

Diplomarbeit

Automatisierte Herdsuche in DVT Datensätzen mit Hilfe eines deep Convolutional Neural Network (CNN) – Eine retrospektive Studie

eingereicht von

Jan-Niclas Aaron Press

zur Erlangung des akademischen Grades

Doktor der Zahnmedizin

(Dr. med. dent.)

an der

Medizinischen Universität Graz

ausgeführt an der

Universitätsklinik für Zahnmedizin und Mundgesundheit

unter der Anleitung von

Univ. OÄ Priv.-Doz. Dr.scient.med. Dr.med.dent Barbara Kirnbauer

und

Univ.-Prof. Dr.med.univ. Dr.med.dent Norbert Jakse

Graz, am 17.04.2024

Eidesstattliche Erklärung

Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe, andere als die angegebenen Quellen nicht verwendet habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am 17.04.2024

Jan-Niclas Aaron Press eh.

Danksagungen

Ein herzliches Dankeschön gilt meiner Betreuerin, Priv. Doz. Kirnbauer, die jederzeit bei Fragen sofort zur Verfügung stand. Dank ihrer fachlichen Expertise und Unterstützung konnte ich diese Diplomarbeit erfolgreich abschließen.

Ebenfalls möchte ich mich bei meinen Freunden und meiner Freundin bedanken, die auch zu diesem Erfolg beigetragen haben. Ihr wart während meines Studiums stets ein wichtiger Bestandteil und habt mich immer unterstützt.

Ich möchte mich von Herzen bei meiner Familie, insbesondere meinen Eltern, bedanken. Ihr habt mich auf jedem Schritt meines Weges begleitet, unterstützt und stets motiviert. Ohne euch wäre ich nie so weit gekommen.

Diese Arbeit widme ich daher voller Liebe und Dankbarkeit meinen Eltern.

Zusammenfassung

Hintergrund:

Zu den häufigsten pathologischen Veränderungen in der Zahnmedizin zählen osteolytische Kieferläsionen im Bereich der Wurzelspitze als Folge einer bakteriellen, zahnassozierten Infektion. Solche periapikalen osteolytischen Läsionen erscheinen im Röntgen als radioluzente Areale mit unterschiedlichem Durchmesser. Zur Vermeidung von Schmerzen und Komplikationen über Abszessbildung bis zum Zahnverlust ist eine sichere Diagnostik essenziell. Konventionell findet diese mittels 2D-Bildgebungsverfahren statt, doch mittlerweile wird diese Technik durch die dreidimensionale digitale Volumentomographie (DVT) ersetzt. Eine detaillierte DVT-Befunderhebung ist im Vergleich zu einer 2D-Diagnostik zeitaufwendiger, der Arbeitsaufwand nimmt mit der steigenden Anzahl an DVT-Untersuchungen zu. Hier kann künstliche Intelligenz ansetzen. Computergestützte Befunderhebung auf Grundlage von Deep Learning von DVT-Datensätzen kann die Arbeitslast von Zahnärzt*innen reduzieren und zusätzlich zu einer Qualitätssteigerung der Befundung führen.

Das Ziel dieser Arbeit besteht darin, einen trainierten KI-Algorithmus aus einer Vorläuferstudie (<https://doi.org/10.1016/j.joen.2022.07.013>) auf Nichtunterlegenheit zu testen.

Methode:

Die Studie beinhaltet 195 klinisch repräsentative Datensätze, die an der Univ.-Klinik für Zahnmedizin und Mundgesundheit in Graz entsprechend den Einschlusskriterien untersucht wurden. Ein Expertinnenteam stellte im Konsens den Goldstandard dar. Insgesamt wurden 2947 Zähne in die Studie eingeschlossen, von denen 300 periapikale osteolytische Läsionen vorwiesen. Anhand des definierten Goldstandards konnte die Sensitivität und Spezifität des KI-Algorithmus ermittelt werden. Um eine indexbasierte Auswertung zu ermöglichen, wurde jeder Läsion ein Index entsprechend ihrem Durchmesser zugeordnet.

Ergebnisse:

Der verwendete Algorithmus erreichte eine Sensitivität von 86,7 % und eine Spezifität von 82,9 %. Die Nicht-Unterlegenheitsprüfung basierend auf der Proof-of-Concept-

Evaluierung lieferte ausreichend Evidenz, um die Nullhypothese für die Spezifität, nicht aber für die Sensitivität zurückzuweisen. Werden aus dem Datensatz Läsionen des Indexscore 1 (< 1mm) exkludiert, steigt die Sensitivität auf 90,4 %.

Schlussfolgerung:

Trotz des klinisch anspruchsvollen Datensatzes zeigt der verwendete KI-Algorithmus vielversprechende Ergebnisse. In nachfolgenden Studien muss der Algorithmus weiter präzisiert werden, um gegenüber anatomischen Variationen, kleinen Läsionen und Artefakten robuster zu werden, damit er in der klinischen Praxis unterstützend eingesetzt werden kann.

Abstract

Background:

The most common pathological changes in dentistry include osteolytic jaw lesions in the area of the root apex as a result of a bacterial, tooth-associated infection. Such periapical osteolytic lesions appear on X-ray as radiolucent areas of varying diameters. Reliable diagnosis is essential to avoid pain and complications ranging from abscess formation to tooth loss. Conventionally, this is done using 2D imaging techniques, but this technology is now being replaced by three-dimensional digital volume tomography (DVT). A detailed CBCT assessment is more time-consuming than 2D diagnostics, which is why the workload increases exponentially with the increasing number of CBCT examinations. This is where artificial intelligence can come in. Computer-aided diagnosis based on deep learning of CBCT data sets can reduce the workload of dentists and also lead to an increase in the quality of the diagnosis.

The aim of this work is to test a trained AI algorithm from a previous study (<https://doi.org/10.1016/j.joen.2022.07.013>) for non-inferiority.

Method:

The study includes 195 clinically representative data sets that were analysed at the University Clinic for Dentistry and Oral Health in Graz in accordance with the inclusion criteria. A team of experts represented the gold standard by consensus. A total of 2947 teeth were included in the study, 300 of which had periapical osteolytic lesions. The sensitivity and specificity of the AI algorithm was determined on the basis of the defined gold standard. To enable an index-based evaluation, each lesion was assigned an index according to its diameter.

Results:

The algorithm used achieved a sensitivity of 86.7 % and a specificity of 82.9 %. The non-inferiority test based on the proof-of-concept evaluation provided sufficient evidence to reject the null hypothesis for specificity, but not for sensitivity. If lesions of index score 1 (< 1mm) are excluded from the data set, the sensitivity increases to 90.4 %.

Conclusions:

Despite the clinically challenging data set, the AI algorithm used shows promising results. In subsequent studies, the algorithm needs to be further refined to become more robust against anatomical variations, small lesions and artefacts so that it can be used to support clinical practice.

Inhaltsverzeichnis

1	Abkürzungen und deren Erklärungen	1
2	Abbildungsverzeichnis	2
3	Tabellenverzeichnis	3
	Einleitung	4
3.1	<i>Künstliche Intelligenz</i>	4
3.1.1	Schlüsselaspekte	4
3.1.2	Funktionsweise der Künstlichen Intelligenz	5
3.1.3	Künstliche neuronale Netzwerke	6
3.1.4	Faltungsnetzwerke	7
3.2	<i>Künstliche Intelligenz und ihre Anwendung</i>	9
3.2.1	Künstliche Intelligenz im Alltag	9
3.2.2	Künstliche Intelligenz in der Medizin	10
3.2.3	Künstliche Intelligenz in der Radiologie	11
3.2.4	Künstliche Intelligenz in der Pathologie	12
3.2.5	Künstliche Intelligenz in der Arzneimittelforschung	13
3.2.6	Künstliche Intelligenz in der Zahnmedizin	14
3.2.7	Neuronale Netzwerke in der konservierenden Zahnheilkunde	15
3.2.8	Neuronale Netzwerke in der Endodontie.....	16
3.2.9	Neuronale Netzwerke in der oralen Chirurgie	19
3.3	<i>Künstliche Intelligenz: Gegenwart und Zukunft</i>	24
3.3.1	KI in der zahnärztlichen Ausbildung	24
3.3.2	KI im Patientenmanagement.....	26
3.3.3	Probleme im Zusammenhang mit der KI	26
3.3.4	Integration von KI in die klinische Versorgung.....	28
3.3.5	Ethik und Künstliche Intelligenz	29
3.3.6	Weitere ethische Prinzipien	31
3.4	<i>Die pathologische Veränderung „Periapikale osteolytische Läsion“</i>	33
3.4.1	Ätiologie.....	33
3.4.2	Indizes für die periapikale Läsion	33
3.4.3	Komplikationen.....	35
3.5	<i>Digitale Volumentomographie</i>	35
3.5.1	Technischer Hintergrund	35
3.5.2	Nachteil einer DVT	38
4	Methodik	39
4.1.1	Datensatz	39
4.1.2	Zielgrößen	41
4.1.3	Hauptzielgrößen	41
4.1.4	Nebenzielparameter	41
4.1.5	Statistische Auswertung	41
4.1.6	Fallzahlplanung	42

4.1.7	CNN-Architektur	43
5	Ergebnisse	45
5.1.1	Anzahl der ausgewerteten Datensätze	45
5.1.2	Verteilung der beurteilten Zähne	45
5.1.3	Verteilung nach Estrela.....	46
5.1.4	Diagnostik im Vergleich	48
5.1.5	Ergebnisse nach Position	49
5.1.6	Sensitivität und Spezifität	52
5.1.7	Auswertungszeit der KI.....	52
6	Diskussion.....	53
7	Conclusio	59
8	Literaturverzeichnis	60

1 Abkürzungen und deren Erklärungen

2D	<i>Zweidimensional</i>
3D	<i>Dreidimensional</i>
AR	<i>Augmented Reality</i>
cm	<i>Centimeter</i>
CNN	<i>Konvolutionelles neuronales Netz</i>
CT	<i>Computertomographie</i>
DSGVO	<i>Datenschutz-Grundverordnung</i>
DVT	<i>digitale Volumetomographie</i>
FOV	<i>Field-of-view</i>
FP	<i>Falsch-positiv</i>
KI	<i>Künstliche Intelligenz</i>
KIS	<i>Knochen-Implantat-Grenzfläche</i>
KNN	<i>Künstliche neuronale Netzwerke</i>
kV	<i>Kilovolt</i>
lp/mm	<i>Linienpaar pro Millimeter</i>
mA	<i>Milliampere</i>
MRT	<i>Magnetresonanztomographie</i>
OK	<i>Oberkiefer</i>
PALen	<i>Periapikale Osteolysen</i>
SCN	<i>SpatialConfiguration-Net</i>
Se	<i>Sensitivität</i>
Sp	<i>Spezifität</i>
Sv	<i>Sievert</i>
UK	<i>Unterkiefer</i>
VR	<i>Virtual Real</i>

2 Abbildungsverzeichnis

Abbildung 1: Schlüsselaspekte der künstlichen Intelligenz (2)	5
Abbildung 2: Schematische Darstellung der Funktionsweise von KI-Modellen. In diesem Beispiel soll die KI eine Katze im Eingabebild erkennen. Dazu erstellt das Modell eine Heatmap vom Eingabebild. Anhand von Pixeln die als Eingabevariable definiert wurden, kann die KI eine Vorhersage erstellen. Zur Unterscheidung von sicheren und relevanten Vorhersagekriterien konzentriert sich die KI in diesem Beispiel auf die Ohren und Nase der Katze. (4)	6
Abbildung 3: Darstellung eines tiefen KNN mit zwei versteckten Schichten (3)	7
Abbildung 4: Darstellung eines CNN mit einer Faltungsschicht, einer Pooling-Schicht und einer vollständig verknüpften Schicht. (3).....	9
Abbildung 5: lokale Tomographie (a) vs. globale Tomographie (b) (89).....	37
Abbildung 6: Schematische Darstellung der CNN-Architektur; SCN= Spatial-Konfigurationsnetzwerk (5).....	44
Abbildung 7: Schematische Darstellung des Zusammengeführten Algorithmus. SCN= Spatial-Konfigurationsnetzwerk (99)	44
Abbildung 8: Beispielhafte KI-Detektion von FPs in den Datensätzen.	56

3 Tabellenverzeichnis

Tabelle 1: Charakteristika der Studien, die Präzision wird mit der Formel: $\text{Präzision} = (\text{Sensitivität}) (\text{Prävalenz}) + (\text{Spezifität}) (1 - \text{Prävalenz})$ beschrieben (31).....	16
Tabelle 2: Charakteristika der Studien, SEN = Sensitivität, SPE = Spezifität und PRÄZ = Präzision, wobei die Präzision mit der Formel; $\text{Präzision} = (\text{Sensitivität}) (\text{Prävalenz}) + (\text{Spezifität}) (1 - \text{Prävalenz})$ beschrieben wird (31).....	19
Tabelle 3: Charakteristika der Studien, SEN = Sensitivität, SPE = Spezifität und PRÄZ = Präzision, wobei die Präzision mit der Formel; $\text{Präzision} = (\text{Sensitivität}) (\text{Prävalenz}) + (\text{Spezifität}) (1 - \text{Prävalenz})$ beschrieben wird (31).....	23
Tabelle 4: Verteilung der Datensätze; OK= Oberkiefer, UK= Unterkiefer.....	45
Tabelle 5: Verteilung der beurteilten Zähne	46
Tabelle 6: Verteilung der Läsionen nach Estrela et al. (85).....	47
Tabelle 7: Verteilung der Läsionen auf den Indexwert 1 für jede Zahngruppe (99)...	47
Tabelle 8: Verteilung der vorhergesagten Läsionen auf die genannten periapikale Indexwerte nach Estrela et al. (85). Blaue Felder stellen wahr-positive Vorhersagen dar, während rote Felder falsch-negative Vorhersagen für einen bestimmten periapikalen Index darstellen. (99)	48
Tabelle 9: Diagnose von Expertinnen und KI im Vergleich.....	48
Tabelle 10: Vorhersagen der Expertinnen im Vergleich zu der KI. 0 steht für keine Läsion und 1 für eine detektierte Läsion. Beim Ausbleiben eines Feldes haben die Expertinnen bzw. die KI den Zahn als fehlend erkannt.	49
Tabelle 11: Ergebnisse der Expertinnen und der KI im Vergleich nach Position.....	50
Tabelle 12: Ergebnisse der Expertinnen und der KI im Vergleich nach Kiefer	51
Tabelle 13: Sensitivität und Spezifität für alle vorhandenen Zahngruppen im OK und UK (99).....	51
Tabelle 14: Sensitivität und Spezifität der trainierten Künstlichen Intelligenz an vorhandenen und fehlenden Zähnen.....	52
Tabelle 15: Mittelwert der Anwendungszeit der KI mit ihrer Standardabweichung	52

Einleitung

3.1 Künstliche Intelligenz

3.1.1 Schlüsselaspekte

Künstliche Intelligenz (KI) beschreibt eine Technologie, die es einer Software ermöglicht, menschliche Intelligenz nachzuahmen und definierte Tätigkeiten auszuführen. Somit ist KI ein Oberbegriff und beschreibt sämtliche Technologien, die in der Lage sind, kognitive Fähigkeiten, wie das Lösen von Problemen oder Aufgaben, zu imitieren. Des Weiteren kann KI aus Merkmalen Lernen und den Ausführungsprozess selbständig präzisieren.

Maschinelles Lernen ist ein wichtiger Bestandteil der KI. Es basiert auf Algorithmen zur Vorhersage von Ergebnissen eingegebener Daten. Dadurch kann eine Software aus Daten lernen und neue Aufgaben ohne menschliches Eingreifen lösen.

Ein neuronales Netz ist ein Netzwerk von Algorithmen, die Signale mit Hilfe künstlicher Neuronen berechnen. Dieses neuronale Netz ist dem menschlichen Gehirn nachempfunden.

Deep Learning ist eine Kombination aus maschinellem Lernen und neuronalen Netzen. Es hat verschiedene Rechenschichten, die hintereinandergeschaltet und miteinander verbunden sind. Auf diese Weise können Eingabedaten analysiert werden.

Das Ziel von Deep Learning ist es, ein neuronales Netzwerk aufzubauen, welches automatisch Muster erkennt, um Merkmale besser zu identifizieren.

Deep Learning wird auch als konvolutionelles neuronales Netz (CNN; Faltungsnetzwerk) bezeichnet. Ein solcher Algorithmus wird verwendet, um Merkmale aus verschiedenen Schichten zu filtern. Dadurch können große und komplexe Bilder analysiert werden. (1, 2)

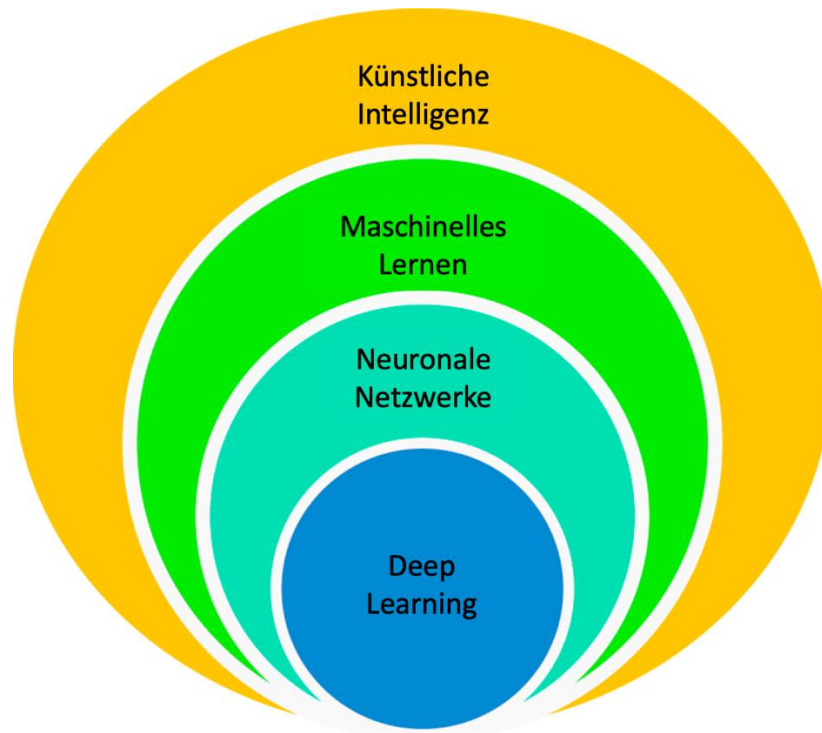


Abbildung 1: Schlüsselaspekte der künstlichen Intelligenz (2)

3.1.2 Funktionsweise der Künstlichen Intelligenz

Die KI-Entwicklung beschreibt zwei Phasen. In der ersten Phase wird die KI "trainiert" und in der zweiten Phase "getestet". Die Parameter des Modellsatzes werden durch die erstellten Trainingsdaten bestimmt. Retrospektiv verwendet das KI-Modell Daten aus früheren Beispielen, wie Patient*innen- oder Röntgendatensätze. Diese Parameter werden dann auf die Testdaten angewendet. Die KI nimmt einen Input und erstellt daraus eine sogenannte Heatmap, die eine Vorhersage liefert. Eine Heatmap ist als eine Wahrscheinlichkeit definiert, mit der sich Orientierungspunkte an einer bestimmten Position in einem Bild befinden. (3) Die erzeugte Heatmap visualisiert grafisch, welche Eingabevariablen (z.B. Pixel) die Vorhersage bestimmt hat. Auf diese Weise können sichere und relevante Vorhersagetechniken unterschieden werden (vgl. Abbildung 2). (4)

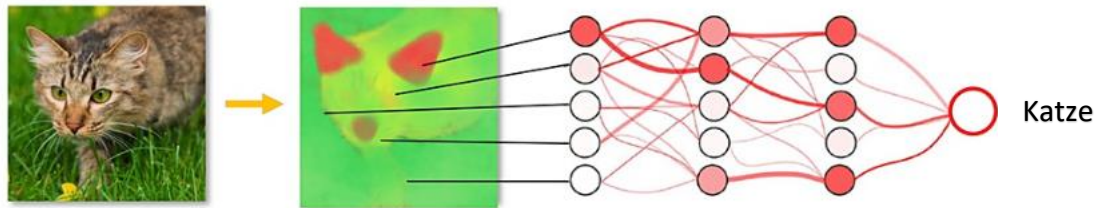


Abbildung 2: Schematische Darstellung der Funktionsweise von KI-Modellen. In diesem Beispiel soll die KI eine Katze im Eingabebild erkennen. Dazu erstellt das Modell eine Heatmap vom Eingabebild. Anhand von Pixeln die als Eingabevariable definiert wurden, kann die KI eine Vorhersage erstellen. Zur Unterscheidung von sicheren und relevanten Vorhersagekriterien konzentriert sich die KI in diesem Beispiel auf die Ohren und Nase der Katze. (4)

3.1.3 Künstliche neuronale Netzwerke

Wie bereits erwähnt sind künstliche neuronale Netzwerke (KNN) ein Teilgebiet des maschinellen Lernens, dessen Architektur von der Struktur des menschlichen Gehirns inspiriert ist. Die Grundeinheit des KNN ist ein künstliches Neuron, das eine mathematische Funktion eines biologischen Neurons darstellt. Um das Interagieren biologischer Neuronen zu simulieren, werden künstliche Neuronen in einer Reihe von Schichten angeordnet, wobei die erste Schicht (Eingabeschicht) die Information erhält, welche vom Netzwerk weiterverarbeitet werden soll. Jede nächste Schicht (versteckte Schicht) erhält Informationen von der vorgeschalteten Schicht, verarbeitet diese und leitet das Ergebnis an die nächste Schicht weiter. Dies wird bis zur letzten Schicht (Ausgabeschicht) weitergeführt. KNNs mit mehr als einer versteckten Schicht werden als deep KNNs beschrieben. Jedes künstliche Neuron hat Eingangs- und Ausgangsverbindungen, wobei jede Verbindung mit einem numerischen Code, dem sogenannten Gewicht, multipliziert wird. Das Gewicht ist ein lernbarer Wert, der den Einfluss eines Neurons auf ein anderes beeinflusst. Um eine Ausgabe zu erzeugen, berechnet jedes künstliche Neuron die Summe seiner gewichteten Eingangswerte, auf die eine nichtlineare Aktivierungsfunktion angewendet wird. Je nach Stärke der gewichteten Eingangssignale entscheidet die Aktivierungsfunktion, ob das Neuron aktiviert und als Ausgabe weitergeleitet wird. (3)

