Hybrides Maschinelles Lernen im Kontext der Produktion

Isabelle Gauger[1(✉)], Tobias Nagel[2], und Marco Huber[1,2]

[1] Institut für Industrielle Fertigung und Fabrikbetrieb IFF, Universität Stuttgart, Stuttgart, Deutschland
{isabelle.gauger,marco.huber}@iff.uni-stuttgart.de
[2] Zentrum für Cyber Cognitive Intelligence (CCI), Fraunhofer IPA, Stuttgart, Deutschland
{tobias.nagel,marco.huber}@ipa.fraunhofer.de

Zusammenfassung. Der Einsatz von Maschinellem Lernen (ML) zur Gewinnung von Erkenntnissen aus großen Datenmengen und der Lösung anspruchsvoller Aufgaben bietet ein enormes Potenzial für die industrielle Produktion. Allerdings ist der Einsatz von ML-Modellen auch mit Herausforderungen verbunden. Dazu gehören die meist großen benötigten Datenmengen sowie die mangelnde Transparenz und Erklärbarkeit dieser Modelle, was ihren Einsatz in der Praxis erschwert. Ein Ansatz, um diesen Herausforderungen zu begegnen, ist, in der Industrie vorhandenes Fachwissen zu nutzen und mit ML-Modellen zu verbinden. Dieser Ansatz wird als *hybrides maschinelles Lernen* bezeichnet. Im Folgenden werden die Vor- und Nachteile von klassischen wissens- und datenbasierten Modellen diskutiert sowie Ansätze vorgestellt, um beide Welten miteinander zu verbinden. Anhand von Beispielen aus der Regelungstechnik und Industrierobotik wird erläutert, wie sich diese anschließend im Produktionskontext umsetzen lassen. Im Fokus stehen dabei besonders Aspekte der Zuverlässigkeit, wie die Erklärbarkeit und Robustheit solcher Systeme, die wichtige Elemente der digitalen Souveränität des Anwenders sind.

Schlüsselwörter: Maschinelles Lernen · Hybride Modelle · Informed Machine Learning

1 Motivation

Der Einsatz von Künstlicher Intelligenz (KI) und ihrem Teilgebiet des Maschinellen Lernens (ML) bietet ein enormes Potenzial für die industrielle Produktion. Angefangen bei der Erhöhung des Automatisierungsgrades sowie der Kapazität und Flexibilität einer Produktion über die Steigerung der Produktqualität bis zur Senkung der Kosten und der Belastung für Mensch und Umwelt.

© Der/die Autor(en) 2022
E. A. Hartmann (Hrsg.): *Digitalisierung souverän gestalten II*, S. 64–79, 2022.
https://doi.org/10.1007/978-3-662-64408-9_6

Die große Stärke von ML-Algorithmen liegt darin, dass sie dazu in der Lage sind, anhand von Daten automatisch zu lernen, eine bestimmte Aufgabe zu lösen. Voraussetzung ist dabei zum einen das Vorhandensein der benötigten Daten und zum anderen die Verfügbarkeit von hohen Rechenkapazitäten. Durch die zunehmende Digitalisierung und Vernetzung im Zusammenhang mit Industrie 4.0 und stark gestiegener Rechenleistung haben sich diese Voraussetzungen in den letzten Jahren zunehmend erfüllt (Bauernhansl 2014; Shang und You 2019; Wuest u. a. 2016).

Die Möglichkeiten, ML in der Produktion einzusetzen, sind vielfältig und umfassen alle Phasen der Produktion. In der Produktentwicklung können ML-Methoden dafür eingesetzt werden, das Design von Produktvarianten zu automatisieren oder Produkteigenschaften vorherzusagen. In der Produktionsplanung lassen sich mithilfe von ML-Methoden unter anderem zukünftige Kundenbedarfe vorhersagen sowie die Arbeitsplanung optimieren und flexibilisieren. In der eigentlichen Produktion können mittels ML der Zustand von Produktionsmaschinen überwacht und eine vorausschauende Instandhaltung umgesetzt sowie Prozessparameter optimiert oder die Produktqualität überwacht werden. Darüber hinaus lassen sich mit ML die Fähigkeiten von Industrierobotern verbessern und deren Programmierung vereinfachen, hin zu intelligenten, hochautomatisierten Systemen (Krauß u. a. 2019; Wagner u. a. 2021; J. Wang u. a. 2018; Wuest u. a. 2016).

Neben dem enormen Potenzial von ML beim Einsatz in der Produktion gibt es auch einige Herausforderungen, die ML mit sich bringt. Wie bereits genannt, muss bei ML-Methoden generell eine große Menge an geeigneten Daten für das Training der Modelle vorhanden sein. Eine weitere Herausforderung ist, dass es sich bei modernen ML-Verfahren in der Regel um Modelle handelt, die der Mensch nicht nachvollziehen kann.

Ein Weg, diesen Herausforderungen zu begegnen, an dem aktiv geforscht wird, ist es, ML-Methoden mit vorhandenem Wissen zu verbinden. In der Industrie existiert eine große Menge an Domänenwissen und wissensbasierte Systeme wie etwa Simulationen oder mathematische Modelle kommen an verschiedensten Stellen zum Einsatz. Dieses Wissen zu nutzen und mit ML-Modellen zu verbinden, ist ein vielversprechender Weg, um die genannten Hürden beim Einsatz moderner ML-Methoden zu überwinden. Man bezeichnet diesen Ansatz als *hybrides maschinelles Lernen* oder auch Informed Machine Learning. Durch das Einbringen von Domänenwissen lässt sich unter anderem die Menge der benötigten Daten für das Training der ML-Modelle reduzieren sowie deren Robustheit und Erklärbarkeit erhöhen (von Rueden, Katharina Beckh u. a. 2021; Karpatne u. a. 2017).

2 Wissensbasiert vs. datenbasiert

Im Produktionskontext werden wissensbasierte Modelle in vielfältiger Weise eingesetzt. Angefangen bei analytischen Modellen über regelbasierte Systeme bis zu aufwendigen multiphysikalischen Simulationen. Die Basis solcher Modelle

bilden in der Regel eine Reihe von Gleichungen oder logischen Regeln, die das System beschreiben. Die Tatsache, dass die Funktionsweise von wissensbasierten Modellen für den Menschen somit nachvollziehbar und die Modelle erklärbar sind, ist eine große Stärke dieser Ansätze. Dies ermöglicht unter anderem, bei sicherheitskritischen Anwendungen zu kontrollieren und sicherzustellen, dass geltende Sicherheitsstandards eingehalten werden. Gleiches gilt für die Einhaltung gesetzlicher Vorgaben, z. B. zum Datenschutz oder zur Gleichstellung. Die Beschreibung des Systems durch eine überschaubare Zahl von Gleichungen oder Regeln bedeutet allerdings auch, dass diese Modelle gewissen Limitationen bezüglich ihrer Genauigkeit unterliegen. In der Realität beeinflussen in der Regel eine Vielzahl bekannter aber auch teilweise unbekannter Faktoren das Verhalten eines Systems. Je höher die Genauigkeitsanforderungen an das Modell sind, desto mehr dieser Zusammenhänge müssen identifiziert und im Modell korrekt berücksichtigt werden. Dies kann schnell dazu führen, dass die Erstellung des Modells sehr aufwendig und das Modell zunehmend kompliziert wird. Alle Zusammenhänge explizit zu modellieren ist im Allgemeinen schlicht nicht möglich.

