkeit auf den ersten Blick klar geregelt: Die Ärztin, der Arzt, die/der eine Untersuchung oder eine Behandlung durchführt, ist hierfür vollumfänglich verantwortlich. Hieraus ergibt sich zwangsläufig, dass sämtliche Modelle der Künstlichen Intelligenz und des maschinellen Lernens bestenfalls „beratend“ zur

Seite stehen können. Zum jetzigen Zeitpunkt wird eine Diagnose niemals ohne einen Arzt bzw. eine Ärztin gestellt, und auch eine Therapie kann nicht durch Maschinen indiziert werden. Die häufig geäußerte Angst von PatientInnen, einer künstlichen Intelligenz ausgeliefert zu sein, entspricht demnach zum jetzigen Zeitpunkt nicht den Tatsachen. Diese scheinbare Sicherheit darf aber nicht darüber hinwegtäuschen, dass sich die Situation innerhalb kürzester Zeit nicht ändern kann. Bereits jetzt sind KI-Systeme menschlichen Experten in vielen Bereichen überlegen. Die bekanntesten Beispiele hierfür sind die Brettspiele Schach und Go. Hier gibt es keinen Menschen, der auch nur annähernd die erreichte Spielstärke der besten Algorithmen erreicht. Werden die zugrundeliegenden Algorithmen als „Beratungssysteme“ genutzt, dann versetzt dies den eigentlichen Spieler/die eigentliche Spielerin in eine bemerkenswerte Situation. Ohne Zweifel werden die vom Algorithmus vorgeschlagenen Züge denen des jeweiligen Spielers/der jeweiligen Spielerin überlegen sein. Jede menschliche Abweichung wird also in den allermeisten Fällen dazu führen, dass ein schlechteres Ergebnis erreicht wird, und aus diesem Grund wird kaum jemand die Empfehlung des Algorithmus ignorieren. Übertragen auf die klinische Situation führt dies faktisch dazu, dass die Ärztin bzw. der Arzt die Verantwortung für eine Entscheidung trägt, die eigentlich autonom von einem Algorithmus getroffen wurde, aber erst im Nachhinein legitimiert wird. Rein rechtlich mag daran nichts auszusetzen sein, dennoch ist aber leicht einzusehen, dass eine derartige Situation zwangsläufig zu Unbehagen führen muss. Hier wird es in Zukunft wichtig sein, dass der Gesetzgeber im medizinischen Umfeld einen klaren Handlungsspielraum vorgibt, der diesem Umstand Rechnung trägt. Ob im Rahmen einer solchen Regelung dann die Hersteller der Algorithmen, diejenigen, die die Daten bereitstellen, oder jemand anderes auch in die Verantwortung genommen werden können, lässt sich zum jetzigen Zeitpunkt noch nicht sagen. Auch wird es nur für Teilbereiche im medizinischen Umfeld möglich sein, klar zu definieren, wann die Entscheidungen eines Algorithmus denen der betreuenden ÄrztInnen überlegen sind.

3.4. Emotionale Hürden

Wenige wirtschaftliche Geschäftssituationen sind so emotional besetzt, wie das Verhältnis zwischen Ärztin oder Arzt und PatientIn. Typischerweise vertrauen PatientInnen auf die ärztliche Heilkunst, die sich im Allgemeinen aus Ausbildung, Erfahrung, Empathie und Intuition zusammensetzt. Viele der getroffenen Entscheidungen im medizinischen Bereich sind mit gravierenden Konsequenzen verbunden. So verwundert es nicht, dass die Suche nach der besten Ärztin oder dem besten Arzt für viele Patientinnen und Patienten eine wichtige Rolle einnimmt.

PatientInnen erhoffen sich, im Rahmen der Behandlung nach neuesten medizinischen Erkenntnissen und mit neuesten Therapieschemata behandelt zu werden. Im besten Fall schließt diese Behandlung eine Berücksichtigung der psychosozialen Situation mit ein, sodass ein ganzheitlicher Heilungsprozess initiiert werden kann. Die Erwartungen an die MedizinerInnen sind groß. Heutzutage wird nahezu als selbstverständlich vorausgesetzt, dass sämtliche Therapien nach der bestmöglichen Evidenz erfolgen sollen, außerdem wird erwartet, dass sämtliche Informationen für die Behandlung in Betracht gezogen werden und zu jeder Tages- und Nachtzeit die optimale Synthese aus allen Informationsquellen erfolgen kann. Darüber hinaus legen viele PatientInnen einen großen Wert auf ein empathisches Vertrauensverhältnis. Das Phänomen des Doctor-Hoppings ist nicht selten die Folge eines enttäuschten Vertrauensverhältnisses oder einer Kränkung, die nicht im fachlichen, sondern im persönlichen Bereich liegt.

