Der Radiologe

Leitthema

Radiologe 2020 · 60:24–31
https://doi.org/10.1007/s00117-019-00616-x
Online publiziert: 6. Dezember 2019



**Jens Kleesiek[1,4] · Jacob M. Murray[1,2] · Christian Strack[1,2] · Georgios Kaissis[3] · Rickmer Braren[3,4]**

[1] AG Computational Radiology, Abteilung Radiologie, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Deutschland

[2] Universität Heidelberg, Heidelberg, Deutschland

[3] Department of Diagnostic and Interventional Radiology, School of Medicine, Technical University of Munich, München, Deutschland

[4] German Cancer Consortium (DKTK), Heidelberg, Deutschland

# Wie funktioniert maschinelles Lernen?

*„I predict that within 10 years no medical imaging study will be reviewed by a radiologist until it has been pre-analyzed by a machine." R. Nick Bryan [1].*

Bereits um 1900 gab es die Technologie für automatische Aufzüge, es dauerte aber bis in die 1950er Jahre, bis diese angewendet wurde. Bis dahin hatten sich die Fahrstuhlnutzer mit einem Fahrstuhlführer aus Fleisch und Blut wohler gefühlt [14]. Heute ist es kaum vorstellbar, dass ein Fahrstuhl nicht automatisch funktioniert. Eine ähnliche Umstellung steht uns auch in der Radiologie bevor. Die Anwendungsbereiche des maschinellen Lernens sind schon jetzt sehr vielfältig und werden eine immer prominentere Rolle in der Radiologie einnehmen. Daher ist es wichtig, dass Ärzte sich mit der Technologie auseinandersetzen, ein grundlegendes Verständnis dafür entwickeln und sich vor allem auch mit den Grenzen und Limitationen beschäftigen.

## (Künstliche) Intelligenz

Was ist Intelligenz? Diese Frage ist nicht leicht zu beantworten. In der Psychologie unterscheidet man verschiedene Kategorien der kognitiven Leistungsfähigkeit. Es gibt jedoch keine einheitliche Definition. Dies erschwert auch die Definition der künstlichen Intelligenz (KI), die man als maschinelle Variante der allgemeinen Intelligenzdefinition betrachten kann (**Abb. 1**). Im technischen Sinne handelt es sich um einen Zweig der Informatik, der sich mit der Simulation von intelligentem Verhalten in Computern beschäftigt. Häufig wird damit die Fähigkeit einer Maschine, intelligentes menschliches Verhalten nachzuahmen, gemeint. Dies wird auch durch das Zitat von Elaine Rich unterstrichen: „Artificial intelligence is the study of how to make computers do things at which, at the moment, people are better" [12].

In der Vergangenheit hat sich gezeigt, dass wir als Menschen dazu neigen, die Messlatte immer höher zu legen, sobald ein als intelligente Leistung definiertes Ziel erreicht wurde. Beispiele hierfür sind die Spiele Schach und Go, bei denen Computer inzwischen die besten menschlichen Spieler schlagen („super-human performance") und daraufhin neue Ziele definiert werden. Im Fall des Spiels Go ist dies nun, dass der Computeralgorithmus[1] nur so viele Trainingspartien spielen darf, wie ein menschlicher Go-Spieler durchschnittlich in seinem Leben spielt [6].

Brettspiele wie Schach, und auch aktuelle Anwendungen in der medizinischen Domäne, sind Beispiele für schwache KI-Verfahren („narrow/weak AI"). Es handelt sich hierbei um Verfahren, die genau ein Problem lösen können, aber versagen, wenn es um abgewandelte Aufgabenstellungen geht, und seien die Abweichungen auch nur gering. Ein Beispiel hierfür ist ein Computerprogramm, das den Lungentumor im Röntgenthorax erkennt, aber den Pneumothorax übersieht, weil es nicht für diese Aufgabe trainiert wurde. Wichtig ist hier zu erwähnen, dass die Algorithmen nicht zwingend „narrow" sind, sondern die Art und Weise, wie sie eingesetzt werden. Es gibt Untergruppen von Algorithmen, die besonders gut für die Erkennung von Mustern in Bilddaten geeignet sind. Hierbei ist es unwichtig, ob es sich um Tumoren in radiologischen oder histologischen Bildern, um Fußgänger in Videosequenzen oder entfernte Galaxien in Teleskop-Bildern handelt. Die Güte hängt dann primär von der Validität der Trainingsdaten ab, die zur Verfügung stehen.

Angestrebt wird die Entwicklung von starken bzw. generellen KI-Verfahren („general/strong AI"), die nicht nur eine Aufgabe, z. B. die Mustererkennung, gut lösen können, sondern eine Kombination von menschlichen Charakteristika aufweisen sollen, wie die Fähigkeit zu schlussfolgern, Sinn und Bedeutung zu entdecken, zu verallgemeinern und aus früheren Erfahrungen zu lernen.

Symbolische KI („good old fashioned AI", GOFAI) ist ein Sammelbegriff für verschiedene Methoden, die sich mit der abstrakten *symbolischen* (menschenlesbaren) Darstellung von Problemen beschäftigen. Hierbei werden Zeichenket-

[1] Bei einem Algorithmus handelt es sich um eine eindeutige Handlungsvorschrift, eine Art *Rezept*, das aus endlich vielen, wohldefinierten Einzelschritten besteht. Zur Lösung einer Aufgabe wird mit Hilfe eines Algorithmus eine definierte Eingabe in eine definierte Ausgabe überführt.