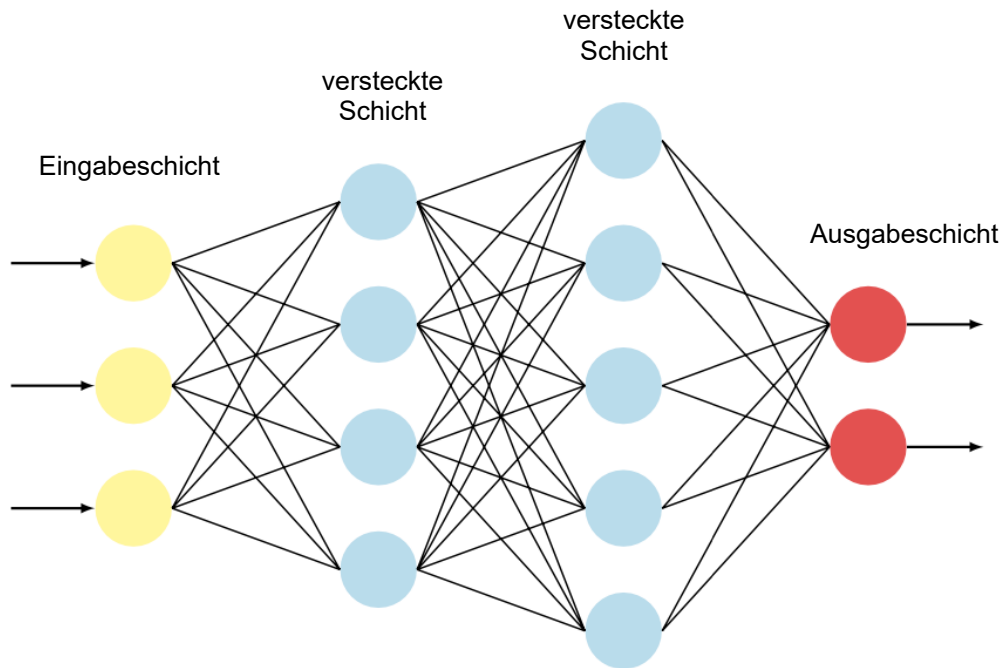


Abbildung 3: Darstellung eines deep KNN mit zwei versteckten Schichten (3)

3.1.4 Faltungsnetzwerke

CNNs sind Teil von künstlichen neuronalen Netzen (s.o.). Diese werden überwiegend für die Verarbeitung von Daten, die Gittermuster aufweisen, wie z.B. räumlichen Strukturen oder Bildern, entwickelt und bestehen aus mehreren Schichten künstlicher Neuronen. (5) Die Schichten werden in drei Haupttypen unterschieden. Dazu gehören die Faltungsschicht, die Pooling-Schicht und die vollständig verknüpfte Schicht (vgl. Abbildung 4). In jeder Faltungsschicht wird ein Filter auf jedes Bildpixel gelegt. Ein Filter ist eine kleine 3x3 Pixel große Matrix aus Gewichten, die dazu dient, spezielle Merkmale in Bildern zu finden. Die Größe des Filters bestimmt die Größe des rezeptiven Feldes, also des Bildbereiches, auf den ein Filter geschaltet wird. Der Filter bewegt sich über das Bild, wobei das Punktprodukt zwischen den Filterwerten und den Bildpixeln berechnet wird. Dieser Vorgang wiederholt sich, bis das Bild vollständig betrachtet wurde. Dabei wird von einer Faltung gesprochen. Die endgültige Ausgabe wird als Feature-Map bezeichnet.

Die Pooling-Schicht befindet sich hinter der Faltungsschicht und dient dazu, Merkmale zu extrahieren und somit die Komplexität des Netzes zu verringern. (5) Dabei wird auch die Rechenkomplexität verringert. Dazu wird die räumliche Größe der Feature-Maps reduziert. Dieser Vorgang wird als Downsampling bezeichnet. Die verkleinerten

Feature-Maps stellen die Eingabe für die nächstniedrigere Ebene dar. In der letzten Faltungsschicht jeder Ebene werden die Heatmaps um den entsprechenden Faktor erhöht und zu den Ausgaben der nächsthöheren Ebene hinzugefügt, bis die ursprüngliche Auflösung erreicht ist. Dies wird als Upsampling bezeichnet (vgl. Abbildung 4).

Die vollständig verbundene Schicht ist am Ende der Netzwerkarchitektur angebracht. Diese klassifiziert die extrahierten Merkmale und erzeugt daraus die Ausgaben. Dazu werden Wahrscheinlichkeitswerte zwischen 0 und 1 generiert.

Dieser Algorithmus macht CNNs für die Bildverarbeitung hoch effizient, da ein Merkmal im gesamten Bild auftreten kann und jede Schicht ihren Output in die nächsthöhere Schicht einspeist und somit Merkmale hierarchisch und progressiv komplexer werden. (5)

Dementsprechend dienen die ersten Schichten zur Erkennung von einfachen Merkmalen wie z.B. Kanten oder Linien. Tiefe Schichten der Netzarchitektur sind in der Lage, komplexe Formen und bestimmte Objekte zu definieren, wie z.B. Zähne und Läsionen. (3, 6)

Beispiele für CNNs sind das Räumliche-Konfigurationsnetzwerk (SpatialConfiguration-Net = SCN) und das U-Netzwerk. Das SCN ist ein Landmarken-Lokalisierungsnetz, bestehend aus zwei Komponenten, dem lokalen Erscheinungsbild und der räumlichen Konfiguration. Es wurde entwickelt, um Heatmap-Bilder in einer konsistenten Weise darzustellen. Die jeweiligen Landmarken des lokalen Erscheinungsbildes und der räumlichen Konfiguration werden zusammengeführt. Daraus ergibt sich die endgültige Heatmap. Durch diese pixelweise Multiplikation werden falsch-positive Vorhersagen aus dem lokalen Erscheinungsbild eliminiert, was eine Stabilisierung gegenüber der fehlerhaften Identifizierung von Orientierungspunkten gewährleistet. (3, 6, 7)

Die entsprechend den vorhergesagten Koordinaten beschnittenen Heatmap-Bilder werden in ein modifiziertes U-Netzwerk eingespeist. Dieses Netzwerk kann über Faltungsoperationen, die Downsampling- und Upsamplingschichten enthalten, die finale Vorhersage erzeugen. (6, 8)

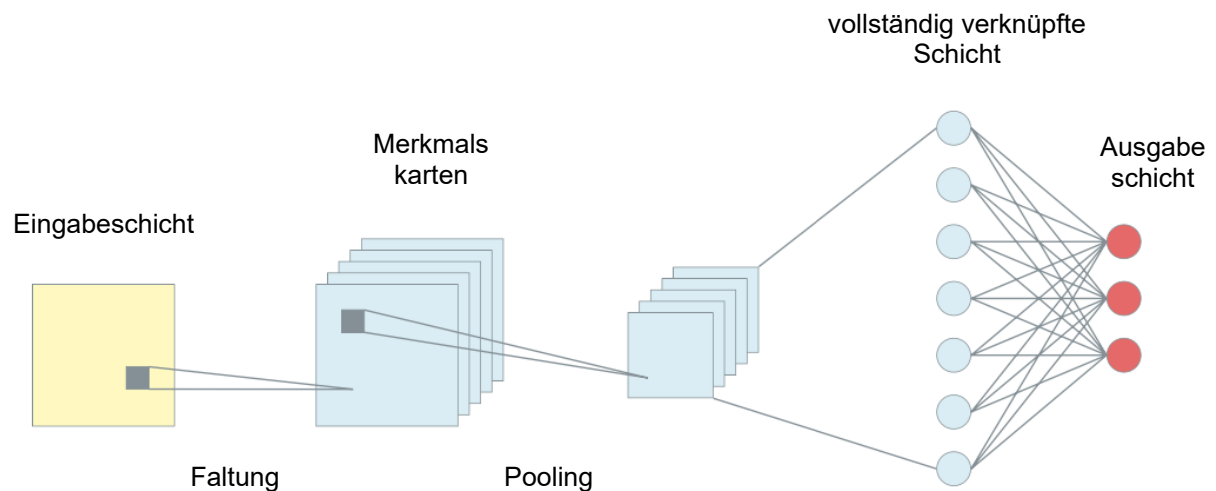


Abbildung 4: Darstellung eines CNN mit einer Faltungsschicht, einer Pooling-Schicht und einer vollständig verknüpften Schicht. (3)

3.2 Künstliche Intelligenz und ihre Anwendung

3.2.1 Künstliche Intelligenz im Alltag

KI kann in vielen Bereichen des täglichen Lebens eingesetzt werden. Nicht immer sind KI-Algorithmen wahrnehmbar. Oft laufen KI-basierte Programme im Hintergrund und wirken nicht direkt auf die Nutzer*innen ein. Zum Beispiel wird KI verwendet, um Empfehlungsmaschinen für Kund*innen im Internet zu erstellen. Dabei werden Daten aus dem Browserverlauf gespeichert und analysiert, um ein Interessenprofil zu erstellen. So können produktplatzierte Werbungen oder Angebote auf Kund*innen abgestimmt werden. (9)

Des Weiteren werden KI-Modelle auch zur Gewährleistung von Datenschutz eingesetzt. Neue Generationen von Mobiltelefonen und Computern nutzen Gesichtserkennungstechniken, um einen sicheren Zugang auf mobile Geräte zu ermöglichen. Abgesehen von persönlicher Nutzung, ist die Gesichtserkennung auch in Hochsicherheitsbereichen eine weit verbreitete Anwendung von KI-Modellen. (9)

Mit der Entwicklung von KI-Modellen in der Automobilindustrie wird das autonome Fahren zu einem präsenten Thema. Automobilhersteller nutzen maschinelles Lernen, um Computer zu entwickeln, die während der Fahrt die Umgebung analysieren, Objekte und Fahrspuren erkennen und Unfällen präventiv entgegenwirken können. Auch das Nutzen von intelligenter Navigation kann Nutzer*innen genaue, zeitnahe und

detaillierte Informationen über den Straßenverkehr zur Verbesserung der Sicherheit liefern. Ebenso können beim Design neuer Infrastruktur KI-basierte Verkehrsmanagementtechnologien eingesetzt werden und als Grundlage für die nächste Generation intelligenter Städte dienen. (9)

Doch KI kann auch das Gegenteil von nützlich oder unterstützend sein. KI-generierte Fake-News beziehen sich aktuell nicht nur auf schriftliche Berichterstattung, sondern beinhalten auch authentische Bilder, die auf Knopfdruck generiert werden können. Solche Fehlinformationen können durch KI personalisiert und individuell auf Religion, Geschlecht oder politische Haltung abgestimmt werden. Dadurch werden KI-generierte Fake-News in den Sozialen Medien noch schneller verbreitet. Verlässliche Anhaltspunkte zur Erkennung von medialen Fake-News gibt es mittlerweile kaum noch. Sprache, Rechtschreibung und Formulierungen von KI-generierten Berichten sind aktuell weitestgehend denen des Menschen entsprechend. Auch das unabhängige Recherchieren kann zukünftig kein eindeutiges Ausschlusskriterium mehr sein, da KI-Algorithmen eigenständig Webseiten und Quellen erstellen und intelligent platzieren können. (10)

3.2.2 Künstliche Intelligenz in der Medizin

Der technologische Fortschritt hat entscheidende Aspekte unserer Gesellschaft und Industrie, darunter auch des Gesundheitswesens, verändert. Innovative digitale Technologien wie Computer Aided Design, Rapid Prototyping, Augmented/Virtual Reality und "Omics"-Analyse werden vermehrt in verschiedenen human- und zahnmedizinischen Disziplinen für diagnostische und therapeutische Zwecke eingesetzt. Der Begriff Künstlicher Intelligenz wurde in den 1950er Jahre etabliert und bezog sich auf die Idee, Maschinen zu bauen, die das menschliche Verhalten imitieren können, um komplexe Aufgaben zu erfüllen. Seit den 1950er Jahren gab es zwei "KI-Winter" in den 1970er Jahren und den späten 1980er Jahren. In dieser Zeit konnten die Erwartungen nicht mit den potenziellen Ergebnissen mithalten auf Grund mangelnder Rechenleistung und Ressourcen. In den späten 2000er Jahren förderte das Aufkommen von modernen und leistungsfähigen Grafikprozessoren den Enthusiasmus für die Entwicklung verschiedener KI-Modelle in den Sektoren autonomes Fahren, maschinelles Lernen, Computer Vision und Intelligente Robotik. (11) Dank jüngster Fortschritte in der digitalen Datenerfassung, dem maschinellen

Lernen und der Erweiterung der Computerinfrastruktur, können die Anwendungsbereiche von KI in der Medizin ausgeweitet werden. KI dringt in Bereiche vor, von denen man glaubte, dass sie nur von Expert*innen abgedeckt werden könnten. (12) Frühere KI-Modelle erbrachten eher unterdurchschnittliche Leistungen. Doch neue Algorithmen sind nun in der Lage, bei problemspezifischen Anwendungen mit dem Menschen gleichzuziehen und ihn sogar zu übertreffen. Zu verdanken ist dies den riesigen Mengen an digitalen Daten, die den Algorithmen zum Training zur Verfügung stehen. Im Gesundheitswesen wird KI in vielen Subbereichen zu einem wichtigen Bestandteil. Darunter fallen zum Beispiel Diagnostik und Bildgebung, Pathologie, Arzneimittelforschung, Fernüberwachung von Patienten, Risikomanagement und virtuelle Assistenten und Krankenhaus- bzw. Praxismanagement. (13)

3.2.3 Künstliche Intelligenz in der Radiologie

Dank Deep Learning haben die Bilderkennungsalgorithmen große Fortschritte gemacht, wodurch radiologische Daten effizienter interpretiert werden können. Das Hauptziel dabei ist eine höhere Effektivität und Effizienz in der klinischen Diagnostik. (14) Im radiologischen Alltag werden medizinische Bilder üblicherweise von Expert*innen manuell ausgewertet, um Befunde zu erkennen, einzuordnen und zu kontrollieren. Eine solche Beurteilung beruht auf Erfahrung und einer entsprechenden Ausbildung. Dadurch erhält sie ein gewisses Maß an Subjektivität. KI-Modelle sind in der Lage, komplexe Muster automatisch in bildgebenden Daten zu erkennen und liefern objektive statt subjektive Bewertungen von radiologischen Merkmalen. Vor allem raschere, genauere und reproduzierbarere Ergebnisse könnten in der Radiologie erzielt werden, wenn die KI als Hilfsmittel herangezogen und in den klinischen Alltag integriert würde. (13) So kann z.B. die Früherkennung von Lungenkrebs unterstützt werden: KI erkennt pulmonale Raumforderungen automatisch und kann diese als benigne oder maligne einstufen. Eine Früherkennung von malignen Läsionen ist u.U. prognosebestimmend und kann somit für Patienten lebensrettend sein. Zusätzlich zeigen Studien über das nicht-kleinzellige Lungenkarzinom, dass die KI in der Lage ist, Aussagen zur Prognostizierung von Fernmetastasen und histologischen Tumorsubtypen zu treffen. Zusätzlich können KI-Algorithmen Krankheitsrückfälle und somatische Mutationen erkennen, sowie bei der Erstellung von

Genexpressionsprofilen und bei der Prognose der Gesamtüberlebenszeit unterstützen. Auch beim Screening von Mammographien kann die KI als Hilfsmittel herangezogen werden. Sie kann Mikroverkalkungen in der Brust erkennen und charakterisieren. (15-19)

Ein weiteres Beispiel für den Einsatz von KI in der Radiologie ist die Bildgebung des Abdomens und des Beckens. Durch die Entwicklung der medizinischen Bildgebung, besonders im Bereich der Computertomographie (CT) und der Magnetresonanztomographie (MRT), werden immer mehr Zufallsbefunde gestellt, darunter auch Leberkarzinome. Auch hier kann die KI die Läsionen in benigne oder maligne unterscheiden und bei Patienten eine Priorisierung der Therapie vornehmen. Aufgrund der anhaltenden Entwicklung im Bereich der radiologischen Bildgebung in Bezug auf Qualität, Empfindlichkeit und Auflösung, können kleinste Unterschiede in der Gewebedichte erkannt und dokumentiert werden. Solche Merkmale sind für KI-Methoden, die in der Klinik eingesetzt werden, schwer zu erkennen. Somit sind Sie noch nicht auf dem Stand des technisch Möglichen. Sie dienen aber als Motivation, diesen Paradigmenwechsel hin zu leistungsfähigen KI-Werkzeugen zu verstehen und zu fördern. (13)

3.2.4 Künstliche Intelligenz in der Pathologie

In der modernen klinischen Praxis spielt die digitale Pathologie eine entscheidende Rolle und wird damit zu einer technologischen Voraussetzung für eine moderne Laborumgebung. Dank Ganz-Objektträger-Bildgebungen und dem technologischen Fortschritt können digitale Objektträgerbilder besser verwaltet werden. Zusätzlich stehen die Bilder der Telepathologie und zum klinischen Austausch zur Verfügung. Auch hier können die Fortschritte der KI, ähnlich wie in der Radiologie, für Bildanalysen genutzt werden. Die Synergie von KI und digitaler Pathologie ermöglicht eine rasche und effiziente bildbasierte Diagnostik. (20) Patholog*innen sind bereits in der Lage, eindeutige Marker zu identifizieren, die mit Krankheitsprozessen assoziiert sind, um die Früherkennung von Pathologien und die Prognose zu verbessern und die geeignetsten Behandlungsmöglichkeiten auszuwählen. Dadurch können in der Pathologie mehr pathologische Anomalien diagnostiziert werden, während gleichzeitig die diagnostische und prognostische Genauigkeit beibehalten wird. Nachteil einer KI-basierten histologischen Diagnose ist es, dass nicht nachvollzogen werden kann, wie

genau die Entscheidungen getroffen wurden. Diese werden als „Black-Boxes“ beschrieben. Auch wenn die KI basierte Algorithmen Vorteile bieten, müssen in Zukunft klinische, rechtliche und regulatorische Fragen geklärt werden. Gleichzeitig müssen die Algorithmen für den Menschen leichter interpretierbar werden, um einen besseren Einblick in ihre Funktionsweise zu erhalten. Es ist daher sinnvoll, die spezifischen Merkmale anzugeben, auf die der Algorithmus fokussiert, um eine bessere Verständlichkeit zu gewährleisten. (21)

3.2.5 Künstliche Intelligenz in der Arzneimittelforschung

Die Erforschung und Entwicklung von Arzneimitteln ist ein wichtiger Forschungszweig für Pharmaunternehmen und Chemiker*innen. Geringe Effizienz, Off-Target Interaktionen, Zeitaufwand und hohe Kosten stellen große Hindernisse und Herausforderungen dar, die die Entwicklung von Medikamenten erheblich beeinträchtigen und den Fortschritt verlangsamen können. Darüber hinaus schaffen große Datenmengen aus Genomik, Proteomik, Mikroarray und klinischen Studien weitere Hürden für die Arzneimittelforschung. KI und maschinelles Lernen können die Erforschung und Entwicklung von Arzneimitteln positiv beeinflussen. Dabei werden sie z. B. in der Peptidsynthese, dem strukturbasierten und ligandenbasierten visuellen Screening, der Toxizitätsvorhersage, der Arzneimittelüberwachung und -freigabe, der Neupositionierung von Arzneimitteln und der physiochemischen Aktivität eingesetzt. (22-24) Darüber hinaus haben neue Techniken der Datenerhebung, Kuratierung und des Datenmanagements die neu entwickelten Modellierungsalgorithmen entscheidend unterstützt. Dank der KI wird die Suche nach Verbindungen in chemischen Bibliotheken, die mehr als 106 Millionen Verbindungen umfassen, einfacher und zeitsparender. Zusätzlich eliminieren KI-Modelle die Toxizitätsgrenze, die aufgrund von Off-Target Interaktionen entsteht. (25)

Somit bietet KI die Möglichkeit, die Kosten zu senken, die Effizienz zu steigern und während des Entdeckungs- und Entwicklungsprozesses Zeit zu sparen. Die zunehmende Datendigitalisierung in den Pharmaunternehmen und im Gesundheitswesen motiviert die Implementierung von KI, um die Probleme bei der Prüfung komplexer Daten zu überwinden. Doch die Entwicklung von KI in der Arzneimittelforschung steht noch vor einigen Hürden. Dazu gehört z.B. eine qualitativ hochwertige Datenerfassung, bei der es zwei entscheidende Probleme gibt. Zum einen

kann die Kennzeichnung nicht binär sein, da die Wirkung eines Arzneimittels auf den menschlichen Organismus wesentlich komplexer ist. Zum anderen ist die Menge der bei der Arzneimittelforschung verfügbaren Daten deutlich geringer im Vergleich zu der enormen Menge an global vorhandenen Informationen. Daher ist eine interdisziplinäre Zusammenarbeit erforderlich, um nicht nur die Quantität, sondern auch die Qualität der Daten zu verbessern. (24)

3.2.6 Künstliche Intelligenz in der Zahnmedizin

Die Röntgenuntersuchung ist ein elementarer Bestandteil der Diagnose und Behandlung der meisten Zahn-, Mund und Kiefererkrankungen, sodass in den Archivierungssystemen und Datenbanken der Krankenhäuser und Kliniken eine große Menge digitaler Röntgendatensätze gespeichert wird. Daher basieren die meisten aktuellen KI-Modelle auf Röntgenbildern, um Zahnärzt*innen bei der Diagnose, Therapie und Prognose verschiedener Zahn- und Kiefererkrankungen zu unterstützen. Dreidimensionale (3D) Bildgebungsverfahren wie die digitale Volumentomographie (DVT) und die CT sind die am häufigsten verwendeten 3D-Bildgebungsmodalitäten in zahnmedizinischen Fachgebieten, insbesondere in der Mund-, Kiefer- und Gesichtschirurgie. Im Vergleich zur konventionellen zweidimensionalen (2D)-Röntgenaufnahmen (d. h. Intraorales Kleinbildröntgen, Bissflügel-, Panoramaröntgen) ermöglicht die DVT/CT die Visualisierung anatomischer Strukturen und Pathologien in 3D und bietet damit nicht nur eine höhere diagnostische Genauigkeit, sondern auch detailliertere Informationen über die tatsächliche Morphologie, das Volumen und die Lokalisation von Läsionen. DVT hat sich in der zahnmedizinischen Praxis etabliert und ist eine ideale Datenquelle für die Entwicklung klinisch anwendbarer KI-Modelle zur Verbesserung der Genauigkeit und Steigerung der Effizienz der Diagnose und Behandlung von Zahn-, Mund- und Kiefererkrankungen. Seit Mitte der 2010er Jahre wurde ein starker Anstieg in der Anzahl der KI-Modelle dokumentiert, die auf Basis von DVT-Datensätzen für zahnmedizinisch-maxillofaziale Anwendungen entwickelt wurden. (11) Reviews von Khanagar et al. (2) und Schwendicke et al. (26) haben in der Vergangenheit gezeigt, dass die Forschung im Bereich der KI in der Zahnmedizin voranschreitet.

