

Künstliche Intelligenz in der zahnärztlich-röntgenologischen Diagnostik

Teil 2

Falk Schwendicke, Joachim Krois



KI in der (Zahn-)Medizin

Eine besondere Form von NN wird beim sog. maschinellen Sehen („Computer Vision“) angewendet. Hierbei werden Bilder beispielsweise klassifiziert (z. B. Bild enthält einen Stuhl), Objekte auf Bildern detektiert (Stuhl steht in der Mitte des Bildes) oder sogar ausgeschnitten (Stuhl hat diesen Umriss) (Abb. 3). Für diese Aufgabe kommen sog. konvolutionäre NN (Convolutional Neural Network, CNN) zur Anwendung. CNNs nutzen Bildfilter, die in der Lage sind, Farben, Kanten, Ecken, Rundungen, Strukturen und komplexere Formen zu erlernen. Kombiniert man diese Bildfilter mit Elementen aus klassischen NN, ist es möglich, u. a. menschliche Gesichter auf Bildern oder fahrende Autos in Videos zu erkennen (LeCun et al. 2015).



Abb. 3: Klassische Aufgaben im Bereich des maschinellen Sehens. Bildklassifikation (l.), Detektion und Bildklassifikation (2. v. l.), Objektdetektion (3. v. l.) und pixelweise Bildsegmentierung (r.). Von links nach rechts nimmt der Komplexitätsgrad der Aufgabe zu. (Quelle: NVIDIA Deep Learning Institute)

CNNs werden zunehmend erfolgreich in der Medizin, z. B. der Ophthalmologie und Dermatologie, eingesetzt (Becker et al. 2017, Goodfellow et al. 2016, Gulshan et al. 2016, Mazurowski et al. 2018). In diesen Fachgebieten konnten diese Techniken teilweise gleich gute oder sogar bessere Genauigkeiten realisieren als Experten des jeweiligen Faches. CNNs werden dafür teils auf mehreren Zehntausend Bildern trainiert und erfüllen überwiegend eng definierte Spezialaufgaben (z. B. Detektion von Retinopathien, nicht jedoch optometrische Vermessung): Für jede „neue“ Aufgabe muss ein eigenes CNN auf neuen Daten entwickelt, trainiert und validiert werden. Um wirklich umfangreich Bildmaterial auf verschiedenste Pathologien hin befunden zu können, ist demnach initial ein großer Befundungsaufwand (auf der Basis großer Datensätze) und ein entsprechender Datenverarbeitungsaufwand notwendig. Sind diese CNNs jedoch einmal trainiert, können schnell auch große Bilddatensätze analysiert werden. Auch in der Zahnmedizin bieten sich zahlreiche Anwendungsmöglichkeiten an (Abb. 4).

