Aus der

Poliklinik für Zahnerhaltung und Parodontologie

Klinikum der Ludwig-Maximilians-Universität München

Erkennung von parodontalem Knochenabbau auf periapikalen Röntgenbildern mittels künstlicher Intelligenz

Dissertation

zum Erwerb des Doktorgrades der Zahnmedizin

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

vorgelegt von

Patrick Hoss

aus

Albstadt

Jahr

2024

Mit Genehmigung der Medizinischen Fakultät der

Ludwig-Maximilians-Universität München

Erstes Gutachten:	Prof. Dr. Jan Kühnisch
Zweites Gutachten:	Prof. Dr. Roswitha Heinrich-Weltzien
Drittes Gutachten:	Prof. Dr. Thomas Pfluger
Dekan:	Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung:	01.07.2024

Affidavit

Eidesstattliche Versicherung

Hoss, Patrick
Name, Vorname

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel:

„Erkennung von parodontalem Knochenabbau auf periapikalen Röntgenbildern mittels künstlicher Intelligenz“

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 14.02.2024
Ort, Datum

Patrick Hoss
Unterschrift Doktorandin bzw. Doktorand

Inhaltsverzeichnis

Publikationsliste

Hoss, P., Meyer, O., Wölfle, U. C., Wülk, A., Meusburger, T., Meier, L., Hickel, R., Gruhn, V., Hesenius, M., Kühnisch, J., Dujic, H. (2023). Detection of Periodontal Bone Loss on Periapical Radiographs—A Diagnostic Study Using Different Convolutional Neural Networks. *Journal of Clinical Medicine*, 12(22), 7189. https://doi.org/10.3390/jcm12227189

Dujic, H., Meyer, O., Hoss, P., Wölfle, U. C., Wülk, A., Meusburger, T., Meier, L., Gruhn, V., Hesenius, M., Hickel, R., Kühnisch, J. (2023). Automatized Detection of Periodontal Bone Loss on Periapical Radiographs by Vision Transformer Networks. *Diagnostics*, 13(23), 3562. https://doi.org/10.3390/diagnostics13233562

Abkürzungsverzeichnis

CAL	Klinischer Attachmentverlust
KI	Künstliche Intelligenz
ML	Machine Learning
DL	Deep Learning
NN	Neural Network
CNN	Convolutional Neural Network
ACC	Accuracy
AUC	Fläche unter der Receiver Operating Characteristic Curve

1. Beitrag zu den Veröffentlichungen

1.1 Beitrag zu Veröffentlichung I

Publikation "Detection of Periodontal Bone Loss on Periapical Radiographs—A Diagnostic Study Using Different Convolutional Neural Networks"

	Patrick Hoss	**Prof. Dr. Jan Kühnisch**	**Co-Autoren**
Ethikantrag	-	100 %	-
Literaturrecherche	100 %	-	-
Projektidee/ Studiendesign	30 %	50 %	20 %
Röntgenbild-Auswahl	80 %	-	20 %
Befundung der Röntgenbilder	80 %	-	20 %
Modell-Training & Datenanalyse	-	-	100%
Dateninterpretation und Auswahl veröffentlichungswürdiger Daten	90 %	-	10 %
Manuskript Management bis zur Veröffentlichung	80 %	-	20 %
Verfassen und Einreichen der Dissertation	100%	-	-

1.2 Beitrag zu Veröffentlichung II

Publikation "Automatized Detection of Periodontal Bone Loss on Periapical Radiographs by Vision Transformer Networks"			
	Patrick Hoss	**Prof. Dr. Jan Kühnisch**	**Co-Autoren**
Ethikantrag	-	100 %	-
Literaturrecherche	-	-	100 %
Projektidee/ Studiendesign	10 %	50 %	40 %
Röntgenbild-Auswahl	20 %	-	80 %
Befundung der Röntgenbilder	20 %	-	80 %
Modell-Training & Datenanalyse	-	-	100%
Dateninterpretation und Auswahl veröffentlichungswürdiger Daten	10 %	-	90 %
Manuskript Management bis zur Veröffentlichung	20%	-	80 %

2. Einleitung

In der deutschen Bevölkerung ist neben einer rückläufigen Anzahl fehlender Zähne in der Gruppe der jüngeren Senioren (65- bis 74-Jährige) auch eine Halbierung der vollständigen Zahnlosigkeit zwischen 1997 und 2014 auf 12,4% in besagter Altersgruppe festzustellen (Jordan und Micheelis, 2016). Dies lässt darauf schließen, dass letztlich mehr Zähne bis ins hohe Alter erhalten bleiben können. Diese Entwicklung im Bereich der Zahngesundheit wird somit dazu führen, dass der Behandlungsbedarf im Feld parodontaler Erkrankungen steigen wird und dementsprechend einer adäquaten Diagnostik und Therapie parodontal geschädigter Zähne enorme Bedeutung entgegenkommt.

Unter dem Krankheitsbild der Parodontitis versteht man eine multifaktoriell bedingte, entzündliche Erkrankung, die insbesondere mit der Besiedelung der Wurzeloberfläche mit einem dysbiotischen Biofilm in Verbindung gebracht wird (Abdulkareem et al., 2023). Wie auch unter Berücksichtigung der Empfehlungen des World Workshops 2017 zur Klassifizierung von Parodontalerkrankungen zu erkennen ist, ist die Parodontitis in der initialen Phase vornehmlich klinisch unter anderem über Messung der Sondierungstiefen und Beurteilung des klinischen Attachmentverlusts (CAL) zu identifizieren (Papapanou et al., 2018; Tonetti et al., 2018), da die röntgenologische Einschätzung des parodontalen Knochenabbaus speziell bei Defekten geringer bukkolingualer Breite und Tiefe eingeschränkt ist (Pepelassi et al., 2000). Schreitet der Krankheitsverlauf jedoch voran, nimmt die Relevanz der röntgenologischen Beurteilung von parodontalen Knochendefekten deutlich zu, da diese dann röntgenologisch präziser veranschaulicht werden können (Fiorellini et al., 2021).

Sowohl periapikale Röntgenbilder als auch Panoramaschichtaufnahmen werden als röntgenologische Diagnostikmethoden zur Beurteilung von parodontalem Knochenabbau verwendet. Dabei ist zu betonen, dass es sich bei der röntgenologischen Bewertung des parodontalen Knochenabbaus um eine herausfordernde Aufgabe für den Zahnarzt handelt, da diese durch mögliche Überlagerungen anatomischer Strukturen und verschiedene Projektionswinkel erschwert sein kann. Dies könnte auch eine Erklärung dafür sein, warum die

röntgenologische Einschätzung des parodontalen Knochenabbaus durch Zahnärzte beispielsweise im Vergleich zu kariösen Läsionen oder auch Paro-Endo-Läsionen zu uneinheitlicheren Ergebnissen führt und somit weniger zuverlässig erscheint (Meusburger et al., 2023). Um solche diagnostischen Differenzen zu verringern, wäre es sinnvoll, die röntgenologische Diagnostik von parodontalem Knochenabbau durch die Verwendung von Algorithmen, die auf künstlicher Intelligenz (KI) basieren, zu objektivieren.