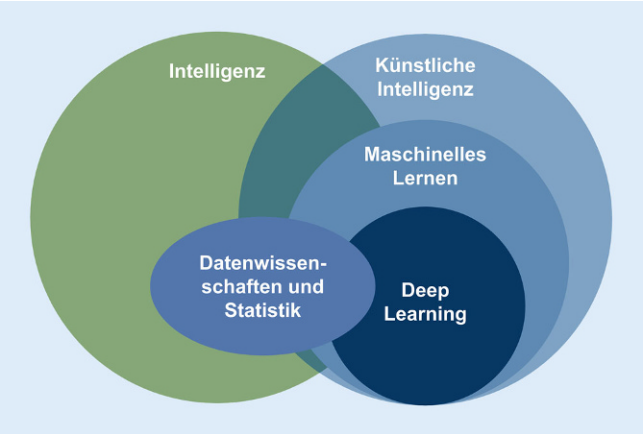


**Abb. 1** ◀ Künstliche Intelligenz – maschinelles Lernen – Deep Learning

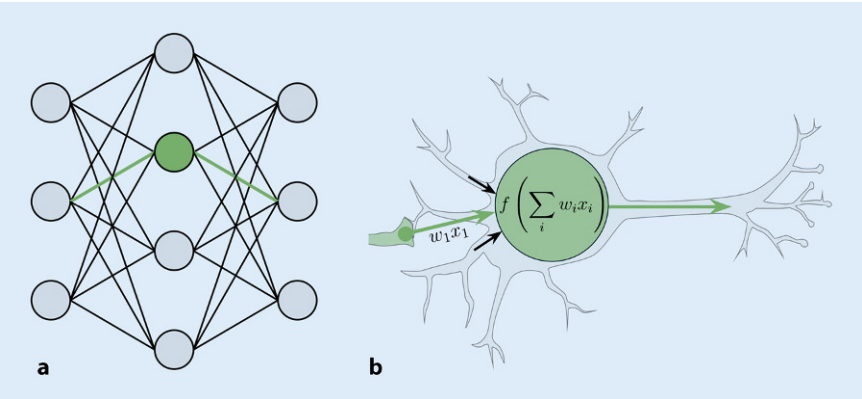


**Abb. 2** ▲ Künstliches neuronales Netz und biologisches Vorbild. **a** Einfaches Netzwerk mit 3 Neuronenschichten, die durch senkrechte Reihen von Kreisen schematisch dargestellt sind: Eingabe- (*links* 3 Neurone), Zwischen- (*Mitte* 4 Neurone, „hidden layer") und Ausgabeschicht (*rechts* 3 Neurone). Soll beispielsweise vorhergesagt werden, ob ein Erguss in einem Röntgenbild vorliegt, könnte die Eingabeschicht der Anzahl der Bildpunkte entsprechen und die Ausgabeschicht würde 2 Neurone, eins für jede Klasse (Erguss ja/nein), enthalten. **b** Schematische Darstellung eines Neurons als Ausschnitt von **a**. Die Aktivierung des Neurons wird durch eine gewichtete Summe der vorgeschalteten Neurone ($w_ix_i$) berechnet, die über eine Aktivierungsfunktion *f* an die nachgeschalteten Neuronen weitergegeben wird. Exemplarisch für einen Ein- und einen Ausgang des grünen Neurons dargestellt. Der Schwellenwert ist durch $x_0 = 1$ modelliert. Während des Trainings des KNN werden die Gewichte *w* gelernt. Weitere Details im Abschnitt „Künstliche neuronale Netze"

ten (Symbole) und deren Relationen, die menschlichen Einheiten oder Konzepten entsprechen, in einer Wissensdatenbank gespeichert. Wenn beispielsweise Konzepte von Städten und Bundesländern vorgegeben werden (Heidelberg ist eine Stadt, Baden-Württemberg ist ein Land) sowie deren Relationen (Heidelberg liegt in Baden-Württemberg), dann sind Abfragen nach Städten und wo diese liegen trivial. Synonym werden daher für symbolische KI auch die Begriffe regelbasierte Systeme oder Expertensysteme verwendet. Diese Systeme erlauben es, Schlussfolgerungen und Entscheidungen auf der Grundlage der gespeicherten Fakten und Regeln zu treffen.

GOFAI war das dominierende Paradigma der KI-Forschung bis Ende der 1980er Jahre. Aktuell sind andere Methoden vorherrschend, die Daten analysieren und so Wissen aufbauen können. Maschinelles Lernen (ML) ist ein Sammelbegriff für diese Methoden, der häufig synonym zum Ausdruck künstliche Intelligenz verwendet wird. Es werden Computeralgorithmen verwendet, um autonom aus Daten und Informationen zu lernen. Im Gegensatz zu den oben angeführten Expertensystemen werden Computer nicht explizit für die Lösung der Aufgaben programmiert, sondern die eingesetzten Algorithmen können selbstständig lernen und sich an die Aufgabe anpassen. Man unterscheidet verschiedene Lernparadigmen, die überwachtes („supervised"), unüberwachtes („unsupervised") und verstärkendes („reinforcement") Lernen sowie Mischformen und graduelle Abstufungen beinhalten. Beim verstärkenden Lernen sucht ein Agent eine Strategie, die eine Belohnung maximiert. Ein Agent ist in diesem Fall eine virtuelle Komponente, die die Entscheidung („policy") lernt, welche Handlung („action") in einer bestimmten Situation („state") durchgeführt wird. Soll beispielsweise der kürzeste Weg aus einem Labyrinth gefunden werden, kann der Agent lernen, wie er an Wegkreuzungen („state") abbiegen muss („action"). Im Fall des unüberwachten Lernens werden durch Algorithmen Muster in den Eingabedaten identifiziert. Beispiele hierfür sind Cluster-Algorithmen, bei denen basierend auf Ähnlichkeitsmustern in den Daten ähnliche Datenpunkte in *Haufen* gruppiert werden. Beim überwachten Lernen spricht man auch von *Lernen mit Lehrer*, da während des Trainings neben den Eingabedaten auch die gewünschten Lösungen zur Verfügung stehen. Hierzu werden im Verlauf noch Beispiele vorgestellt.