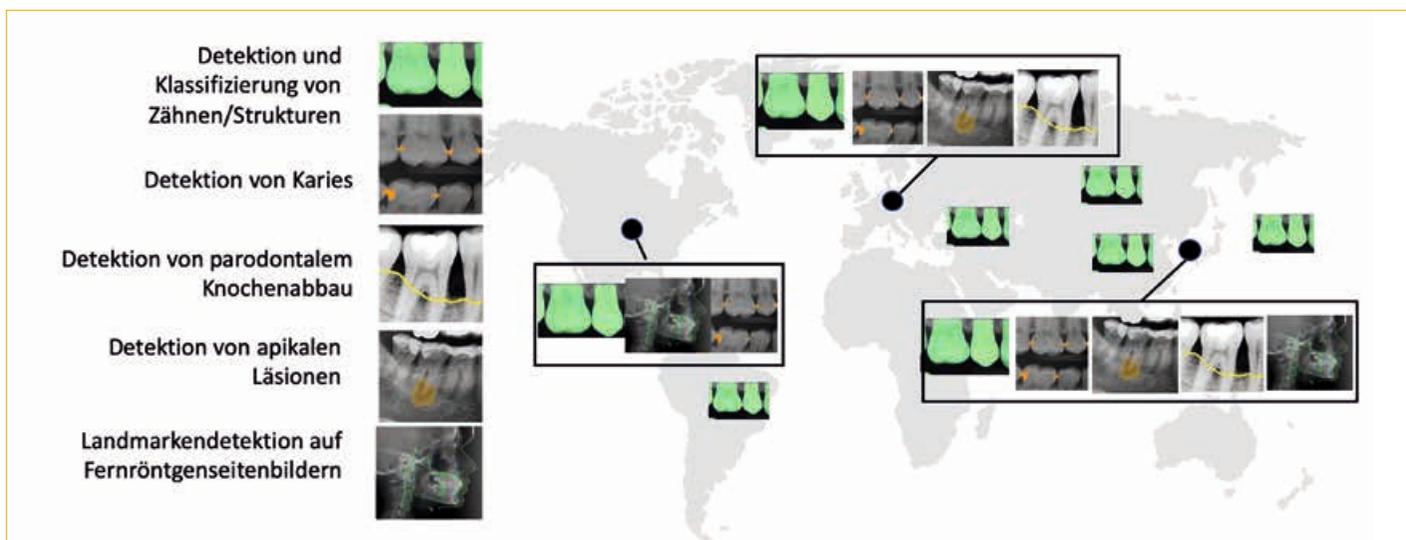


Abb. 4: Beispielhafte Einsatzfelder von KI zur zahnärztlichen Röntgendiagnostik (I.) und publizierte Arbeiten in verschiedenen Feldern, kartografisch dargestellt (Daten aus Schwendicke et al. 2019). Zahnärztliche KI-Forschung hat sich bisher auf die Detektion und Klassifikation von Zähnen und anderen anatomischen Strukturen fokussiert; wenige Zentren (Seoul / Korea, Berlin / Deutschland, USA) weltweit arbeiten intensiv wissenschaftlich an der Detektion von zahnmedizinischen Pathologien.

Die meisten Studien arbeiteten zunächst an der Detektion und Klassifikation von Zähnen auf zahnmedizinischen Röntgenbildern, u. a. Panoramaschichtaufnahmen (Krois 2019b, Schwendicke et al. 2019), bzw. genereller an der Detektion und Klassifikation von anatomischen Strukturen (Kieferknochen, Kanal des N. mandibularis etc.). Hier scheinen CNNs in der Lage zu sein, Zähne auf solchen Bildern zu detektieren und mit hoher Genauigkeit den korrekten Zahnnamen zuzuordnen, auch die Segmentierung von anatomischen Strukturen gelingt relativ gut sowohl in 2D als auch 3D (bei letzteren sind diese Techniken, nicht unbedingt unter Nutzung von CNNs, oftmals bereits kommerziell erhältlich). Die Detektion und Klassifikation von Zähnen mutet zunächst simpel an und hat per se keinen direkten Nutzen, ist aber die Grundlage für die anschließende zahnbezogene Befundung und Dokumentation von Restaurationen oder Pathologien – und technisch nicht unbedingt weniger anspruchsvoll als die Detektion und Klassifikation von Pathologien (was die Unterschiede zwischen menschlichem und maschinellem Lernen und Erkennen verdeutlicht – Zahnärzte sehen die Zahnerkennung auf solchen Bildern sicher nur selten als Schwierigkeit an!). Die Detektion und Klassifikation von Restaurationen fällt ebenso in diese Kategorie – sie ist für das menschliche Auge einfach und für den Zahnarzt keine schwierige Aufgabe. Ein vollständiger Röntgenbefund umfasst jedoch auch die Dokumentation von Opazitäten i. S. von Restaurationen; durch den Einsatz von KI gelingt dies verlässlich in wenigen Sekunden.

Zunehmend werden CNNs auch eingesetzt, um dentale Pathologien (Karies, parodontaler Knochenabbau, apikale Läsionen), aber auch andere medizinisch-relevante Pathologien (z. B. Sinusits, Osteoporose) zu erkennen (Ekert et al. 2019a, Krois et al. 2019a, Lee et al. 2018b, Lee et al. 2018a). Die CNN sind oftmals ähnlich erfolgreich bei der Detektion der Pathologien wie erfahrene Vergleichszahnärzte; nicht immer wird das CNN jedoch gegen unabhängige Zahnärzte getestet. Bemerkenswert ist, dass die CNNs bisher selten besser als Zahnärzte sind und zahnärztliche Spezialisten momentan mindestens genauso genau oder sogar genauer als CNNs sind, vor allem wenn sie ausreichend Zeit (z. B. eben im Rahmen einer Studie) haben.