Unter KI versteht man die Möglichkeit, dass Maschinen bzw. Computer bestimmte Aufgaben ausführen können, die normalerweise mit menschlicher Intelligenz und Verhaltensweise assoziiert sind (Schwendicke et al., 2020). Es gibt verschiedene Teilbereiche der KI. Hierbei zu nennen sind unter anderem „machine learning“ (ML) sowie „deep learning“ (DL). Beim ML werden Datensätze angewendet, die zum Trainieren des Modells verwendet werden. Nach Identifizierung und Erlernung der statistischen Datenmuster kann die Leistung des ML-Modells anhand eines noch nicht zur Verfügung gestellten Testdatensatzes beurteilt werden und auf Grundlage des Erlernten Vorhersagen über noch nicht gesehene Daten getroffen werden (Schwendicke et al., 2021). Je nachdem, welche Art an Datensätzen den ML-Modellen zur Verfügung gestellt werden, können beispielsweise Aussagen über Diagnosen, Therapieempfehlungen und auch zukünftige Krankheitsverläufe getroffen werden (Pethani, 2021). Die Genauigkeit der seitens des ML-Modells getätigten Vorhersagen sind aber maßgeblich von der Qualität der Daten abhängig, die zum Training des ML-Modells verwendet wurden (Pethani, 2021). Als eine Weiterentwicklung von ML kann DL angesehen werden. Beim DL werden Algorithmen in zahlreichen Schichten organisiert, um ein künstliches neuronales Netz (engl. „neural network“, NN) aufzubauen, welches selbstständig lernen und eigene Entscheidungen treffen kann (Greenhill et al., 2020). Dabei werden unter anderem selbstlernende Rückwärtspropagierungs-Algorithmen verwendet, die durch Fehlerrückführung durch das Netz den Einfluss von neuronalen Strukturen auf den vorliegenden Fehler ermitteln. Durch Anpassung des Einflusses der entsprechenden Neuronenverbindungen können dann die Vorhersagen schrittweise präzisiert werden (Hwang et al., 2019).

Tatsächlich wird der KI in der Zahnmedizin zunehmend Aufmerksamkeit gewidmet und wird so unter anderem in Zusammenhang mit der Diagnostik von zahnmedizinischen Pathologien gebracht. In der Literatur wird vereinzelt der KI-Einsatz bei der Auswertung klinischer, dentaler Fotografien dokumentiert. Stellvertretend seien hier die Studien zur KI-unterstützten, automatisierten Kariesdetektion (Kühnisch et al., 2022) sowie zur Erkennung von Zähnen, die von einer zunehmend prävalenten Molaren-Inzisiven-Hypomineralisation betroffen sind (Schönewolf et al., 2022), zu nennen. Mehrheitlich befassten sich jedoch in jüngster Vergangenheit zahlreiche Arbeitsgruppen mit der KI-basierten Diagnostik an Röntgenbildern, um mittels DL-basierten, faltenden neuronalen Netzen (engl. „convolutional neural network“, CNN) verschiedene Pathologien zu detektieren. So sind Studien mit dem Ziel der automatisierten, röntgenologischen Karieserkennung mittels KI zu finden, bei welchen sowohl intraorale Röntgenbilder (Chen et al., 2022; Lee et al., 2018a) als auch Panoramaschichtaufnahmen (Lian et al., 2021) verwendet wurden. Des Weiteren wurde die KI bereits dafür eingesetzt, um in zahnärztlichen Röntgenbildern apikale Läsionen aufzufinden (Moidu et al., 2022; Li et al., 2021a; Ekert et al., 2019). Auch die KI-unterstützte Ermittlung von parodontalem Knochenabbau wurde bereits auf intraoralen Röntgenbildern (Chen et al., 2023; Alotaibi et al., 2022; Lee et al., 2022; Tsoromokos et al., 2022; Lee et al., 2018b) und auf Panoramaschichtaufnahmen (Jiang et al., 2022; Widyaningrum et al., 2022; Li et al., 2021b; Kim et al., 2019; Krois et al., 2019) analysiert. Auffällig dabei ist, dass im Vergleich zur röntgenologischen Beurteilung von anderen pathologischen Befunden die des parodontalen Knochenabbaus mittels KI bisher oftmals geringere diagnostische Genauigkeitswerte verzeichnet. Dies könnte eventuell dadurch begründet sein, dass das Ausmaß des Knochenabbaus über das gesamte Röntgenbild hinweg variieren kann und somit schwierig zu beurteilen ist. Deshalb ist es gerade im Hinblick auf die KI-basierte Einschätzung von parodontalem Knochenabbau auf Röntgenbildern umso wichtiger, Studien mit möglichst großen und für alle Zahngruppen repräsentativen Datensätzen durchzuführen, um diesbezüglich die diagnostische Genauigkeit zu erhöhen.

3. Zielstellung

Die vorliegende Dissertation soll einen Beitrag zur KI-basierten, automatisierten Erkennung von parodontalem Knochenabbau auf periapikalen Röntgenbildern leisten. Ein Vergleich von bislang vorliegenden Studien zu dieser Thematik ist durch Verwendung unterschiedlicher Datensätze, durch Ausschluss bestimmter Zahngruppen oder auch durch variierende Auswertungsmethoden nur erschwert möglich (Scott et al., 2023). Deshalb verfolgte das Dissertationsprojekt das Ziel, die diagnostische Leistung von fünf verschiedenen CNNs bei der Beurteilung des parodontalen Knochenabbaus auf periapikalen Röntgenbildern aller Zahngruppen mit standardisierten Variablen zu erfassen und zu vergleichen. Nach bestem Wissensstand ist in der Literatur bis auf die hier vorliegende Studie (Hoss et al., 2023) bisher noch keine andere Studie publiziert worden, die verschiedene CNNs mit dem Ziel der automatisierten Erkennung von parodontalem Knochenabbau auf periapikalen Röntgenbildern vergleicht.

Als Hypothese wurde dabei aufgestellt, dass für die verschiedenen CNNs und anatomischen Regionen übereinstimmende diagnostische Genauigkeitswerte erreicht werden, die darüber hinaus bei mindestens 90% liegen würden.

4. Material und Methodik

In der hier vorliegenden Studie (Hoss et al., 2023) wurde ein Datensatz mit insgesamt 21.819 anonymisierten, periapikalen Röntgenbildern verwendet. Dabei verteilten sich die Bilder über alle Zahngruppen hinweg und variierten darüber hinaus auch in der Ausprägung des parodontalen Knochenabbaus. Inadäquate Röntgenbilder, die beispielsweise durch Verzerrungen oder unvollständig abgebildete Zähne nur eingeschränkt beurteilbar sind, wurden im Datensatz nicht berücksichtigt.

Im Rahmen eines mehrtätigen Seminars, das von Prof. Dr. Kühnisch geleitet wurde, wurden die teilnehmenden Zahnärzte (N = 7) zunächst im Hinblick auf die röntgenologische Beurteilung von parodontalem Knochenabbau geschult. Infolgedessen wurde ein Kalibrierungskurs durchgeführt, bei dem für die Beurteilung des parodontalen Knochenabbaus auf Röntgenbildern moderate bis substanzielle Kappa-Werte von 0,516 bis 0,796 für Inter- und 0,455 bis 0,889 für Intra-Untersucher-Reproduzierbarkeit erreicht wurde (Meusburger et al., 2023).