Bei datenbasierten Modellen, wie künstlichen neuronalen Netzen, werden die Zusammenhänge hingegen rein aus vorhandenen Daten gelernt und müssen nicht explizit modelliert werden. Mit solchen Modellen konnten unter anderem in der Objekterkennung anhand von Bilddaten große Fortschritte erzielt werden (Howard u. a. 2017; Redmon u. a. 2016). Anstatt dass der Mensch explizit definieren muss, anhand welcher Merkmale bestimmte Objekte identifiziert werden, lernt das Modell dies implizit anhand von Beispieldaten. Dabei extrahiert das Modell auch Zusammenhänge aus den Daten, die dem Menschen verborgen bleiben, und erreicht Genauigkeiten, die wissensbasierte Verfahren teils deutlich übertreffen. Für die derzeit sehr häufig verwendeten künstlichen neuronale Netze gilt, dass sie dazu in der Lage sind, jede stetige Funktion beliebig genau zu approximieren, d. h., sie können jeden beliebigen Zusammenhang lernen (Hornik, Stinchcombe und White 1989). Dadurch können mit diesen Verfahren sehr hohe Genauigkeiten erreicht werden.

Allerdings setzt der Einsatz von ML-Modellen voraus, dass Daten für das Trainieren der Modelle in ausreichender Menge und Qualität vorhanden sind. Auch wenn die bei Unternehmen vorhandene Datenmenge im Allgemeinen immer weiter zunimmt, liegen für den konkreten Anwendungsfall häufig keine Daten in geeigneter Form und ausreichender Menge vor, sodass deren Beschaffung mitunter sehr aufwendig sein kann (Krauß u. a. 2019). Viele ML-Modelle extrapolieren auch sehr schlecht, d. h., im Bereich der Daten liefern sie gute Ergebnisse, aber außerhalb dieses Bereichs können die Vorhersagen des Modells beliebig falsch sein. Bei wissensbasierten Ansätzen hat man dieses Problem meist nicht.

Darüber hinaus sind die mitunter leistungsfähigsten ML-Modelle zugleich auch am wenigsten nachvollziehbar. Bei diesen Modellen ist es sogar für Experten nicht möglich, die innere Logik des Modells zu verstehen und nachzuvollziehen, warum das Modell eine bestimmte Entscheidung getroffen hat. Die mangel-

nde Transparenz und Erklärbarkeit dieser Modelle ist insbesondere dann kritisch, wenn Menschen zu Schaden kommen könnten, wie z. B. bei medizinischen Anwendungen oder beim autonomen Fahren.

In der Vergangenheit hat sich gezeigt, dass ML-Modelle nicht unbedingt mit bekannten naturwissenschaftlichen Gesetzen konsistent sind (Karpatne u. a. 2017). Darüber hinaus wurde festgestellt, dass ML-Modelle mitunter auch ein unerwünschtes Verhalten lernen, was zu Fehldiagnosen in der Medizin oder zu Diskriminierung führen kann (Angwin u. a. 2016; Zech 2018). Außerdem ist bekannt, dass ML-Modelle häufig nicht besonders robust gegen gezielte Versuche sind, sie in die Irre zu führen. So kann eine kleine mutwillige Veränderung der Eingaben, z. B. die Veränderung von Pixeln in einem Bild, welche dem Menschen gar nicht auffällt, dazu führen, dass das ML-Modell eine falsche Entscheidung trifft. Man spricht in diesem Fall von sogenannten Adversarial Attacks (Goodfellow u. a. 2017). Vor diesem Hintergrund stellt sich die Herausforderung, wie solche Modelle verifiziert werden können (Huang u. a. 2017). Dazu gehört auch die Tatsache, dass ML-Modelle Einschränkungen durch naturwissenschaftliche Gesetze, Sicherheitsrichtlinien oder gesetzliche Vorgaben nicht automatisch einhalten.

Zusammenfassend lässt sich sagen, dass der große Vorteil wissensbasierter Modelle darin liegt, dass die Modelle erklärbar, verifizierbar und kontrollierbar sind und auf etabliertem Fachwissen fußen. Darüber hinaus müssen keine großen Trainingsdatensätze gesammelt werden. Können die Zusammenhänge allerdings nur mit sehr viel Aufwand oder gar nicht explizit modelliert werden oder sind sie nicht bekannt, kommt man mit wissensbasierten Modellen unter Umständen nicht mehr weiter. Moderne ML-Modelle, die die Zusammenhänge anhand von Daten lernen, haben in solchen Fällen großes Potenzial demonstriert und zu Durchbrüchen bei schwierigen Aufgaben wie der Objekterkennung oder Sprachverarbeitung geführt. Aufgrund der mangelnden Erklärbarkeit dieser Modelle ergeben sich jedoch Herausforderungen in Bezug auf ihre Verifizierbarkeit und Kontrollierbarkeit.

3 Das Beste aus beiden Welten – hybrides maschinelles Lernen

Im vorherigen Abschnitt wurden die Vor- und Nachteile von wissens- und datenbasierten Modellen diskutiert. Die Stärken beider Ansätze zu verbinden erscheint dabei sehr attraktiv. Die Idee ist, auf der einen Seite unbekannte Zusammenhänge anhand von Daten zu lernen und auf der anderen Seite bestehendes Vorwissen über bekannte Zusammenhänge zu nutzen. Die dabei verfolgten und unter dem Begriff hybrides maschinelles Lernen zusammengefassten Ansätze sind sehr vielfältig und hängen von der konkreten Anwendung und der Form des wissensbasierten Modells ab. In Abb. 1 wird hybrides maschinelles Lernen schematisch in den Kontext von wissens- und datenbasierten Modellen eingeordnet. Die Ansätze variieren hinsichtlich der Verwendung von Fachwissen und Daten, zeichnen sich aber dadurch aus, dass immer beide Komponenten einfließen.