3.2.7 Neuronale Netzwerke in der konservierenden Zahnheilkunde

Karies ist eine der häufigsten Zahnerkrankungen, was die frühzeitige Entdeckung dieser pathologischen Veränderung sehr wichtig macht. Standardisiert werden zur klinischen Diagnose Zahnsonden eingesetzt. Diese Methode setzt ein hohes Maß an Erfahrung voraus, weswegen sie sehr subjektiv ist. Besonders im Approximalbereich ist eine kariöse Läsion schwer zu beurteilen. (27) Die Verwendung von Röntgenbildern zur Kariesdiagnose hat sich im Laufe der Jahre zu einem Routinewerkzeug entwickelt. Dabei werden für das Karies-Screening am häufigsten Bissflügel-, intraorale Kleinbildröntgenaufnahmen und Panoramarröntgenaufnahmen verwendet. Dreidimensionale Röntgenaufnahmen werden hierbei seltener eingesetzt. Die Karieserkennung in radiologischen Datensätzen kann durch neuronale Netzwerke unterstützt werden. Dadurch kann die Untersuchung schneller und präziser durchgeführt werden. (28) Ein Beispiel eines entwickelten neuronalen Netzwerkes zur Kariesdiagnose wird in der Studie von Geetha et al. (29) (vgl. Tabelle 1) beschrieben. Das hier verwendete KI-Modell wies eine Präzision von 97,1 % auf, wobei die Rate der falsch positiven Ergebnisse bei 2,8 % lag.

In der konservierenden Zahnheilkunde kann KI auch unterstützend bei der Planung von Zahnrestorationen und Kavitätenpräparationstechniken eingesetzt werden. In der Studie von Javed et al. (30) (vgl. Tabelle 1) wurde gezeigt, dass der Ausbreitungsgrad der Karies prognostiziert werden konnte. Hierbei wurden neuronale Netzwerke verwendet, um auf Grundlage eines Prä-*Streptococcus mutans* die Aktivität eines Post-*Streptococcus mutans* vor der Kariesexkavation vorherzusagen. Anhand dieser Ergebnisse kann eine patientenindividuelle und suffiziente Kariesexkavation bestimmt werden. Die Präzision der verwendeten KI betrug 99,03 % und wurde mikrobiologisch bestätigt. Eine suffiziente Kariesexkavation verhindert eine erneute Exkavation und reduziert das Risiko eines Traumas der Pulpa.

<i>Studie</i>	<i>Art der Daten</i>	<i>Anzahl der Datensätze</i>	<i>der Präzision</i>
<i>Geetha et al. (29)</i>	<i>Periapikale Röntgenbilder</i>	<i>105 Bilder</i>	<i>97,1%</i>
<i>Javed et al. (30)</i>	<i>Molare</i>	<i>45 Zähne</i>	<i>99,03%</i>

Tabelle 1: Charakteristika der Studien, die Präzision wird mit der Formel: Präzision = (Sensitivität) (Prävalenz) + (Spezifität) (1 - Prävalenz) beschrieben (31)

3.2.8 Neuronale Netzwerke in der Endodontie

Auch in der Endodontie gewinnt die KI zunehmend an Bedeutung. Sie kann bei der Detektion von periapikalen Läsionen und Wurzelfrakturen, bei der Bewertung der Anatomie des Wurzelkanalsystems, bei der Bestimmung der Arbeitslänge und bei der Erfolgsprognose eingesetzt werden. (32) In einer Studie von Saghiri et al. (33) wurde der Ansatz entwickelt, das Foramen apikale mithilfe von neuronalen Netzwerken auf Grundlage von Merkmalsextraktionen aus Röntgenbildern zu lokalisieren und darzustellen. Anschließend wurden endodontische Feilen verwendet, um die Länge der Kanäle auf den dargestellten Röntgenbildern mit und ohne KI zu bestimmen. Kontrolliert wurden die Ergebnisse mithilfe von Stereomikroskopie. Die Ergebnisse zeigten, dass in 76 % der Fälle konventionell die Länge des Kanals richtig bestimmt wurde. Unter Verwendung der KI konnten 96 % der Foramen apikale richtig lokalisiert werden (vgl. Tabelle 2). Allerdings war die Qualität des Datensatzes gering, da es sich um extrahierte, nicht erhaltungswürdige Zähne handelte. Die Bestimmung der Arbeitslänge ist eine der entscheidenden klinischen Phasen, die das Ergebnis der Wurzelkanalbehandlung beeinflussen. Durch eine genaue Lokalisierung kann das Risiko verringert werden, den Kanal unzureichend zu reinigen. Zusätzlich kann gewährleistet werden, dass das Wurzelfüllmaterial im Kanal verbleibt und nicht in das periapikale Gewebe eindringt, was zu einer schlechteren Prognose der Behandlung führen könnte. Dies zeigt, dass neuronale Netzwerke eingesetzt werden können, um die Lokalisation des Foramen apikale besser zu bestimmen. (34)

Wie erwähnt kann KI auch zur Erkennung von periapikalen Läsionen verwendet werden. Eine periapikale Läsion kann beispielsweise durch eine apikale Parodontitis entstehen. Die apikale Parodontitis ist eine Entzündung, die hauptsächlich durch einen

bakteriellen Entzündungsprozess im Wurzelkanalsystem ausgelöst wird. Detektiert werden solche Läsionen durch Röntgenbilder, da sich diese Läsionen als periapikale Transluzenz darstellen. Dafür wird in den meisten Fällen ein zweidimensionales Panoramaröntgenbild oder eine dreidimensionale digitale Volumentomographie eingesetzt. Hier kann ein deep Learning Algorithmus eingesetzt werden, um periapikale Läsionen in DVT-Datensätzen zu detektieren. Setzer et al. (35) (vgl. Tabelle 2) erzielten dabei eine Sensitivität von 93 % und eine Spezifität von 88 %. Sie beschrieben einen KI-Algorithmus, der zur Erkennung von periapikalen Läsionen trainiert wurde. Die Studie enthielt jedoch eine Limitation, da die Leistungen der trainierten KI mit der Segmentierung von Kliniker*innen verglichen wurden, welche auf subjektiven Fehlern beruhen kann. Eine weitere Limitation bestand in den niedrigen Dice-Index-Verhältnissen für die Segmentierung der Läsionen, die durch eine Erhöhung der Trainingsgröße behoben werden müssten. (34) Auch Kirnbauer et al. (6) (vgl. Tabelle 2) untersuchten die Wirksamkeit von deep Learning Algorithmen zur automatischen Erkennung von periapikalen Läsionen in DVT-Datensätzen. Dabei wurden 144 Datensätze in die Studie eingeschleust. Es wurde eine Sensitivität von 97,1 % und eine Spezifität von 88,0 % erreicht.

Orhan et al. (36) (vgl. Tabelle 2) verwendeten ebenfalls deep Learning Algorithmen, um periapikale Läsionen in DVT-Datensätzen zu detektieren. Mithilfe von Faltungsnetzwerken konnten 142 von 153 Läsionen erkannt werden (Präzision = 92,8 %). Neuronale Faltungsnetzwerke sind eine spezielle Art von künstlichen neuronalen Netzen, die zur Erkennung von Merkmalen in Bildern verwendet werden. Darüber hinaus zeigt diese Studie, dass die KI in der Lage ist, Zähne korrekt entsprechend ihrer anatomischen Position einzuordnen. Nur ein Zahn wurde in der Studie falsch nummeriert. Die Studie zeigt, dass das Vorhandensein von Paro-Endo-Läsionen und parodontalen Defekten das Ergebnis negativ beeinflussen kann. Zusätzlich müssen anatomische Strukturen, wie z.B. das Foramen mentale, segmentiert werden, was sich auf die Analyse der Messung der KI-Modelle auswirken kann. (34)

In der Studie von Ekert et al. (37) (vgl. Tabelle 2) wurden 85 Panoramaröntgen mithilfe von neuronalen Faltungsnetzwerken auf periapikale Läsionen untersucht. Dabei wurde erkannt, dass die Position der Zähne Einfluss auf das Ergebnis des Screenings hatte und verschiedene Zahntypen unterschiedlich schwer zu beurteilen waren. Dies ist auf den Entstehungsprozess des zweidimensionalen Röntgenbildes zurückzuführen. Aus diesem Grund kann die Diagnose unsicher sein und bedarf

genauerer Präzision bei der Bildgebung. Bei Molaren war die Sensitivität (87 %) höher als bei anderen Zähnen, während die Spezifität (84 %) geringer war.

Des Weiteren kann KI eingesetzt werden, um vertikale Wurzelfrakturen zu lokalisieren. Dies sind Risse, die klinisch vollständig oder unvollständig erscheinen können und bei endodontisch behandelten oder unbehandelten Zähnen auftreten können. Wurzelfrakturen sind jedoch bei endodontisch vorbehandelten Zähnen häufiger. Solche Frakturen sind schwer zu identifizieren. Meist werden sie erst wahrgenommen, wenn signifikante periapikale Veränderungen auftreten. Um die diagnostische Effizienz von Kliniker*innen verbessern zu können, kann KI unterstützend eingesetzt werden. Kositbowornchai et al. (38) (vgl. Tabelle 2) entwickelten ein Modell zur Diagnose von vertikalen Wurzelfrakturen und dokumentierten hervorragende Leistungen.

Zur Vorhersage der Lebensfähigkeit von Stammzellen kann KI ebenfalls eingesetzt werden. Bindal et al. (39) evaluierten aus Zahnpulpa isolierte dentale Stammzellen in verschiedenen regenerativen Therapien mit Hilfe des Neuro-Fuzzy-Inferenzsystems. Dabei wurden sowohl neuronale Netze als auch Fuzzy-Logik verwendet, um komplexe Probleme zu lösen und Muster in Daten zu analysieren. Das Ergebnis stellt eine Hybridansatz dar, der die Stärken beider Systeme kombiniert. In einem simulierten klinischen Szenario konnte das KI-Modell die Lebensfähigkeit von Stammzellen nach einer bakteriellen Infektion mit Lipopolysacchariden vorhersagen.

Die genannten Modelle könnten als ergänzendes Instrument in der Praxis eingesetzt werden, um den klinischen Entscheidungsprozess zu beschleunigen und die Effektivität von Behandlungen zu verbessern. In den meisten Studien wurden die KI-Algorithmen jedoch unter Verwendung einer geringen Anzahl an Datensätzen für Training und Auswertung entwickelt. Die gesammelten Daten wurden hauptsächlich an einer einzigen Klinik und von einem Röntgengerät ermittelt. Deswegen können diese Ergebnisse aufgrund der fehlenden Heterogenität nicht verallgemeinert werden. Um diese Einschränkungen zu überwinden, müssen zukünftige Studien mehr Daten auswerten und interdisziplinäre Daten austauschen, um noch bessere und genauere Ergebnisse zu erhalten. (34)

<i>Studie</i>	<i>Art der Daten</i>	<i>Anzahl der Ergebnisse</i>	<i>Datensätze</i>
<i>Saghiri et al. (33)</i>	<i>Kleinbilder</i>	<i>50 Bilder</i>	<i>PRÄZ= 96 %</i>
<i>Setzer et al. (35)</i>	<i>DVT-Datensätze</i>	<i>20 Bilder</i>	<i>SEN= 93,0 % SPE= 88,0 %</i>
<i>Orhan et al. (36)</i>	<i>DVT-Datensätze</i>	<i>153 Bilder</i>	<i>SEN= 92,8 %</i>
<i>Ekert et al. (37)</i>	<i>Panoramaröntgenbilder</i>	<i>85 Bilder</i>	<i>SEN= 65 % SPE= 87 %</i>
<i>Zheng et al. (40)</i>	<i>DVT-Datensätze</i>	<i>50 Bilder</i>	<i>SPE= 90.0 %</i>
<i>Kositbowornchai et al. (38)</i>	<i>Kleinbilder</i>	<i>200 Bilder</i>	<i>SEN= 98 % SPE= 90,5 %</i>
<i>Kirnbauer et al. (6)</i>	<i>DVT-Datensätze</i>	<i>144 Bilder</i>	<i>SEN= 97,1 % SPE= 88,0 %</i>

Tabelle 2: Charakteristika der Studien, SEN = Sensitivität, SPE = Spezifität und PRÄZ = Präzision, wobei die Präzision mit der Formel; Präzision = (Sensitivität) (Prävalenz) + (Spezifität) (1 - Prävalenz) beschrieben wird (31)

3.2.9 Neuronale Netzwerke in der oralen Chirurgie

Die Extraktion des dritten Molaren ist einer der häufigsten chirurgischen Eingriffe in der Zahnmedizin. Dabei können Komplikationen wie die Parästhesie des Nervus alveolaris inferior entstehen, da dieser Nerv anatomisch in unmittelbarer Nähe des Zahnes liegt. Die Studie von Kim et al. (41) (vgl. Tabelle 3) untersuchte, ob KI verwendet werden kann, um vorherzusagen, ob die Extraktion eines unteren Weisheitszahnes zu Parästhesie führen kann. Dabei wurden 300 präoperative, zweidimensionale Panoramaröntgenaufnahmen von Patient*innen untersucht, bei denen eine Weisheitszahnentfernung geplant war. Hier betrug die durchschnittliche Genauigkeit 82,7 %. Dies zeigt, dass Faltungsnetzwerke bei der Vorhersage von Parästhesien des Nervus alveolaris inferior verwendet werden könnten. Doch wurden bei der Studie etwa 18 % falsch-positive und 16 % falsch-negative Ergebnisse festgestellt. Aus diesem Grund führen die Autoren an, dass die Ergebnisse dieser Studie, auf Grundlage von zweidimensionalen Panoramaröntgenaufnahmen, keine absoluten Kriterien für die Vorhersage von Parästhesien darstellen. Es bedarf weiterer

Forschung, um diese Technik als zuverlässiges Mittel für die Vorhersage verwenden zu können. (31)

Des Weiteren findet KI auch in der Differenzierung von Läsionen im Kieferbereich Anwendung. Das Ameloblastom und die odontogene Keratozyste sind klinisch häufig vorkommende benigne Läsionen, welche sich ubiquitär im Kiefer manifestieren können. Prädilektionsstellen sind der Ramus ascendens und der Corpus mandibulae. (42) Aufgrund ihrer Unterschiede im biologischen Verhalten, liegen zwei verschiedene Behandlungsstrategien vor. Dementsprechend ist es für die Prognose der Behandlung wichtig, die Krankheit richtig zu diagnostizieren und zu unterscheiden. Wegen der Notwendigkeit einer schnelleren und präziseren Diagnose von Kieferzysten und -tumoren wurden in mehreren Studien KI-Modelle, insbesondere Deep-Learning-Modelle, auf 2D-Panoramaröntgenbildern zur automatischen Diagnose von Zysten und Tumoren entwickelt. In der Studie von Liu et al. (43) (vgl. Tabelle 3) wurden zweidimensionale Panoramaröntgenbilder verwendet, um die benignen Tumore zu untersuchen. Um ein Ameloblastom von einer odontogenen Keratozyste zu differenzieren, wurde ein auf einem Transfer-Learning-Algorithmus basierendes Faltungsnetzwerk verwendet. Dieser beruht auf einem schon bestehenden Algorithmus, welcher auf die neuen Daten angepasst wurde. Für eine bessere Beurteilung der Regionen wurden die Röntgendatensätze speziell bearbeitet. Dabei wurde jedes Bild von Radiolog*innen zugeschnitten und eine inverse logarithmische Transformation sowie eine Histogramm-Entzerrung durchgeführt, um den Kontrast der Region zu erhöhen. Alle Läsionen wurden durch eine histopathologische Untersuchung bestätigt. Die Präzision der verwendeten KI betrug 90,36 %.

Zur Unterscheidung von Kieferzysten und -tumoren kann auch die DVT herangezogen werden. Lee et al. (44) (vgl. Tabelle 3) entwickelten CNN-Modelle für die automatisierte Erkennung, Segmentierung und Klassifizierung von odontogenen Keratozysten und periapikalen Zysten in 2D-Panorama- und DVT-Aufnahmen. KI-Modelle, die auf DVT-Datensätzen trainiert wurden, zeigten eine bessere Funktionalität als solche, die auf Panoramadatensätzen trainiert wurden. Dies ist darauf zurückzuführen, dass DVT-Datensätze quantitativ mehr Merkmale in jedem Voxel der Läsionsregion darstellen können. Bispo et al. (45) und Chai et al. (46) (vgl. Tabelle 3) entwickelten CNN-Modelle zur automatisierten Klassifizierung von Ameloblastomen und odontogenen Keratozysten auf CT-Datensätzen. Chai et. al (46) verglichen in ihrer Studie die Leistung des verwendeten KI-Modells mit sieben erfahrenen und dreißig jungen Mund-

Kiefer-Gesichtschirurg*innen. Das trainierte Modell übertraf sowohl die erfahrenen als auch die jüngeren Mund-Kiefer-Gesichtschirurg*innen mit einer Genauigkeit von 84,6 %. Die Präzision von erfahrenen und jüngeren Mund-Kiefer-Gesichtschirurg*innen betrug 65,7 % bzw. 58,5 %.

Dieses Ergebnis zeigt, dass neuronale Netzwerke auch hier unterstützend eingesetzt werden können. Darüber hinaus kann die rechtzeitige Überweisung an ein Zentrum für Mund-, Kiefer- und Gesichtschirurgie unterstützt werden, um eine möglichst frühzeitige Behandlung zu gewährleisten. (11, 47)

Auch in der Implantologie könnten KI-Systeme eingesetzt werden. Die Planung von Zahnimplantaten unter Verwendung dreidimensionaler DVT-Datensätze könnte durch eine trainierte KI unterstützt werden. In der Studie von Kwak et al. (48) wurde eine Methode zur Bewertung der Weichgewebedicke an der Knochen-Implantat-Grenzfläche (KIS) entwickelt, die auf der Analyse ihrer Ultraschallreaktion unter Verwendung einer simulationsbasierten CNN basiert. Die Qualität der Osseointegration an der KIS ist entscheidend für die Langlebigkeit der Implantate. Die Schwierigkeiten bei der Sicherstellung der klinischen Zuverlässigkeit ergeben sich aus der komplexen Beschaffenheit der Grenzfläche, welche mit der Rauigkeit der Implantatoberfläche und einer Weichgewebsschicht an der KIS zusammenhängt. Der Einsatz von KI in der Implantologie entwickelt sich schnell weiter. In der täglichen Praxis kann dies sehr hilfreich sein, da hier hohe Präzision und sorgfältige Planung erforderlich sind, um eine langfristig suffiziente Behandlung durchzuführen. Des Weiteren kann KI eingesetzt werden, um die Platzierung von Implantaten virtuell zu simulieren und dabei die Faktoren Knochendichte, Platzangebot und Ästhetik zu analysieren. Zahnärzt*innen könnten so mehrere verschiedene Behandlungsoptionen visualisieren und die Implantatpositionen vor dem operativen Eingriff optimieren. KI-Modelle können auch hier eingesetzt werden, um Komplikationen während der Implantatinsertion zu minimieren. Dabei analysieren diese die Position und Ausrichtung der chirurgischen Instrumente. Außerdem könnten KI gesteuerte Algorithmen postchirurgisch die Osseointegration bewerten. Theoretisch könnten dabei auch Veränderungen in der Knochendichte und Periimplantitis frühzeitig erkannt werden. Es muss jedoch betont werden, dass die KI in der Implantologie zwar vielversprechend ist, aber nur als zusätzliches Hilfsmittel zur Unterstützung des zahnärztlichen Personals eingesetzt werden sollte und nicht als Ersatz für dessen Fachwissen. (49)

Auch die radiologische Beurteilung der Kieferhöhle ist bei apikalen Eingriffen an den Oberkieferseitenzähnen und bei Sinusaugmentationen für die Implantation essenziell. Die genaue Diagnose und Klassifizierung von Kieferpathologien vor chirurgischen Eingriffen ist für das Behandlungsergebnis entscheidend. Bei vorliegenden pathologischen Veränderungen erhöht sich das Risiko einer ostiomeatalen Dysfunktion und Sinusinfektion. Jung et al. (50) (vgl. Tabelle 3) entwickelten ein CNN-Modell basierend auf DVT-Datensätzen zur Segmentierung von Kieferhöhlenpathologien wie Schleimhautverdickungen und Schleimhautretentionszysten. Bei internen Datensätzen erzielte das Modell gute Ergebnisse, während bei externen Bildern die Leistung schlechter war. Hung et al. (51) (vgl. Tabelle 3) zeigten in ihrer Studie, dass ihr entwickeltes CNN-Modell hohe Leistungen bei der automatischen Erkennung, Segmentierung und Messung von Schleimhautverdickungen und Schleimhautretentionszysten in der Kieferhöhle erzielte.