Relevant werden diese Techniken vor allem aus einem weiteren Grund: Sie werden Teil einer „digitalen Bewegung“ in der Zahnmedizin. Bilddaten aus professionellen Quellen (z. B. eben Röntgenbilder, aber auch Oberflächenscans, Fotos etc.), aber auch der Patienten selbst (Handyfotos o. ä.) sowie weitere klinische Daten aus der Praxis oder dem Patientenalltag (Zahnbürstendaten, Ernährungsdaten, Social Media-Daten) können zunehmend vernetzt werden und zur Diagnostik und Profilierung des Patienten eingesetzt werden. Gerade letzteres klingt ein Stück weit bedrohlich, werden doch von dem Patienten Datensätze kompiliert, deren Inhalte ihm selbst vielleicht nicht einmal bewusst waren. Zudem entsteht ein sog. „Digitaler Zwilling“; ein unberechtigter Zugriff auf diese sensiblen personenbezogenen Daten kann demnach massiv schädlich sein. Das Nutzen dieser Daten kann jedoch eine völlige andere Art der Medizin befördern: Statt Patienten „nur“ in Risikogruppen wie „hohes“ oder „niedriges“ Kariesrisiko einzugruppieren (stratifizierte Medizin), kann ein viel individuelleres Bild jedes einzelnen Patienten gewonnen werden (personalisierte Medizin), wodurch Patienten sicherer, risikoärmer und erfolgreicher behandelt werden können (Präzisionsmedizin). Ebenso wird es dem individuellen Patienten zukünftig vermehrt möglich sein, seinen Gesundheitszustand zu „erleben“ und zu steuern; Patienten werden zunehmend Teil von Medizin, statt sie passiv zu „empfangen“ (partizipatorische Medizin). Die neuen Technologien sollen zudem ermöglichen, Erkrankungen früher, teilweise vor ihrem „Ausbruch“ zu erkennen und verhindern zu können (präventive statt vor allem kurative Medizin). Die Umsetzung einer solchen sog. „P4-Zahnmedizin“ wird mit Herausforderungen, aber auch großen Chancen kommen (Hood 2018). Entscheidend wird dabei sein, KI-Technologien robust und vertrauenswürdig zu machen.

Robuste KI = Vertrauen

Die zahlreichen Anwendungsbeispiele demonstrieren das Potenzial dieser Techniken, schlussendlich durch eine bessere Diagnostik und eine sichere, wirksamere und effizientere Therapiezuordnung nützlich zu sein. Der Arzt und Medizinforscher Eric Topol geht sogar davon aus, dass KI in der Medizin Ärzte entlasten und das Arzt-Patientenverhältnis wieder menschlich machen könnte (Topol 2019).

Eine große Schwäche der allermeisten Anwendungen von KI in der Medizin ist allerdings die Robustheit dieser Modelle und die möglichen Verzerrungen, denen ihre Ergebnisse unterliegen. Anders als beim autonomen Fahren oder der Gesichtserkennung steht KI in der Medizin erst seit 2014 im Fokus – also gerade einmal fünf Jahre! Viele Entwicklungen und Studienergebnisse sind schlicht zu „jung“, um umfänglich validiert und im Alltag erprobt sein zu können. Umgekehrt ist der technologische Fortschritt zurzeit so rasant, dass es nahezu unethisch erscheint, diese Technologien zunächst für weitere drei bis fünf Jahre zu erproben und bis dahin dem medizinischen Alltag vorzuenthalten. Was gilt es also zumindest zu tun, um ein Grundvertrauen in KI zu befördern?