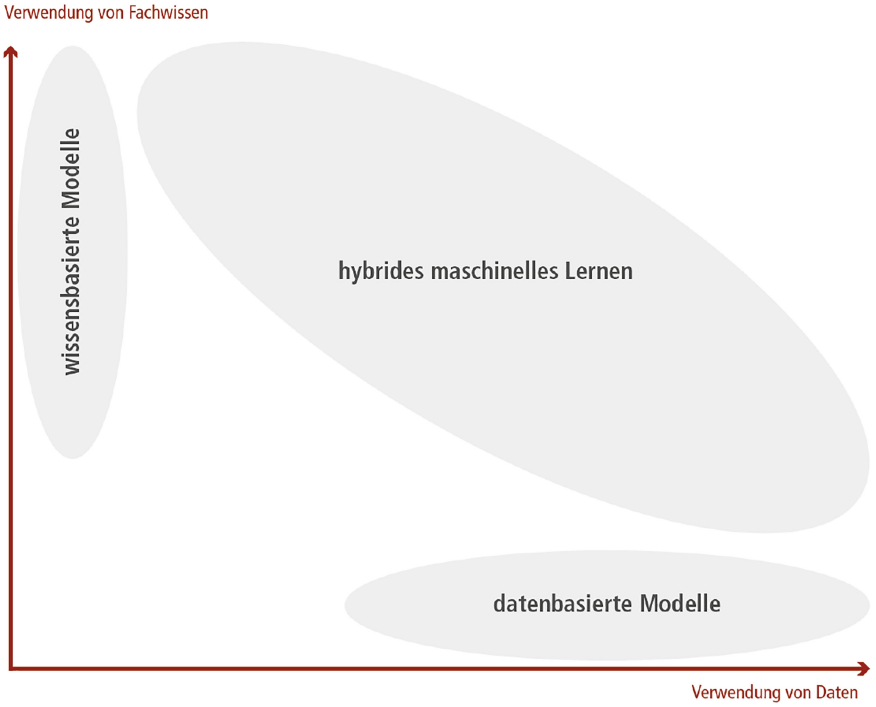

Abb. 1. Einordnung von hybridem maschinellem Lernen in den Kontext von wissens- und datenbasierten Modellen. In Anlehnung an Karpatne u. a. 2017

Residuen-Ansatz: Ein allgemeiner Ansatz sind hybride Modelle, die ein wissensbasiertes und ein datenbasiertes Modell additiv kombinieren (Karpatne u. a. 2017). In vielen Fällen sind wissensbasierte Modelle vorhanden, die das reale System näherungsweise beschreiben. Reicht die Genauigkeit dieser Modelle für die geplante Anwendung nicht aus, können sie durch datenbasierte Modelle ergänzt werden. Die Idee ist hier, dass ein wissensbasiertes Modell einen Teil der Zusammenhänge beschreibt und der Rest von einem datenbasierten Modell gelernt wird. Dies ist attraktiv, weil mit diesem hybriden Modell eine höhere Genauigkeit erreicht werden kann, während das Modell zum Teil erklärbar bleibt. Darüber hinaus weist ein derartiges Modell eine gewisse Robustheit gegenüber Störungen auf. Eine falsche Vorhersage des datenbasierten Modells führt lediglich zu einer weniger genauen, aber nicht zu einer völlig falschen Vorhersage des hybriden Modells und es kann besser extrapolieren. In Abschn. 4.1 wird diese Variante anhand des Anwendungsbeispiels Roboterkalibrierung ausführlich erläutert.

Stellvertreteransatz: Im Gegensatz dazu gibt es auch Fälle, in denen die vorhandenen wissensbasierten Modelle durchaus sehr genau sind. Es handelt sich dabei dann in der Regel um aufwendige physikalische Simulationen. Hier liegt das Problem häufig darin, dass diese Simulationen mit einem sehr großen Rechenaufwand verbunden sind. In diesem Fall können datenbasierte Modelle

genutzt werden, um anhand vorhandener Simulationsergebnisse näherungsweise rechenarme Stellvertretermodelle der ursprünglichen Modelle zu lernen (von Rueden, Mayer, Sifa u. a. 2020). Bei der Optimierung der Prozessparameter von Produktionsprozessen können solche datenbasierten Stellvertretermodelle beispielsweise eingesetzt werden, um in einem großen Suchraum vielversprechende Kandidaten für die Prozessparameter zu finden. Für diese Kandidaten werden die Ergebnisse des datenbasierten Modells dann durch eine genaue, aber rechenintensive Simulation validiert, bevor Testexperimente auf dem realen System durchgeführt werden (Pfrommer u. a. 2018).

Reihenschaltung: Eine weitere Variante, wissensbasierte und datenbasierte Modelle zu kombinieren, besteht darin, die Ausgabe eines datenbasierten Modells in ein wissensbasiertes Modell einfließen zu lassen. Hier kann das datenbasierte Modell beispielsweise Größen vorhersagen, die bisher nur ungenau geschätzt werden und so die Vorhersage des wissensbasierten Modells verbessern (Karpatne u. a. 2017). Die umgekehrte Variante, dass die Ausgabe eines wissensbasierten Modells als Eingabe für ein datenbasiertes Modell genutzt wird, ist ebenfalls verbreitet.

Synthetische Trainingsdaten: Ein anderer allgemeiner Ansatz besteht darin, wissensbasierte Modelle oder ihre Ergebnisse direkt in die Pipeline des maschinellen Lernens zu integrieren (von Rueden, Mayer, Sifa u. a. 2020). Eine häufig gewählte Variante ist dabei, die Ergebnisse wissensbasierter Modelle als zusätzliche Trainingsdaten für datenbasierte Modelle zu verwenden (von Rueden, Katharina Beckh u. a. 2021). Das wissensbasierte Modell, hier in der Regel eine Simulation, liefert bei dieser Variante lediglich zusätzliche synthetische Daten für das Training des ML-Modells. In der Anwendung wird dann ausschließlich das ML-Modell verwendet. Dies bietet sich insbesondere an, wenn wenig reale Trainingsdaten zur Verfügung stehen, weil das Sammeln realer Daten zeitintensiv und teuer ist. Eine anspruchsvolle Robotikanwendung ist der Griff in die Kiste (engl. Bin Picking), bei der es darum geht, ungeordnete Objekte aus einem Behälter zu vereinzeln. Für die Erkennung der Objekte und die Bestimmung ihrer 6D-Pose, d. h. ihrer Position und Orientierung im Raum, eignet sich der Einsatz datenbasierter ML-Methoden. Allerdings stellt die Menge der benötigten Trainingsdaten eine Hürde dar, da das Sammeln und Annotieren dieser Daten mit großem Zeitaufwand verbunden ist. Durch die Kombination von realen Daten mit synthetischen Daten, die mithilfe einer Simulation erzeugt wurden, lassen sich hier hinreichend große Datensätze für den Einsatz von ML-Methoden erzeugen (Kleeberger, Landgraf und Huber 2019).