Die vorliegenden Röntgenbilder wurden sodann nach folgenden diagnostischen Kriterien klassifiziert: gesundes Parodontium (0), beginnender parodontaler Knochenabbau bis zu 15% (1), moderater parodontaler Knochenabbau von 15-33% (2) und schwerer parodontaler Knochenabbau von mehr als 33% der Wurzellänge (3). Die diagnostischen Entscheidungen wurden daraufhin nochmals von klinisch erfahreneren Zahnärzten der Arbeitsgruppe unabhängig voneinander kontrolliert. Bei eventuell vorliegenden Unstimmigkeiten wurden die entsprechenden Röntgenbilder innerhalb der Arbeitsgruppe nochmals analysiert, bis letztlich eine einheitliche Diagnose für jedes Röntgenbild festgelegt werden konnte, die zum Referenzstandard für das zyklische CNN-Training erklärt wurde. Dabei ist zu erwähnen, dass die finalen Diagnosen im Rahmen eines binären Entscheidungsprozesses (Kategorie 0 versus 1-3) getroffen wurden.

Der Datensatz (N = 21.819) wurde daraufhin in einen Trainingsdatensatz (N = 18.819) und einen Testdatensatz (N = 3000) unterteilt. Die Variabilität des Trainingsdatensatzes wurde mittels Python (Version 3.8.5) gesteigert, indem die Röntgenbilder beispielsweise durch zufällige Rotationen und Änderungen des Kontrastes bearbeitet wurden. Die nun vorliegenden Röntgenbilder wurden dann für das Training von fünf verschiedenen, vortrainierten CNNs (ResNet-18,

MobileNet V2, ConvNeXT/small, ConvNeXT/base und ConvNeXT/large) verwendet, die durch den Projektpartner, der der Universität Duisburg – Essen angehörig ist, programmiert wurden. Der Testdatensatz wurde nicht für das Training der Netzwerke verwendet und diente ausschließlich zur Prüfung des Lernprozesses sowie zu Evaluierungszwecken der entwickelten CNN-Netzwerke.

Im Rahmen der statistischen Analyse wurden für die verschiedenen entwickelten CNNs neben diagnostischen Genauigkeiten (engl. „accuracy“, ACC) zudem auch Sensitivitäts- und Spezifitätswerte, positive bzw. negative prädiktive Werte sowie Flächen unter der Receiver Operating Characteristic Curve (AUC) angegeben.

5. Ergebnisse

Bei der Beurteilung des unabhängigen Testdatensatzes im Hinblick auf vorliegenden parodontalen Knochenabbau auf periapikalen Röntgenbildern erzielten die entwickelten CNNs ähnliche diagnostische Genauigkeiten (ACC) von 82,0% bis 84,8%. Die korrespondierenden AUC-Werte reichten von 0,884 bis 0,913. Die diagnostische Leistung variierte jedoch in Abhängigkeit von der untersuchten Zahngruppe. Unabhängig vom verwendeten CNN wurden generell bessere ACC-Werte für Unterkieferzähne im Vergleich zu Oberkieferzähnen und für Frontzähne im Vergleich zu Seitenzähnen dokumentiert. Im Unterkiefer wurden ACC-Werte zwischen 94,9% und 96,0% für Frontzähne sowie zwischen 82,2% und 86,1% für Seitenzähne erreicht. Im Vergleich dazu lagen entsprechende ACC-Werte im Oberkiefer bei 86,0% bis 88,6% (Frontzähne) und 78,0% bis 80,7% (Seitenzähne).

6. Diskussion

Insgesamt konnte im Rahmen des Dissertationsprojektes gezeigt werden, dass die Beurteilung des parodontalen Knochenabbaus auf periapikalen Röntgenbildern mithilfe von verschiedenen CNNs möglich ist. Dennoch musste die initial aufgestellte Hypothese verworfen werden, obwohl die getesteten CNNs eine ähnliche diagnostische Leistung aufweisen. Grund dafür ist, dass zum einen die primär erwartete diagnostische Leistung von mindestens 90% von keinem entwickelten CNN erreicht wurde, zum anderen aber auch Unterschiede in der korrekten Bewertung des parodontalen Knochenabbaus je nach vorliegender Zahngruppe dokumentiert werden konnten. Möglich erscheint, dass die Projektionstechnik sowie überlagernde anatomische Strukturen wie Nasenhöhle oder Kieferhöhlen die diagnostische Leistung für Oberkieferzahngruppen negativ beeinträchtigt haben. Interessanterweise haben einige bereits vorliegende Studien zu derselben Thematik den Aspekt der variierenden diagnostischen Leistung in Abhängigkeit von der anatomischen Region durch Ausschluss bestimmter Zahngruppen (Alotaibi et al., 2022; Tsoromokos et al., 2022; Lee et al., 2018b) nicht aufgefasst, so dass ein Vergleich bisheriger Studienergebnisse teilweise nur begrenzt möglich ist (Scott et al., 2023). Nichtsdestotrotz haben sich CNNs auch in den bereits veröffentlichten Studien als nützlich erwiesen, um parodontalen Knochenabbau auf periapikalen Röntgenbildern zu evaluieren (Chen et al., 2023; Alotaibi et al., 2022; Lee et al., 2022; Tsoromokos et al., 2022; Lee et al., 2018b). Lee et al. (2022) entwickelten beispielsweise ein DL-basiertes CNN, das den parodontalen Knochenabbau auf periapikalen Röntgenbildern gemäß der neuesten Stadieneinteilung (Papapanou et al., 2018) klassifizierte. Für die Stadien I, II und III wurden in diesem Zusammenhang hohe ACC-Werte von 0,91, 0,88 und 0,99 beschrieben. Auch Chen et al. (2023) dokumentierten mit einer diagnostischen Genauigkeit von 97% beeindruckende Ergebnisse für die Erkennung von parodontalem Knochenabbau mittels CNN.

Es ist also folglich nicht von der Hand zu weisen, dass sich die CNNs als effektiv in der Bildanalyse erwiesen haben. Dank ihrer hierarchischen Struktur sind CNNs dazu in der Lage, mithilfe von lernbaren Filtern zunächst einfachere Merkmale zu erkennen und darauf aufbauend komplexere Strukturen innerhalb eines Bildes zu identifizieren (Krizhevsky et al., 2012). Dahingegend nutzen Transformer-Netzwerke, die im medizinischen Fachbereich ursprünglich zur Sprachen- und

Texterkennung bzw. -verarbeitung genutzt wurden (Roshanzamir et al., 2021), einen sogenannten Attention-Mechanismus. Dieser Mechanismus ermöglicht dem Netzwerk, die Aufmerksamkeit gezielt auf relevante Bildteile zu lenken und somit eine detaillierte Auswertung durchzuführen, sodass Transformer-Netzwerke mittlerweile auch zur Verarbeitung von Bildern genutzt werden (Dosovitskiy et al., 2020). Da auch hier im Bereich der Objekterkennung auf Röntgenbildern sehr zufriedenstellende Ergebnisse erzielt werden konnten (Felsch et al., 2023; Zhou et al., 2023; Gao et al., 2022; Sheng et al., 2022; Ying et al., 2022), ließ sich vermuten, dass auch mithilfe von Transformer-Netzwerken eine gute diagnostische Genauigkeit bei der Einschätzung des parodontalen Knochenabbaus auf periapikalen Röntgenbildern zu erzielen ist. Deshalb wurde im Rahmen des Dissertationsprojektes neben der Untersuchung von Hoss et al. (2023) eine weitere Studie durchgeführt, bei der der parodontale Knochenabbau auf periapikalen Röntgenbildern durch fünf verschiedene DL-basierte Transformer-Netzwerke beurteilt wurde (Dujic et al., 2023). Bei der Evaluierung des unabhängigen Testdatensatzes (N = 3000) erreichten die Transformer-Netzwerke ACC-Werte von 83,4% bis 85,2% (Dujic et al., 2023). Ebenso wie bei Hoss et al. (2023) konnten auch hier Unterschiede in der diagnostischen Leistung je nach vorliegender anatomischer Region festgestellt werden (Dujic et al., 2023).