Es gibt diverse und stetig wachsende Methoden, die im Bereich des ML eingesetzt werden. Die sicherlich aktuell am häufigsten in der Presse erwähnten Verfahren umfassen Varianten sog. künstlicher neuronaler Netze (KNN; ◘ **Abb. 2**). Wenn diese besonders viele Neuronenschichten aufweisen, spricht man auch von Deep Learning (DL). Zusammenfassend kann man vereinfacht sagen, dass die Verfahren eine Abbildung von Eingabedaten auf eine Ausgabe *lernen*. Das Lernen wird, je nach Verfahren, unterschiedlich realisiert und beinhaltet eine lineare oder nichtlineare Kombination (Gewichtung) der Eingabedaten, um auf die gewünschte Ausgabe abzubilden. Im Fall der KNN werden die Gewichte basierend auf den Trainingsdaten gelernt. Bei der Ausgabe kann es sich um eine Klassifikation, z. B. Unterscheidung von Tumor- und normalem Gewebe, oder Regression, Vorhersage eines kontinuierlichen Werts für ein Merkmal, beispielsweise eines Laborwerts, handeln.

Ein wesentlicher Unterschied zwischen KNN und anderen ML-Verfahren ist die Art und Weise der Merkmalsextraktion, die häufig ursächlich für ein besseres Abschneiden der neuronalen Netze ist. Bei klassischen ML-Ansätzen definiert der Mensch die Merkmale, die aus den Daten extrahiert werden und dann als Eingabe für den Algorithmus dienen. Hierbei handelt es sich beispielsweise um die Form und Farbe eines Objekts, die aus den Bilddaten berechnet werden. Im Gegensatz dazu bilden sich in den KNN während des Trainings Merkmalsdetektoren aus; sie lernen also, auf welche Merkmale geachtet werden muss, um ein optimales Ergebnis zu erzielen (weitere Details unten). In Bezug auf das vorherige Beispiel kann dies, muss aber nicht, so etwas wie Form und Farbe sein. Es können aber auch Merkmale bzw. hochdimensionale Kombinationen von Merkmalen sein, die für die menschliche Auffassung nicht intuitiv begreiflich (zu komplex) sind.

Die Datenmenge und -qualität ist von äußerster Bedeutung für das Training der Algorithmen und beeinflusst die Güte der Verfahren wesentlich. Insbesondere im medizinischen Bereich sind aus ethischen und datenschutzrechtlichen Gründen häufig weniger Daten verfügbar als in anderen Domänen. Dies ist ein Grund, warum hier die Verfahren noch nachhinken. Da es sich um hochsensible Gesundheitsdaten handelt, müssen – zurecht – Sicherheitsmaßnahmen und Schutzmechanismen greifen. Um dennoch den technologischen Fortschritt auch für medizinische Verfahren nutzen zu können, gibt es eine Vielzahl von Projekten, die sich mit speziellen Lösungsansätzen beschäftigen [5, 13, 16, 18]. Ein besonderes Projekt in Bezug zur onkologischen Bildgebung ist die Joint Imaging Platform (JIP) des Deutschen Konsortiums für translationale Krebsforschung (DKTK) [8]. Hier werden u. a. Methoden des verteilten Lernens („federated learning“) entwickelt, um Algorithmen anstatt der Patientendaten zwischen den teilnehmenden Standorten auszutauschen. So ist gewährleistet, dass die Patientendaten sicher innerhalb der lokalen Klinik-Infrastruktur verbleiben.

Radiologe 2020 · 60:24–31 https://doi.org/10.1007/s00117-019-00616-x


J. Kleesiek · J. M. Murray · C. Strack · G. Kaissis · R. Braren

# Wie funktioniert maschinelles Lernen?


## Zusammenfassung

**Hintergrund.** Die Methoden des maschinellen Lernens und der künstlichen Intelligenz (KI) etablieren sich langsam aber sicher im medizinischen Alltag. Zukünftig werden sie uns bei Diagnose und Therapie unterstützen und so die Behandlung zum Wohl des Patienten verbessern. Es ist daher wichtig, sich mit diesem Thema auseinanderzusetzen und ein Grundverständnis dafür zu entwickeln.
**Ziel der Arbeit.** Dieser Artikel soll einen Überblick über das spannende und dynamische Feld des maschinellen Lernens geben und als Einführung in Methoden, v. a. des überwachten Lernens, dienen. Neben Definitionen und einfachen Beispielen werden auch Limitationen diskutiert.
**Schlussfolgerung.** Die Grundprinzipien der Methoden sind einfach. Dennoch sind die Gründe für eine Entscheidung häufig durch ihre hochdimensionale Natur nicht oder nur schwer durch den Menschen nachvollziehbar. Um Vertrauen in die neuen Technologien aufzubauen und eine sichere Anwendung zu garantieren, benötigen wir nachvollziehbare Algorithmen und prospektive *Wirksamkeitsstudien*.



## Schlüsselwörter

Neue Technologien · Maschinelles Lernen · Deep Learning · Künstliche neuronale Netzwerke · Digitale Kompetenz


# A primer on machine learning


## Abstract

**Background.** The methods of machine learning and artificial intelligence are slowly but surely being introduced in everyday medical practice. In the future, they will support us in diagnosis and therapy and thus improve treatment for the benefit of the individual patient. It is therefore important to deal with this topic and to develop a basic understanding of it.
**Objectives.** This article gives an overview of the exciting and dynamic field of machine learning and serves as an introduction to some methods primarily from the realm of supervised learning. In addition to definitions and simple examples, limitations are discussed.
**Conclusions.** The basic principles behind the methods are simple. Nevertheless, due to their high dimensional nature, the factors influencing the results are often difficult or impossible to understand by humans. In order to build confidence in the new technologies and to guarantee their safe application, we need explainable algorithms and prospective *effectiveness* studies.



## Keywords

New technologies · Machine learning · Deep learning · Artificial neural networks · Digital literacy


## Geschichte des Feldes

Eine Übersicht wichtiger Schritte in der Entwicklung von KNN ist in **Tab. 1** angeführt. Die Geburtsstunde des KI-Feldes wird auf den Zeitraum 1952–1956 datiert. Gefolgt von einer optimistischen Periode in den 60er und 70er Jahren des letzten Jahrhunderts, die durch den ersten sog. KI-Winter gebremst wurde, in dem die Erfolge des Feldes hinter den Erwartungen zurückblieben. Danach folgte ein weiterer Zyklus aus Aufschwung und Rückschlägen, der von den 80er bis in das erste Drittel der 90er Jahre reichte. Seitdem geht es ständig aufwärts, und KI-Methoden finden in immer mehr Bereichen Einzug, was durch deren Verwendung in realen Anwendungen unterstrichen wird.