Ergebnisvalidierung: Eine andere, bereits genannte Variante, die am Ende der ML-Pipeline ansetzt, ist, wissensbasierte Modelle dafür zu nutzen, die Vorhersagen rein datenbasierter Modelle zu validieren (von Rueden, Mayer, Sifa u. a. 2020). Beim Stellvertreteransatz wird das ML-Modell mit synthetischen Daten von einer Simulation darauf trainiert, diese möglichst gut zu approximieren. Anschließend werden in der Anwendung Vorhersagen des ML-Modells mit der aufwendigeren Simulation validiert. Darüber hinaus kann ein wissensbasiertes Modell aber auch dazu genutzt werden, ein ML-Modell zu validieren,

das ausschließlich mit realen Trainingsdaten trainiert wurde. Bei dieser Variante kommt das wissensbasierte Modell nur in der Anwendung zum Einsatz, um Vorhersagen des ML-Modells auf ihre Konsistenz mit bekanntem Wissen zu prüfen.

Integrierte Lernalgorithmen und Modellarchitekturen: Darüber hinaus wird auch an Möglichkeiten geforscht, bekannte Gesetzmäßigkeiten aus wissensbasierten Modellen direkt in den Lernalgorithmus oder die Modellarchitektur von datenbasierten Modellen zu integrieren (von Rueden, Katharina Beckh u. a. 2021). Durch das Einbringen bekannter Gesetzmäßigkeiten wird der Suchraum bei der Optimierung des datenbasierten Modells auf Lösungen eingeschränkt, die diesen Gesetzmäßigkeiten gehorchen. Ein Beispiel aus der Additiven Fertigung ist die Vorhersage der Bondqualität und Porosität von gedruckten Teilen. Hier wurden bekannte physikalische Gesetze in den Lernalgorithmus integriert. Dadurch konnte erreicht werden, dass auch bei geringer Datenmenge ein Modell gelernt wird, das gut generalisiert und genaue, physikalisch konsistente Vorhersagen liefert (Kapusuzoglu und Mahadevan 2020).

Im Allgemeinen können nicht nur wissensbasierte Modelle wie Simulationen oder analytische Modelle in datenbasierte Modelle integriert werden, sondern auch andere Formen von Domänenwissen. Die Integration von Domänenwissen aller Art in ML-Modelle wird auch als Informed Machine Learning oder Theory-guided Data Science bezeichnet (Karpatne u. a. 2017; von Rueden, Katharina Beckh u. a. 2021). Domänenwissen kann hier auch zum Beispiel durch logische Regeln, Wissensgraphen oder menschliches Feedback eingebracht werden. Dies zu betrachten würde hier allerdings zu weit führen.

4 Hybrides maschinelles Lernen für die Produktion

Die bereits aufgeführten Nachteile rein datengetriebener Modelle stellen in der Produktionstechnik ein besonderes Hindernis dar. Eine Robotiksteuerung könnte theoretisch hervorragend über aktuelle Reinforcement-Learning-Ansätze trainiert oder eine Produktionsplanung über KI-Methoden optimiert werden. Es ist aus finanziellen Gründen jedoch nicht möglich, mit einem zeitintensiven Training Produktionsressourcen zu belegen. Daher bleiben nur zwei Möglichkeiten: Den Trainingsprozess beschleunigen oder ein Lernen in der Simulation. Da Letzteres häufig nicht vorhanden ist bzw. nur mit großem Aufwand erzeugt werden kann, stellt das Beschleunigen des Trainingsprozesses mit hybriden KI-Ansätzen einen validen Ansatz dar. Vorwissen von Produktionsplanern kann den Suchraum der optimalen Lösung z. B. stark eingrenzen oder bereits existierende Dynamikmodelle von diversen Robotern können mit KI-Methoden befähigt werden, weitere, komplexere Aufgaben durchzuführen. Auf rein datengetriebene Ansätze zu verzichten macht das gefundene Modell besser interpretierbar, sodass dem Algorithmus ein größeres Vertrauen geschenkt werden kann und Informationen über Robustheit direkt abgeleitet werden können. Die folgenden Abschnitte beleuchten die zuletzt genannten Aspekte genauer.

4.1 Hybrides maschinelles Lernen für die Industrierobotik

In der Produktion kommen Industrieroboter für Aufgaben wie Schweißen, Lackieren oder Montieren zum Einsatz. Für die zuverlässige und genaue Ausführung dieser Aufgaben müssen Industrieroboter hohe Anforderungen an die Genauigkeit erfüllen. Allerdings können verschiedene Faktoren wie Fertigungstoleranzen, Getriebespiel oder Verschleiß zu signifikanten Abweichungen zwischen dem realen Roboter und seinem kinematischen Modell führen, was eine geringe absolute Positionsgenauigkeit des Roboters zur Folge hat. Dies ist insbesondere bei der Offline-Programmierung von Industrierobotern problematisch (Chen-Gang u. a. 2014). Um dem entgegenzuwirken und eine hohe Genauigkeit sicherzustellen, müssen Roboter daher regelmäßig nachkalibriert werden. Unter Kalibrierung versteht man hier den Vorgang, die vorhandene Ungenauigkeit eines Roboters in Bezug auf seine Absolutposition mithilfe von Software zu reduzieren. An der Hardware wird dabei nichts verändert.

Gängig sind dabei verschiedene Varianten der sogenannten modellbasierten Kalibrierung (Everett, Driels und Mooring 1987; Okamura und F. Park 1996; Stone 1987). Bei diesem Verfahren werden die Parameter des kinematischen Modells, welches zur Bahnplanung und Steuerung des Roboters verwendet wird, so angepasst, dass das Modell den realen Roboter besser beschreibt. Hierfür wird zunächst für einige geplante Endeffektor-Posen im Arbeitsraum die tatsächliche Endeffektor-Pose des Roboters mit einem externen Messsystem vermessen. Diese Messdaten werden dann dazu verwendet, die Parameter des Modells so anzupassen, dass die geplanten Endeffektor-Posen im Anschluss mit höherer Genauigkeit erreicht werden (Chen-Gang u. a. 2014; Roth, Mooring und Ravani 1987).