- 1. robuste, generalisierbare Trainingsdaten einsetzen und validieren:** Viele in der Zahnmedizin angewandte CNNs sind auf kleinen Datensätzen entwickelt worden (Schwendicke et al. 2019); generell sind medizinische Datensätze verglichen mit anderen Domänen von KI oftmals klein. Stammen diese Datensätze dann nur aus einer Klinik oder beinhalten Daten nur einer Ethnie, eines Geräteherstellers o. ä. ist eine Generalisierbarkeit nicht gegeben. Zahlreiche Beispiele für einen sog. Selektionsbias in KI-Technologien sind bekannt. Um

diese Generalisierbarkeit zu prüfen, sollten die trainierten Modelle nicht nur intern „naiv“ überprüft werden (wie gut hat das Modell den vorhandenen Datensatz gelernt?) oder durch sog. Kreuzvalidierung getestet werden (hierbei wird ein immer neuer Teil des Datensatzes zufällig nicht in das Training mit einbezogen, sondern „beiseite“ gelegt und dann zum Testen eingesetzt) sondern durch externe Validierung (also zumindest an einem Testdatensatz als Teil des Gesamtdatensatzes, den das Modell jedoch nie vorher gesehen hat, sog. „Hold-Out-Testset“). Ideal ist eine Validierung in einem komplett anderen Datensatz (z. B. aus einer anderen Klinik).

2. **Definition nützlicher Goldstandards, Einsatz nützlicher Metriken, Nützlichkeitsbeweis:** Viele Studien in der medizinischen KI-Forschung entwickeln Modelle, die gegen einen „Goldstandard“ beziehungsweise Referenztest trainiert und validiert werden. Da die meisten Studien Datensätze mit hunderten, tausenden oder zehntausenden Bildern, die in der klinischen Routine entstanden, einsetzen, sind diese Goldstandards allerdings weniger „hart“ als gewöhnlich: So wird in der Kariesdiagnostikforschung z. B. eine neue Kariesdetektionsmethode gegen einen histologischen oder mikroradiografischen Standard geprüft, unter der Annahme, diese Methoden würden mit hoher Sicherheit anzeigen, ob Karies wirklich vorhanden ist. Eine solche Validierung ist für KI-Forschung nur selten möglich. Auch das Erheben zusätzlicher Tests (andere Bildgebungsverfahren wie z. B. DVT / CT, Biomarker etc.) verbietet sich oft aus ethischen oder Kostengründen. Stattdessen wird der Goldstandard anders hergestellt: Oftmals bewerten (annotieren) mehrere (teilweise deutlich mehr als fünf) erfahrene Ärzte bzw. Zahnärzte dasselbe Bild, wodurch die Bewertungsfehler des Einzelnen reduziert und das Modell genauer als der „Durchschnittsbefunder“ werden soll (ein Modell, das nur gegen die Befunde eines einzelnen Zahnarztes trainiert wird, kann nie besser sein als dieser einzelne Zahnarzt). Die Logik ist also: Umso mehr Zahnärzte ein Bild befundet haben, umso besser die Qualität des daraus erwachsenden Referenztests und damit des zu entwickelnden Modells. Allerdings stellen sich auch hier diverse Fragen. U. a. ist unklar, nach welchen Regeln aus diesen Mehrfachbefunden nun die „Wahrheit“ (Zahn krank oder gesund) generiert werden soll (Soll es eine Mehrheitsentscheidung geben? Einen Schwellenwert? Was passiert mit Daten, bei denen dieser unterschritten wird oder die Mehrheit knapp ist?). Um die Schwierigkeiten dieser „ungenauen“ (englisch „fuzzy“) Goldstandards abzuschwächen, wird empfohlen, möglichst weitere Tests zur sog. Triangulation einzusetzen (Walsh 2018). Diese können idealerweise auch der klinischen Routine entstammen: So kann bei der Kariesbefundung zusätzlich auf Daten aus der klinischen Untersuchung zurückgegriffen werden. Bei der Detektion parodontalen Knochenabbaus können klinische Attachmentverlustmessungen o. ä. eingesetzt werden. Allerdings sind auch diese Tests ungenau und mit Fehlern behaftet; zudem ist die Dokumentationsqualität in vielen Kliniken oder Praxen unzureichend; auch sind diese Daten nicht immer ohne größeren Aufwand exportierbar. Des Weiteren sollten Metriken zur Beschreibung der Modellgenauigkeit eingesetzt werden, die nicht nur technisch relevant sind, sondern auch Informationen über den klinischen Nutzen liefern (Schwendicke et al. 2019). So sind beispielsweise Modelle, die teilweise hohe Genauigkeiten („accuracy“) aufweisen, mitunter nicht nützlicher als einfaches Raten (wenn z. B. bestimmte Erkrankungen sehr selten sind, kann es passieren, dass ein Modell lernt, immer „gesund“ vorherzusagen und so relativ hohe Genauigkeiten erreicht; nützlich ist dieses Modell jedoch nicht!). Der schlussendliche Nützlichkeitsbeweis sollte zudem in der Praxis erbracht werden: Führt der Einsatz von KI zu besseren Entscheidungen? Diagnostische Validität hat für den Patienten oftmals nur wenig direkten Nutzen oder Schaden – diese entstehen durch die anschließende Therapie! KI sollte die Therapiewahl verbessern, also wirksamer, sicherer und effizienter machen oder zumindest die Abläufe verbessern (Prozessparameter optimieren).