Insgesamt ist anzumerken, dass sowohl die dokumentierten Ergebnisse von Hoss et al. (2023) als auch von Dujic et al. (2023) als akzeptabel einzuordnen sind. Dennoch ist unter Berücksichtigung der Schwächen der hier vorliegenden Studien davon auszugehen, dass die diagnostische Leistung bei der KI-basierten, röntgenologischen Beurteilung des parodontalen Knochenabbaus weiter gesteigert werden kann. So wurde zwar in den beiden Studien Hoss et al. (2023) und Dujic et al. (2023) ein großer Datensatz (N = 21.819) verwendet, der nahezu gleich viel Röntgenbilder von Oberkiefer- und Unterkieferzähnen aufwies, dennoch konnte ein Ungleichgewicht im Datensatz festgestellt werden. Denn zum einen stammten deutlich mehr Röntgenaufnahmen vom Seitenzahnbereich als von der anterioren Region, zum anderen waren Röntgenbilder mit moderatem und schwerem parodontalen Knochenabbau unterrepräsentiert. Solch ein Ungleichgewicht im Datensatz könnte die diagnostische Leistung der CNNs negativ beeinflusst haben. Darüber hinaus kann als weitere Einschränkung angeführt werden, dass durch die Verwendung von anonymisierten

Röntgenbildern keine ergänzenden patientenspezifischen, klinischen Daten zur Verfügung standen. Doch gerade bei initialen Stadien der Parodontitis wären klinische Informationen wie Sondierungstiefen, CAL oder Blutung auf Sondierung hilfreich für die Diagnosefindung, da wie eingangs bereits erwähnt ossäre Defekte geringer Ausprägung röntgenologisch nur eingeschränkt zu detektieren sind (Pepelassi et al., 2000). Nicht zuletzt sollte als Schwäche auch angeführt werden, dass bei der Diagnosefindung je nach An- oder Abwesenheit von parodontalem Knochenabbau lediglich binär entschieden wurde (Kategorie 0 versus 1-3), ohne dass bestimmte Bildmerkmale annotiert wurden.

7. Zusammenfassung und Ausblick

Hinsichtlich der vorliegenden Ergebnisse lässt sich abschließend festhalten, dass die röntgenologische Diagnostik von parodontalem Knochenabbau mittels KI durchaus realisierbar ist. Während bei der Verwendung von CNNs zur Beurteilung des parodontalen Knochenabbaus auf periapikalen Röntgenbildern in der hier vorliegenden Studie ACC-Werte von 82,0% bis 84,8% erzielt wurden, wurden beim Einsatz von Transformer-Netzwerken ACC-Werte von 83,4% bis 85,2% erreicht (Dujic et al., 2023). Zusammenfassend lässt sich deshalb sagen, dass die Ergebnisse zwar als akzeptabel anzusehen sind, sich die initial formulierte Hypothese mit einer zu erwartenden diagnostischen Genauigkeit von mindestens 90% jedoch nicht bestätigt hat.

Durch bestimmte Anpassungen des methodologischen Ansatzes sollte es in der Zukunft sehr wohl möglich sein, die diagnostische Leistung der KI weiter zu verbessern. So wird es einerseits darauf ankommen, einen großen Datensatz zu verwenden, der gleichermaßen repräsentativ für alle Zahngruppen und Ausmaße des Knochenabbaus ist. Andererseits wird es von enormer Bedeutung sein, diejenigen Merkmale auf Röntgenbildern exakt zu annotieren, die mit parodontalem Knochenabbau in Relation stehen. Dadurch können der KI im Rahmen des Trainingsprozesses äußerst präzise Vorgaben gegeben werden. Natürlich ist das exakte Markieren spezifischer Strukturen in Röntgenbildern als zeitaufwendig anzusehen. Dennoch ist anzunehmen, dass sich gerade die exakte Annotierung zum entscheidenden Faktor für die Verbesserung der diagnostischen Genauigkeit bei der KI-basierten Beurteilung des parodontalen Knochenabbaus entwickeln könnte. Veröffentlichte Studien, die einen exakten Annotierungsprozess aufweisen, zeigten dementsprechend bereits hohe diagnostische Genauigkeiten der entwickelten KI (Chen et al., 2023; Lee et al., 2022). Folglich wird zukünftig die präzise Annotierung von spezifischen Bildmerkmalen von der Arbeitsgruppe umgesetzt.

8. Abstract (English)

The present dissertation project focused on the assessment of periodontal bone loss (PBL) using artificial intelligence (AI) and accordingly aimed to determine and compare the diagnostic performance of five different convolutional neural networks (CNNs) in detecting PBL on periapical radiographs. A dataset of 21.819 anonymized periapical radiographs was classified by calibrated dentists according to the degree of PBL into the following categories: healthy periodontium (0), mild (1), moderate (2) or severe PBL (3). A binary diagnosis decision (0 versus 1-3) was made for each image, which served as a reference standard for training and evaluation of the AI. The data set (N = 21.819) was then divided into a training set (N = 18.819) and an independent test set (N = 3.000). After training the CNNs, the AI-based algorithm evaluated the test set for the assessment of PBL. Here, the overall diagnostic accuracy (ACC) of the CNNs ranged between 82.0% and 84.8%. In a related study by Dujic et al. (2023), transformer networks were used to detect PBL on periapical radiographs. In this context, the transformer networks achieved an overall ACC between 83.4% and 85.2%. Interestingly, both studies showed differences in the diagnostic performance of the AI depending of the anatomical region. Predominantly, better results were reported for mandibular teeth compared to maxillary teeth and for the anterior region compared to the posterior region. Considering the documented outcomes, an AI-based evaluation of PBL using CNNs or transformer networks seems feasible. Nevertheless, the originally formulated hypothesis with an expected accuracy of 90% had to be rejected.

However, by making certain adjustments to the methodological approach, it should be possible to further improve the diagnostic performance of the AI. Therefore, it will be of enormous importance to use a large data set that is equally representative of all tooth groups and dimensions of bone loss. In addition, by exactly annotating the features on X-rays associated with periodontal bone loss, the AI could be given extremely precise guidelines as part of the training process. It can´t be denied that the exact marking of specific structures in X-ray images is time-consuming. Nevertheless, it can be assumed that precise annotation could become the decisive factor for increasing diagnostic accuracy in the AI-based assessment of PBL. Accordingly, published studies showing an accurate annotation process have already demonstrated high diagnostic accuracies of the

developed AI (Chen et al., 2023; Lee et al., 2022). Consequently, the precise annotation of specific image features will be implemented by the working group in the future.