Der Vorteil der modellbasierten Kalibrierung ist, dass die Methode einfach und nachvollziehbar ist und nur eine überschaubare Menge an Messdaten benötigt wird. Allerdings lassen sich mit dieser Methode nur Geometriefehler korrigieren. Viele andere Fehlerquellen, wie z. B. ein Getriebespiel oder das Durchbiegen von Gelenken und Verbindungen, bleiben dabei unberücksichtigt. Es gibt zwar Ansätze, um das Durchbiegen der Gelenke unter dem Eigengewicht des Roboters explizit zu modellieren und in das Modell einzubauen (Zhou, Nguyen und Kang 2014), es ist aber nicht möglich, dies für alle Fehlereinflüsse zu tun.

Eine Alternative besteht darin, die klassische modellbasierte Kalibrierung mit einem tiefen neuronalen Netz, sprich einem datenbasierten Modell, zu einem hybriden Modell zu verbinden. Hierfür eignet sich ein Residuen-Ansatz (siehe Kap. 3), wie er in Abb. 2 dargestellt ist und im Folgenden beschrieben wird.

Ausgangspunkt ist ein analytisches Modell f, welches das betrachtete System näherungsweise beschreibt. D. h. für eine Eingabe X liefert das analytische Modell die Ausgabe $f(X)$, welche ungefähr dem wahren Wert Y entspricht. Die Differenz zwischen wahrem Wert und Modell $\Delta(X) = Y - f(X)$ wird als Residuum bezeichnet. Nun wird ein künstliches neuronales Netz anhand von Messdaten darauf trainiert, für eine Eingabe X das Residuum Δ vorherzusagen. Das vorhergesagte Residuum $\Delta(X)$ wird dann zur Ausgabe des analytischen Modells $f(X)$ hinzuaddiert. Das neuronale Netz liefert also einen kleinen Korrekturterm,

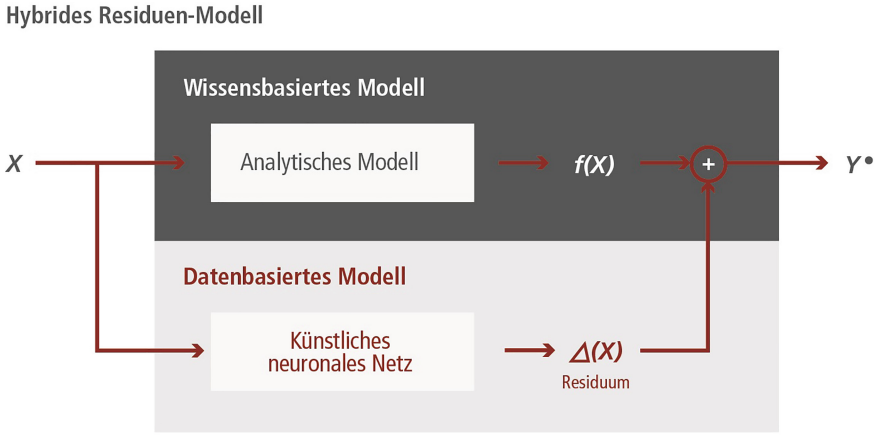

Abb. 2. Schematische Darstellung eines hybriden Residuen-Modells, das ein wissensbasiertes, analytisches mit einem datenbasierten ML-Modell kombiniert

um die Ausgabe des analytischen Modells zu verbessern. Ziel ist, so ein genaueres Modell des realen Systems zu erhalten und eine möglichst genaue Vorhersage Y^*.

Im Fall der Roboterkalibierung wird das tiefe neuronale Netz darauf trainiert, die Differenz zwischen dem angepassten kinematischen Modell, welches ausschließlich geometrische Fehler berücksichtigt, und dem realen Roboter zu beschreiben. Dafür muss eine entsprechende Zahl zusätzlicher Messdaten gesammelt werden. Die Idee ist, dass das Netz anhand der Messdaten lernt, die nicht-geometrischen Fehler zu korrigieren und man so ein sehr genaues Modell des realen Roboters erhält. Aktuelle Forschung auf diesem Gebiet zeigt, dass sich mit einem solchen Hybridansatz die absolute Positionsgenauigkeit von Industrierobotern gegenüber klassischen Verfahren deutlich verbessern lässt (Gadringer u. a. 2020; Landgraf u. a. 2021; Nguyen, Le und Kang 2019).

Prinzipiell ist es auch möglich, anstatt des Hybridmodells ein rein datenbasiertes Modell zu verwenden und die komplette Kinematik des Roboters z. B. mit einem tiefen neuronalen Netz anhand von Messdaten zu lernen (Csiszar, Eilers und Verl 2017). Fehler werden hier sozusagen direkt mitgelernt. Allerdings besteht beim Lernen einer inversen Kinematik, die für eine gegebene Endeffektor-Pose die zugehörigen Gelenkwinkel des Roboters liefert, die Problematik, dass es in der Regel mehrere oder im Fall von Singularitäten auch unendlich viele Lösungen gibt. ML-Modelle können damit im Allgemeinen nicht gut umgehen. Aufgrund der mangelnden Erklärbarkeit und Verifizierbarkeit tiefer neuronaler Netze ergeben sich außerdem Schwierigkeiten in sicherheitstechnischer Hinsicht. Die Hybridlösung, bei der die Kinematik des Roboters durch ein nachvollziehbares, analytisches Modell beschrieben wird und das Netz lediglich einen kleinen Korrekturterm liefert, bietet hier Vorteile hinsichtlich Robustheit und Sicherheit. Darüber hinaus vermeidet es die beschriebene Problematik beim Lernen einer inversen Kinematik mit einem ML-Modell.

Abb. 3. Darstellung eines klassischen Regelkreises, bestehend aus Regelstrecke und Regler (Lunze 2010)

4.2 Hybrides maschinelles Lernen für die Regelungstechnik

Die Regelungstechnik befasst sich als Teilbereich der Automatisierungstechnik mit der Aufgabe, einem gegebenen System ein gewünschtes Verhalten einzuprägen. Dazu wird ein Regler ausgelegt, welcher die im System beeinflussbaren Größen entsprechend berechnet. Ein einfaches Beispiel stellt ein Heizungssystem dar. Das als Regler fungierende Thermostat berechnet anhand der Raumtemperatur das notwendige Warmwasser, welches in den Heizkörper fließt, sodass der Raum auf eine gewünschte Temperatur gebracht wird. Häufig wird das gesamte Konstrukt in Form eines Regelkreises (siehe Abb. 3) dargestellt. Das zu regelnde System wird als Regelstrecke bezeichnet und hat zwei Eingangsgrößen: die Stellgröße, die der Nutzer direkt beeinflussen kann, sowie die Störgröße, die sich aus der Systemumgebung ergibt und nicht manipulierbar ist. Die Ausgangsgröße, hier als Regelgröße bezeichnet, stellt den zu regelnden Istwert dar. Der Regler erhält die Differenz zwischen Führungs- und Regelgröße, sodass auf Basis dessen eine Stellgröße berechnet werden kann.