Diese Studien zeigen, dass KI das Potential besitzt, Zahnärzt*innen bei der Identifizierung von Kieferhöhlenläsionen, der Bewertung des Ausmaßes der Sinustrübung und der Planung von chirurgischen Eingriffen in der Sinusregion zu unterstützen.

Generell ist zu sagen, dass der Einsatz von neuronalen Netzwerken in der oralen Chirurgie das Auftreten von Komplikationen minimieren kann. Die Modelle haben das Potential, als klinisch praktikable Instrumente für eine frühestmögliche Diagnose zu dienen, was zu einem präziseren Ansatz für die Behandlung führt. (11, 31)

Studie	Art der Daten	Anzahl der Datensätze	Ergebnisse
Kim et al. (41)	Panoramaröntgenbilder	300 Bilder	SEN= 84 % SPE= 82 %
Liu et al. (43)	Panoramaröntgenbilder	420 Bilder	SEN= 92,88 % SPE= 87,80 %
Lee et al. (44)	Panoramaröntgenbilder und DVT-Datensätze	228 Panorama- und 197 DVT-Bilder	Panorama/DVT SEN= 88 %/96 % SPE= 77 %/77 %
Bispo et al. (45)	CT	2500 Bilder aus 40 Scans	PRÄZ= 90-92 %
Chai et al. (46)	DVT	272 Bilder	SEN= 87,2 % SPE= 82,1 %
Jung et al. (50)	DVT	83 Bilder	PRÄZ Intern= 76 % PRÄZ Extern= 54 %
Hung et al. (51)	DVT	347 Bilder	SEN= 79%-93% SPE= 89%-93%

Tabelle 3: Charakteristika der Studien, SEN = Sensitivität, SPE = Spezifität und PRÄZ = Präzision, wobei die Präzision mit der Formel; $Präzision = (Sensitivität) (Prävalenz) + (Spezifität) (1 - Prävalenz)$ beschrieben wird (31)

3.3 Künstliche Intelligenz: Gegenwart und Zukunft

Künstliche Intelligenz hat in vielen Bereichen des Gesundheitswesens erheblich an Präsenz und Bedeutung gewonnen. Sie ist in der Lage, kognitive Fähigkeiten des Menschen nachzuahmen, um komplexe Vorhersagen und Entscheidungen im Gesundheitswesen zu treffen. Sie stellt einen Paradigmenwechsel im Gesundheitswesen dar, der durch die zunehmende Verfügbarkeit von Daten und Fortschritte in der Analysetechnologie vorangetrieben wird. Häufig verwendete KI-Modelle sind neuronale Faltungsnetze, Deep Learning und künstliche neuronale Netze. Diese Modelle finden Anwendung in der Terminplanung, der Patientenbetreuung, der Analyse von Arzneimittelwechselwirkungen, der prognostischen Diagnostik und der robotergestützten Behandlung. In der Endodontie hat die KI ihre Genauigkeit und Präzision bei der Erkennung, Bewertung und Vorhersage von Pathologien unter Beweis gestellt. Dies zeigt, dass die Erwartungen an KI und die tatsächlichen Möglichkeiten in Einklang gebracht werden können. Bevor KI-Modelle jedoch in den klinischen Routinebetrieb integriert werden können, müssen ihre Kosteneffizienz, Zuverlässigkeit und Anwendbarkeit weiter überprüft werden. (52, 53)

3.3.1 KI in der zahnärztlichen Ausbildung

Die klinische Zahnmedizin ist ein komplexer Bereich der zahnmedizinischen Ausbildung. Die Entwicklung klinischer Kompetenzen geht Hand in Hand mit der Aneignung großer Wissensmengen und dem Erwerb klinischer Fertigkeiten, anatomischer Kenntnisse und Problemlösungsfähigkeiten. (54) Die klinischen Fertigkeiten umfassen unter anderem die Patientenberatung und die Durchführung einer klinischen Untersuchung. Derzeit erfolgt die Ausbildung klinischer Kompetenzen nach einem Lehrlingsansatz, bei dem die Studierenden bei der Behandlung von Patient*innen eng überwacht werden. Dies kann zu Unannehmlichkeiten, Komplikationen und langen Behandlungszeiten für die Patient*innen führen. Bisher bestand die klinische Ausbildung der Zahnmedizinstudent*innen aus einer Kombination von theoretischem Unterricht und praktischen Übungen. Diese Methode ist kostspielig und zeitaufwendig. Zudem kann der Zugang zu einer professionellen Ausbildung in komplexeren Szenarien erschwert sein. In den 1980er Jahren wurde ein

intelligentes Tutoriensystem entwickelt. Dieses System basiert auf Virtual Reality (VR) und Augmented Reality (AR) und wird in der zahnmedizinischen Ausbildung eingesetzt. AR definiert eine Technologie, die ein computergeneriertes Bild über die Ansicht des Benutzers von der realen Welt legt und so eine zusammengesetzte Ansicht bietet, wobei VR eine räumliche Umgebung eigenständig simuliert, mit der auf scheinbar reale oder physische Weise interagiert werden kann. (55) Inzwischen gibt es eine gut akzeptierte Standardarchitektur für die Zusammenführung der genannten Systeme in die zahnmedizinische Ausbildung und eine Reihe gut entwickelter und getesteter Techniken zur Benutzermodellierung. (56) Dabei werden Szenarien generiert, die die klinische Arbeit an Patient*innen simulieren und alle Risiken, die mit dem Training in der realen Klinik verbunden sind, minimieren. Diese Ausbildungstechnik hat das präklinische virtuelle Patientenfeedback verbessert. Die interaktive Phase schafft eine qualitativ hochwertige Lernumgebung, indem sie den Studierenden die Möglichkeit gibt, ihre Arbeit zu bewerten und mit dem Goldstandard zu vergleichen. Mit diesen neuen Werkzeugen kann der Zugang zum Lernen erweitert und die Qualität der Interaktionen verbessert werden. (56) Zum Beispiel kann mit einem VR-Simulator die Kariesentfernung mittels Bohrer geübt werden. Studiengruppen, welche ihre Fähigkeiten virtuell erlangten, zeigten vergleichbare Resultate gegenüber der konventionellen Trainingsgruppe. Doch konnten durch VR die iatrogenen Schäden und Aufsichts- bzw. Lehrzeiten reduziert werden. (57) Doch Mubay et al. (58) und Al-Saud et al. (59) weisen darauf hin, dass AR nicht isoliert eingesetzt werden sollte und eine solche Technologie nicht mit der klinischen Umgebung mit Tutor*innen vor Ort gleichzusetzen ist. Doch könne gezeigt werden, dass solche Systeme ein nützliches Tool in der präklinischen sowie klinischen Ausbildung darstellen. (60)

Virtual Reality und Augmented Reality können nicht nur in der Ausbildung von Zahnmedizinstudenten eingesetzt werden. VR kann auch den Prozess der Bereitstellung ästhetischer Prothesen vereinfachen und somit die Erwartungen der Patient*innen erfüllen. Dank KI-Systemen und erweiterter Realität können Patient*innen virtuelle Prothesen ausprobieren, die so lange adaptierbar sind, bis die Vorstellungen der Patient*innen erfüllt sind. Die finale Prothese kann dann genau nach diesem Abbild hergestellt werden. (53)

3.3.2 KI im Patientenmanagement

Virtuelle Assistenten, die auf künstlicher Intelligenz basieren, können alltägliche Aufgaben in der Zahnarztpraxis präziser, fehlerresistenter und mit weniger Personalaufwand erledigen. KI kann dazu verwendet werden, Termine entsprechend den Wünschen der Patient*innen und Ärzt*innen zu koordinieren. Außerdem können Patient*innen via KI-Algorithmen automatisch auf erforderliche Kontrolluntersuchungen aufmerksam gemacht werden. Des Weiteren können virtuelle Assistenten die Zahnärzt*innen über genetische Dispositionen, Lebensstilinformationen und Vorerkrankungen informieren. Dazu gehört z.B. das Parodontal-Screening für Patient*innen mit Diabetes, Mundkrebs-Screening für Patient*innen mit Nikotinabusus, die prophylaktische Einnahme von Antibiotika bei Patient*innen mit kardiovaskulären Erkrankungen, sowie Erinnerungen an bestehende Allergien. Zusätzlich können solche KI-Modelle unterstützend bei der Verwaltung, Versicherung, Teleassistenz und der Behandlungsplanung eingesetzt werden. (61)

Solche KI-Modelle sind in der Lage eine detaillierte virtuelle Patientendatenbank zu erstellen, die eine optimale Behandlung der Patient*innen ermöglicht. Dazu gehört z.B. die Sammlung aller notwendigen zahnärztlichen Unterlagen, zusätzlicher oraler Fotografien und Röntgenaufnahmen, die für die Diagnose einer Zahnerkrankung erforderlich sind. (53)

KI-Modelle, die in der klinischen Medizin Anwendung finden, könnten jedoch einige Nachteile aufweisen. Die Algorithmen basieren in der Regel auf einigen ungerechtfertigten und nicht vollständig nachvollziehbaren Annahmen, was zu einer relativen Skepsis hinsichtlich ihrer Zuverlässigkeit in einem realen klinischen Szenario führen kann. KI kann in der Zahnmedizin eine große Hilfe sein, aber biologische Systeme sind viel komplexer und KI-Systeme können menschliches Wissen, Kompetenz und Entscheidungsfähigkeit nicht vollständig ersetzen. (62)

3.3.3 Probleme im Zusammenhang mit der KI

Trotz des großen Potenzials haben KI-Lösungen in der Regel noch keinen Eingang in die medizinische Routinepraxis gefunden. In der Zahnmedizin beispielsweise werden neuronale Faltungsnetzwerke erst seit 2015 in der Forschung zu dentalen

Röntgenbildern eingesetzt. (63) Verschiedene Gründe könnten erklären, warum die KI noch nicht vollständig in die Zahnmedizin integriert wurde.

Zum einen sind medizinische und zahnmedizinische Daten aufgrund von Datenschutzgesetzen und organisatorischen Barrieren nicht in der genormten Form verfügbar. Die benötigten Daten sind oft in getrennten, individualisierten und nur begrenzt zugänglichen interoperablen Systemen gespeichert. Im Vergleich zu Datensätzen, die für die Entwicklung von KI-Modellen in anderen Bereichen verwendet werden können, sind medizinische Datensätze oft unstrukturiert und relativ klein. Ferner sind medizinische Datensätze komplex, mehrdimensional und sensibel. Es ist eine Herausforderung, sie zu triangulieren oder zu validieren. Die Auswahl von Stichproben führt häufig zu einer Selektionsverzerrungen, bei denen übermäßig kranke Patient*innen (z. B. Krankenhausdaten), übermäßig gesunde Patient*innen (z. B. Daten aus tragbaren Geräten) oder übermäßig wohlhabende Patient*innen (z. B. Daten von Personen, die sich Zahnbehandlungen in Ländern ohne universelle Gesundheitsversorgung leisten können) überrepräsentiert sind. KI-Anwendungen, die auf der Grundlage solcher Daten entwickelt werden, sind naturgemäß verzerrt. (64)

Zum anderen sind die Datenverarbeitungen, Messungen sowie Validierungen der Ergebnisse in der zahnmedizinischen KI-Forschung oft nicht ausreichend replizierbar und robust. In den meisten Fällen sind die Ergebnisse nicht reproduzierbar und es ist unklar, wie die Datensätze ausgewählt, kuratiert und aufbereitet wurden. Die Daten werden häufig sowohl für das Training als auch für das Testen verwendet, was zu einem "Data Snooping Bias" führt. (62) Walsh (65) hat in ihrer Studie gezeigt, dass es viele herausfordernde Aspekte bei der Planung, Durchführung und Interpretation einer Studie zur diagnostischen Genauigkeit von Tests gibt. Es sei in der Regel nicht möglich, einen "harten" Goldstandard zu definieren, und es bestehe keine Einigkeit darüber, wie viele Expert*innen erforderlich sind, um einen Datenpunkt zu markieren und wie verschiedene Markierungen solcher "unscharfen" Goldstandards zusammengeführt werden können.

Darüber hinaus sind die KI-Ergebnisse in der Zahnmedizin nicht uneingeschränkt anwendbar. Die bereitgestellten Informationen könnten die notwendigen und komplexen Entscheidungen im klinischen Betrieb nur teilweise unterstützen. Ferner bleiben Fragen der Verantwortlichkeit und Transparenz offen. (66)

Zu beachten ist auch, dass KI-basierte Software derselben Konformitätsprüfung gemäß Art. 61 VO (EU) 745/2017 nach EU-Recht unterliegt wie ein Medizinprodukt.

Dabei wird das Medizinprodukt bzw. die KI-basierte Software einem Konformitätsbewertungsverfahren unterzogen, welches aus mehreren Schritten besteht, beispielsweise der technischen Dokumentation und der Garantie der Produktionsqualität. So wird gewährleistet, dass das Produkt langfristig sicher und leistungsfähig ist. (67)

3.3.4 Integration von KI in die klinische Versorgung

Auch wenn geeignete Aufgaben, Daten und Evidenzen für KI vorhanden sind, kann ihr Potenzial nur durch eine suffizient geführte Integration in die klinische Versorgung voll ausgeschöpft werden. Kliniker*innen müssen das Know-how entwickeln, KI-gestützte Ergebnisse zu interpretieren und zu verstehen. Sie müssen auch lernen, diese in ihre Routinearbeit zu integrieren. "Black-Boxes" erschweren die Einführung dieser Instrumente in die Praxis. Bei der Verwendung eines deep Learning Algorithmus mit einem "Black-Box"-Ansatz können die Benutzer*innen beispielsweise Daten als Eingabe eingeben, um ein Ergebnis als Ausgabe zu erhalten. Die Benutzer*innen können jedoch nicht zurückverfolgen, wie dieses Ergebnis zustande gekommen ist. Im Vergleich zu traditionellen Entscheidungshilfen in der klinischen Praxis generieren diese Techniken ihre Ergebnisse aus unbeobachteten Ergebnissen. Dadurch können Kliniker*innen ihre Gültigkeit nicht vollständig überprüfen und nachvollziehen.

KI hat zwar das Potenzial, Prozesse effizienter zu gestalten, die Ergebnisse sind jedoch eher unterstützend als ersetzend. Im Idealfall kann KI, wenn sie angemessen in den klinischen Arbeitsablauf integriert wird, klinische Muster und Erkenntnisse liefern, die über die derzeitigen menschlichen Fähigkeiten hinausgehen, und die Kliniker*innen von einem Teil der Last befreien, die riesigen und wachsenden Mengen an Gesundheitsdaten und Wissen in den klinischen Arbeitsablauf und die klinische Praxis zu integrieren. Kliniker*innen können sich dann effizienter der Aufgabe widmen, sich um die Bedürfnisse und Werte der Patient*innen zu kümmern, um deren optimale Gesundheit zu erreichen. Die Kombination von KI und menschlicher Intelligenz könnte wahrscheinlich den leistungsfähigsten Ansatz darstellen, um diese grundlegende Aufgabe der Gesundheitsversorgung zu erfüllen. (66)

3.3.5 Ethik und Künstliche Intelligenz

Wie in den vorangegangenen Abschnitten erläutert, hat sich die KI in den letzten Jahren weiterentwickelt und in vielen Bereichen unseres Lebens Einzug gehalten. Während sich dieser Trend weiter fortsetzt, sind die langfristigen Auswirkungen von KI schwer vorhersehbar. So reicht der aktuelle Diskurs über KI und ihre Auswirkungen von Utopien einer besseren und gerechteren Welt bis hin zu Dystopien einer den Menschen beherrschenden Superintelligenz. Während die Zukunft der KI weitestgehend offen ist, haben wir als Menschheit die Möglichkeit, die Entwicklung von KI zu lenken und diese unter menschenrechtlicher und ethischer Perspektive zu beeinflussen. Die UNESCO hat sich diesem Thema angenommen und im Jahr 2021 eine Empfehlung zur Ethik Künstlicher Intelligenz verabschiedet. Zu diesem Gremium gehörten 193 Mitgliedstaaten sowie mehrere Stakeholder. Die UNESCO-Empfehlung ist die erste global formulierte Antwort auf eine der größten Herausforderungen im Zusammenhang mit KI und ihrer Zukunft. (68)

Angesichts der Vielfalt der wirtschaftlichen, sozialen, kulturellen, bürgerlichen und politischen Rechte, deren Achtung, Schutz und Durchsetzung durch die zunehmende Nutzung von KI berührt werden, besteht ein Bedarf an Grundsätzen und Regeln, um das Potenzial von KI zu nutzen und gleichzeitig die individuellen Freiheiten zu schützen und den sozialen Zusammenhalt zu gewährleisten.

KI muss grundsätzlich Menschenrechte achten, sichern und fördern. Weltweit muss KI im Prozess der Planung und der Umsetzung an ethischen Werten, Menschenrechten und Völkerrecht ausgerichtet werden. Dabei müssen alle KI-Akteur*innen den gesamten Lebenszyklus von KI-Systemen kontrollieren. Es muss sichergestellt werden, dass die Wissenschaft unter Wahrung von Datenschutz Zugang zu Forschungsdaten erhält und diese nicht von Unternehmen monopolisiert werden. Dieser Zugang muss neben der Wissenschaft auch ausgewählten zivilgesellschaftlichen Akteur*innen ermöglicht werden, die anderen Interessen als der Wissenschaft folgen und eine öffentliche "Watchdog-Funktion" wahrnehmen können. Hier soll speziell die Qualität der Daten erhöht und der Zugang erleichtert werden. Um den Schutz von Daten sicherzustellen, wurde die Dateninfrastruktur "GAIA-X" geschaffen, welche in regelmäßigem Austausch mit verschiedenen Datenschutzaufsichtsstellen steht.

Mögliche Einschlüsse von Diskriminierung im gesamten Zyklus von KI-Modellen müssen vermieden werden. Dazu wird empfohlen, Gender, Inclusiveness und Diversity Mainstreaming in allen Entwicklungsprozessen zu berücksichtigen. Dabei sollte eine entsprechende Dokumentationspflicht eingeführt werden, um solche Stereotypen und Heterogenitäten auszuschließen.

Zudem kann KI einem "Western-Bias", also eine Verzerrung durch den verstärkten westlichen Entwicklungseinfluss, unterliegen. Daher müssen neben der nationalen Offenlegung auch globale KI-Entwickler*innen im Rahmen einer technischen Entwicklungskooperation einbezogen werden. Gerade der fehlende Zugang zu Daten in einkommensschwachen bzw. Low-Developed-Countries kann die globale Verzerrung verstärken.

KI sollte neben der staatlichen auch einer öffentlichen Kontrolle unterliegen. Klare Rahmenbedingungen für die Durchführung von ethischen Folgenabschätzungen und Kontrollmechanismen müssen entwickelt und umgesetzt werden. Zudem müssen KI-Entscheidungsprozesse verständlich, erklärbar sowie rechtfertigbar sein. Nutzende dürfen sich gegenüber der KI nicht schutzlos oder überfordert fühlen. Daher müssen die Rolle und das Potenzial von KI in der Öffentlichkeit durch Wissensformate und Referenzen dargestellt werden.

Ethische Folgenabschätzungen können Vorteile, Bedenken und Risiken von KI-Modellen identifizieren und bewerten. Je nach Bewertung müssen dann Maßnahmen zur Risikovermeidung, -minderung und -überwachung vorgeschlagen werden.

Nachhaltigkeit ist ebenfalls ein Thema, welches in der Entwicklung von KI-Systemen berücksichtigt werden sollte. KI muss immer im Hinblick auf die Förderung der Sustainable Development Goals (= Ziele für nachhaltige Entwicklung) betrachtet werden. Generell ist der Einsatz von KI im Hinblick auf die ökologische Nachhaltigkeit ebenfalls regelrecht einzuordnen. (68)

Ziel der UNESCO KI-Ethik-Empfehlung ist es, weltweit akzeptierte, ethische Standards für KI-Systeme zu formulieren, die in vollem Einklang mit dem Völkerrecht, insbesondere den Menschenrechten, stehen. Wichtig ist auch die Fokussierung auf definierte "blind-spots" bisheriger KI-Regelungen. Dazu gehören z.B. der Umweltschutz, der nachhaltige und ressourcenschonende Einsatz von KI und der Einsatz von KI im Bildungsbereich unter der Prämisse "Bildung für alle".

Im Zusammenhang mit der UNESCO-Empfehlung ist zu erwähnen, dass es sich zwar nicht um ein formell verbindliches Recht handelt, die Empfehlung aber in einem

mehrjährigen Prozess so detailliert wie ein völkerrechtlicher Vertrag ausgehandelt wurde, so dass sie sich qualitativ nicht von einfachen Resolutionen und Erklärungen unterscheidet. Auch wenn die Empfehlung nicht gerichtlich durchsetzbar ist, so entfaltet sie doch Wirkung gegenüber den Staaten. (68)

3.3.6 Weitere ethische Prinzipien

Neben der UNESCO-Empfehlung haben sich im Laufe der Etablierung von KI-Systemen weitere ethische Prinzipien entwickelt, um auf bestehende Missstände und mögliche Chancen aufmerksam zu machen. Dazu gehören z.B. das PAPA-Modell, das AI4People-Modell sowie weitere KI-Ethik-Prinzipien. (69-71) Zusammengefasst ergeben sich daraus sechs zentrale ethische Prinzipien: Wohltätigkeit, Transparenz, Nicht-Böswilligkeit, Autonomie, Gerechtigkeit und Datenschutz.