3. KI erklärbar und vertrauenswürdig machen: Modelle der KI sind oft hochkomplex; die zugrundeliegende mathematische Struktur erlaubt nicht, ohne weiteres zu verstehen, welche Bildkomponenten (sog. „Features“) von dem Modell bei seinen Entscheidungen (z. B. Karies vorhanden / nicht vorhanden) berücksichtigt werden. KI-Modelle sind oft „Black Box Modelle“, die anfällig für unbeobachtete Verzerrungen sind: Möglicherweise werden nicht jene Features, die Ärzte und Zahnärzte zur Entscheidung heranziehen eingesetzt, sondern Artefakte. Dies hätte gravierende Konsequenzen, wenn das Modell an anderen Daten als den Trainingsdaten (wo es offensichtlich mit dem Lernen von Artefakten erfolgreich war) eingesetzt würde. Hierfür gibt es zahlreiche Beispiele (Paul et al. 2019, Ho et al. 2019, Tang et al. 2018) und einige, nichtmedizinische Fälle haben es in die Presse geschafft (z. B. die gescheiterte Vorhersage des Ausgangs der Fußballweltmeisterschaft 2018, Amazons KI-gestütztes Einstellungsverfahren zeigt eine klare Bevorzugung männlicher Bewerber u. a.). Mittlerweile gibt es eine Reihe von Techniken, die die zugrundeliegende „Logik“ von KI erklärbar machen. Ein ganzer Forschungsbereich widmet sich dieser Aufgabe („Explainable AI“). So können heute die Bereiche, die ein Modell zur Vorhersage nutzt, visualisiert werden oder bestimmte Tests zur Anfälligkeit von Modellen für Verzerrungen eingesetzt werden. Gerade bei der Visualisierung der vom Modell eingesetzten Features sollte aber berücksichtigt werden, dass diese Modelle möglicherweise wirklich Muster erkennen, die dem menschlichen Auge nicht zugänglich sind: Eine mangelnde Logik aus unserer Sicht muss deshalb nicht zwingend für eine Verzerrung des Modells sprechen, sondern kann eine „superhumane“ Genauigkeit anzeigen. Nichtsdestotrotz sollten solche Befunde stets kritisch hinterfragt werden.

Alle drei Punkte – bessere Daten und Validierungen, Definition von Goldstandards und Metriken, Verstehen und Vertrauen in KI – sind Subjekt intensiver Forschungsbemühungen. Eine Initiative der WHO zusammen mit der International Telecommunication Union (ITU) spricht sich beispielsweise für die Definition von Standards zur Entwicklung und Erprobung von KI in der Medizin aus; eine entsprechende Fachgruppe Zahnmedizin in der beratenden „Fokusgruppe AI4Health“ wird zurzeit etabliert. Nur durch Anwendung einer rigorosen Methodik kann aus dem jetzigen Hype eine stetige Erfolgsgeschichte werden und KI sich das Vertrauen der Nutzer – Ärzte, Patienten, Kostenträger – in der Medizin verdienen.