Betrachtet man das Auf und Ab des Feldes in der Vergangenheit, stellt sich die Frage, was diesmal anders ist, und warum jetzt mit einem Durchbruch der KI-Anwendungen gerechnet werden kann. Zum einen ist das Feld methodisch gereift, und so hat auch die Rechenleistung der Computer in den letzten Jahren deutlich zugenommen, insbesondere durch die Verfügbarkeit immer leistungsfähiger Grafikprozes-

| Tab. 1 | Wichtige Meilensteine in der Entwicklung der künstlichen neuronalen Netzwerke |
|---|---|
| 1943 | Vorstellung eines Neuronenmodells auf Basis eines elektrischen Schaltkreises durch Warren McCulloch und Walter Pitts |
| 1949 | Donald Hebb stellt seine Lernregel vor, die auch das Grundgerüst für viele der heutigen KNN-Lernregeln darstellt |
| 1950 | Alan Turing schlägt den „Turing-Test" vor, mit dem bestimmt werden soll, ob ein Computer das Denkvermögen eines Menschen besitzt. Um den Test zu bestehen, darf ein Mensch, der mit zwei ihm unbekannten Gesprächspartnern, einer ein Computer, der andere ein Mensch, interagiert, nicht unterscheiden können, welcher der Computer und welcher der Mensch ist. Es gibt immer wieder Meldungen, dass der Test bestanden worden sei, zuletzt durch Google Duplex in 2018. Doch bisher scheint es noch ein offenes Problem zu sein, wenngleich eine Lösung nicht weit entfernt zu sein scheint |
| 1955 | Der Ausdruck *Artificial Intelligence* wird durch John McCarthy geprägt |
| 1956 | Die Dartmouth Sommerkonferenz findet statt. Sie gilt allgemein als Geburtsstunde der künstlichen Intelligenz als akademisches Fachgebiet |
| 1958 | Frank Rosenblatt stellt mit dem Perzeptron das erste künstliche neuronale Netzwerk vor, dessen Neurone in abgewandelter Form auch heute noch zum Einsatz kommen |
| 1979 | Fukushima stellt das Neocognitron vor. Dies ist die Geburtsstunde der Faltungsnetzwerke („convolutional neural networks", CNN), die gerade im Bereich der Mustererkennung in Bildern heute den Goldstandard darstellen |
| 1982 | Vorstellung der ersten Anwendung des Backpropagation-Algorithmus für das effiziente Training von neuronalen Netzen durch Paul Werbos [17] |
| 1986 | Erste Verwendung des Ausdrucks *Deep Learning* [3] |
| 1997 | Das Computerschachsystem Deep Blue besiegt den amtierenden Schachweltmeister Garri Kasparow. Dies gibt dem ML-Feld einen großen Aufschwung. Deep Blue verwendet jedoch keine KNN |
| 2015 | Dass CNN AlexNet gewinnt mit Abstand den ImageNet-Wettkampf, bei dem es um die Klassifikation von Bilddaten geht (die Datenbank enthält mehr als 14 Mio. Bilder mit über 20.000 Kategorien). Dies gab der Verwendung und Entwicklung von Deep Learning noch einmal enormen Aufwind. Im Jahr 2017 war dann erstmals die Fehlerrate der Algorithmen zur Bildklassifikation dieser Daten besser als die der menschlichen Referenz |

soren (GPUs). Die GPUs sind für die arithmetischen Operationen, die für das Training von KNN benötigt werden, optimiert, und es ist nun möglich, sehr tiefe Netze (mit vielen Neuronenschichten) zu trainieren. Diese haben sich als besonders leistungsfähig herausgestellt. Zum anderen wurde erkannt, dass durch die Verfügbarkeit von frei zugänglichen Algorithmen und Datensätzen das Feld profitiert.

## Lineare Separierbarkeit

Ein ML-Verfahren wandelt Eingabedaten in Ausgaben um. Diese Transformation wird im Fall des überwachten Lernens mittels bekannter Beispiele von Ein- und Ausgabedaten-Paaren *gelernt*. Für ein einfaches Klassifikationsbeispiel ist dies in **Abb. 3a** gezeigt. In der Abbildung sind Kreise und Quadrate zu sehen, die jeweils der Klassenzugehörigkeit entsprechen. Es kann eine Gerade bestimmt (gelernt) werden, die diese beiden Klassen voneinander trennt (lineare Separierbarkeit). Mit Hilfe dieser Gerade kann dann eine Zuordnung für unbekannte Datenpunkte (exemplarisch als rotes und grünes Fragezeichen dargestellt) vorgenommen werden. Mathematisch lässt sich dies an einem einfachen Beispiel erklären. Eine Gerade hat die allgemeine Form:

$$y = mx + b$$

$y$ und $x$ entsprechen Variablen, $m$ der Steigung und $b$ dem Y-Achsenschnittpunkt. Nehmen wir an, dass $m$ mit 1,5 und $b$ mit −1 bestimmt wurden. Dann kann die Gleichung aufgestellt und umgeformt werden als:

$$y = 1{,}5x - 1$$

$$1{,}5x - 1 - y = 0$$

Dargestellt als Ungleichung, kann die Klassenzugehörigkeit bestimmt werden:

$$\text{Klasse} = \begin{cases} \text{Kreis } 1{,}5x - 1 - y < 0 \\ \text{Quadrat } 1{,}5x - 1 - y \geq 0 \end{cases}$$

Wenn für das rote Fragezeichen die Merkmale $x = 6$ und $y = 1$ angenommen werden, dann erhalten wir 9-1-1 = 7 (Quadrat). Für das grüne Fragezeichen mit den Merkmalen $x = 4$ und $y = 6$ ergibt sich −1 (Kreis).