Unter Wohltätigkeit wird die Entwicklung einer „guten“ KI definiert, die an menschlichen Werten ausgerichtet ist. Für eine gesellschaftlich nutzenstiftende KI müssen Mehrwerte auf ökonomischer, sozialer und ökologischer Ebene verfolgt werden. (71) Für das Prinzip der Transparenz, müssen Informationen einer KI-Lösung verständlich und ausreichend zugänglich vorgestellt werden. Frühestmöglich sollten Verantwortlichkeiten, Datenquellen, -arten und -auswertung erläutert und dauerhaft festgehalten werden. Interessengruppen müssen langfristig und barrierefrei Informationen über die KI erhalten können. (71) Für öffentlich zugängliche KI-Systeme müssen Sicherheit und Schadensvermeidung bei der Entwicklung und Nutzung stets im Vordergrund stehen. Eine eingebaute Bewertung der Datenqualität kann hier Abhilfe schaffen. Auch staatliche Interventionen im Bereich der KI-Forschung, -Entwicklung und -Nutzung tragen zur Sicherheit bei. Unternehmen können aktiv um genannte Kontrollen bitten und somit der Gesellschaft methodisches Vorgehen signalisieren und eine gesellschaftliche Annäherung an KI ermöglichen. (71) In der Interaktion mit KI sollte die Entscheidungsbefugnis sowie der Handlungsfreiraum den Fähigkeiten des Menschen entsprechend zugeteilt werden. Die Autonomie der KI-Anwender*innen muss stets gewahrt bleiben, damit sie nicht von der KI abhängig und in ihren Entscheidungs- und Handlungsfeldern eingeschränkt werden. Aktuelle Fortbildungen erlauben es den Nutzer*innen, weiterhin die Entscheidungshoheit zu behalten. (72) Um Fairness bei der Entwicklung und Nutzung von KI zu gewährleisten und zu fördern, sind technische Standards, transparente Verfahren und eine Sensibilisierung der Öffentlichkeit erforderlich. Die diverse Zivilbevölkerung sollte in

ihrer Breite bei der KI-Entwicklung miteinbezogen und in den Daten abgebildet sein. (69, 71) Im Zusammenhang mit Datensicherung müssen sich Anbieter von KI z.B. in der Europäischen Union bzw. in Österreich an die Datenschutz-Grundverordnung (DSGVO) und in anderen Regionen an ähnliche Verordnungen halten. Dabei wird unterschieden, ob sensible Daten, also personenbezogene Daten, oder nicht-sensible Daten verarbeitet werden und welches Risiko mit der folgenden Datenverarbeitung verbunden ist. Gesundheitsdaten werden in Österreich laut DSGVO als eine besondere Kategorie personenbezogener bzw. sensibler Daten eingestuft. Solche Daten stehen unter strengen Regularien. Die Einsichtnahme in solche sensiblen Daten oder deren Verwendung bedarf besonderer Gründe, um das Risiko eines Missbrauchs oder Schadens zu vermeiden. Die Erlaubnistatbestände werden ebenfalls in der DSGVO genau geregelt. Dazu gehören z.B. Zwecke der Gesundheitsvorsorge, aber auch wissenschaftliche Zwecke. (73)

Das "EU-AI-Act" stellt eine umfassende Verordnung der EU über KI dar, welche im Juni 2024 in Kraft treten soll. Dabei ordnet das Gesetz die Anwendungsgebiete von KI drei Risikokategorien zu. So werden KI-Modelle, die ein inakzeptables Risiko repräsentieren, wie z.B. staatliche Social-Scoring-Systeme, also eine Bewertung des sozialen Verhaltens von Personen zum Zwecke der Verhaltensprognose bzw. -steuerung, verboten. KI-Modelle, die einem hohen Risiko unterliegen, wie z.B. Algorithmen, die persönliche Daten verwenden und analysieren, unterliegen strengen Regularien und rechtlichen Anforderungen. Als drittes bleiben KI-Anwendungen, die weder als verboten noch risikoreich eingestuft werden, weitgehend unreguliert. (76)

Am Ende ist es wichtig, eine Balance zwischen den Interessen von Einzelpersonen, Unternehmen und der Gesellschaft zu wahren. Zu große, exzessive Restriktionen haben auf der anderen Seite die Charakteristik, die Entwicklung von KI-Systemen zu verlangsamen oder gar zu verhindern. (74, 75)

3.4 Die pathologische Veränderung „Periapikale osteolytische Läsion“

3.4.1 Ätiologie

Das Bindegewebe des Zahnes befindet sich in der Zahnpulpa und wird von Hartgewebe, wie Zahnschmelz, Dentin und Zement, umgeben. Durch diesen strukturellen Aufbau kann gewährleistet werden, dass die Zahnpulpa vor Mikroorganismen geschützt wird. Traumata, Karies oder Zahnabnutzungen können periapikale osteolytische Läsionen auslösen. Bei einer Pulpaexposition können Mikroorganismen das Pulpagewebe besiedeln, was zu einer Pulpanekrose und periradikulären pathologischen Veränderungen führen kann. (77) Möller et. al. zeigen in ihrer Studie, dass Mikroorganismen und ihre Produkte eine zentrale Rolle in der Entstehung und Etablierung von periapikalen osteolytischen Läsionen spielen. (78) Dazu zählen fakultativ anaerobe Streptokokken, coliforme Stäbchen und obligat anaerobe Bakterien. Tritt eine Pulpanekrose aufgrund einer Invasion von Mikroorganismen auf, können weder das körpereigene Immunsystem noch systemische Antibiotika die Infektion wirksam aufhalten, da eine suffiziente Blutversorgung fehlt. (79) Ohne Behandlung breitet sich die Entzündung über das Foramen apikale aus und eine apikale Parodontitis kann entstehen. (80) Die Mikroorganismen resorbieren den Knochen und ersetzen diesen durch granulomatöses Gewebe und eine dichte Schicht aus polymorphkernigen neutrophilen Leukozyten. Die dadurch entstandene periapikale Radioluzenz ist das häufigste klinische Anzeichen einer periapikalen osteolytischen Läsion. (81)

3.4.2 Indizes für die periapikale Läsion

Zur radiologischen Beurteilung des Zustands im Periapikalbereich wurden spezifische Indizes entwickelt, die sich auf zweidimensionale sowie dreidimensionale Bildgebungstechniken beziehen.

Der Strindberg-Index beurteilt eine periapikale Läsion nach einer endodontischen Behandlung. Weist der Zahn nach der Behandlung radiologisch normales parodontales Ligament und eine intakte Lamina dura auf, wird die Behandlung als erfolgreich bewertet. Kann keine Reduktion der periapikalen Läsion festgestellt

werden, hat sich eine neue Läsion entwickelt oder kann eine Unterbrechung in der Lamina dura beobachtet werden, gilt die Behandlung als erfolglos. Zusätzlich wird noch in den Parameter "unsicher" unterschieden, wenn der Zustand unklar ist. Dieser Index kann nur erhoben werden, wenn der Status der Zähne vor und nach der Behandlung bekannt ist. (82) Reith et al. (83) beschreiben einen Index über das Vorhandensein oder Nichtvorhandensein einer periapikalen Parodontitis. Dabei wird in die Kriterien periapikale Knochenzerstörung sicher nicht vorhanden, periapikale Knochenzerstörung wahrscheinlich nicht vorhanden, unsicher, periapikale Knochenzerstörung wahrscheinlich vorhanden und eindeutige periapikale Knochenzerstörung unterschieden. Orstavik et al. (84) beschreiben den aktuellsten und aussagekräftigsten Index für die Beurteilung von periapikalen osteolytischen Läsionen in 2D-Datensätzen. Der periapikale Index beschreibt ebenfalls fünf Kategorien. Score 1 beschreibt eine radiologisch normale periapikale Anatomie, leichte Änderungen in der Knochenstruktur werden im Score 2 beschrieben. Läsionen mit leichtem Knochenverlust werden in den Indexscore 3 eingeordnet, wobei apikale Parodontitiden mit deutlicher Radioluzenz im apikalen Bereich in Score 4 eingeteilt werden. Indexscore Klasse 5 beschreibt eine Parodontitis mit Merkmalen einer Exazerbation.

Dank neuer Errungenschaften in der radiologischen Bildgebung, sind auch 3D Bildgebungstechniken wie die Digitale Volumentomographie ein elementarer Bestandteil der zahnmedizinischen Diagnostik. Estrela et. al. entwickelten einen neuen Index zur Klassifizierung von periapikalen osteolytischen Läsionen. Ziel dieses Indexes ist die Beschreibung einer Methode, die auf der Analyse und Interpretation von hochauflösenden dreidimensionalen Datensätzen basiert. Dabei wird die Läsionsgröße und die vorliegende Zerstörung der Kortikalis durch eine periapikale Läsion beschrieben. Indexscore 0 beschreibt eine intakte periapikale Struktur, wobei der Indexscore 1 Läsionen mit einer Größe von 0.5-1mm erfasst. Der Indexscore 2 definiert eine Ausdehnung von 1-2mm, der Indexscore 3 eine von 2-4mm, Indexscore 4 eine von 4-8mm und der Indexscore 5 eine Ausdehnung von über 8mm. Zusätzlich zu den Läsionsgrößen wird ein weiterer Parameter in die Evaluierung miteinbezogen. Der Buchstabe "E" beschreibt eine Expansion der Kortikalis, während "D" sich auf eine Destruktion bezieht. (85). Esposito et al. (86) kritisierten die Reproduzierbarkeit dieses Index. Somit modifizierten sie diesen und legten drei feste und reproduzierbare

Dimensionen zur Messung der Läsionsgröße fest. Dementsprechend wurden die Messlängen mesiodistal, bukkolingual und koronoapikal festgelegt.

3.4.3 Komplikationen

Unbehandelte apikale osteolytische Läsionen können chronisch werden und zu Schmerzen, Komplikationen wie Abszessbildung und Zahnverlust führen. Zusätzlich können die von den Bakterien produzierten Toxine in den Blutkreislauf gelangen, was zu einer erhöhten systemischen Entzündung führt. Eine systemische Bakteriämie kann mit koronalen Herzerkrankungen, Diabetes mellitus, Leberzirrhose und dem akuten Myokardinfarkt in Verbindung stehen. Daher ist eine frühzeitige Diagnose essenziell. (79, 87)

3.5 Digitale Volumentomographie

3.5.1 Technischer Hintergrund

Ein 3D DVT-Datensatz wird aus hunderten 2D Projektionsbildern zusammengesetzt. Die Projektionsbilder werden während einer Rotationsbewegung um ein Isozentrum aufgenommen. Meistens befindet sich das Isozentrum in der Mitte des Volumens. Später wird durch einen Rückprojektionsalgorithmus ein 3D-Volumen erzeugt.

Die Bildaufnahme bei einem DVT unterscheidet sich von der konventionellen CT. Bei einer CT ist die Strahlengeometrie fächerförmig und das Volumen wird aus einer Reihe von Schnitten rekonstruiert, welche in mehreren Rotationen entstehen.

Im Normalfall arbeitet ein DVT-Gerät mit einer Spannung von 80 - 120 Kilovolt (kV) und einer Stromstärke von 3 - 10 Milliampere (mA). Im Vergleich zu CT-Geräten beziehen DVT-Geräte weniger Strom, wodurch sich der Rauschpegel erhöht. Aus diesem Grund sind DVT-Bilder im Allgemeinen qualitativ schlechter als jene von CT-Geräten, was sich durch eine geringere Kontrastauflösung und ein höheres Rauschen auszeichnet. (88, 89) Dies führt zu einer schlechten Weichteilauflösung im DVT. (90) Mittlerweile ist es möglich, das Rauschen durch neue Rekonstruktionsalgorithmen zu reduzieren. Diese können jedoch noch nicht flächendeckend für dentale DVT-Geräte verwendet werden. (91)

Die Scanzeit liegt in der Regel zwischen 10 und 20 Sekunden. Es ist zu beachten, dass die genannte Scanzeit von der Belichtungszeit aufgrund des gepulsten Strahlenbündels zu unterscheiden ist. (88)

Die räumliche Auflösung der DVT-Bilder wird in der Literatur zwischen 0,6 und 2,8 Linienpaar pro Millimeter (lp/mm) angegeben. Die effektive räumliche Auflösung in DVT-Bildern wird durch den zweidimensionalen Detektor, den dreidimensionalen Rekonstruktionsprozess, lange Scanzeiten, Patient*innenbewegungen während des Scans und verschiedene andere Parameter beeinflusst. Unter der Berücksichtigung der genannten Punkte liegt die effektive räumliche Auflösung knapp über 1 lp/mm. (88, 92)

Ein weiteres wichtiges Unterscheidungsmerkmal von einem DVT sind die verschiedenen field-of-view (Sichtfeldgrößen/ FOV). Durch diese Adjustierungen können verschiedene radiologische Aufgaben optimiert werden. Die FOVs reichen von 4 Centimetern (cm) im Durchmesser sowie in der Höhe bis zu großen FOVs mit über 20 cm Durchmesser und ungefähr derselben Höhe. DVTs wenden somit lokale Tomographien an. CTs wenden größere FOVs an und können gesamte Körperteile scannen. Sie führen eine globale Tomographie durch (vgl. Abbildung 7). Bei der globalen Tomographie liegt das gesamte Objekt dauerhaft im Röntgenstrahl. Dabei wird auch das gesamte Objekt rekonstruiert. Bei einer lokalen Tomographie wird ein Teil des Objektes belichtet. Nur ein Teil des belichteten Objektes wird entsprechend der durch das FOV dargestellten Regionen rekonstruiert und visualisiert. Gewebe außerhalb des FOVs trägt zu Abschwächungen bei. Die Aufgabe besteht darin, nur im FOV enthaltenes Gewebe zu rekonstruieren. Findet dies nicht kontrolliert statt, kann es zu teilweise fehlerhaften Werten führen. Dies liegt daran, dass die Rekonstruktionen im Allgemeinen aus Rückprojektionen aus den Werten des FOV besteht.(88)

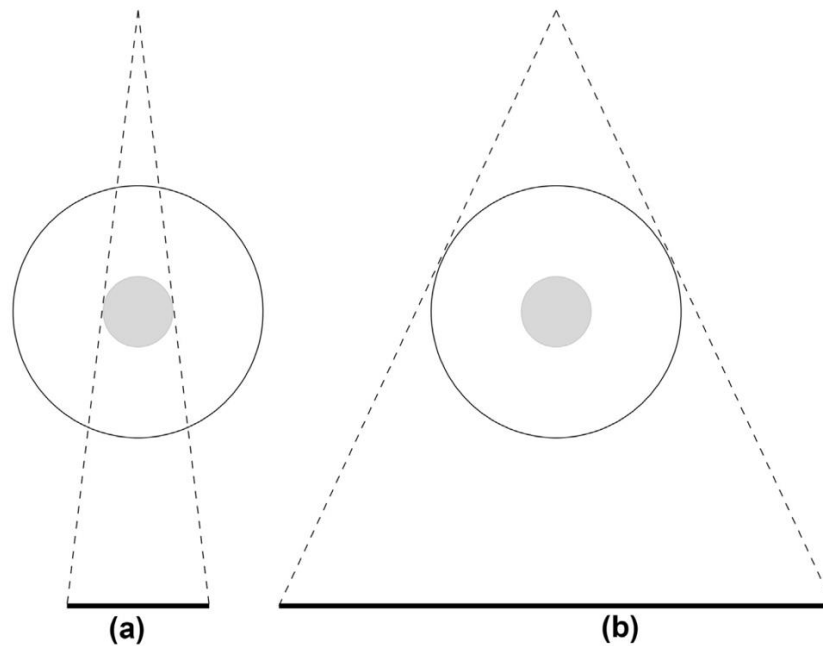


Abbildung 5: lokale Tomographie (a) vs. globale Tomographie (b) (88)

Die effektive Dosis für eine DVT ist im Allgemeinen höher als bei anderen konventionellen Röntgenverfahren. Bei einer intraoralen Röntgenaufnahme mit Filmdetektoren beträgt die effektive Dosis zwischen 3,5 und 8,2 μ Sievert (Sv) und bei Geräten mit digitalen Detektoren zwischen 1,2 und 2,5 μ Sv. (93) Beim Panoramaröntgen liegt die effektive Dosis je nach verwendetem Gerät zwischen 19 und 75 μ Sv. (94) Es ist wichtig zu erwähnen, dass die effektiven Dosen stark variieren können. (88, 94) Bei der DVT kann die effektive Dosis zwischen 5 und 1073 μ Sv liegen, wobei hier der Bereich von der Größe des FOV abhängig ist. Bei einem CT-Gerät wird die effektive Dosis zwischen 685 und 1410 μ Sv beschrieben. (88) Die effektiven Dosen von CBCT und CT-Geräten bei optimierten Belichtungsniveaus können ähnlich sein. Die Optimierung der effektiven Dosis ist eine ethische Verpflichtung und wird durch die Europäische Union gesetzlich vorgeschrieben. Somit muss die Strahlendosis so gering wie möglich gehalten werden und gleichzeitig die diagnostische Qualität der Bilder aufrechterhalten. Bei einer Optimierung kann durch Änderungen der Volumengröße, der Belichtungsfaktoren/Voxelgröße und des Rotationsbogens die Dosis angepasst werden. (95)

3.5.2 Nachteil einer DVT

Trotz des breiten Anwendungsspektrums einer DVT ist darauf zu achten, dass Fehlinterpretationen vermieden werden. Z.B. kann das Vorhandensein von intrakanalen Metallstiften Artefakte im Datensatz hervorrufen. Diese Artefakte können die Bildinterpretation beeinflussen. Dabei könnten Metallartefakte im zu untersuchenden Zahn und im Nachbarzahn die Analyse und Diagnose des Bildes erschweren. In einigen Fällen ist eine fundierte Interpretation nicht möglich. In solchen Fällen wird empfohlen auf ein konventionelles 2D-Röntgenverfahren zurückzugreifen. Artefakte entstehen infolge von Problemen im Rekonstruktionsalgorithmus beim Übergang von Metall zu Gewebe. Dieses Problem ist auch bei konventionellen CT-Verfahren darstellbar. (96)

4 Methodik

4.1.1 Datensatz

Bei dieser Studie handelt es sich um eine Nichtunterlegenheitsstudie der Abteilung für Orale Chirurgie und Kieferorthopädie der Univ.-Klinik für Zahnmedizin und Mundgesundheits Graz.

Die Ergebnisse dieser Studie wurden bereits publiziert und sind unter folgender Referenz zu finden:

Hadzic, A.; Urschler, M.; Press, J.-N.A.; Riedl, R.; Rugani, P.; Štern, D.; Kirnbauer, B. Evaluating a Periapical Lesion Detection CNN on a Clinically Representative CBCT Dataset—A Validation Study. *J. Clin. Med.* 2024, 13, 197. <https://doi.org/10.3390/jcm13010197>

Diese Studie stellt eine Folgearbeit der Studie von Kirnbauer et al. (6) (<https://doi.org/10.1016/j.joen.2022.07.013>) dar.

Für die Durchführung dieser Studie wurden DVT-Datensätzen von Patient*innen der Univ.-Klinik für Zahnmedizin und Mundgesundheits in Graz gesammelt, die im Jahr 2018 zur DVT-Untersuchung vorstellig waren. Den Einschlusskriterien entsprechende Datensätze wurden in die Studie eingeführt.

Folgende Einschluss- bzw. Ausschlusskriterien wurden angewandt:

- Das DVT-Aufnahmevervolumen soll den Ober- und/oder Unterkiefer umfassen.
- Geräteeinstellung: Planmeca ProMax [®] 3D Max (Planmeca, Helsinki, Finnland) mit einem Aufnahmevervolumen von 10.0 x 5.9 cm oder 10.0 x 9.3cm, 200-µm Voxelgröße und Normalmodus (96 kV, 5.6-9.0 mA, 12s).
- Restbezahnung von mindestens 3 bestehenden Zähnen pro Kiefer.
- Abgeschlossenes Wurzelwachstum/ Patient*innenalter > 20 Jahre.
- Akzeptable Bildqualität (Ausschluss radiologisch nicht beurteilbarer Datensätze mit z.B. ausgeprägten Metallartefakten und Bewegungsartefakten, welche eine suffiziente radiologische Beurteilung verhindern).

Bei der Kollektion der DVT-Datensätze wurde nicht geschlechterspezifisch unterschieden. Die Einschleusung der DVT-Datensätze in diese Studie und die dabei

entstandene Geschlechterverteilung ist Zufall und wurde nicht beeinflusst. Eine daraus resultierende Altersverteilung wurde ebenfalls nicht beeinflusst.

Eine klinische Untersuchung von Patient*innen war nicht Teil dieser Studie. Es wurden lediglich die DVT-Datensätze nach einem definierten Schema beurteilt.

Den Einschlusskriterien entsprechende dreidimensionale Röntgenbilder wurden in einer Excel-Tabelle dokumentiert und gespeichert. Um dem Datenschutz der Patient*innen zu gewährleisten, wurden die Datensätze pseudonymisiert gespeichert und mit einer fortlaufenden Nummer codiert (001, 002, 003, ...). Diese Pseudonymisierung und elektronische Archivierung führte die Studienleiterin durch, welche alleinigen Zugriff auf die Originaldatensätze und die dazugehörigen Patient*innendaten hatte.

Nach Pseudonymisierung wurden die Datensätze einer manuellen Befundung und digitalen Markierung im Sinne einer Segmentierung der pathologisch veränderten Bereiche (Periapikale Osteolysen, PALen) durch 1 Studenten und 2 Expertinnen im Konsensverfahren unterzogen.

Zur Bildbetrachtung wurde die open-source Software ITK-Snap herangezogen. (97) Die Bildbetrachtung erfolgte auf einem Röntgenbefundmonitor (MDNC-2221 Monitor, Auflösung 1600 × 1200; Größe 432 × 324 mm; 59.9 Hz; Barco Control Rooms GmbH, Karlsruhe, Germany). Die ausgewählten DVT- Datensätze wurden parallel dazu vom programmierten deep CNN hinsichtlich PALen gescreent und ausgewertet. Die statistische Auswertung erfolgte mit der statistischen Analysesoftware SPSS (IBM SPSS statistics 27.0, IBM Corporation, New York, NY, USA) gemäß der unten definierten Zielgrößen. In der Tabelle wurde in Ober- und/oder Unterkiefer unterschieden. Zähne mit apikal pathologischen osteolytischen Läsionen wurden entsprechend dem Universal Numbering System gekennzeichnet und aufgezeigt (vgl. Abbildung 5). Röntgenologisch beherdete Zähne wurden mit einer 1 gekennzeichnet, gesunde Zähne mit einer 0 und fehlende Zähne mit dem Auslassen der Spalte.