Zusammenfassung

Bildgebung ist aus der zahnmedizinischen Diagnostik nicht wegzudenken. Die Möglichkeiten, die sich aus der Digitalisierung und „Künstlicher Intelligenz“ ergeben, werden auch in der Zahnmedizin Diagnoseassistenzsysteme entstehen lassen; diese werden zunehmend auch Therapieentscheidungen unterstützen. Der Zahnarzt bzw. die Zahnärztin werden jedoch weiterhin die Verantwortlichkeit tragen und demnach auch schlussendlich gemeinsam mit dem Patienten und unterstützt durch KI diese Entscheidungen treffen müssen. Bei aller Begeisterung sind jedoch zurzeit noch viele Unwägbarkeiten und Unsicherheiten vorhanden, die überwunden werden müssen, bevor KI in die Routineversorgung einzieht. Langfristig können die besseren Möglichkeiten, zunehmend große Datenmengen sinnvoll zu verarbeiten, diagnostische und therapeutische Entscheidungen sicherer, zuverlässiger und wirksamer machen. Hierbei werden Techniken der KI (z. B. CNN in der Bildanalytik) eine große Rolle spielen.



**PD Dr. med. dent.
Falk Schwendicke**

- 2003–2008 Studium der Zahn-, Mund- und Kieferheilkunde, Charité – Universitätsmedizin Berlin
- 2009 Approbation und Promotion
- 2009–2012 Zahnarzt, Banbury, Oxfordshire, Vereinigtes Königreich
- 2012–2013 Wissenschaftlicher Mitarbeiter, Klinik für Zahnerhaltungskunde und Parodontologie der Universitätsklinik für Zahn-, Mund- und Kieferheilkunde, Universitätsklinikum Schleswig-Holstein, Christian-Albrechts-Universität zu Kiel

- 2013 Ernennung zum Oberarzt für Kariologie und Präventivzahnmedizin, Abteilung für Zahnerhaltung und Präventivzahnmedizin, Charité – Universitätsmedizin Berlin
- 2014–2015 Master of Dental Public Health (MDPH), University of Manchester (UK)
- 2015 Lehrbefugnis und Venia Legendi für das Fach Zahnmedizin
- 2016 Vorsitzender des Fachbereichs Zahnmedizin im Deutschen Netzwerk für Evidenzbasierte Medizin
- Mitglied in zahlreichen Verbänden, Arbeitsgemeinschaften und Vereinigungen
- Fachautor zahlreicher Publikationen in nationalen und internationalen Journals
- Forschungsschwerpunkte: Selektive Kariesexkavation und Kariesversiegelung tiefer Läsionen; Vitalerhaltung der Pulpa; Probiotika in der Kariesprävention; Kosten und Nutzen zahnmedizinischer Diagnostik und Therapien; Epidemiologie und zahnmedizinische Versorgung



Dr. rer. nat. Joachim Krois

- seit 2017 Datenanalyst und wiss. Mitarbeiter an der Charité – Universitätsmedizin Berlin
- 2016–2017 freiberuflicher Datenwissenschaftler
- 2016 Promotion (Dr. rer. nat.)
- 2013–2016 wiss. Projektkoordination an der FU Berlin
- 2008–2012 wiss. Mitarbeiter an der FU Berlin (Geowissenschaften)
- Trainer und Lehrbeauftragter für Statistik, Data Science, Machine Learning und Coding

Kontakt:

Abteilung für Digitale und Allgemeine Zahnheilkunde, Versorgungsforschung und Gesundheitsökonomie CharitéCentrum 3 für Zahn-, Mund- und Kieferheilkunde Charité – Universitätsmedizin Berlin Aßmannshäuser Straße 4–6 D-14197 Berlin
Tel. +49 (0)30 / 45 06 62 55 6
falk.schwendicke@charite.de

Referenzen

Becker, A. S., Marcon, M., Ghafoor, S., Wurnig, M. C., Frauenfelder, T. & Boss, A. (2017) Deep Learning in Mammography: Diagnostic Accuracy of a Multipurpose Image Analysis Software in the Detection of Breast Cancer. *Invest Radiol* 52, 434–440. doi:10.1097/rli.0000000000000358.