9. Veröffentlichung I

Journal of Clinical Medicine
Volume 12, Issue 22, 7189
November 2023

Detection of Periodontal Bone Loss on Periapical Radiographs - A Diagnostic Study Using Different Convolutional Neural Networks

Patrick Hoss, Ole Meyer, Uta Christine Wölfle, Annika Wülk, Theresa Meusburger, Leon Meier, Reinhard Hickel, Volker Gruhn, Marc Hesenius, Jan Kühnisch, Helena Dujic

Journal of *Clinical Medicine*

MDPI

Article

Detection of Periodontal Bone Loss on Periapical Radiographs—A Diagnostic Study Using Different Convolutional Neural Networks

Patrick Hoss [1], Ole Meyer [2], Uta Christine Wölfle [1], Annika Wülk [1], Theresa Meusburger [1], Leon Meier [1], Reinhard Hickel [1], Volker Gruhn [2], Marc Hesenius [2], Jan Kühnisch [1,*] and Helena Dujic [1]

[1] Department of Conservative Dentistry and Periodontology, LMU University Hospital, LMU Munich, 80336 Munich, Germany; patrick.hoss@t-online.de (P.H.); uta.woelfle@med.uni-muenchen.de (U.C.W.); annika.wuelk@gmx.de (A.W.); theresa.meusburger@hotmail.com (T.M.); meierl.leon@gmail.com (L.M.); hickel@dent.med.uni-muenchen.de (R.H.); h.dujic@med.uni-muenchen.de (H.D.)

[2] Institute for Software Engineering, University of Duisburg-Essen, 45127 Essen, Germany; ole.meyer@uni-due.de (O.M.); volker.gruhn@paluno.uni-due.de (V.G.); marc.hesenius@uni-due.de (M.H.)

* Correspondence: jkuehn@dent.med.uni-muenchen.de; Tel.: +49-89-4400-59343 (ext. 59301)

Abstract: Interest in machine learning models and convolutional neural networks (CNNs) for diagnostic purposes is steadily increasing in dentistry. Here, CNNs can potentially help in the classification of periodontal bone loss (PBL). In this study, the diagnostic performance of five CNNs in detecting PBL on periapical radiographs was analyzed. A set of anonymized periapical radiographs (N = 21,819) was evaluated by a group of trained and calibrated dentists and classified into radiographs without PBL or with mild, moderate, or severe PBL. Five CNNs were trained over five epochs. Statistically, diagnostic performance was analyzed using accuracy (ACC), sensitivity (SE), specificity (SP), and area under the receiver operating curve (AUC). Here, overall ACC ranged from 82.0% to 84.8%, SE 88.8–90.7%, SP 66.2–71.2%, and AUC 0.884–0.913, indicating similar diagnostic performance of the five CNNs. Furthermore, performance differences were evident in the individual sextant groups. Here, the highest values were found for the mandibular anterior teeth (ACC 94.9–96.0%) and the lowest values for the maxillary posterior teeth (78.0–80.7%). It can be concluded that automatic assessment of PBL seems to be possible, but that diagnostic accuracy varies depending on the location in the dentition. Future research is needed to improve performance for all tooth groups.

Keywords: artificial intelligence; bone loss; convolutional neural networks; deep learning; dental radiography; machine learning; periodontitis

Citation: Hoss, P.; Meyer, O.; Wölfle, U.C.; Wülk, A.; Meusburger, T.; Meier, L.; Hickel, R.; Gruhn, V.; Hesenius, M.; Kühnisch, J.; et al. Detection of Periodontal Bone Loss on Periapical Radiographs—A Diagnostic Study Using Different Convolutional Neural Networks. *J. Clin. Med.* **2023**, *12*, 7189. https://doi.org/10.3390/jcm12227189

Academic Editor: Agostino Guida

Received: 3 November 2023
Revised: 14 November 2023
Accepted: 18 November 2023
Published: 20 November 2023

1. Introduction

Periodontitis is a prevalent dental health problem and can be classified as a major global challenge that affects developed and developing countries [1–3]. Triggered by bacterial colonization of the root surface, the host's immune system reacts with inflammatory processes to the microbial transition from a symbiotic bacterial environment to that of dysbiotic pathogens, leading to loss of supporting tooth tissue, pocket formation, and ulceration of the pocket epithelium [4,5]. If the condition advances, periodontal bone loss (PBL) can occur as the principal pathological characteristic of periodontitis [6]. Moreover, severe periodontitis is a major cause of missing teeth in adults, leading to reduced oral functioning and ultimately having an adverse effect on general health [7,8]. In this context, the link between periodontal disease and various systemic diseases such as cardiovascular diseases [9], diabetes [10], and respiratory diseases [11] should be emphasized. Considering the mostly irreversible consequences of periodontal disease, frequent periodontal screening is essential for the treatment of all patients and should be part of routine oral inspection [12]. According to the new guidelines introduced by the workshop on the classification of periodontal

and peri-implant diseases and conditions [13,14], the evaluation of clinical attachment loss as well as the radiographic assessment of PBL has become critical in categorizing periodontitis into specific stages and subsequently in indicating optimal disease management. Nevertheless, both the clinical measurements and the radiographic assessment of PBL remain controversial in terms of their reliability. The measurement of clinical attachment loss by periodontal probing varies due to individual probing force, probe angulation, and varying probe tip diameter [15,16]. In addition, radiographic PBL evaluation represents a challenging task for a clinician due to possible variations in contrast and exposure angle as well as structural overlap, so that the interpretation of dental radiographs may lead to inconsistencies among dentists [17–19]. Here, the use of artificial intelligence (AI)-based diagnostics could reduce these diagnostic discrepancies. Consequently, several work groups have investigated the use of AI-based methods for automatized PBL detection on periapical radiographs [19–29] and panoramic X-rays [18,30–40]. In these studies, on the one hand, convolutional neural networks (CNNs) have shown potential in accurately detecting PBL on radiographs. However, due to differing CNNs and varying data sets, the existing studies show significant heterogeneity and, therefore, are difficult to compare [41–43]. In addition, little is known about whether different CNNs or anatomical regions influence diagnostic performance. Therefore, the aim of this study was to evaluate the diagnostic performance of five commonly used CNNs for automated PBL detection on periapical radiographs representing all sextants (upper and lower posterior teeth and upper and lower anterior teeth) and to statistically report their diagnostic performance with standardized variables, avoiding non-comparable results. In detail, it was first hypothesized that the diagnostic performance of the tested CNNs would have an accuracy of at least 90%. Secondly, diagnostic accuracy was hypothesized to be the same between all CNNs and anatomical regions.

2. Materials and Methods

2.1. Study Design

The Ethics Committee of the Medical Faculty of the Ludwig-Maximilians University of Munich approved this study protocol with project number 020-798. The recommendations of the Standard for Reporting of Diagnostic Accuracy Studies (STARD) steering committee [44] and the recommendations for the reporting of AI studies in dentistry [45] were followed in the study report.