4.1.2 Zielgrößen

4.1.3 Hauptzielgrößen

- Anzahl der richtig-positiv (TPR=true positive rate / Sensitivität) und richtig-negativ (TNR=true negative rate / Spezifität) erkannten PALen im tiefen CNN im Vergleich zu den Ergebnissen von 2 Expertinnen (2 Zahnärztinnen).

4.1.4 Nebenzielparameter

- Zahnstatus (angelegte Zähne, Zahnlücken, Implantate, Restaurationen, endodontische Versorgung)
- Lokalisation der PALen (Kiefer und Zahnposition)
- Klassifikation der PALen gemäß Estrela et al. (85)
- Auswertungszeit der KI

4.1.5 Statistische Auswertung

Die Auswertung der Hauptzielgrößen Anzahl der richtig positiv erkannten periapikalen osteolytischen Läsionen (Sensitivität) und Anzahl der richtig negativ detektierten periapikalen osteolytischen Läsionen (Spezifität) erfolgte mittels einseitigem Binomialtest (Alpha= 2,5 %).

Für die Studie wurde folgende Hypothesen getestet:

Sensitivität = $p(\text{Se})$

Nullhypothese: $p(\text{Se}) \leq 0.90$

Alternativhypothese: $p(\text{Se}) > 0.90$

Spezifität = $p(\text{Sp})$

Nullhypothese: $p(\text{Sp}) \leq 0.82$

Alternativhypothese: $p(\text{Sp}) > 0.82$

Sensitivität und Spezifität wurde mittels 95 % Konfidenzintervall dargestellt.

Die Auswertungen erfolgten mittels deskriptiver Statistik. Mittelwertsdifferenzen und deren dazugehöriges 95 % Konfidenzintervall wurden berechnet. Je nach Datenlage

erfolgte die Darstellung der Ergebnisse stratifiziert nach Ober- und/oder Unterkiefer und nach Zahnposition (insgesamt 16 mögliche Strata).

4.1.6 Fallzahlplanung

Bei der Vorläuferstudie (<https://doi.org/10.1016/j.joen.2022.07.013>) wurden bei N= 2128 Zähne aus 144 DVT-Datensätzen N=206 (9.7 %) Läsionen detektiert. Das trainierte deep CNN zur Detektion von periapikalen osteolytischen Läsionen definierte eine Sensitivität von 97 % (95 % Konfidenzintervall: 95 % - 99 %) und eine Spezifität von 88 % (95 % Konfidenzintervall: 87 % - 89 %). Als erwartete Sensitivität und Spezifität wurden die unteren Grenzen der jeweiligen 95 % Konfidenzintervall angenommen.

Zur Überprüfung der Nicht-Unterlegenheit des Algorithmus auf Sensitivität im Vergleich zum Goldstandard wird der Non-inferiority margin auf 90 % festgelegt. Bei einem Stichprobenumfang von 243 Läsionen unter Verwendung eines einseitigen exakten Testes ($\alpha = 2.5\%$) und unter der Annahme, dass eine Sensitivität von 95 % beobachtet wird, kann eine Power von $> 80\%$ erreicht werden.

Der Annahme geschuldet, dass ca. 10 % der zu untersuchenden Zähne eine periapikale osteolytische Läsion aufweisen, wurde eine Fallzahl von mindestens 2430 Zähnen definiert.

Bezüglich Spezifität kann mit einem Stichprobenumfang von $n=476$ und unter der Annahme einer beobachteten Spezifität von 87 % sowie bei einem Non-inferiority margin von 82 % eine Power von $> 80\%$ erreicht werden (einseitiger exakter Test, $\alpha=2.5\%$).

Diese Fallzahlplanung erfolgte in Kooperation mit Dipl.-Ing. Dr. Regina Riedl vom Institut für Medizinische Informatik, Statistik und Dokumentation der Medizinischen Universität Graz.

4.1.7 CNN-Architektur

Zur automatischen PAL-Erkennung wurde in mehreren Prozessen ein neuronales Netzwerk entwickelt, trainiert und validiert. (<https://doi.org/10.1016/j.joen.2022.07.013>) In Folge der Arbeit von Kirnbauer et al. (6) (vgl. Abbildung 7) erfolgte in Kooperation mit dem Institut für Medizinische Informatik, Statistik und Dokumentation in der Arbeitsgruppe von Martin Urschler und Arnela Hadzic eine Vereinfachung des digitalen Verarbeitungsprozesses, so dass die Schritte der Zahn- und Läsionserkennung in einem durchgehenden Workflow zusammengeführt wurden (vgl. Abbildung 8).

Für die Testphase wurde ein Computer verwendet, der mit einem AMD Ryzen 5 5600G Prozessor mit 3,9 Gigahertz und einer NVIDIA GeForce GTX TITAN X Grafikkarte ausgestattet war. Als Betriebssystem kam Ubuntu 20.04 zum Einsatz und für die Programmierung wurde Python 3.7 in Verbindung mit TensorFlow 1.15.0 verwendet. Zur Leistungssteigerung der Hardware wurde ein Corsair Vengeance LPX Arbeitsspeicher mit 16 Gigabyte (2x8 GB) eingebaut. Insgesamt verfügte der Computer dank einer A-Data Solid-State-Drive (SSD) über 512 GB Speicherplatz. Dadurch konnte der gesamte Datensatz plus Segmentierung verschlüsselt gespeichert werden.

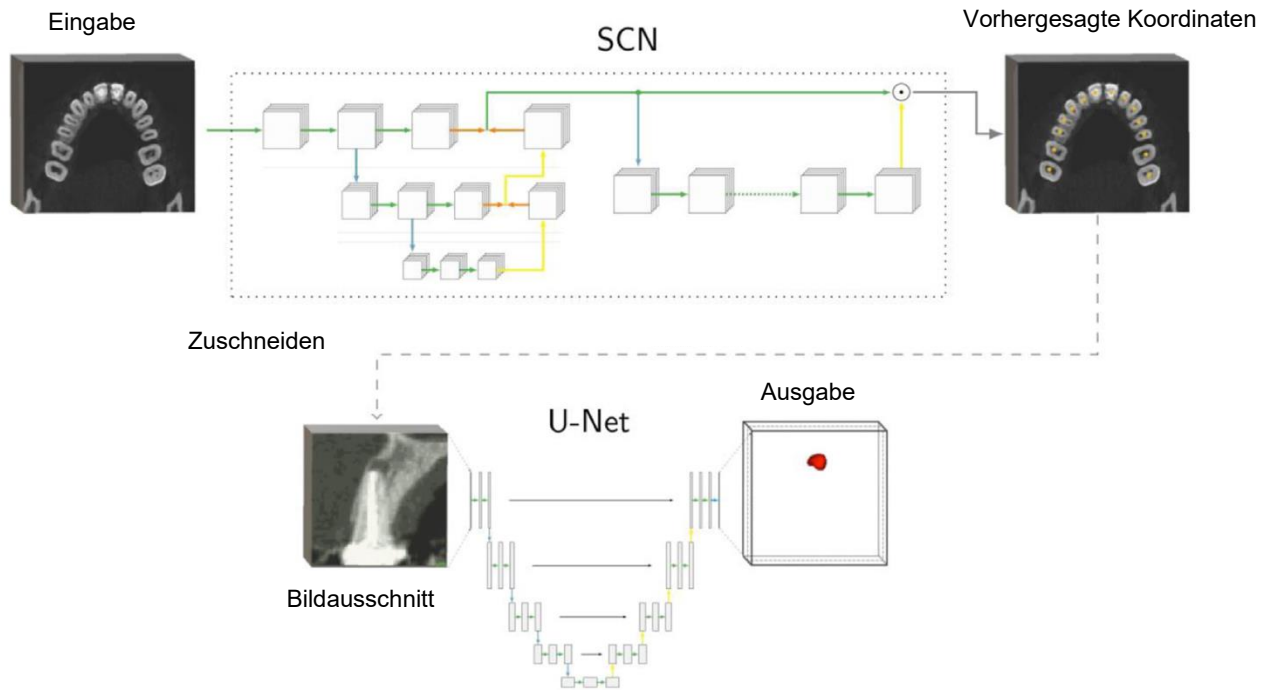


Abbildung 6: Schematische Darstellung der CNN-Architektur; SCN= Spatial-Konfigurationsnetzwerk (6)

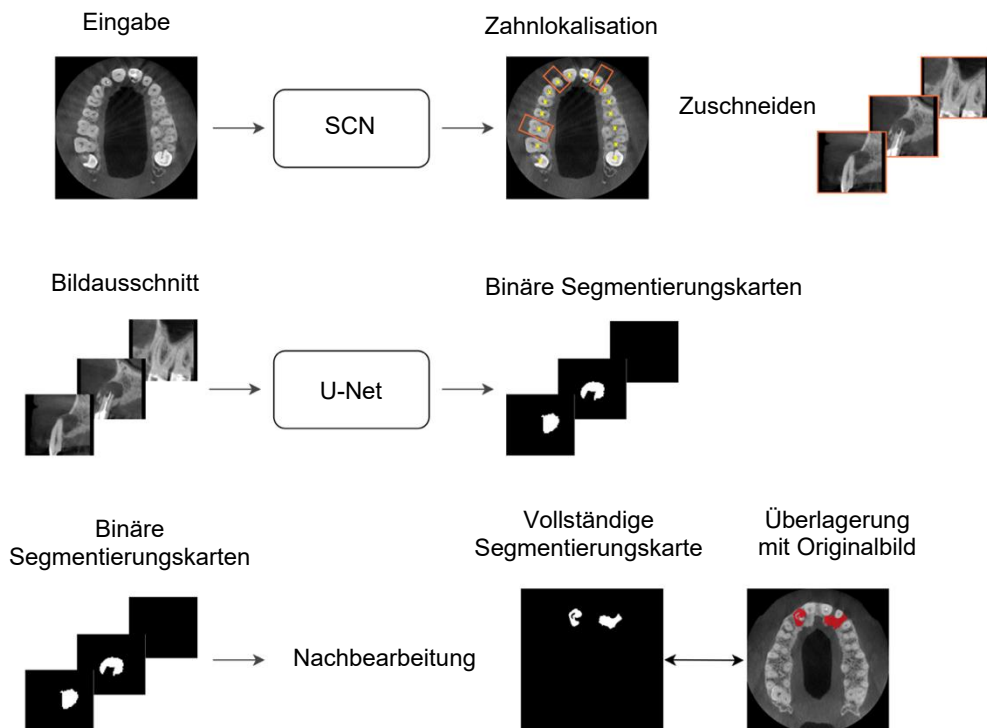


Abbildung 7: Schematische Darstellung des zusammengeführten Algorithmus'. SCN= Spatial-Konfigurationsnetzwerk (98)

5 Ergebnisse

5.1.1 Anzahl der ausgewerteten Datensätze

An der Abteilung für Orale Chirurgie und Kieferorthopädie der Univ.-Klinik für Zahnmedizin und Mundgesundheit Graz wurden 195 DVT-Datensätze aus dem Jahr 2018 entsprechend den Einschlusskriterien gescreent. Die Datensätze unterteilen sich in 164 Datensätze mit einem abgebildeten Kiefer und 31 Datensätze mit zwei Kiefern. Daraus resultierten 226 Kiefer, wobei 125 Oberkiefer und 101 Unterkiefer untersucht wurden.

Patienten	195
Kiefer	226 (OK/UK 125/101 55.3 %/44.6 %)
1 Kiefer (Ober- bzw. Unterkiefer)	164 (84.1 %)
2 Kiefer (Ober- und Unterkiefer)	31 (15,9 %)

Tabelle 4: Verteilung der Datensätze; OK= Oberkiefer, UK= Unterkiefer

5.1.2 Verteilung der beurteilten Zähne

Aus den 195 DVT-Aufnahmen wurden 2947 Zähne generiert, die in der Studie von den Expertinnen untersucht wurden. Aus dieser Zahl ergab sich eine gewisse Verteilung der erfassten Zähne. Diese Verteilung wurde nicht beeinflusst. In Tabelle 5 ist ersichtlich, dass jeder Zahn in der Studie vertreten ist und keine Zahngruppe signifikant häufiger oder seltener vertreten ist. Es war jedoch zu erwarten, dass die Frontzähne stärker vertreten sein würden als die Seitenzähne. Dementsprechend wurden 1297 Frontzähne (1er - 3er) und nur 860 Seitenzähne (6er - 8er) in der Studie berücksichtigt.

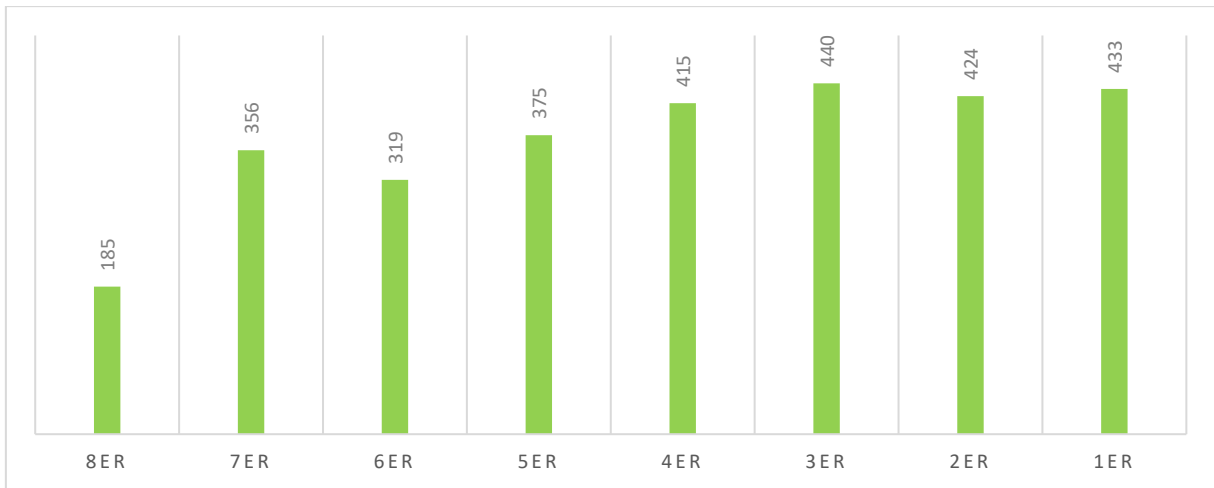


Tabelle 5: Verteilung der beurteilten Zähne

5.1.3 Verteilung nach Estrela

Als Nebenzielparamester wurden die detektierten Läsionen nach Estrela klassifiziert. (85) In der Studie wurden von den Expertinnen 300 periapikale osteolytische Läsionen in den DVT-Datensätzen diagnostiziert. Tabelle 6 zeigt die Verteilung der Estrela-Klassifikation. Es wurden 28 Läsionen der Klasse 1, 61 der Klasse 2, 67 der Klasse 3, 85 der Klasse 4 und 61 der Klasse 5 diagnostiziert.

Neben der allgemeinen Estrela-Klassifikation wurde auch die erweiterte Estrela-Klassifikation herangezogen, um Expansionen oder Destruktionen des kortikalen Knochens bei periapikalen osteolytischen Läsionen darzustellen. In 114 Fällen von insgesamt 303 Läsionen wurden solche Merkmale von den Expertinnen festgestellt. In der Tabelle 17 beschreibt der Buchstabe "E" eine Expansion, der Buchstabe "D" eine Destruktion der Kortikalis. Estrela Klasse 1 zeigt in keinem DVT-Datensatz eine Expansion oder Destruktion. Im Vergleich dazu wurde die Estrela-Klasse 2 55-mal diagnostiziert, wobei keine Expansion und 4-mal eine Destruktion nachgewiesen werden konnte. Eine Klasse 3 mit Expansion wurde 4-mal und eine mit Destruktion 10-mal nachgewiesen. Ab Klasse 4 zeigt sich, dass Expansionen und Destruktionen des Knochens häufiger auftreten. Hier liegt in 14 Fällen eine Expansion und in 33 Fällen eine Destruktion vor. Klasse 5 weist die höchste Anzahl an Destruktionen auf, wobei 9-mal eine Expansion diagnostiziert wurde. In 12 Fällen konnte weder eine Expansion noch eine Destruktion festgestellt werden. Tabelle 7 zeigt die 28 Läsionen des Indexscore 1 verteilt auf die jeweiligen Zahngruppen.

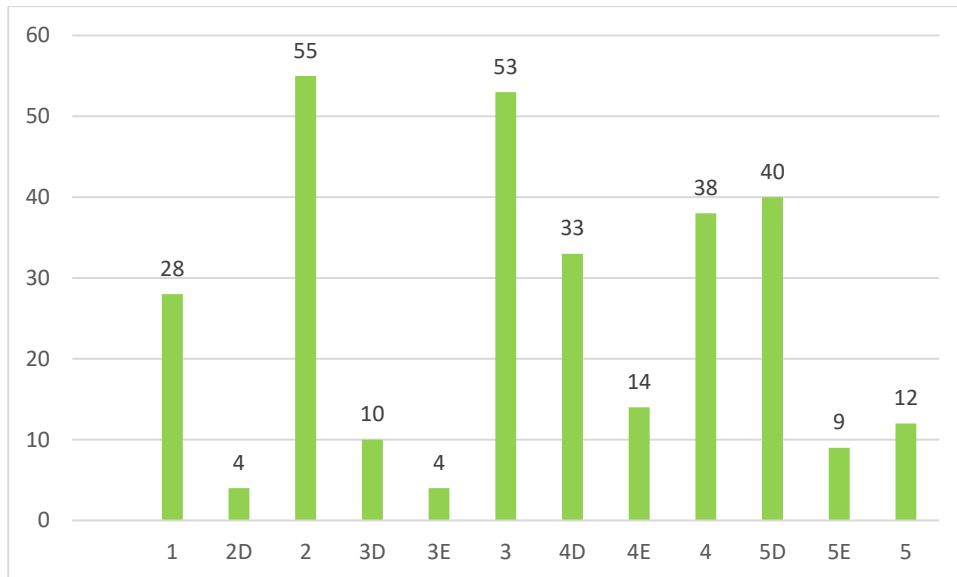


Tabelle 6: Verteilung der Läsionen nach Estrela et al. (85)

	Indexscore 1 (<1mm)
dritter Molar	3 (27.3 %)
zweiter Molar	4 (6.0 %)
erster Molar	3 (3.5 %)
zweiter Prämolare	6 (14.0 %)
erster Prämolare	2 (5.7 %)
Eckzähne	1 (7.7 %)
Seitliche Schneidezähne	3 (21.4 %)
Zentrale Schneidezähne	6 (18.8 %)
Gesamt	28 (9.3 %)

Tabelle 7: Verteilung der Läsionen auf den Indexwert 1 für jede Zahngruppe (98)

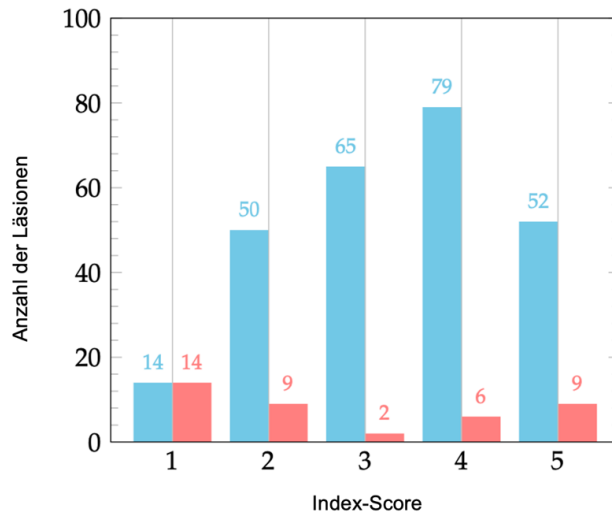


Tabelle 8: Verteilung der vorhergesagten Läsionen auf die genannten periapikalen Indexwerte nach Estrela et al. (85) Blaue Felder stellen wahr-positive Vorhersagen dar, während rote Felder falsch-negative Vorhersagen für einen bestimmten periapikalen Index darstellen. (98)

5.1.4 Diagnostik im Vergleich

Von den 2947 Zähnen, die von den Expertinnen beurteilt wurden, wiesen 2647 Zähne keine Läsionen auf und 300 Zähne zeigten eine periapikale Läsion. Im Gegensatz dazu erkannte die KI bei 2270 Zähnen keine Läsion und bei 719 Zähnen eine periapikale osteolytische Läsion. Dementsprechend beurteilte die KI insgesamt 2989 Zähne.

Tabelle 9 vergleicht das Verhältnis der Vorhersagen der Expertinnen und KI zueinander.

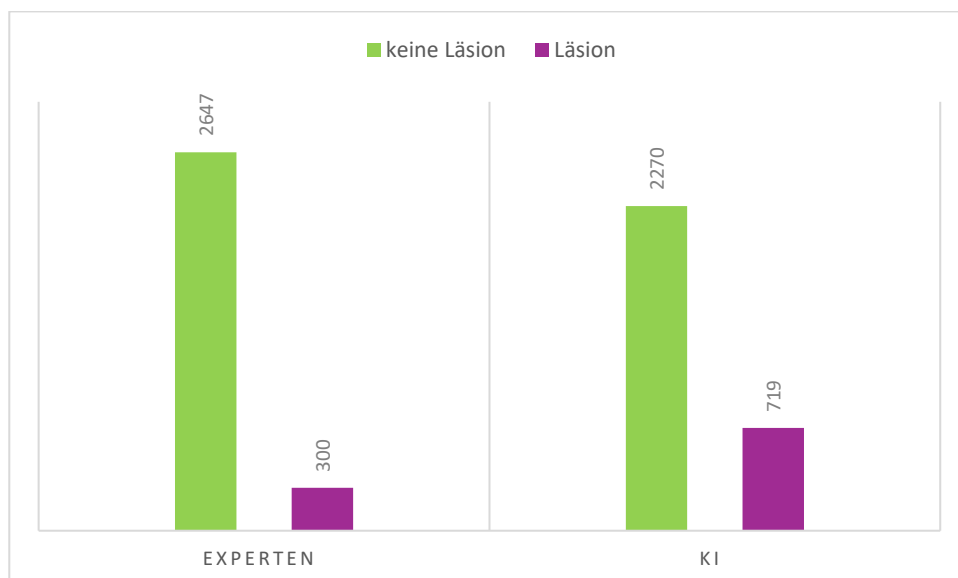


Tabelle 9: Diagnose von Expertinnen und KI im Vergleich

Expertinnen	KI	Häufigkeit
-	-	3251
-	1	42
0	0	2230
0	1	417
1	0	40
1	1	260

Tabelle 10: Vorhersagen der Expertinnen im Vergleich zu der KI. 0 steht für keine Läsion und 1 für eine detektierte Läsion. Beim Ausbleiben eines Feldes haben die Expertinnen bzw. die KI den Zahn als fehlend erkannt.