Aber wie können die Parameter $m$ und $b$ gelernt werden? Hierzu gibt es verschiedene Verfahren, die in der Regel auf Minimierung einer Fehlerfunktion („error or loss function") beruhen. So könnte ein trivialer Algorithmus darauf beruhen, dass $m$ und $b$ zufällig gewählt werden und dann evaluiert wird, wie viele der Trainingsbeispiele richtig klassifiziert wurden. Dies wird so lange wiederholt, bis eine ausreichende Genauigkeit erreicht wird. Es gibt viele Verfahren, die das Klassifikationsproblem effizienter und besser lösen.

Eine lineare Separierbarkeit ist nicht immer möglich. Einige Daten lassen sich nicht durch eine Gerade trennen, und es sind kompliziertere, nichtlineare Funktionen notwendig, beispielsweise eine *geschlängelte* Linie, welche die Klassen voneinander abgrenzt (**Abb. 3b**). Natürlich gibt es auch Daten, die sich überhaupt nicht trennen lassen, weil entweder nicht die richtigen Merkmale ausgewählt wurden, z. B. Anzahl der Räder, wenn man versucht PKW voneinander abzugrenzen, oder weil schlichtweg eine Trennung nicht möglich ist. Dies kann ggf. auch durch die Datenqualität bedingt sein, z. B. durch Rauschen, welches die Messwerte verfälscht.

## Künstliche neuronale Netze

KNN können u. a. zur Klassifikation und Regression[2] eingesetzt werden. Auch hierbei werden Parameter anhand von Beispieldaten optimiert, bis ein zufriedenstellendes Ergebnis erzielt wird. Durch eine Aktivierungsfunktion wer-

[2] Regression bezeichnet die Vorhersage eines quantitativen Werts, beispielsweise eines Laborparameters.

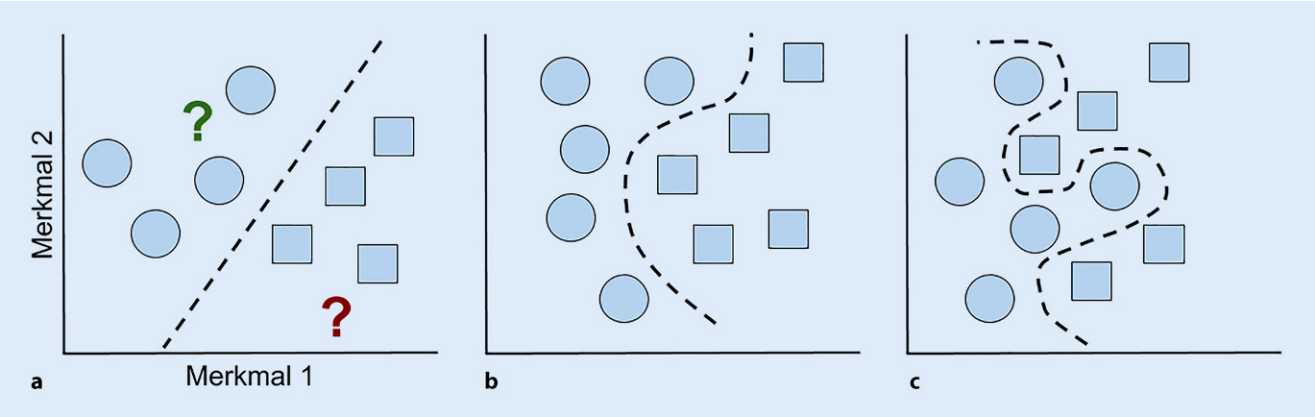


**Abb. 3** ▲ Klassifikation. **a** *Kreise* (Klasse 1) und *Quadrate* (Klasse 2) sind durch Merkmal 1 (X-Achse) und Merkmal 2 (Y-Achse) charakterisiert. Dies könnten beispielsweise Größe und Gewicht sein. Die beiden unbekannten Eingabedaten, das *rote* bzw. *grüne Fragezeichen*, können einer Klasse zugeordnet werden, wenn die Trenngerade bekannt ist. Diese kann durch ML-Verfahren gelernt werden. **b** Klassifikation mittels nichtlinearer Trennungslinie **c** Overfitting durch zu starke Anpassung an die Trainingsdaten

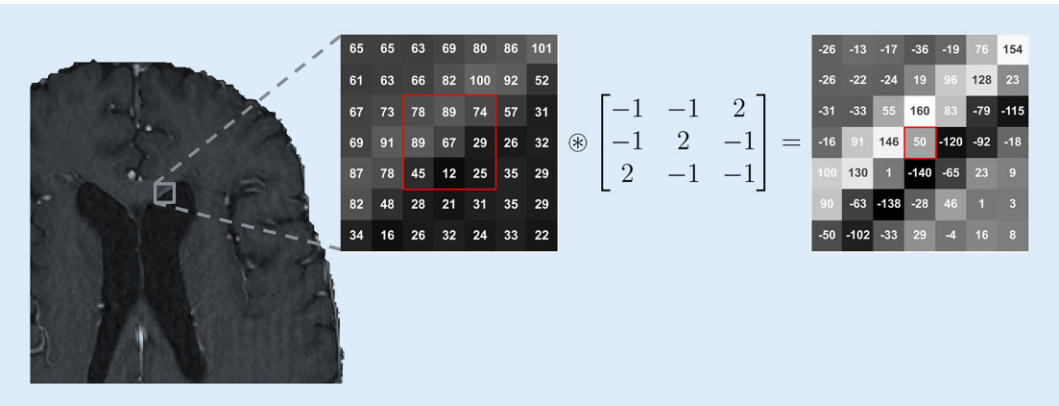


**Abb. 4** ▲ Ausschnitt einer Aktivierungskarte. Der vergrößerte Ausschnitt eines Gehirns (MRT) zeigt die Grauwerte der einzelnen Pixel (linke Matrix). Die Faltung mit einem Filter speziell für diagonale Kanten (mittlere Matrix) ergibt das *rechte Bild*. Die Kante wurde *erkannt* (rechte Matrix). Exemplarisch dargestellt ist der Wert 50, der sich durch die Summe der Multiplikation der *rot markierten* Werte mit dem Filter ergibt. Während des Trainings werden die Filter durch das Faltungsnetzwerk (CNN) gelernt