2.2. Periapical Radiographs

For this study, anonymized periapical radiographs taken at the Department of Conservative Dentistry and Periodontology (Dental School of the LMU) and other dental practices were used. A high-quality image sample was secured by excluding inadequate X-rays, e.g., distorted images, images with incomplete teeth, or radiographs with implants. Following these exclusion criteria, a data set with 21,819 periapical radiographs stored in jpg format was assembled.

2.3. Categorization of Periodontal Bone Loss (Reference Standard)

Prior to the start of the study, a two-day workshop was held by the principal investigator (J.K.), during which the group of participating dentists ($N = 7$) was trained. In addition, the efficiency of the training was determined during a calibration course. Reproducibility of PBL within and between investigators was assessed using 150 periapical radiographs, and the corresponding inter- and intra-examiner reliability showed substantial kappa values [17]. The detailed kappa values are specified in Table 1. A group of graduated dentists (P.H., T.M., A.W., L.M.) then pre-categorized all X-rays by differentiating between healthy periodontium and mild, moderate, or severe PBL [13,14]. Following this, more clinically experienced examiners (H.D., U.W., J.K.) independently counterchecked each diagnostic decision. More specifically, these diagnostic criteria and ratings were applied: 0—healthy periodontium, PBL not detectable, 1—mild radiographic PBL up to 15% in the coronal third of the tooth, 2—moderate radiographic PBL between 15% and 33% of the root

length, and 3—severe radiographic PBL beyond the coronal third of the tooth (Figure 1). In case of differing diagnostic opinions, each image was subject to continued discussion until consensus was achieved. The use of anonymized periapical radiographs meant that no further clinical information could have been acquired to make a diagnostic decision. One dichotomized diagnosis decision (0 vs. 1–3) was made for each X-ray, which consequently became the reference standard for the cyclic training and the repeated evaluation of the AI-based CNN.

Table 1. Cohen's kappa values for inter- and intra-examiner reliability for the detection of PBL, calculated among participating dentists ($N = 7$) in relation to the reference standard.

Examiner	Inter-Examiner	Intra-Examiner
P.H.	0.601–0.650	0.889
T.M.	0.620–0.658	0.554
A.W.	0.762–0.796	0.779
L.M.	0.516–0.565	0.797
U.W.	0.658–0.699	0.455
J.K.	0.706–0.748	0.579
H.D.	0.529–0.534	0.767

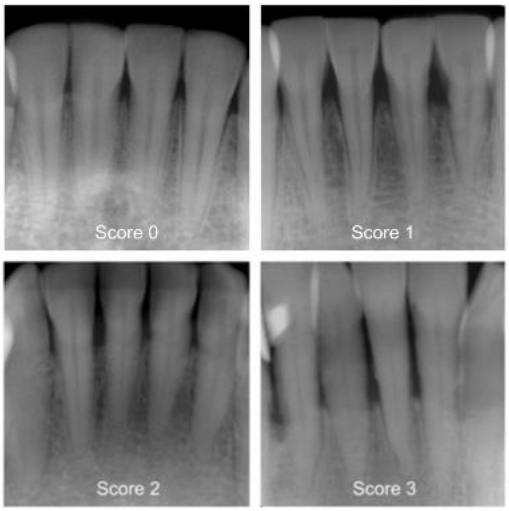

Figure 1. Examples of periapical radiographs for all categories: healthy periodontium, periodontal bone loss (PBL) not detectable (Score 0), mild radiographic PBL up to 15% in the coronal third of the tooth (Score 1), moderate radiographic PBL between 15% and 33% of the root length (Score 2), and severe radiographic PBL beyond the coronal third of the tooth (Score 3).

2.4. Training of the Deep-Learning-Based CNNs (Test Method)

Hereafter, the utilized pipeline of well-established methods for developing the AI-based algorithm is explained. Initially, the whole image set of 21,819 periapical radiographs was subdivided into a training set (N = 18,819) and a test set (N = 3000). The latter was randomly selected from the entire data set, ensuring that all sextants were equally represented. This served as an independent test set for evaluation purposes only and was not included in the model training.

By using Python (version 3.8.5, https://www.python.org accessed on 17 November 2023) in conjunction with the PyTorch library (version 1.12.0, https://pytorch.org accessed on 17 November 2023), the training set was augmented so that the variability of the included

radiographs could be improved. Therefore, images were modified using different transformations: random rotation up to 180 degrees, random changes in brightness, contrast, and saturation up to 20% with color jitter, and random affine transformation (translation up to 30% of the image size and zooming out up to 70%). As a result, a new, unique, and virtual grayscale image (RGB format) was created.

The augmented images were used to train the following pretrained CNNs: ResNet-18 [46], MobileNet V2 [47], ConvNeXT/small, ConvNeXT/base, and ConvNeXT/large [48]. The batch size amounted to 16 randomly selected images. The random selection of the respective images into batches was done using PyTorch's built-in DataLoader class. The learning performance was repeatedly verified with the test set after 30 training steps. All CNNs were trained using backpropagation to determine the gradient for learning. Furthermore, the training was accelerated using Floating Point 16 and a university-based computer (i9 10850K 10 × 3.60 GHz, Intel Corp., Santa Clara, CA, USA) equipped with 48 GB RAM and a professional graphic card (GeForce RTX 3060, Nvidia, Santa Clara, CA, USA). Each CNN was trained over 5 epochs, with cross entropy loss as an error function and an application of the Adam optimizer (Betas 0.9 and 0.999, Epsilon $\times 10^{-8}$).

2.5. Statistical Analysis

The data were analyzed using Python (version 3.8.5). By computing the number of true positives (TPs), false positives (FPs), true negatives (TNs) and false negatives (FNs), the diagnostic accuracy (ACC = (TN + TP)/(TN + TP + FN + FP)) was identified. The sensitivity (SE), specificity (SP), positive predictive values (PPVs), negative predictive values (NPVs), and the area under the receiver operating characteristic (ROC) curve (AUC) were calculated with respect to the utilized CNN [49].

3. Results

For the purpose of this study, a total of 21,819 periapical radiographs were selected and divided into sextants (upper and lower posterior teeth as well as upper and lower anterior teeth). The image distribution in relation to the anatomical region and the PBL can be taken from Table 2. While the number of radiographs from the upper jaw was found to be comparable to that from the lower jaw, the overwhelming majority of images originated from posterior teeth compared to anterior teeth. Moreover, most included periapical radiographs showing teeth affected by mild PBL (42.6%). In contrast, radiographs with severe PBL had a notably lower proportion (6.9%) in the total data set.

Table 2. Overview of the included periapical radiographs (N = 21,819) in relation to the corresponding sextants and periodontal diagnosis.