5.1.5 Ergebnisse nach Position

Tabelle 11 zeigt die Ergebnisse der Expertinnen und der KI getrennt nach der jeweiligen Position. Hier wird dargestellt, in welchen Fällen die KI keine Läsion detektiert hat und wie oft die Expertinnen im Vergleich dazu ebenfalls keine Läsion erkannten oder eine Läsion diagnostizierten. Ebenso wird dargestellt, wie oft die KI Läsionen detektierte und im Vergleich dazu die Expertinnen ebenfalls eine oder keine Läsionen detektierten. Zum Beispiel segmentierte die KI in 387 Fällen den 3er als gesund und in 383 Fällen bestätigten die Expertinnen diese Diagnose. Nur in 4 Fällen diagnostizierten die Expertinnen eine Läsion an dieser Stelle, an der die KI nicht reagierte. In 53 Fällen diagnostizierte die KI eine periapikale Läsion. In 44 Fällen erkannten auch die Expertinnen an dieser Stelle eine periapikale Läsion und in 9 Fällen wurden die 3er als gesund eingestuft. In Tabelle 11 sind die falsch-positiv detektierten leeren Alveolen (n=42) exkludiert. Somit hat die KI 677 falsch-positive Läsionen markiert.

Daten	Zähne	Beschreibung	Expertinnen haben keine Läsion detektiert	Expertinnen haben eine Läsion detektiert
KI	8er	keine Läsion	142	4
KI	8er	Läsion	32	7
KI	7er	keine Läsion	204	6
KI	7er	Läsion	85	61
KI	6er	keine Läsion	165	7
KI	6er	Läsion	69	78
KI	5er	keine Läsion	271	5
KI	5er	Läsion	61	38
KI	4er	keine Läsion	332	6
KI	4er	Läsion	48	29
KI	3er	keine Läsion	383	4
KI	3er	Läsion	44	9
KI	2er	keine Läsion	380	5
KI	2er	Läsion	30	9
KI	1er	keine Läsion	353	3
KI	1er	Läsion	48	29

Tabelle 11: Ergebnisse der Expertinnen und der KI im Vergleich nach Position

In Tabelle 12 werden ebenfalls die Ergebnisse der KI mit denen der Expertinnen verglichen. Hier wird jedoch zwischen Oberkiefer und Unterkiefer unterschieden. Es wurden im Unterkiefer 1076 Zähne als richtig gesund erkannt hat. Nur 16 Zähne wurden als falsch-negativ diagnostiziert. Im Oberkiefer wurden 1152 Zähne als richtig-

negativ erkannt und 24 Zähne als falsch-negativ beschrieben. Die KI diagnostizierte jedoch im Unterkiefer 169 Zähne als falsch-positiv. Im Oberkiefer wurden dagegen 248 Zähne als falsch-positiv erkannt.

Daten	Kiefer	Beschreibung	Expertinnen haben keine Läsion detektiert	Expertinnen haben eine Läsion detektiert
KI	OK	keine Läsion	1152	24
KI	OK	Läsion	248	172
KI	UK	keine Läsion	1076	16
KI	UK	Läsion	169	88

Tabelle 12: Ergebnisse der Expertinnen und der KI im Vergleich nach Kiefer

Tabelle 13 zeigt die Sensitivität und Spezifität für alle Zahngruppen und die Läsionen im OK und UK. Dabei wurden die Zahngruppen vom OK und UK zusammengefasst.

	Anzahl der Läsionen	Sensitivität (%)	Spezifität (%)
Gesamt	300	86.7	84.3
OK	169	87.76	82.31
UK	104	84.62	86.43
dritter Molar	11	63.64	81.61
zweiter Molar	67	91.04	70.59
erster Molar	85	91.76	70.51
zweiter Prämolare	43	88.37	81.63
erster Prämolare	35	82.86	87.37
Eckzähne	13	69.23	89.70
Seitliche Schneidezähne	14	64.29	92.68
Zentrale Schneidezähne	32	90.63	88.03

Tabelle 13: Sensitivität und Spezifität für alle vorhandenen Zahngruppen im OK und UK (98)

5.1.6 Sensitivität und Spezifität

Bei der Berechnung der Sensitivität und Spezifität wurden die Ergebnisse mit denen der Expertinnen verglichen. Die Ergebnisse der Expertinnen lieferten die richtig-positiven bzw. richtig-negativen Werte der periapikalen osteolytischen Läsionen in DVT-Datensätzen für die Studie und bildeten somit den Goldstandard. Wie in Tabelle 14 erkennbar ergab sich daraus eine Sensitivität von 86.7 % und die ermittelte Spezifität lag bei 82.9 %.

Sensitivität = 86.7 % (95 % CI exact: 82.3 % - 90.3 %) Nicht-Unterlegenheit p=0.975

Spezifität= 82.9 % (95 % CI exact: 81.5 % - 84.3 %) Nicht-Unterlegenheit p=0.109

Tabelle 14: Sensitivität und Spezifität der trainierten Künstlichen Intelligenz an vorhandenen und fehlenden Zähnen (inkludiert N=42 Zähne die von der KI als Läsion gewertet wurden, jedoch fehlten)

5.1.7 Auswertungszeit der KI

Der Mittelwert der Analysezeit der trainierten KI betrug 01:38 Minuten bei einer Standardabweichung von 00:16 Minuten.

Mittelwert (min)	00:01:38
Standardabweichung (min)	00:00:16

Tabelle 15: Mittelwert der Anwendungszeit der KI mit ihrer Standardabweichung

6 Diskussion

Modernste Methoden des maschinellen Lernens, insbesondere deep neuronale Netzwerke zur Unterstützung von Zahnärzt*innen bei der Erkennung und Segmentierung von Läsionen in zahnmedizinischen Bilddaten, haben weitreichende Fortschritte erzielt. Dennoch kämpfen diese mit Problemen aufgrund von mangelnder Transparenz und Generalisierung der Datensätze aus der klinischen Praxis. Außerdem ist es schwierig, bei der Entwicklung einen Goldstandard zu definieren, der auf alle Screening-Parameter anwendbar ist. In dieser Studie wurde ein kürzlich veröffentlichter Algorithmus zur automatischen Segmentierung periapikaler Läsionen aus DVT-Datensätzen gründlich evaluiert und auf Nichtunterlegenheit getestet. Der verwendete Algorithmus verbarg sich hinter einer grafischen Benutzeroberfläche, die eine Läsionssegmentierung aus einem Eingabebild des in der Studie verwendeten Testdatensatzes erzeugte. Der untersuchte Datensatz bestand aus 196 Proband*innen aus der klinischen Praxis. Im Vergleich zum Datensatz, der für die Entwicklung des Algorithmus verwendet wurde, spiegelte dieser Datensatz ein anspruchsvolleres, sehr repräsentatives Szenario aus der realen klinischen Praxis wider. Hier wurden im Vergleich zum Entwicklungsdatensatz entscheidende Parameter einbezogen, die im vorherigen Datensatz nicht berücksichtigt wurden und die Ergebnisse der Sensitivität und Spezifität beeinflussten. Dazu gehörte die Einschließung von teilweise und vollständig impaktierten Zähnen, leeren Alveolen, Zahnimplantaten und Augmentationen.

Darüber hinaus war das Screeningprotokoll sehr streng bei der Definition von falsch-positiven (FP) Ergebnissen, da bereits ein falsch-positives Voxel in der Segmentierung zu einer FP-Aussage führte. Auch Segmentierungen, welche nicht zahnbezogen waren, zum Beispiel außerhalb des Knochens, wurden als eine falsch-positive Läsion einem Zahn zugeordnet. Dies resultierte in einer sehr strengen Bewertung, lieferte aber ein realistisches Testszenario für das KI-Modell. Tabelle 5 zeigt die genaue Verteilung der erfassten Zähne, was ebenfalls für ein realistisches klinisches Szenario spricht. Eine Limitation dieses Datensatzes ist die geringe Anzahl der Weisheitszähne im Vergleich zu den anderen Zahngruppen (n=185). Dies ist darauf zurückzuführen, dass Weisheitszähne in einigen Fällen nicht angelegt oder im Laufe des Lebens extrahiert wurden.

Das verwendete KI-Modell konnte erfolgreich auf die vorhandenen und fehlenden Zähne von 195 Proband*innen angewendet werden (99,5 % der Probanden*innen des gesamten Datensatzes). Ein Datensatz konnte jedoch nicht gescreent werden. Der genaue Grund ist unbekannt, wahrscheinlich war die gespeicherte Datei defekt.

Die definierte Hauptzielgröße dieser Studie war die Anzahl der richtig-positiv (TPR=true positive rate / Sensitivität) und richtig-negativ (TNR=true negative rate / Spezifität) erkannten PALen im tiefen CNN im Vergleich zu den Ergebnissen von 2 Expertinnen.

Dabei ergab sich eine Sensitivität von 86,7 % und eine Spezifität von 82,9 % für die Segmentierung von PALen an vorhandenen und fehlenden Zähnen in diesem anspruchsvollen Szenario. Die Nicht-Unterlegenheitstests, die auf der Grundlage der Sensitivitäts- und Spezifitätsschätzungen aus der Proof-of-Concept-Evaluierung entwickelt wurden, lieferten ausreichend Evidenz, um die Nullhypothese für die Spezifität, nicht aber für die Sensitivität zurückzuweisen. Trotz der Verringerung bei der Sensitivität, erscheint der KI-Algorithmus vielversprechend, da in den Ergebnissen seine Qualität gezeigt werden konnte. Dennoch sind die Sensitivitäts- und Spezifitätsergebnisse in Vergleichsstudien höher. In der Studie von Setzer et al. (15) wurden im Gegensatz lediglich 20 ausgewählte DVT-Datensätze mit 61 Wurzeln analysiert, was nicht einem realen klinischen Szenario entspricht. In der Studie von Orhan et al. (16) wurden 109 DVT-Datensätze mit 153 PALen gescreent. Allerdings wurde der spezifische Kiefer- und Zahnbereich, der mit den periapikalen Läsionen assoziiert war, von einem menschlichen Beobachter bestimmt, was die Anzahl der FP-Ergebnisse in dieser Studie einschränkt. In der Studie von Saghiri et al. (33), der eine Sensitivität ihres KI-Modells von 93 % beschrieb, untersuchten nur 50 ausgewählte einkanalige Zähne. Der Vergleich mit anderen Studien und deren Ergebnissen ist schwierig, da sich jede Studie in der Anzahl der Datensätze, dem Auswertungsprotokoll und der Eingabemethode unterscheidet. In dieser Studie können das strenge Auswertungsprotokoll und der Testdatensatz aus der klinischen Praxis hervorgehoben werden.

Mögliche Ursachen für FP-Ergebnisse lassen sich gut darstellen. Es konnte festgestellt werden, dass die KI mehr falsch-positive Läsionen im OK (N=248) als im UK (N=169) detektierte. Der Algorithmus zeigte eine signifikant höhere Spezifität für Zähne im Unterkiefer, während der Unterschied in der Sensitivität zwischen Unter- und Oberkiefer statistisch nicht signifikant war. Dieser Rückgang der falsch-positiven

Befunde im Unterkiefer ist wahrscheinlich auf die bessere radiologische Beurteilbarkeit des Unterkiefers im Vergleich zum Oberkiefer zurückzuführen, da der Kontrast zwischen röntgentransparenten Läsionen und Alveolarknochen oder Zähnen im Unterkiefer höher ist. Zusätzlich spielte die Kieferhöhle eine entscheidende Rolle. Zähne, deren Wurzeln in die Kieferhöhle ragten oder eine enge anatomische Beziehung zur Kieferhöhle aufwiesen, wurden von der KI häufiger als krank markiert als Zähne, die weit von der Kieferhöhle entfernt standen. Des Weiteren waren FPs auf die Nähe zum Nervus incisivus, Nervus alveolaris inferior und Nervus mentalis zurückzuführen.

Artefakte, die durch Metallstifte oder Zahnimplantate verursacht wurden, können unabhängig vom Kiefer ebenfalls Probleme bei der Erkennung von Läsionen verursachen. Hier ist zu erwähnen, dass dreidimensionale Röntgenverfahren anfällig für solche metallischen Versorgungen sind. Selbst für Expert*innen können solche Artefakte die Interpretation und Analyse des Datensatzes erschweren. (96) Auffallend war auch, dass die KI in bestimmten Fällen augmentiertes Knochenmaterial als Läsion markierte, was auf die veränderte Struktur des entsprechenden Knochentransplants zurückzuführen ist. Ein weiterer Faktor, der häufig zu FPs führte, waren leere Alveolen. Tabelle 10 zeigt, dass in 42 Fällen die KI eine leere Alveole als FP segmentiert hatte. Werden diese aus der Studie exkludiert, steigt die Spezifität auf 84,3 % (vgl. Tabelle 13). Mittelwertig haben die Expert*innen pro Datensatz eine Läsion detektiert, während die KI pro Datensatz median drei Läsionen markierte.

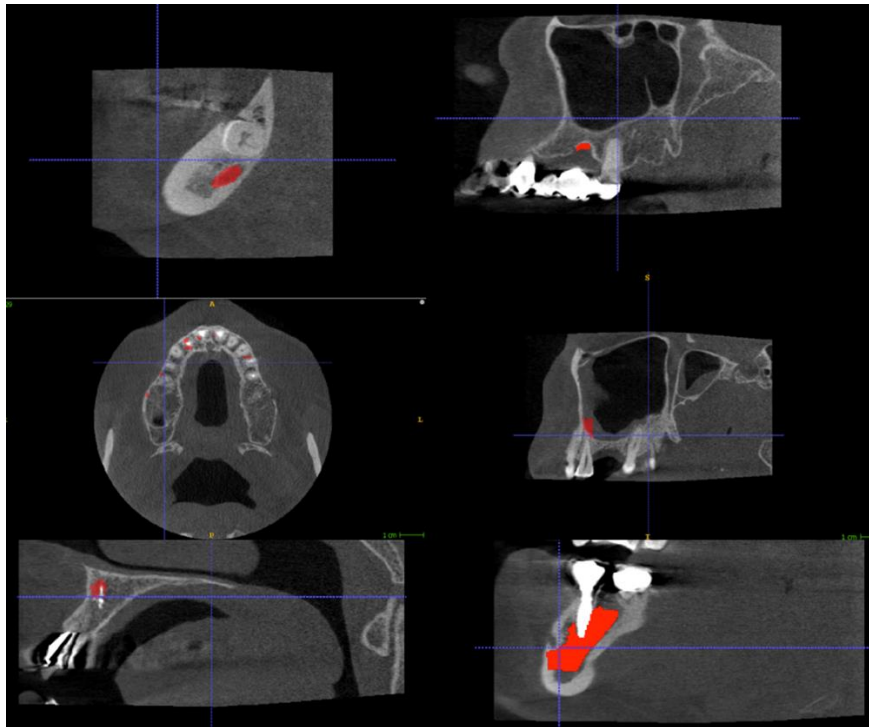


Abbildung 8: Beispielhafte KI-Detektion von FPs in den Datensätzen

Zu den Nebenzielparametern gehörte unter anderem auch die Auswertungszeit der KI. Zu den Hauptindikationen für die DVT-Bildgebung gehört wie oben erwähnt die Implantatplanung, die operative Entfernung von Weisheitszähnen und die Diagnostik oraler Pathologien. PALen sind oft Sekundärbefunde und bilden daher einen wichtigen Teil der DVT-Befundung. Die Dokumentation von Sekundärbefunden fordert viel Zeit, weswegen eine detaillierte Befundung oft nicht möglich ist. Der gezeigte KI-Algorithmus kann in solchen Fällen Abhilfe schaffen und unterstützend eingesetzt werden. Die durchschnittliche Auswertungszeit der KI betrug 01:38 Minuten. Dies bedeutet eine erhebliche Zeitersparnis für die Zahnärzt*innen und reduziert die Wahrscheinlichkeit, dass PALen übersehen oder Zähne als falsch-positiv bewertet werden. In wenigen Fällen betrug die Auswertungszeit (n=4) mehr als 3 Minuten, was auf die Komplexität des Datensatzes zurückgeführt werden kann. Die Software benötigt mehr Zeit, um größere Bilder mit mehr Zähnen bzw. zwei Kiefern zu verarbeiten, da jeder Zahn vor dem Segmentierungsschritt gecropped wird. Der Faktor Zeit ist ebenso abhängig vom Betriebssystem und der Rechenleistung der Hardware. Ein weiterer Nebenzielparameter war die Klassifizierung nach Estrela et al. (85) Auf Basis der Expert*innen wurde jede Läsion nach ihrem Läsionsdurchmesser in fünf verschiedene periapikale Index-Score-Kategorien unterteilt. Zusätzlich wurde in die Eigenschaften „E“ und „D“ unterschieden, was auf die Expansion und die Destruktion

des Knochens durch eine PAL hinweist. Tabelle 8 zeigt, dass je höher der Indexscore, desto geringer die falsch-negative Detektion der KI war. Kleine Läsionen mit einem Durchmesser von < 1mm (Indexscore 1) wurden in 50 % der Fälle als falsch-negativ detektiert. Im Vergleich dazu erreicht die Sensitivität nach Stratifizierung der Indexklassen 2-5 (> 1mm Läsionsdurchmesser) 90,4 %. Dementsprechend wird der Einsatz des vorgeschlagenen KI-Modells ab einer Läsionsgröße von 1 mm empfohlen. Auch in der Studie von Tsai et al. (99) wurde gezeigt, dass die DVT erst ab einer Läsionsgröße von ca. 1,4mm bei der Diagnostik von PALen konventionellen zweidimensionalen Röntgenbildern überlegen ist.

Werden die im Datensatz vertretenen Zahngruppen auf Sensitivität und Spezifität verglichen, fällt eine Abnahme der Spezifität bei den ersten und zweiten Molaren auf (vgl. Tabelle 13). Dies zeigt, dass eine enge anatomische Beziehung zur Kieferhöhle die Spezifität beeinflussen kann. Darüber hinaus kann gezeigt werden, dass die Spezifität der lateralen Schneidezähne höher ist als die der zentralen Schneidezähne, was auf einen möglichen Einfluss des Nervus incisivus auf die Ergebnisse hinweist. In der Tabelle 13 ist zusätzlich erkennbar, dass die dritten Molaren und die lateralen Schneidezähne die geringste Sensitivität aufweisen. Betrachtet man nun die Verteilung der Läsionsgrößen auf die jeweiligen Zahngruppen, wird deutlich, dass der Index-Score 1 prozentual häufiger bei diesen Zahngruppen vorhanden ist (vgl. Tabelle 7). Wie bereits erwähnt ist die Sensitivität bei dieser Läsionsgröße geringer, was bei der geringen Anzahl an Läsionen in diesen Zahngruppen großen Einfluss auf das Gesamtergebnis hat. Die Zahngruppe der Eckzähne weist ebenfalls eine geringe Anzahl von Läsionen auf (n=13). Auffällig ist, dass die Zahngruppen mit den wenigsten Läsionen auch die geringste Sensitivität besitzen. Eine geringe Anzahl an FPs haben dementsprechend einen großen Einfluss auf die Sensitivität in diesen Zahngruppen. Betrachtet man nur die Zahngruppen mit einer hohen Anzahl von Läsionen, so ergibt sich eine mittlere Sensitivität von 88,93 % (vgl. Tabelle 13).

Limitationen in dem verwendeten Datensatz sind zu erwähnen. Zum einen sind, wie bereits erwähnt, die Zahngruppen nicht gleichmäßig vertreten. Die dritten Molaren sind im Vergleich unterrepräsentiert. Zum anderen wurden Datensätze von Patient*innen unter 20 Jahren ausgeschlossen, damit sichergestellt werden konnte, dass das Wurzelwachstum von Zähnen vollständig abgeschlossen ist. Somit ist der getestete Algorithmus nicht vollständig auf jede Altersgruppe und nicht auf Zahnkeime anwendbar. Außerdem muss erwähnt werden, dass es bei der Erstellung eines

Goldstandards als klinischem Referenzstandard zu Fehlern kommen kann. (65) Diese Fehlerquelle lässt sich durch unser Studienprotokoll und die interdisziplinäre Untersuchung durch zwei Expertinnen minimieren. Völlig auszuschließen ist diese Fehlerquelle jedoch nicht. Des Weiteren stammen die Datensätze aus einem einzigen Krankenhaus und sind dementsprechend naturgemäß verzerrt. (64) Um das KI-Modell weiter zu modifizieren, müssen auch Datensätze von anderen Regionen und Krankenhäuser eingeschlossen werden, damit auch andere Ethnien berücksichtigt werden können. Darüber hinaus bietet sich an, in zukünftigen Studien Datensätze aus niedergelassen Praxen heranzuziehen, um eine praxisorientierte Entwicklung und Modifizierung zu gewährleisten und eine Heterogenität zu minimieren. Eine weitere Limitation besteht darin, dass eine geringe Anzahl an Läsionen mit dem Indexscore 1 repräsentiert sind. Zur Herstellung des Datensatzes wurde die DVT mit einer bestimmten Geräteeinstellung verwendet. Wie oben erwähnt handelt es sich um ein Gerät der Firma Planmeca. Um sicherzustellen, dass der trainierte Algorithmus auch für andere Hersteller bzw. Geräteeinstellungen zuverlässig anwendbar ist, muss auch hier ein Vergleichsdatensatz erstellt und analysiert werden. Zur Verbesserung des Algorithmus, ist der Datensatz zu erweitern und es muss sichergestellt werden, dass genannte Limitationen ausgeschlossen werden.