den Nichtlinearitäten eingeführt. Aktivierungsfunktionen stellen eine Abstraktion des Aktionspotenzials biologischer Zellen dar und werden im englischen Sprachgebrauch auch als „squashing" oder „transfer functions" bezeichnet, da sie die Eingabe durch Transformation in einen kleineren Wertebereich *quetschen*. Sie erlauben es, in Kombination mit der hierarchischen Verschachtelung durch zusätzliche neuronale Schichten beliebig komplizierte Funktionen zu erhalten [2]. Die Funktionsweise eines künstlichen Neurons ist in **Abb. 2** dargestellt. Die eingehenden Verbindungen werden gewichtet aufaddiert und durch eine nichtlineare Aktivierungsfunktion transformiert. Diese Transformation führt in der Regel zu einer Abbildung in den Wertebereich [0, 1] bzw. [−1,1]. Das Ergebnis kann dann als Eingabe für ein nachgeschaltetes Neuron dienen, wo es wieder gewichtet in die Summation einfließt, usw. Sieht man sich die Formeln an, so erinnern auch diese an die oben erwähnte Geradengleichung (**Abb. 2**).

Das Training der neuronalen Netze basiert ebenfalls auf Mechanismen, die man möglicherweise aus dem Schulunterricht kennt, z. B. Ableitungen mit der Kettenregel. Initial werden den Gewichten des Netzwerks Zufallswerte zugewiesen. Propagiert man eine Eingabe durch das Netzwerk, resultiert dies in einer Reihe von Transformationen, die zunächst zu einer zufälligen Ausgabe führen. Da die richtige Ausgabe bekannt ist, kann die Abweichung zu dieser, – also der Fehler, – berechnet werden. Wenn die Fehlerfunktion so gewählt wird, dass sie differenzierbar ist, kann deren Ableitung berechnet und über die Kettenregel der Anteil, den die einzelnen Neuronen am Gesamtfehler haben, durch das Netzwerk zurück propagiert („backpropagation") und die Gewichte entsprechend angepasst werden. Dies macht man mit tausenden von Trainingsbeispielen, bis die Gewichte des Netzwerks so eingestellt sind, dass der Fehler minimal ist. Man spricht dann auch davon, dass „das Netzwerk konvergiert hat". Dieses Optimierungsverfahren wird als Gradientenabstieg bezeichnet, da die Richtung des Gradienten, bestimmt durch die Ableitungen, verwendet wird, um die Gewichte proportional anzupassen.

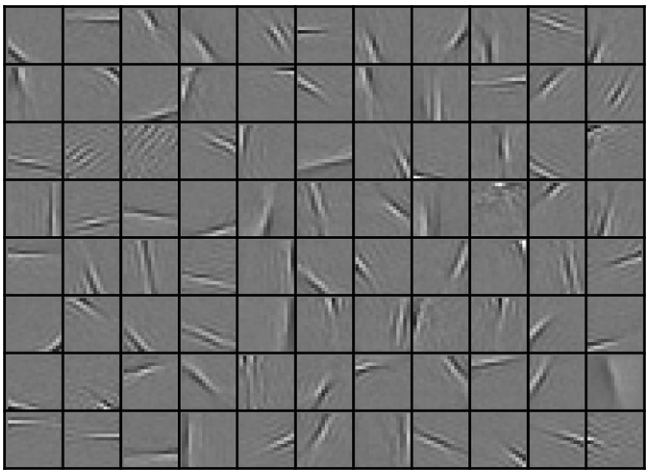

**Abb. 5** ◀ Durch ein künstliches neuronales Netz gelernte Filter (*Ausschnitt*). Diese dienen u. a. der Kantendetektion und ähneln rezeptiven Feldern des visuellen Kortex. Wie in ◘ **Abb. 4** dargestellt, könnte eines dieser Quadrate einem Filter für diagonale Kanten entsprechen

Backpropagation ist das zentrale Prinzip bei den meisten Deep-Learning-Algorithmen und wird auch für das Training von Faltungsnetzwerken (CNN) angewendet, die sich insbesondere im Bereich der Bildverarbeitung als De-facto-Standard etabliert haben – beispielsweise für die Segmentierung von Hirntumoren in MRT-Aufnahmen [7]. Um zu verstehen, warum dies der Fall ist, muss man sich zunächst vergegenwärtigen, was ein Computer *sieht*. Die einzelnen Bildpunkte eines Bildes entsprechen Grauwerten, die man in einer Matrix darstellen kann (◘ **Abb. 4**). Bei der Faltungsoperation handelt es sich nun ebenfalls um kleine Matrizen, bezeichnet als Filter, die über das Bild geschoben werden. Dabei werden die Werte multipliziert und aufaddiert (◘ **Abb. 4**). Für jede dieser Operationen entsteht ein Zahlenwert. Zusammengenommen entsprechen dies Zahlenwerte neuen *Bilder* (eins für jeden Filter), die als Aktivierungskarten bezeichnet werden. Sie stellen die Antworten eines Filters an jeder räumlichen Position dar. Wie auch zuvor dienen diese Aktivierungskarten, die Ausgaben einer Schicht, als Eingabe für die nächste Schicht. Während des Lernens bildet das Netzwerk immer bessere Filter aus. Die Zahlenwerte der Filter entsprechen also den Gewichten $w$, die während des Trainings mit einer Fehlerfunktion optimiert werden. Man spricht davon, dass Filter aktiviert werden, wenn sie ein visuelles Merkmal erkennen, z. B. eine Kante einer bestimmten Ausrichtung (◘ **Abb. 4** links) oder eine bestimmte Farbe. Erstaunlicherweise ähneln die durch den Computer gelernten Filter Mustern, die auch im visuellen Cortex V1 gefunden werden (◘ **Abb. 5**). Im biologischen Analogon spricht man von rezeptiven Feldern von Neuronen. Ebenso wie im biologischen Gegenstück finden sich in den unteren Schichten der künstlichen Netzwerke Low-level-Merkmale, beispielsweise die Kantendetektion, die zu höheren Konzepten in den oberen Schichten kombiniert werden [9].