Expert Classification		Healthy Periodontium (Score 0)		Mild PBL (Score 1)		Moderate PBL (Score 2)		Severe PBL (Score 3)		Total	
		N	%	*N*	%	*N*	%	*N*	%	*N*	%
Upper jaw	Anteriors	653	3.0	661	3.0	433	2.0	197	0.9	1944	8.9
	1st Quadrant	1701	7.8	1826	8.4	851	3.9	367	1.7	4745	21.8
	2nd Quadrant	1231	5.6	2080	9.5	1093	5.0	312	1.5	4716	21.6
Lower jaw	Anteriors	202	0.9	676	3.1	786	3.6	325	1.5	1989	9.1
	3rd Quadrant	1477	6.8	2033	9.3	593	2.7	157	0.7	4260	19.5
	4th Quadrant	1282	5.9	2027	9.3	713	3.3	143	0.6	4165	19.1
	Total	6546	30.0	9303	42.6	4469	20.5	1501	6.9	21,819	100

The overall diagnostic performance for automatized detection of PBL on periapical radiographs in relation to the CNNs used are specified in Tables 3 and 4. The CNNs achieved an overall ACC between 82.0% and 84.8%. The associated AUC values ranged from 0.884 to 0.913. Moreover, all tested CNNs showed consistently higher SE values varying between 88.8% and 90.7% compared to the SP values, which ranged from 66.2% to 71.2%.

Table 3. Overview of the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) distribution for the independent test set (N = 3000 radiographs), which was evaluated by the AI-based algorithm for the assessment of periodontal bone loss.

CNN	True Positive (TP)		True Negative (TN)		False Positive (FP)		False Negative (FN)	
	N	%	*N*	%	*N*	%	*N*	%
ResNet-18	1876	62.5	609	20.3	294	9.8	221	7.4
MobileNetV2	1863	62.1	598	19.9	305	10.2	234	7.8
ConvNeXT/s [1]	1877	62.6	639	21.3	264	8.8	220	7.3
ConvNeXT/b [2]	1901	63.4	643	21.4	260	8.7	196	6.5
ConvNeXT/l [3]	1890	63.0	637	21.2	266	8.9	207	6.9

[1] small, [2] base, [3] large.

Table 4. Overview of the overall diagnostic performance of the developed convolutional neural network (CNN), where the independent test set (N = 3000 radiographs) was evaluated by the AI-based algorithm for the assessment of periodontal bone loss. The overall diagnostic accuracy (ACC), sensitivity (SE), specificity (SP), negative predictive value (NPV), positive predictive value (PPV), and area under the receiver operating characteristic curve (AUC) were predicted.

CNN	Diagnostic Performance					
	ACC	SE	SP	NPV	PPV	AUC
ResNet-18	82.8	89.5	67.4	73.4	86.5	0.884
MobileNetV2	82.0	88.8	66.2	71.9	85.9	0.884
ConvNeXT/s [1]	83.9	89.5	70.8	74.4	87.7	0.903
ConvNeXT/b [2]	84.8	90.7	71.2	76.6	88.0	0.911
ConvNeXT/l [3]	84.2	90.1	70.5	75.5	87.7	0.913

[1] small, [2] base, [3] large.

When investigating the diagnostic performance of the CNNs depending on the anatomical region (Tables 5 and 6), better results were mainly documented for mandibular teeth compared to maxillary teeth. In the anterior region, ACC values from 94.9% to 96.0% were observed for mandibular teeth and from 86.0% to 88.6% for maxillary teeth. When considering posterior teeth only, the ACC ranged from 82.2% to 86.1% for mandibular teeth and varied between 78.0% and 80.7% for maxillary teeth. In principle, the same tendency was also observed for the AUC values (Table 6).

Table 5. Overview of the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) distribution for the independent test set (N = 3000 radiographs) in different sextants, which was evaluated by the AI-based algorithm for the assessment of periodontal bone loss.

CNN	True Positive (TP)		True Negative (TN)		False Positive (FP)		False Negative (FN)	
	N	%	*N*	%	*N*	%	*N*	%
Radiographs with maxillary anterior teeth								
ResNet-18	155	58.7	72	27.3	27	10.2	10	3.8
MobileNetV2	154	58.3	79	29.9	20	7.6	11	4.2
ConvNeXT/s [1]	155	58.7	79	29.9	20	7.6	10	3.8
ConvNeXT/b [2]	157	59.5	77	29.2	22	8.3	8	3.0
ConvNeXT/l [3]	158	59.8	74	28.0	25	9.5	7	2.7
Radiographs with maxillary posterior teeth								
ResNet-18	786	59.1	263	19.8	151	11.4	129	9.7
MobileNetV2	798	60.0	239	18.0	175	13.2	117	8.8
ConvNeXT/s [1]	783	58.9	275	20.7	139	10.5	132	9.9
ConvNeXT/b [2]	794	59.7	278	20.9	136	10.2	121	9.1
ConvNeXT/l [3]	794	59.8	266	20.0	148	11.1	121	9.1

Table 5. *Cont.*

CNN	True Positive (TP)		True Negative (TN)		False Positive (FP)		False Negative (FN)	
	N	%	*N*	%	*N*	%	*N*	%
Radiographs with mandibular anterior teeth								
ResNet-18	244	89.7	14	5.2	11	4.0	3	1.1
MobileNetV2	239	87.9	19	7.0	6	2.2	8	2.9
ConvNeXT/s [1]	242	89.0	19	7.0	6	2.2	5	1.8
ConvNeXT/b [2]	244	89.7	17	6.3	8	2.9	3	1.1
ConvNeXT/l [3]	243	89.3	18	6.6	7	2.6	4	1.5
Radiographs with mandibular posterior teeth								
ResNet-18	691	60.9	260	22.9	105	9.3	79	6.9
MobileNetV2	672	59.2	261	23.0	104	9.2	98	8.6
ConvNeXT/s [1]	697	61.4	266	23.4	99	8.7	73	6.4
ConvNeXT/b [2]	706	62.2	271	23.9	94	8.3	64	5.6
ConvNeXT/l [3]	695	61.2	279	24.6	86	7.6	75	6.6

[1] small, [2] base, [3] large.

Table 6. Overview of the diagnostic performance of the developed convolutional neural networks (CNNs) for different sextants, where the independent test set (*N* = 3000 radiographs) was evaluated by the AI-based algorithm for the assessment of periodontal bone loss. The overall diagnostic accuracy (ACC), sensitivity (SE), specificity (SP), negative predictive value (NPV), positive predictive value (PPV), and area under the receiver operating characteristic curve (AUC) were predicted.

	Diagnostic Performance					
	ACC	SE	SP	NPV	PPV	AUC
Radiographs with maxillary anterior teeth						
ResNet-18	86.0	93.9	72.7	87.8	85.2	0.925
MobileNetV2	88.3	93.3	79.8	87.8	88.5	0.935
ConvNeXT/s [1]	88.6	93.9	79.8	88.8	88.6	0.951
ConvNeXT/b [2]	88.6	95.2	77.8	90.6	87.7	0.959
ConvNeXT/l [3]	87.9	95.8	74.7	91.4	86.3	0.950
Radiographs with maxillary posterior teeth						
ResNet-18	78.9	85.9	63.5	67.1	83.9	0.844
MobileNetV2	78.0	87.2	57.7	67.1	82.0	0.839
ConvNeXT/s [1]	79.6	85.6	66.4	67.6	84.9	0.858
ConvNeXT/b [2]	80.7	86.8	67.1	69.7	85.4	0.868
ConvNeXT/l [3]	79.8	86.8	64.3	68.7	84.3	0.866
Radiographs with mandibular anterior teeth						
ResNet-18	94.9	98.8	56.0	82.4	95.7	0.942
MobileNetV2	94.9	96.8	76.0	70.4	97.6	0.960
ConvNeXT/s [1]	96.0	98.0	76.0	79.2	97.6	0.969
ConvNeXT/b [2]	96.0	98.8	68.0	85.0	96.8	0.978
ConvNeXT/l [3]	96.0	98.4	72.0	81.8	97.2	0.980
Radiographs with mandibular posterior teeth						
ResNet-18	83.8	89.7	71.2	76.7	86.8	0.895
MobileNetV2	82.2	87.3	71.5	72.7	86.6	0.893
ConvNeXT/s [1]	84.8	90.5	72.9	78.5	87.6	0.916
ConvNeXT/b [2]	86.1	91.7	74.2	80.9	88.3	0.921
ConvNeXT/l [3]	85.8	90.3	76.4	78.8	89.0	0.930

[1] small, [2] base, [3] large.