7 Conclusio

Zusammenfassend kann gesagt werden, dass KI in der modernen medizinischen Praxis unterstützend vielfach eingesetzt wird und die Einsatzbereiche stetig erweitert werden. Vor allem radiologische und andere Bilddaten können mittlerweile leicht und rasch bearbeitet und KI-basiert analysiert werden. Eine Vielzahl unterschiedlicher KI-Algorithmen werden eingesetzt, um optimale Ergebnisse in 2D- und 3D-Bildgebungsverfahren zu erzielen. Somit können KI-Algorithmen beispielsweise in der Endodontie, oralen Chirurgie und der Implantologie eingesetzt werden. Des Weiteren werden trainierte Module auch in der medizinischen Ausbildung und im Patientenmanagement in Kliniken und Praxen eingesetzt. Bei der Entwicklung von KI-Modellen müssen stets ethische Faktoren berücksichtigt werden um Schutz, Transparenz und Nachhaltigkeit im Umgang mit KI zu gewährleisten.

Das getestete KI-Modell in dieser Studie zeigt trotz des strengen Beurteilungsprotokolls und dem realen klinischen Datensatz vielversprechende Ergebnisse bei der Segmentierung von periapikalen osteolytischen Läsionen. Dabei erreicht die KI eine Sensitivität von 86,7 % und eine Spezifität von 82,9 %. Der Test auf Nichtunterlegenheit konnte für die Spezifität zurückgewiesen werden, jedoch nicht für die Sensitivität.

Zwar muss der trainierte Algorithmus weiter verfeinert werden, um tatsächlich Anwendung im klinischen Alltag finden zu können, doch konnten dank dieser Studie Fehler in der Segmentierung aufgezeigt werden, die in Folge eliminiert werden können. Zukünftige Datensätze müssen mehr Läsionen unter 1mm aufweisen, um eine Verzerrung zu vermeiden. Außerdem müssen mehr Läsionen an dritten Molaren sowie Eckzähnen und seitlichen Schneidezähnen inkludiert werden, damit FPs keine zu große Auswirkung auf das Ergebnis haben und dieses im Gegensatz zu stärker vertretenen Zahngruppen ausreißen. Wird das verwendete KI-Modell weiter verbessert und zeigt eine höhere Robustheit gegenüber Artefakten und Ausreißen, kann es im klinischen Alltag eingesetzt werden. Dafür wäre jedoch eine Zertifizierung als Medizinprodukt notwendig, was eine teure und aufwändige Konformitätserklärung gemäß Art 61 VO (EU) 745/2017 erforderlich machen würde. Dabei ist jedoch zu beachten, dass die KI nur unterstützend eingesetzt werden sollte und niemals den Faktor Mensch im medizinischen Alltag ersetzen kann.

8 Literaturverzeichnis

References

1. Akst J. A primer: artificial intelligence versus neural networks. Inspiring Innovation: The Scientist Exploring Life. 2019;65802.
2. Khanagar SB, Al-Ehaideb A, Maganur PC, Vishwanathaiah S, Patil S, Baeshen HA, et al. Developments, application, and performance of artificial intelligence in dentistry - A systematic review. J Dent Sci. 2021 January 01;16(1):508-22.
3. Hadzic A, Supervisor D, Štern, Advisor B, Kirnbauer. Automatic Detection of Teeth and Periapical Osteolytic Lesions in 3D CBCT Images with Deep CNNs MASTER'S THESIS to achieve the university degree of.
4. Schwendicke F, Samek W, Krois J. Artificial Intelligence in Dentistry: Chances and Challenges. J Dent Res. 2020 July 01;99(7):769-74.
5. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights into imaging. 2018;9:611-29.
6. Kirnbauer B, Hadzic A, Jakse N, Bischof H, Stern D. Automatic Detection of Periapical Osteolytic Lesions on Cone-beam Computed Tomography Using Deep Convolutional Neuronal Networks. J Endod. 2022;48(11):1434-40.
7. Payer C, Štern D, Bischof H, Urschler M. Integrating spatial configuration into heatmap regression based CNNs for landmark localization. Med Image Anal. 2019;54:207-19.
8. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18; Springer; 2015.
9. 18 Cutting-Edge Artificial Intelligence Applications in 2024 [Internet].; 2023 [updated 27.12.; cited 17.02.2024]. Available from: <https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/artificial-intelligence-applications>.
10. Künstliche Intelligenz: Die Flut gefälschter Fakten [Internet].; 2023 [updated 20.11.; cited 17.02.2024]. Available from: <https://www.lmu.de/de/newsroom/newsuebersicht/news/kuenstliche-intelligenz-die-flut-gefaelschter-fakten.html>.
11. Hung KF, Ai QYH, Wong LM, Yeung AWK, Li DTS, Leung YY. Current Applications of Deep Learning and Radiomics on CT and CBCT for Maxillofacial Diseases. Diagnostics (Basel). 2022 December 29;13(1):110. doi: 10.3390/diagnostics13010110.

12. Yu K, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng.* 2018 October 01;2(10):719-31.
13. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer.* 2018 August 01;18(8):500-10.
14. Thiesse P, Ollivier L, Di Stefano-Louineau D, Négrier S, Savary J, Pignard K, et al. Response rate accuracy in oncology trials: reasons for interobserver variability. Groupe Français d'Immunothérapie of the Fédération Nationale des Centres de Lutte Contre le Cancer. *Journal of Clinical Oncology.* 1997;15(12):3507-14.
15. Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RTH, Hermann G, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother Oncol.* 2015 March 01;114(3):345-50.
16. Wu W, Parmar C, Grossmann P, Quackenbush J, Lambin P, Bussink J, et al. Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology. *Front Oncol.* 2016 March 30;6:71.
17. Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, et al. Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. *Cancer Res.* 2017 July 15;77(14):3922-30.
18. Grossmann P, Stringfield O, El-Hachem N, Bui MM, Rios Velazquez E, Parmar C, et al. Defining the biological basis of radiomic phenotypes in lung cancer. *Elife.* 2017 July 21;6:10.7554/eLife.23421.
19. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. *Sci Rep.* 2015 August 17;5:13087.
20. Farahani N, Parwani AV, Pantanowitz L. Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. *Pathology and Laboratory Medicine International.* 2015:23-33.
21. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol.* 2019 May 01;20(5):e253-61.
22. Smith JS, Roitberg AE, Isayev O. Transforming Computational Drug Discovery with Machine Learning and AI. *ACS Med Chem Lett.* 2018 October 08;9(11):1065-9.
23. Zhong F, Xing J, Li X, Liu X, Fu Z, Xiong Z, et al. Artificial intelligence in drug design. *Science China Life Sciences.* 2018;61:1191-204.
24. Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol Divers.* 2021 August 01;25(3):1315-60.
25. Brown N, Ertl P, Lewis R, Luksch T, Reker D, Schneider N. Artificial intelligence in chemistry and drug design. *J Comput Aided Mol Des.* 2020;34:709-15.

26. Schwendicke F, Samek W, Krois J. Artificial Intelligence in Dentistry: Chances and Challenges. *J Dent Res*. 2020 July 01;99(7):769-74.
27. Olsen GF, Brilliant SS, Primeaux D, Najarian K. An image-processing enabled dental caries detection system. 2009 ICME International Conference on Complex Medical Engineering; IEEE; 2009.
28. Kositbowornchai S, Siriteptawee S, Plermkamon S, Bureerat S, Chetchotsak D. An artificial neural network for detection of simulated dental caries. *International Journal of Computer Assisted Radiology and Surgery*. 2006;1:91-6.
29. Geetha V, Aprameya KS, Hinduja DM. Dental caries diagnosis in digital radiographs using back-propagation neural network. *Health Inf Sci Syst*. 2020 January 03;8(1):8,y. eCollection 2020 Dec.
30. Javed S, Zakirulla M, Baig RU, Asif SM, Meer AB. Development of artificial neural network model for prediction of post-streptococcus mutans in dental caries. *Comput Methods Programs Biomed*. 2020;186:105198.
31. Ossowska A, Kusiak A, Swietlik D. Artificial Intelligence in Dentistry-Narrative Review. *Int J Environ Res Public Health*. 2022 March 15;19(6):3449. doi: 10.3390/ijerph19063449.
32. Aminoshariae A, Kulild J, Nagendrababu V. Artificial intelligence in endodontics: current applications and future directions. *J Endod*. 2021;47(9):1352-7.
33. Saghiri MA, Asgar K, Boukani KK, Lotfi M, Aghili H, Delvarani A, et al. A new approach for locating the minor apical foramen using an artificial neural network. *Int Endod J*. 2012;45(3):257-65.
34. Khanagar SB, Alfadley A, Alfouzan K, Awawdeh M, Alaqla A, Jamleh A. Developments and Performance of Artificial Intelligence Models Designed for Application in Endodontics: A Systematic Review. *Diagnostics (Basel)*. 2023 January 23;13(3):414. doi: 10.3390/diagnostics13030414.
35. Setzer FC, Shi KJ, Zhang Z, Yan H, Yoon H, Mupparapu M, et al. Artificial intelligence for the computer-aided detection of periapical lesions in cone-beam computed tomographic images. *J Endod*. 2020;46(7):987-93.
36. Orhan K, Bayrakdar IS, Ezhov M, Kravtsov A, Özyürek T. Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. *Int Endod J*. 2020;53(5):680-9.
37. Ekert T, Krois J, Meinhold L, Elhennawy K, Emara R, Golla T, et al. Deep learning for the radiographic detection of apical lesions. *J Endod*. 2019;45(7):917,922. e5.
38. Kositbowornchai S, Plermkamon S, Tangkosol T. Performance of an artificial neural network for vertical root fracture detection: an ex vivo study. *Dental traumatology*. 2013;29(2):151-5.

39. Bindal P, Bindal U, Lin CW, Kasim NHA, Ramasamy TSA, Dabbagh A, et al. Neuro-fuzzy method for predicting the viability of stem cells treated at different time-concentration conditions. *Technology and Health Care*. 2017;25(6):1041-51.
40. Z. Zheng, H. Yan, F. C. Setzer, K. J. Shi, M. Mupparapu, J. Li. Anatomically Constrained Deep Learning for Automating Dental CBCT Segmentation and Lesion Detection. *IEEE Transactions on Automation Science and Engineering*. 2021;18(2):603-14.
41. Kim BS, Yeom HG, Lee JH, Shin WS, Yun JP, Jeong SH, et al. Deep Learning-Based Prediction of Paresthesia after Third Molar Extraction: A Preliminary Study. *Diagnostics (Basel)*. 2021 August 30;11(9):1572. doi: 10.3390/diagnostics11091572.
42. Wright JM, Vered M. Update from the 4th Edition of the World Health Organization Classification of Head and Neck Tumours: Odontogenic and Maxillofacial Bone Tumors. *Head Neck Pathol*. 2017 March 01;11(1):68-77.
43. Liu Z, Liu J, Zhou Z, Zhang Q, Wu H, Zhai G, et al. Differential diagnosis of ameloblastoma and odontogenic keratocyst by machine learning of panoramic radiographs. *International Journal of Computer Assisted Radiology and Surgery*. 2021;16:415-22.
44. Lee J, Kim D, Jeong S. Diagnosis of cystic lesions using panoramic and cone beam computed tomographic images based on deep learning neural network. *Oral Dis*. 2020;26(1):152-8.
45. Bispo MS, Pierre Júnior, Mário Lúcio Gomes de Queiroz, Apolinário Jr AL, Dos Santos JN, Junior BC, Neves FS, et al. Computer tomographic differential diagnosis of ameloblastoma and odontogenic keratocyst: classification using a convolutional neural network. *Dentomaxillofacial Radiology*. 2021;50(7):20210002.
46. Chai Z, Mao L, Chen H, Sun T, Shen X, Liu J, et al. Improved diagnostic accuracy of ameloblastoma and odontogenic keratocyst on cone-beam CT by artificial intelligence. *Frontiers in Oncology*. 2022;11:793417.
47. Liu Z, Liu J, Zhou Z, Zhang Q, Wu H, Zhai G, et al. Differential diagnosis of ameloblastoma and odontogenic keratocyst by machine learning of panoramic radiographs. *Int J Comput Assist Radiol Surg*. 2021 March 01;16(3):415-22.
48. Kwak Y, Nguyen V, Hériveaux Y, Belanger P, Park J, Haïat G. Ultrasonic assessment of osseointegration phenomena at the bone-implant interface using convolutional neural network. *J Acoust Soc Am*. 2021;149(6):4337-47.
49. Fuglsig, Joao Marcus de Carvalho E Silva, Reis INRD, Yeung AWK, Bornstein MM, Spin-Neto R. The current role and future potential of digital diagnostic imaging in implant dentistry: A scoping review. *Clin Oral Implants Res*. 2023 November 22.
50. Jung S, Lim H, Lee S, Cho Y, Song I. Deep active learning for automatic segmentation of maxillary sinus lesions using a convolutional neural network. *Diagnostics*. 2021;11(4):688.

51. Hung KF, Ai QYH, King AD, Bornstein MM, Wong LM, Leung YY. Automatic detection and segmentation of morphological changes of the maxillary sinus mucosa on cone-beam computed tomography images using a three-dimensional convolutional neural network. *Clin Oral Investig*. 2022;26(5):3987-98.
52. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017 June 21;2(4):230-43.
53. Agrawal P, Nikhade P. Artificial Intelligence in Dentistry: Past, Present, and Future. *Cureus*. 2022 July 28;14(7):e27405.
54. Dzyuba N, Jandu J, Yates J, Kushnerev E. Virtual and augmented reality in dental education: The good, the bad and the better. *Eur J Dental Education*. 2022 - 11-20.
55. Dictionary OE. Oxford english dictionary. Simpson, Ja & Weiner, Esc. 1989;3.
56. Dută M, Amariei CI, Bogdan CM, Popovici DM, Ionescu N, Nuca CI. An overview of virtual and augmented reality in dental education. *Oral Health Dent Manag*. 2011;10(1):42-9.
57. Vincent M, Joseph D, Amory C, Paoli N, Ambrosini P, Mortier É, et al. Contribution of haptic simulation to analogic training environment in restorative dentistry. *J Dent Educ*. 2020;84(3):367-76.
58. Murbay S, Chang JWW, Yeung S, Neelakantan P. Evaluation of the introduction of a dental virtual simulator on the performance of undergraduate dental students in the pre-clinical operative dentistry course. *European Journal of Dental Education*. 2020;24(1):5-16.
59. Al-Saud LM, Mushtaq F, Allsop MJ, Culmer PC, Mirghani I, Yates E, et al. Feedback and motor skill acquisition using a haptic dental simulator. *European Journal of Dental Education*. 2017;21(4):240-7.
60. Pulijala Y, Ma M, Pears M, Peebles D, Ayoub A. Effectiveness of immersive virtual reality in surgical training—a randomized control trial. *Journal of Oral and Maxillofacial Surgery*. 2018;76(5):1065-72.
61. Alexander B, John S, Aralamoodu PO. Artificial intelligence in dentistry: Current concepts and a peep into the future. *Int J Adv Res*. 2018;6(12):1105-8.
62. Khanna SS, Dhaimade PA. Artificial intelligence: transforming dentistry today. *Indian J Basic Appl Med Res*. 2017;6(3):161-7.
63. Schwendicke F, Golla T, Dreher M, Krois J. Convolutional neural networks for dental image diagnostics: A scoping review. *J Dent*. 2019 December 01;91:103226.
64. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA Intern Med*. 2018 November 01;178(11):1544-7.

65. Walsh T. Fuzzy gold standards: Approaches to handling an imperfect reference standard. *J Dent.* 2018 July 01;74 Suppl 1:S47-9.
66. Maddox TM, Rumsfeld JS, Payne PR. Questions for artificial intelligence in health care. *JAMA.* 2019;321(1):31-2.
67. Verordnung (EU) 2017/745 des Europäischen Parlaments und des Rates vom 5. April 2017 über Medizinprodukte, zur Änderung der Richtlinie 2001/83/EG, der Verordnung (EG) Nr. 178/2002 und der Verordnung (EG) Nr. 1223/2009 und zur Aufhebung der Richtlinien 90/385/EWG und 93/42/EWG des Rates (Text von Bedeutung für den EWR.) [Internet].; 2023 [updated 20.03.; cited 24.02.2024].
68. UNESCO- Empfehlung zur Ethik Künstlicher Intelligenz Bedingungen zur Implementierung in Deutschland.
69. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines.* 2018 -11-26;28(4):689.
70. Mason OR. No title. Four ethical issues of the information age. Retrieved June 28, 2007. 1986.
71. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nature machine intelligence.* 2019;1(9):389-99.
72. McBride N. ACTIVE Ethics: An Information Systems Ethics for the Information Age. . 2014.
73. Schutz S, Daten. Schutz sensibler Daten Position der Gesundheitssektionen VIII und IX des BMASGK.
74. Prinzipien für die ethische Nutzung künstlicher Intelligenz, 59, , 2. Sess. (2022).
75. Sidorenko EL, Khisamova ZI, Monastyrsky UE. The main ethical risks of using artificial intelligence in business. . 2021:423-9.
76. Future of Life Institute. Das EU-Gesetz zur künstlichen Intelligenz. . 2024.
77. Zaleckiene V, Peciuliene V, Brukiene V, Drukteinis S. Traumatic dental injuries: etiology, prevalence and possible outcomes. *Stomatologija.* 2014;16(1):7-14.
78. Moller AJ, Fabricius L, Dahlen G, Ohman AE, Heyden G. Influence on periapical tissues of indigenous oral bacteria and necrotic pulp tissue in monkeys. *Scand.J.Dent.Res.* 1981 December 01;89(6):475-84.
79. Segura-Egea JJ, Gould K, Sen BH, Jonasson P, Cotti E, Mazzoni A, et al. Antibiotics in Endodontics: a review. *Int.Endod.J.* 2017 December 01;50(12):1169-84.
80. Abbott PV. Classification, diagnosis and clinical manifestations of apical periodontitis. . 2004;8(1):36-54.

81. Ramachandran Nair PN. Light and electron microscopic studies of root canal flora and periapical lesions. *J.Endod.* 1987 January 01;13(1):29-39.
82. Strindberg L. The dependence of the results of pulp therapy on certain factors. *Acta Odontol.Scand.* 1956;14(21):1-175.
83. REIT C, GRÖNDAHL H. Application of statistical decision theory to radiographic diagnosis of endodontically treated teeth. *Eur.J.Oral Sci.* 1983;91(3):213-8.
84. Ørstavik D, Kerekes K, Eriksen HM. The periapical index: a scoring system for radiographic assessment of apical periodontitis. . 1986;2(1):20-34.
85. Estrela C, Bueno MR, Azevedo BC, Azevedo JR, Pecora JD. A new periapical index based on cone beam computed tomography. *J.Endod.* 2008 November 01;34(11):1325-31.
86. Esposito S, Cardaropoli M, Cotti E. A suggested technique for the application of the cone beam computed tomography periapical index. *Dentomaxillofac.Radiol.* 2011 December 01;40(8):506-12.
87. Åberg F, Helenius-Hietala J, Meurman J, Isoniemi H. Association between dental infections and the clinical course of chronic liver disease. . 2014;44(3):349-53.
88. Schulze RKW, Drage NA. Cone-beam computed tomography and its applications in dental and maxillofacial radiology. *Clin.Radiol.* 2020 September 01;75(9):647-57.
89. Pauwels R, Beinsberger J, Stamatakis H, Tsiklakis K, Walker A, Bosmans H, et al. Comparison of spatial and contrast resolution for cone-beam computed tomography scanners. . 2012;114(1):127-35.
90. Suomalainen A, Pakbaznejad Esmaeili E, Robinson S. Dentomaxillofacial imaging with panoramic views and cone beam CT. *Insights Imaging.* 2015 February 01;6(1):1-16.
91. Widmann G, Bischel A, Stratis A, Bosmans H, Jacobs R, Gassner E, et al. Spatial and contrast resolution of ultralow dose dentomaxillofacial CT imaging using iterative reconstruction technology. *Dentomaxillofac.Radiol.* 2017 April 01;46(4):20160452.
92. Brullmann D, Schulze RKW. Spatial resolution in CBCT machines for dental/maxillofacial applications-what do we know today? *Dentomaxillofac.Radiol.* 2015;44(1):20140204.
93. Shatskiy I. Effective Doses and Radiation Risks from Common Dental Radiographic, Panoramic and Cbct Examinations. *Radiat.Prot.Dosimetry.* 2021 October 12;195(3-4):296-305.
94. Granlund C, Thilander-Klang A, Ylhan B, Lofthag-Hansen S, Ekestubbe A. Absorbed organ and effective doses from digital intra-oral and panoramic radiography applying the ICRP 103 recommendations for effective dose estimations. *Br.J.Radiol.* 2016 October 01;89(1066):20151052.

95. Pauwels R, Seynaeve L, Henriques JCG, de Oliveira-Santos C, Souza PC, Westphalen FH, et al. Optimization of dental CBCT exposures through mAs reduction. *Dentomaxillofac.Radiol.* 2015;44(9):20150108.
96. Lofthag-Hansen S, Huumonen S, Grondahl K, Grondahl H. Limited cone-beam CT and intraoral radiography for the diagnosis of periapical pathology. *Oral Surg.Oral Med.Oral Pathol.Oral Radiol.Endod.* 2007 January 01;103(1):114-9.
97. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage.* 2006 July 01;31(3):1116-28.
98. Hadzic A, Urschler M, Press JA, Riedl R, Rugani P, Stern D, et al. Evaluating a Periapical Lesion Detection CNN on a Clinically Representative CBCT Dataset-A Validation Study. *J.Clin.Med.* 2023 December 29;13(1):197. doi: 10.3390/jcm13010197.
99. Tsai P, Torabinejad M, Rice D, Azevedo B. Accuracy of cone-beam computed tomography and periapical radiography in detecting small periapical lesions. *J.Endod.* 2012 July 01;38(7):965-70.