BfS (2019) Röntgendiagnostik: Häufigkeit und Strahlenexposition. <https://www.bfs.de/DE/themen/ion/anwendung-medicin/diagnostik/roentgen/haeufigkeit-exposition.html>.

Ekert, T., Krois, J., Meinhold, L., Elhennawy, K., Emara, R., Golla, T. & Schwendicke, F. (2019) Deep Learning for the Radiographic Detection of Apical Lesions. *J Endod.*

Goodfellow, I., Bengio, Y. & Courville, A. (2016) *Deep Learning*. Cambridge, MA: MIT Press.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L. & Webster, D. R. (2016) Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* 316, 2402–2410. doi:10.1001/jama.2016.17216.

Ho, C. W. L., Soon, D., Caals, K. & Kapur, J. (2019) Governance of automated image analysis and artificial intelligence analytics in healthcare. *Clin Radiol* 74, 329–337. doi:10.1016/j.crad.2019.02.005.

Hood, L. (2018) Lessons Learned as President of the Institute for Systems Biology (2000–2018). *Genomics Proteomics & Bioinformatics* 16.

Krois, J., Ekert, T., Meinhold, L., Golla, T., Kharbot, B., Wittemeier, A., Dorfer, C. & Schwendicke, F. (2019a) Deep Learning for the Radiographic Detection of Periodontal Bone Loss. *Sci Rep* 9, 8495. doi:10.1038/s41598-019-44839-3.

Krois, J., Gehrung, S., Garcia-Cantu, A., Golla, T., Dreher, M., Schwendicke, F. (2019b) Instance segmentation of tooth restorations on dental panoramic radiographs using convolutional neural networks. In: *IADR General Session*.

KZBV (2017) *KZBV Jahrbuch 2017*. Berlin. LeCun, Y., Bengio, Y. & Hinton, G. (2015) Deep learning. *Nature* 521, 436–444. doi:10.1038/nature14539.

Lee, J. H., Kim, D. H., Jeong, S. N. & Choi, S. H. (2018a) Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent* 77, 106–111. doi:10.1016/j.jdent.2018.07.015.

Lee, J. H., Kim, D. H., Jeong, S. N. & Choi, S. H. (2018b) Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *J Periodontol Implant Sci* 48, 114–123. doi:10.5051/jpis.2018.48.2.114.

Mazurowski, M., Buda, M., Saha, A. & Bashir, M. (2018) Deep learning in radiology: an overview of the concepts and a survey of the state of the art. *arXiv:1802.08717v1*.

McCulloch, W. S. & Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics* 5, 115–133. doi:10.1007/BF02478259.

Paul, R., Schabath, M., Balagurunathan, Y., Liu, Y., Li, Q., Gillies, R., Hall, L. O. & Goldgof, D. B. (2019) Explaining Deep Features Using Radiologist-Defined Semantic Features and Traditional Quantitative Features. *Tomography* 5, 192–200. doi:10.18383/jtom.2018.00034.

Schwendicke, F., Golla, T., Dreher, M., Krois, J. (2019) Convolutional neural networks for dental image diagnostics: A Scoping Review. *J Dent* 10.1016/j.jdent.2019.103226.

Tang, A., Tam, R., Cadrin-Chenevert, A., Guest, W., Chong, J., Barfett, J., Chepelev, L., Cairns, R., Mitchell, J. R., Cicero, M. D., Poudrette, M. G., Jaremko, J. L., Reinhold, C., Gallix, B., Gray, B. & Geis, R. (2018) Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology. *Can Assoc Radiol J* 69, 120–135. doi:10.1016/j.carj.2018.02.002.

Topol, E. (2019) *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. New York: Basic Books.

Walsh, T. (2018) Fuzzy gold standards: Approaches to handling an imperfect reference standard. *J Dent* 74 Suppl 1, S47–s49. doi:10.1016/j.jdent.2018.04.022.