Um die Regelungsaufgabe zu lösen, müssen nach Lunze 2010 verschiedene Schritte abgearbeitet werden. Nach Formulierung der Regelungsaufgabe werden zunächst die Regel- und Stellgrößen definiert. Anschließend muss die Regelstrecke modelliert und identifiziert werden. In der Praxis stellt dies oftmals die größte und zeitaufwendigste Arbeit dar. Erst wenn ein Modell der Regelstrecke bekannt ist, kann ein Reglerentwurf erfolgen. Anschließend wird der entworfene Regler in Simulationen analysiert, um ihn final umzusetzen. Die Art des Reglerentwurfs steht und fällt mit der Wahl der Modellierung. Einfache Systeme, deren Dynamik bekannt sind, lassen sich mittels Differentialgleichungen modellieren. Dies stellt einen rein wissensbasierten Ansatz dar. Ist die Dynamik zu komplex, werden datengetriebene Ansätze verfolgt. Dabei kann die Regelstrecke über verschiedene Ansätze, wie z. B. Autoregressive Moving Average Modelle, kurz ARMA-Modelle, (Lydia u. a. 2016) oder künstliche neuronale Netze (C. Wang und Hill 2006), modelliert werden. In letzter Zeit werden immer öfter hybride Ansätze verfolgt. Dies hat den Vorteil, dass ein nicht gänzlich unbekanntes System sicher über klassische Verfahren modelliert werden kann, während die unbekannten Komponenten zeiteffizient über datengetriebene Ansätze abgebildet werden. Ein aktuelles Beispiel stellt die Identifikation von Parametern in Lithium-Ionen-Batterien dar. Die Modellstruktur dessen ist hinlänglich bekannt, sodass hier über ein Ersatzschaltbild aus in Reihe geschalteten Widerstands-Kondensator-Schwingkreisen ein hinreichend

genaues Modell erstellt wird. Die Identifikation der zugehörigen Widerstands- und Kapazitätswerte zur Laufzeit stellt jedoch eine Herausforderung dar, wofür Hashemi u. a. 2021 die datengetriebenen Methoden Support-Vector-Machine und Gaußprozessregressor einsetzen.

Datengetriebene Ansätze können jedoch auch direkt in der Modellbildung die wissensbasierten Methoden unterstützen. Dies ist vor allem dann hilfreich, wenn der Arbeitspunkt keine ausreichende Datenbasis besitzt. Wird zum Beispiel ein neuronales Netz um einen Arbeitspunkt mithilfe von Daten trainiert, so ist das Netz in diesem Bereich häufig hinreichend valide. Wird der Arbeitspunkt im Anschluss jedoch verlassen, stellt die schlechte Extrapolationseigenschaft von neuronalen Netzen ein Problem dar. Ein hybrides Modell, welches nicht rein datengetrieben aufgebaut ist, sorgt in diesem Bereich für bessere Resultate, wie in S. Park u. a. 2021 anhand eines Kühlaggregats gezeigt wurde. In diesem Fall erweitert ein neuronales Netz bereits bekannte thermodynamische Zusammenhänge um eine Parameterschätzung.

Um lineare und zeitunabhängige Systeme in aller Gänze zu identifizieren, wird in Nagel und Huber 2021 die Variante eines neuronalen Autoencoders verwendet. Dabei werden Messgleichungen in ein neuronales Netz gegeben, welches eine Schätzung der Systemgleichungen ausgibt. Die Güte der Schätzung wird im Anschluss direkt evaluiert und auf Basis dessen das neuronale Netz trainiert. Ist bereits ein Teil des Systems bekannt, kann dieser von der Identifikation ausgeschlossen werden, was das Training und die Schätzung stark erleichtert, beschleunigt und genauer macht. Einen Schritt weiter gehen Lampert und Martius 2018. Die Autoren stellen eine Möglichkeit vor, direkt nichtlineare Differentialgleichungen als Regelstrecke mithilfe eines neuronalen Netzes zu identifizieren. Dazu werden verschiedene mathematische Operatoren über Gewichtskoeffizienten miteinander verknüpft. Die Methode funktioniert jedoch nur mit einem gewissen physikalischen Hintergrundwissen des zu untersuchenden Systems, sodass effektiv ein hybrides Modell entsteht. Allen voran ist es notwendig zu wissen, welche mathematischen Operatoren die Systemdynamik am besten wiedergeben. Weiterhin muss festgestellt werden, wie komplex das System ist, sodass das identifizierte Modell nicht über- oder unterangepasst ist. Problematisch ist ausschließlich die Tatsache, dass aufgrund der besonderen Eigenschaft neuronaler Netze nur differenzierbare Operatoren ausgewählt werden können und nicht-differenzierbare höchstens in der letzten Neuronenschicht Anwendung finden. Diese kommen in der Natur jedoch relativ häufig vor, sodass nicht alle Systeme korrekt abgebildet werden können.

5 Erklärbarkeit und digitale Souveränität

Mit einer fortschreitenden Verwendung von maschinellen Lernverfahren in der Industrie geht auch das Bedürfnis nach einer Erklärbarkeit ebenjener einher (Belle und Papantonis 2020). Insofern das verwendete Verfahren eine reine Black Box darstellt, also lediglich die Ein- und Ausgangsgrößen bekannt sind, jedoch kein Wissen über die inneren Prozesse und Vorgänge vorhanden ist, stellt die

tatsächliche Verwendung im Alltag ein großes Hindernis dar. Prüfungen über die korrekte Funktionsweise können ausschließlich über Kreuzvalidierungen erfolgen und eine Fehlfunktion nur per Tests ausgeschlossen werden. In Roscher u. a. 2020 werden in diesem Zusammenhang drei Begriffe unterschieden: Die *Transparenz* eines Modells bezieht sich darauf, ob der Prozess, wie Modellparameter aus Trainingsdaten abgeleitet werden, von fachbezogenen Personen beschrieben und motiviert werden kann. Beispielhaft kann ein künstliches neuronales Netz genannt werden. Die durchgeführten Berechnungen sind per se transparent, da die Ein-Ausgangsbeziehung über eine Gleichung ausgedrückt werden kann. Die Hyperparameter, also z. B. die Neuronenanzahl, werden jedoch meist nach Erfahrungswissen oder heuristisch ausgewählt. *Interpretierbarkeit* hingegen bezieht sich darauf, ob die Entscheidung des Modells selbst nachvollzogen werden kann und worauf die durchgeführte Entscheidungsfindung beruht. Dies ist bei künstlichen neuronalen Netzen nicht mehr erfüllt. Die Anzahl der Operationen ist hier viel zu groß, als dass ein Mensch diese Schritt für Schritt durchgehen und daraus eine sinnvolle Erkenntnis über die Entscheidungsfindung des Modells ziehen könnte. Letztlich ist die tatsächliche *Erklärbarkeit* stark von der Nutzerabsicht abhängig. Ist ein Modell in seiner Entscheidungsfindung interpretierbar, so kann ein Nutzer mit Domänenwissen und Dateneinsicht eine Erklärung der Modellresultate erhalten.