## Komplexität

Das gezeigte Beispiel der linearen Separierbarkeit ist zweidimensional. In der Regel handelt es sich jedoch um viele Merkmale, die einen hochdimensionalen Raum aufspannen. In diesem Raum wird nach der optimalen Trennung der Datenpunkte gesucht. Wenn beispielsweise 200 Merkmale aus den Daten extrahiert werden, dann ist jeder Datenpunkt durch einen 200-dimensionalen Vektor charakterisiert. Diese oder eine noch höhere Dimensionalität findet man nicht selten in Radiomics-Studien. (Für Details zu der dort häufig vorzufindenden vordefinierten Merkmalsextraktion siehe den Beitrag von *Murray et al.* [11].)

KNN haben Millionen von Parametern (Gewichten), die während des Trainings optimiert werden. Für den Menschen ist es dann in der Regel nicht mehr nachvollziehbar, wie Merkmale kombiniert werden oder wie Entscheidungen zustande kommen. Auch bedürfen das Training der Netzwerke sowie anderer Methoden einer gewissen Erfahrung des Programmierers. So ist es nicht immer offensichtlich, an welchen Stellschrauben gedreht werden muss, wenn das Netzwerk einmal nicht konvergiert. Neben den Parametern, die durch das Training optimiert werden, gibt es noch eine Reihe von sog. Hyperparametern, die durch den Benutzer festgelegt werden. Hierzu gehören beispielsweise die Anzahl der Neurone pro Schicht, die Anzahl der Schichten des Netzwerks oder die Filtergröße. Diese Parameter werden nicht durch das Training verändert.

## Limitationen und Fallstricke

Viele der häufig verwendeten ML-Verfahren sind sehr mächtig, d. h. sie können eine beliebig komplexe Funktion approximieren. Dies bringt die Gefahr mit sich, dass sie sich nahezu perfekt an die Trainingsdaten anpassen. In dem Fall spricht man von „overfitting" (◘ **Abb. 3c**). Auf den ersten Blick mag es dem entsprechen, was man erreichen möchte – dies ist aber keineswegs der Fall! Ziel muss es sein, ein Modell zu erhalten, das möglichst gut generalisiert, also auch mit unbekannten Daten sinnvolle Resultate ergibt. Um die Güte des trainierten Modells abschätzen zu können, ist es daher eine etablierte Vorgehensweise, die Daten in 3 Gruppen aufzuteilen: Trainings-, Validierungs- und Testdaten. Die Trainingsdaten verwendet man für das Training des Modells, die Validierungsdaten werden dazu verwendet, die Hyperparameter des Models anzupassen, und die Testdaten werden ausschließlich für die Evaluation herangezogen. Beim Lesen von Publikationen sollte man daher darauf achten, ob auch wirklich an einem unabhängigen Testdatensatz evaluiert wurde. Idealerweise sollte dieser aus einer anderen Datenquelle stammen, beispielsweise an einem MRT-Gerät eines anderen Herstellers akquiriert worden sein. Im medizinischen Bereich ist dies allerdings oftmals nicht der Fall, da im Vergleich zu anderen Domänen die Datenmenge den limitierenden Faktor darstellt.

Insgesamt ist die Datenqualität, -quantität und -vorverarbeitung ausschlaggebend für das erfolgreiche Training des ML-Verfahrens. So haben beispielsweise fehlende Werte und die Normalisierung einen großen Einfluss auf das Ergebnis. Aber auch die Entfernung

vermeintlicher Ausreißer, die in Wahrheit zu einer unterrepräsentierten Klasse gehören, kann dazu führen, dass der Algorithmus bei der Klassifizierung unbekannter Daten dieser Klasse eine falsche Zuordnung vornimmt. Allgemein kann die Vorselektion bzw. die Zusammensetzung der Grundgesamtheit, aus der die Daten stammen, zu einem sog. „sampling bias" führen, der das Lernen maßgeblich beeinflusst. Werden beispielsweise Daten von einer anderen ethnischen Population oder nur eines Geschlechts zum Training verwendet, können möglicherweise (verborgene) Eigenheiten nicht aus den Daten gelernt werden, und es kommt zu Fehlern bei der Anwendung auf nicht berücksichtigte Personengruppen. Mit der Erkennung von diesen, nicht zum Algorithmus passenden Daten („out of distribution data"), beschäftigt sich ein ganzes Forschungsfeld.

Im Gegensatz zu KNN, die lernen, auf welche Merkmale zu achten ist, bergen Verfahren, bei denen der Mensch die Merkmale vorgibt, das Risiko, dass nicht die richtigen Merkmale ausgewählt werden. Dies ist ein wesentlicher Grund, warum Deep-Learning-Verfahren im Vergleich zu den klassischen Ansätzen letztendlich häufig bessere Ergebnisse erzielen.

## Ausblick

Die Verwendung von neuen Technologien verläuft häufig nicht linear. Hat sich erst einmal deren Nutzen gezeigt, kommt es dann zu einem rapiden flächendeckenden Einsatz. Für die Anwendung von ML-Werkzeugen zur Unterstützung des Arztes in der Radiologie und der Medizin im Allgemeinen steht uns dieser Wandel noch bevor. Es ist aber klar, dass er kommen wird. Um darauf vorbereitet zu sein, sollte man sich mit den Grundprinzipien vertraut machen – die nächste Generation von Ärzten muss bereits im Studium darauf vorbereitet werden. Es kann aber nicht erwartet werden, dass jeder Arzt auch zum Informatiker, Mathematiker oder KI-Forscher wird. ML muss sich zu einem interdisziplinären Miteinander entwickeln, welches die Konvergenz von Mensch und Maschine begünstigt [15]. Im Fall der Radiologie verspricht uns der *Bionic Radiologist* eine bessere Patientenversorgung bei niedrigeren Kosten [4]. Eine große Hoffnung liegt hierbei auch auf der Erkennung von neuen Zusammenhängen, die aufgrund ihrer Komplexität uns Menschen sonst nicht zugänglich wären [10].