Ob PatientInnen sich lieber Computern anvertrauen, darf daher verzweifelt werden. Die ärztliche Heilkunst besteht eben nicht nur das Abarbeiten von standardisierten Diagnose- und Therapieabläufen, sondern eben auch aus verständnisvollem Zuhören, Eingehen auf die Wünsche und das Weltbild der PatientInnen und ein menschlich annotierter Ratschlag. All das ist zum jetzigen Zeitpunkt wahrscheinlich nicht durch Algorithmen leistbar. Aus diesem Grund gibt es bereits den Ruf nach adäquaten Einsatz von ÄrztInnen bei der Versorgung von PatientInnen.¹⁸ Allerdings kann sich die Einstellung der Bevölkerung in den nächsten Jahren dahingehend ändern, dass auch Computer als Behandler akzeptiert werden.

18 Mittelman 2018

4. Ausblick

Sind die Methoden des Maschinellen Lernens und der Künstlichen Intelligenz bereits breit im klinischen Alltag etabliert? In Anbetracht der theoretischen Möglichkeiten muss diese Frage klar mit „nein“ beantwortet werden. Allerdings haben die zugrundeliegenden Methoden in jedem Fall das Potential, unser Verständnis und unseren Zugang zur Medizin grundlegend zu verändern. Die hierbei existierenden Möglichkeiten erscheinen unüberschaubar, und im Bereich von großen Konzernen existiert eine Art „Goldgräberstimmung“, die zu großen Investitionen und fast unüberschaubaren Veränderungen und Entwicklungsprozessen führt. Diese Entwicklung droht im Moment eine Vielzahl der beteiligten Proponenten zu überrollen, und man hat oft den Eindruck, dass potentielle AnwenderInnen mit der Geschwindigkeit dieser Entwicklungen nicht mithalten können. Auch die nicht ganz unbegründete Angst, überflüssig werden zu können, erleichtert hierbei die Situation nicht.

Unstrittig scheint zu sein, dass sich diese Entwicklung nicht aufhalten lässt. Es wird unsere Aufgabe sein, diese Entwicklung adäquat zu begleiten. Wünschenswert wäre es, wenn diese Entwicklung von Ärztinnen und Ärzten mit angetrieben und gesteuert wird. Auch in Zukunft werden sie die Anbieter der „Ware“ Gesundheit sein, wenngleich der Zugang sich wahrscheinlich grundlegend verändern wird. Im schlechtesten Fall wird die Entwicklung von großen Konzernen übernommen, was in der Folge zu einer verminderten Einflussnahme durch ÄrztInnen und PatientInnen führen wird. Der Schritt dorthin ist sicherlich nicht einfach, da die Grundprinzipien des Maschinellen Lernens und der Künstlichen Intelligenz derzeit nur von wenigen verstanden werden und aus diesem Grund die Beurteilung von neuen Verfahren sehr schwierig sein dürfte. Um aber auch in Zukunft eine adäquate PatientInnenversorgung zu ermöglichen, sind Anstrengungen in diese Richtung unerlässlich.