Interpretierbarkeit kann über die Verwendung einfacher statistischer Modelle erzielt werden. Dazu zählen zum Beispiel verallgemeinerte lineare Modelle, aber auch Entscheidungsbäume oder naive Bayes-Klassifiaktoren (Burkart und Huber 2021; Molnar 2020). Problematisch ist, dass mit steigender Interpretierbarkeit die Klassifikationsgüte stark sinkt, sodass hybride Methoden eine Verbesserung darstellen können. T. Wang und Lin 2019 zum Beispiel verwenden eine Kombination aus einem interpretierbaren Modell und einer Black Box. Das interpretierbare Modell kommt in bestimmten Regionen des Trainingsdatensatzes zum Einsatz, wo die Klassifikationsgüte ausreicht. Dies ist z. B. bei einem Datensatz mit binärer Zielgröße der Fall, wenn die Entscheidungsgrenze selbst nichtlineares Verhalten hat, sie jedoch in einem linearen Streifen bleibt (siehe Abb. 4).

Ähnlich wie in Abschn. 4.2 im Bezug zur Auslegung von Reglern können mathematische Modellkenntnisse generell eingesetzt werden, um eine höhere Interpretierbarkeit des Modells und somit eine höhere Vertrauenswürdigkeit zu erzielen. Die Black Box wird durch die Kombination mit menschlichem Wissen zu einer Grey Box. Dies wird z. B. in Massa Gray und Schmidt 2018 ausgenutzt, indem ein Gaußprozessregressor ausschließlich zur Fehlerkorrektur des mathematischen Modells verwendet wird. Ein weiterer Anwendungsfall wird in Jung 2020 beschrieben, wo ein mathematisches Modell und ein neuronales Netz mit dem Ziel der Anomalieerkennung parallel genutzt werden. Sind die Prädiktionen beider Modelle stark unterschiedlich, kann auf eine Anomalie der aktuellen Systemwerte geschlossen werden.

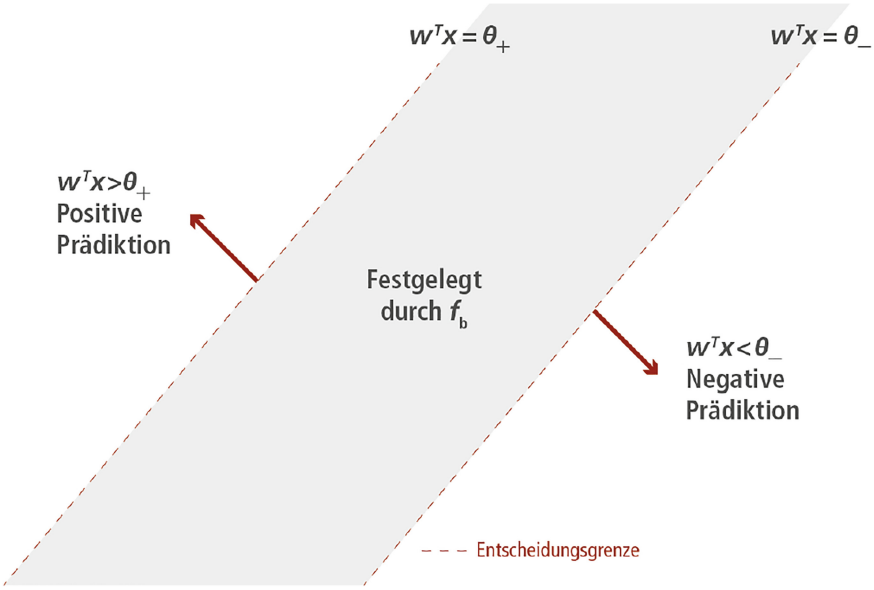

Abb. 4. Darstellung einer einfachen Klassifikation aus (T. Wang und Lin 2019), wobei die genaue Entscheidungsgrenze durch ein Black-Box-Modell f_b genauer wiedergegeben wird, während Bereiche, die weiter entfernt sind, durch ein interpretierbares lineares Modell abgebildet werden, welches durch Parameter w und Eingangsgrößen x ausgedrückt wird. θ legt dabei als Skalar die Entscheidungsgrenze fest

6 Ausblick

Gegenwärtig konzentriert sich die KI-Forschung auf die Entwicklung neuer Methoden im modellfreien und somit rein datengetriebenen Gebiet. Da jeder Anwendungsfall andere Herausforderungen mit sich bringt, ist ein Verwerfen von bereits bestehenden problemspezifischen Kenntnissen Vergeudung. Wie im Artikel beschrieben, zeigen verschiedenste Beispiele, dass modellspezifisches Wissen den Trainingsprozess sowohl beschleunigen als auch die Resultate verbessern kann. Ein weiteres Verallgemeinern der bestehenden hybriden Methoden wird zu einem größeren Anwendungsfeld führen, sodass hybride Modelle das Feld der KI sinnvoll erweitern und weiter zum gesellschaftlichen Fortschritt beitragen.

Literatur

Angwin, J., et al.: Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminalsentencing (2016). Zugegriffen: 30. Juni 2021

Bauernhansl, T.: „Industrie 4.0 – die industrielle Revolution geht weiter". wt Werkstattstechnik online **104**(3), 105 (2014)

Belle, V., Papantonis, I.: Principles and practice of explainable machine learning. arXiv: 2009.11698 [cs.LG] (2020)

Burkart, N., Huber, M.: A survey on the explainability of supervised machine learning. J. Artif. Intell. Res. **70**, 245–317 (2021)

Chen-Gang et al.: Review on kinematics calibration technology of serial robots. Int. J. Precis. Eng. Manuf. **158**, 1759–1774 (2014)

Csiszar, A., Eilers, J., Verl, A.: On solving the inverse kinematics problem using neural networks. Proceedings 2017 24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP) (2017)

Everett, L., Driels, M., Mooring, B.: Kinematic modelling for robot calibration. Proceedings. 1987 IEEE International Conference on Robotics and Automation. Raleigh, NC, USA: Institute of Electrical and Electronics Engineers, S. 183–189 (1987)