Um Vertrauen in KI-Ergebnisses zu entwickeln, wird häufig eine Erklärung der Gründe („explainable AI") für eine Entscheidung gefordert. Dass Vertrauen entsteht, ist wichtig, braucht aber Zeit und positive Erfahrungen beim Einsatz der Methoden. Dass es prinzipiell möglich ist, Vertrauen zu gewinnen, ohne alle Abläufe im Detail zu verstehen, kennen wir aus anderen medizinischen Bereichen: Wir verschreiben Medikamente, deren Wirkmechanismus wir nicht kennen [14] oder verlassen uns auf die automatische Detektion von Extrasystolen im EKG, ohne genaue Kenntnisse der technischen Umsetzung zu haben. Neben der Entwicklung von Algorithmen, die uns Aufschlüsse über die internen Mechanismen geben, brauchen wir daher zwingend prospektive Studien, die die *Wirksamkeit* der KI-Methoden zeigen. Sicherlich wird es schon bald erste prospektiv-randomisierte Studien geben, die die Überlegenheit von Diagnosestellung und Therapiewahl mit Unterstützung durch Algorithmen demonstrieren werden.

## Fazit für die Praxis

- **Aktuell ist der Einsatz von Methoden der künstlichen Intelligenz noch nicht im medizinischen Alltag angekommen.**
- **Dass KI-Methoden zur Unterstützung in Diagnose und Therapie Einzug finden werden, ist sicher. Es könnte schneller gehen, als man denkt.**
- **Die Grundprinzipien des maschinellen Lernens müssen daher zukünftigen Generationen von Ärzten bekannt sein, insbesondere, um Chancen und Limitationen bewerten zu können.**
- **Um Vertrauen in die neuen Technologien aufzubauen, werden nachvollziehbare Algorithmen und prospektive *Wirksamkeitsstudien* benötigt.**


## Korrespondenzadresse

**Dr. Dr. Jens Kleesiek**
AG Computational Radiology, Abteilung Radiologie, Deutsches Krebsforschungszentrum (DKFZ)
Im Neuenheimer Feld 280, 69120 Heidelberg, Deutschland
j.kleesiek@dkfz-heidelberg.de


**Danksagung.** Wir danken Frau Dr. Bettina Beuthien-Baumann für wertvolle Kommentare zum Manuskript.

## Einhaltung ethischer Richtlinien

**Interessenkonflikt.** J. Kleesiek, J.M. Murray, C. Strack, G. Kaissis und R. Braren geben an, dass kein Interessenkonflikt besteht.

Für diesen Beitrag wurden von den Autoren keine Studien an Menschen oder Tieren durchgeführt. Für die aufgeführten Studien gelten die jeweils dort angegebenen ethischen Richtlinien.

## Literatur


1. Bryan RN (2016) Machine learning in radiology. RSNA News 26:4–6
2. Cybenko G (1989) Approximation by superpositions of a sigmoidal function. Math Control Signals Syst 2:303–314. https://doi.org/10.1007/BF02551274
3. Dechter R (1986) Learning while searching in constraint-satisfaction-problems. In: Proc. Fifth AAAI Natl. Conf. Artif. Intell. AAAI Press, S 178–183, https://dl.acm.org/citation.cfm?id=2887799, Philadelphia, Pennsylvania, 11–15 August 1986
4. Dewey M, Wilkens U (2019) The bionic radiologist: avoiding blurry pictures and providing greater insights. Npj Digit Med 2:65. https://doi.org/10.1038/s41746-019-0142-9
5. FeatureCloud https://featurecloud.eu/. Zugegriffen: 17. Sept. 2019
6. Grace K, Salvatier J, Dafoe A et al (2018) Viewpoint: when will ai exceed human performance? Evidence from ai experts. J Artif Intell Res 62:729–754. https://doi.org/10.1613/jair.1.11222
7. Isensee F, Petersen J, Kohl SAA et al (2019) nnU-Net: Breaking the Spell on Successful Medical Image Segmentation
8. Joint Imaging Platform https://jip.dktk.dkfz.de/jiphomepage/. Zugegriffen: 3. Sept. 2019
9. Jones N (2014) Computer science: the learning machines. Nature 505:146. https://doi.org/10.1038/505146a
10. Kleesiek J, Murray JM, Kaissis GA, Braren R (2020) Künstliche Intelligenz und Maschinelles Lernen in der onkologischen Bildgebung. Onkologe. https://doi.org/10.1007/s00761-019-00679-4
11. Murray JM, Kaissis G, Braren R, Kleesiek J (2019) Wie funktioniert Radiomics? Radiologe. https://doi.org/10.1007/s00117-019-00617-w
12. Rich E (1983) Artificial intelligence. McGraw-Hill Inc, US. ISBN 10: 0-070-52261-8

13. van Soest J, Sun C, Mussmann O et al (2018) Using the personal health train for automated and privacy-preserving analytics on vertically partitioned data. Stud Health Technol Inform 247:581–585
14. Topol EJ (2019) Deep medicine: how artificial intelligence can make healthcare human again, 1. Aufl. Basic Books, Hachette Book Group, New York
15. Topol EJ (2019) High-performance medicine: the convergence of human and artificial intelligence. Nat Med 25:44–56. https://doi.org/10.1038/s41591-018-0300-7
16. What is NFDI4Life Umbrella? https://www.nfdi4life.de/. Zugegriffen: 17. September 2019
17. Who Invented Backpropagation? http://people.idsia.ch/~juergen/who-invented-backpropagation.html. Zugegriffen: 18. September 2019
18. Xia Q, Sifah EB, Asamoah KO et al (2017) MeDshare: trust-less medical data sharing among cloud service providers via Blockchain. IEEE Access 5:14757–14767. https://doi.org/10.1109/ACCESS.2017.2730843