Literatur

- Aljuaid A, Anwar M. Survey of Supervised Learning for Medical Image Processing. *SN Comput Sci.* 2022 May 17;3(4):292.
- Chang HY, Jung CK, Woo JI, Lee S, Cho J, Kim SW, et al. Artificial Intelligence in Pathology. *J Pathol Transl Med.* 2019 Jan 15;53(1):1–12.
- Danaher J. Why AI Doomsayers are Like Sceptical Theists and Why it Matters. *Minds Mach.* 2015 Aug;25(3):231–46.
- Fritsch SJ, Blankenheim A, Wahl A, Hetfeld P, Maassen O, Deffge S, et al. Attitudes and perception of artificial intelligence in healthcare: A cross-sectional survey among patients. *Digit Health.* 2022 Jan;8:205520762211167.
- Furnham A. 50 Schlüsselideen Psychologie [Internet]. Heidelberg: Spektrum Akademischer Verlag; 2010 [cited 2023 Dec 5]. Available from: <http://link.springer.com/10.1007/978-3-8274-2379-5>
- Gessl AS, Schlögl S, Mevenkamp N. On the perceptions and acceptance of artificially intelligent robotics and the psychology of the future elderly. *Behav Inf Technol.* 2019 Nov 2;38(11):1068–87.
- Hastie T, Tibshirani R, Friedman J. Overview of Supervised Learning. In: *The Elements of Statistical Learning* [Internet]. New York, NY: Springer New York; 2009 [cited 2023 Dec 5]. p. 9–41. (Springer Series in Statistics). Available from: http://link.springer.com/10.1007/978-0-387-84858-7_2 (Hastie a)
- Hastie T, Tibshirani R, Friedman J. Unsupervised Learning. In: *The Elements of Statistical Learning* [Internet]. New York, NY: Springer New York; 2009 [cited 2023 Dec 5]. p. 485–585. (Springer Series in Statistics). Available from: http://link.springer.com/10.1007/978-0-387-84858-7_14 (Hastie b)
- He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med.* 2019 Jan;25(1):30–6.
- Hogarty DT, Su JC, Phan K, Attia M, Hossny M, Nahavandi S, et al. Artificial Intelligence in Dermatology – Where We Are and the Way to the Future: A Review. *Am J Clin Dermatol.* 2020 Feb;21(1):41–7.
- Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer.* 2018 Aug;18(8):500–10.
- Jiang Z, Bo L, Wang L, Xie Y, Cao J, Yao Y, et al. Interpretable machine-learning model for real-time, clustered risk factor analysis of sepsis and septic death in critical care. *Comput Methods Programs Biomed.* 2023 Nov;241:107772.

- Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med*. 2018 Nov;24(11):1716–20.
- Maslove DM, Tang B, Shankar-Hari M, Lawler PR, Angus DC, Baillie JK, et al. Redefining critical illness. *Nat Med*. 2022 Jun;28(6):1141–8.
- Mearian L, Yes, Google's using your healthcare data – and it's not alone. *Computerworld*. November 15, 2019.
- Mearian L, Yes, Google's using your healthcare data – and it's not alone. *Computerworld*. Retrieved September 11, 2020.
- Mittelman M, Markham S, Taylor M. Patient commentary: Stop hyping artificial intelligence – patients will always need human doctors. *BMJ*. 2018 Nov 7;k4669.
- Pfannstiel MA, editor. Künstliche Intelligenz im Gesundheitswesen: Entwicklungen, Beispiele und Perspektiven [Internet]. Wiesbaden: Springer Fachmedien Wiesbaden; 2022 [cited 2023 Dec 5]. Available from: <https://link.springer.com/10.1007/978-3-658-33597-7>
- Stevenson GA, Kirshner D, Bennion BJ, Yang Y, Zhang X, Zemla A, et al. Clustering Protein Binding Pockets and Identifying Potential Drug Interactions: A Novel Ligand-Based Featurization Method. *J Chem Inf Model*. 2023 Nov 13;63(21):6655–66.
- Sutton RS, Barto AG. Reinforcement learning: an introduction. Second edition. Cambridge, Massachusetts: The MIT Press; 2018. 526 p. (Adaptive computation and machine learning series).
- Wang Z, Katsaros D, Wang J, Biglio N, Hernandez BY, Fei P, et al. Machine learning-based cluster analysis of immune cell subtypes and breast cancer survival. *Sci Rep*. 2023 Nov 3;13(1):18962.
- Xu H, Fang Y, Chou CA, Fard N, Luo L. A reinforcement learning-based optimal control approach for managing an elective surgery backlog after pandemic disruption. *Health Care Manag Sci*. 2023 Sep;26(3):430–46.
- Zeng J, Shao J, Lin S, Zhang H, Su X, Lian X, et al. Optimizing the dynamic treatment regime of in-hospital warfarin anticoagulation in patients after surgical valve replacement using reinforcement learning. *J Am Med Inform Assoc*. 2022 Sep 12;29(10):1722–32.