Gadringer, S., et al.: Robot calibration combining kinematic model and neural network for enhanced positioning and orientation accuracy. IFAC-PapersOnLine **53**(2), 8432–8437 (2020)

Goodfellow, I., et al.: Attacking machine learning with adversarial examples. https://openai.com/blog/adversarial-example-research/ (2017). Zugegriffen: 1. Juli. 2021

Hashemi, S.R., et al.: Machine learning-based model for lithiumion batteries in BMS of electric/hybrid electric aircraft. Int. J. Energy Res. **45**(4), 5747–5765 (2021)

Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. Neural Netw. **2**(5), 359–366

Howard, A., et al.: MobileNets: efficient convolutional neural networks for mobile vision applications. http://arxiv.org/pdf/1704.04861v1 (2017). Zugegriffen: 1. Juli 2021

Huang, X., et al.: Safety verification of deep neural networks. In: Majumdar, R., Kuncak, V. (Hrsg.) Computer Aided Verification. Lecture Notes in Computer Science, S. 3–29. Springer, Cham (2017)

Jung, D.: Residual Generation Using Physically-Based Grey-Box Recurrent Neural Networks For Engine Fault Diagnosis. arXiv: 2008.04644 [eess.SP] (2020)

Kapusuzoglu, B., Mahadevan, S.: Physics-informed and hybrid machine learning in additive manufacturing: application to fused filament fabrication. JOM **72**(12), 4695–4705 (2020)

Karpatne, A., et al. Theory-guided data science: a new paradigm for scientific discovery from data. IEEE Trans. Knowl. Data Eng. **29**(10), 2318–2331 (2017)

Kleeberger, K., Landgraf, C., Huber, M.: Large-scale 6D object pose estimation dataset for industrial bin-picking. In: 2019 IEEE International Conference on Intelligent Robots and Systems (IROS), S. 2573–2578 (2019)

Krauß, J., et al.: Maschinelles Lernen in der Produktion – Anwendungsgebiete und frei verfügbare Datensätze. Ind. 4.0 Manag. **4**, 39–42 (2019)

Landgraf, C., et al.: A hybrid neural network approach for increasing the absolute accuracy of industrial robots. In: 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE) (2021)

Lunze, J.: Regelungstechnik 1, S. 1–4. Springer, Berlin (2010)

Lydia, M., et al.: Linear and non-linear autoregressive models for shortterm wind speed forecasting. Energy Convers. Manage. **112**, 115–124 (2016)
Massa Gray, F., Schmidt, M.:. A hybrid approach to thermal building modelling using a combination of Gaussian processes and grey-box models. Energ. Build. **165**, 56–63 (2018)
Molnar, C.: Interpretable Machine Learning (2020)
Nagel, T., Huber, M.: Autoencoder-inspired Identification of LTI systems. Proceedings of the European Control Conference (ECC) (2021)
Nguyen, H.-N., Le, P.-N., Kang, H.-J. (2019). A new calibration method for enhancing robot position accuracy by combining a robot model–based identification approach and an artificial neural network–based error compensation technique. Adv. Mech. Eng. **11**(1), 1–11
Okamura, K., Park, F.: Kinematic calibration using the product of exponentials formula. Robotica **14**(4), 415–421 (1996)
Park, S., et al.: Machine learning vs. hybrid machine learning model for optimal operation of a chiller. Sci. Technol. Built Environ. **25**(2), 209–220 (2019)
Pfrommer, J., et. al.: Optimisation of manufacturing process parameters using deep neural networks as surrogate models. In: Wang, L. (Hrsg.) Procedia 51st CIRP Conference on Manufacturing Systems, Bd. 72, S. 426–431 (2018)
Redmon, J., et al.: You only look once: Unified, real-time object detection. In: Proceedings 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), S. 779–788 (2016)
Roscher, R., et al.: Explainable machine learning for scientific insights and discoveries. IEEE Access **8**, 42200–42216 (2020). https://doi.org/10.1109/ACCESS.2020.2976199
Roth, Z., Mooring, B., Ravani, B.: An overview of robot calibration. IEEE J. Robot. Autom. **3**(5), 377–385
Sahoo, S., Lampert, C., Martius, G.: Learning equations for extrapolation and control. In: Proc. 35th International Conference on Machine Learning, ICML 2018, Stockholm, Sweden, 2018., Bd. 80. PMLR, S. 4442–4450 (2018)
Shang, C., You, F.: Data analytics and machine learning for smart process manufacturing: Recent advances and perspectives in the big data era. Engineering **5**(6), 1010–1016 (2019)
Stone, H.: "Kinematic Modeling, Identification, and Control of Robotic Manipulators, bd. 29. The Springer International Series in Engineering and Computer Science, Springer, US (1987)
von Rueden, L., Mayer, S., Beckh, K., et al.: Informed Machine Learning – A Taxonomy and Survey of Integrating Knowledge into Learning Systems. IEEE Transactions on Knowledge and Data Engineering (2021)
von Rueden, L., Mayer, S., Sifa, R., et al.: Combining machine learning and simulation to a hybrid modelling approach: Current and future directions. In Berthold, M.R., Feelders, A., Krempl, G. (Hrsg.) Advances in Intelligent Data Analysis XVIII Bd. 12080. Lecture Notes in Computer Science. Springer International Publishing, Cham, S. 548–560 (2020)
Wagner, P., et al.: KI-Anwendungsfälle in der Produktion. Eingereichtes Manuskript (2021)
Wang, C., Hill, D.J.: Learning from neural control. IEEE Trans. Neural Networks **17**(1), 130–146 (2006)
Wang, J., et al.: Deep learning for smart manufacturing: Methods and applications. J. Manuf. Syst. **48**(C), 144–156 (2018)

Wang, T., Lin, Q.: Hybrid Predictive Model: When an Interpretable Model Collaborates with a Black-box Model. arXiv: 1905.04241 [cs.LG] (2019)
Wuest, T., et al.: Machine learning in manufacturing: advantages, challenges, and applications. Prod. Manuf. Res. **4**(1), 23–45 (2016)
Zech, J.: What are radiological deep learning models actually learning?. https://jrzech.medium.com/what-are-radiological-deep-learning-models-actually-learning-f97a546c5b98 (2018). Zugegriffen: 30. Juni 2021
Zhou, J., Nguyen, H.-N., Kang, H.-J.: Simultaneous identification of joint compliance and kinematic parameters of industrial robots. Int. J. Precis. Eng. Manuf. **15**(11), 2257–2264 (2014)

Open Acces Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (http://creativecommons.org/licenses/by/4.0/deed.de) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäßnennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.