All five CNNs, ResNet-18 (ACC 82.8%; AUC 0.884), MobileNetV2 (82.0%; 0.884), ConvNeXT/s (83.9%; 0.903), ConvNeXT/b (84.8%; 0.911) and ConvNeXT/l (84.2%; 0.913), tended to show similar performance data (Table 4). Furthermore, the hierarchy of results

is evident in the receiver operating characteristic (ROC) curves of the five CNNs used to graphically compare diagnostic performance in detecting PBL (Figure 2).

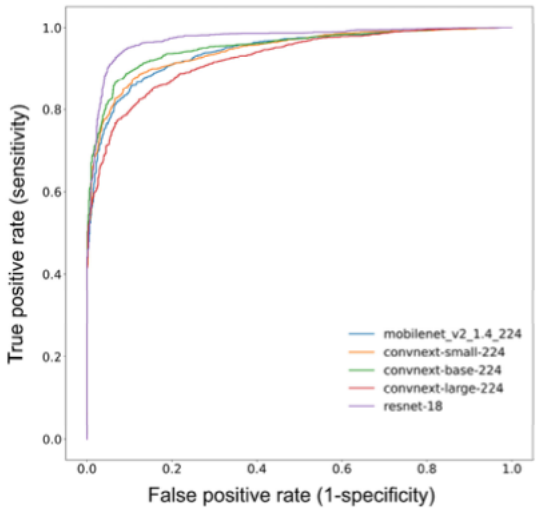

Figure 2. The receiver operating characteristic (ROC) curves graphically visualize the diagnostic performance of the developed convolutional neural networks (CNNs) in detecting PBL.

4. Discussion

The present study was able to demonstrate that different CNN architectures are able to detect PBL on periapical radiographs. However, with an overall accuracy between 82.0% and 84.8%, none of the CNNs tested were able to achieve the primary expected accuracy of 90%. Although the CNNs achieved similar diagnostic performance compared to one another, there were differences for the various sextants. This led to the rejection of the originally formulated hypothesis. Nevertheless, the results obtained provide important information for the discussion.

When considering the ability of the tested CNNs to detect PBL in relation to sextants on periapical radiographs, differences between teeth in the lower and upper jaw were observed (Table 6). Here, the projection technique and overlaying anatomical structures such as the maxillary sinuses or the nasal cavities may have negatively affected the diagnostic performance in the upper jaw. In contrast to the maxilla, mandibular sextants can be captured more accurately by use of the right-angle technique, which results in less distorted images and better diagnostic performance data (Table 6). The previously mentioned factors most likely explain the documented differences in the model performance among sextants, which were found to be similar throughout all included CNNs (Table 6). Such differences are of methodological importance. For example, Tsoromokos et al. [24] included only periapical radiographs with mandibular teeth in their pilot study to avoid data inconsistencies. Additionally, other author groups excluded radiographs from some sextants [20] or vertically rotated maxillary to mandibular teeth [26]. Such procedures may have resulted in biased and/or noncomparable results. Consequently, aiming at increasing the comparability of future studies, it is suggested to provide data for each sextant based on a well-powered image sample.

The diagnostic performance between the included CNNs was found to be similar. In general, our study results are basically in line with recently published studies of similar methodologies for evaluating PBL on periapical radiographs [19–24,26,29]. For example, Lee et al. [26] presented a model that could detect periodontally compromised premo-

lars and molars with a diagnostic accuracy of 82.8% and 73.5%, respectively. As part of the PBL assessment, Chen et al. [25] compared so-called fast and faster R-CNNs and then determined the severity of PBL. Unfortunately, no detailed accuracy values were provided [25]. Lee et al. [23] trained a machine learning model with precisely annotated periapical radiographs, which also classified PBL according to the latest classification [13]. In this context, high AUC values of 0.89, 0.90, and 0.90 were obtained for stages I, II, and III, respectively [23]. Another study with an accurate annotation process was introduced by Chen et al. [29]. Here, the model based on deep CNN algorithms provided an accuracy of 97% for the detection of PBL on periapical radiographs and showed superior performance compared to dentists. To the best of our knowledge, no study has compared multiple CNNs for PBL detection on periapical radiographs. In the literature, there is only one similarly designed study available that tested different CNNs to identify implant characteristics on periapical radiographs [50]. When also considering studies that analyzed panoramic X-rays for the presence of PBL, it can be concluded that the model metrics were found to be similar [18,30–34,36–40]. For instance, Krois et al. [38] presented a deep feed-forward CNN to detect PBL on image segments from panoramic radiographs. They chose binary decision making to distinguish between the presence or absence of PBL by introducing a cut-off value (20%, 25%, and 30%). A mean accuracy of 81% for PBL detection was achieved by the utilized CNN. In addition, the panoramic radiographs were manually cropped, focusing on a single tooth, and the images were flipped vertically by 180 degrees during pre-processing. Subsequently, it can be seen from the results that the diagnostic performance was validated in certain subgroups of teeth, with the highest accuracy value being reported for molars (86%). The deep learning model proposed in the study of Jiang et al. [30] was also applied to detect PBL on panoramic radiographs. The diagnostic performance of the model varied between 71% and 81% for different tooth groups. Interestingly, lower accuracy values were obtained not only for maxillary molars but also for mandibular anterior teeth, suggesting that overlapping anatomical structures may negatively impact the diagnostic performance for the anterior region in panoramic radiographs. Furthermore, the diagnostic performance for each periodontal stage was compared between the model and dentists. At all stages, the model achieved higher accuracy and sensitivity values compared to the dentists. Considering the reported results, it is worth noting that the author groups that accurately annotated PBL or features of PBL on panoramic radiographs generally published more favorable results [29,32,34,37].

This study has strengths and limitations. In view of the significant heterogeneity that previous studies have shown not only in their data sets (e.g., excluding certain tooth groups, the number of radiographs) but also in the evaluation method of diagnostic performance, then the training of commonly used CNNs with a data set representative of all sextants and the representation of their diagnostic performance with standardized variables can be considered a strength of this study [24,41–43]. Establishing a representative image data set for a particular finding with a relevant number of images can be considered a crucial factor. When comparing studies in terms of the total number of periapical radiographs, our study revealed a large data set ($N = 21,819$). Only Kearney et al. [51] utilized a larger data set, with over 100,000 radiographs; however, this study differed from our study methodologically by determining the clinical attachment level instead of PBL. Additionally, studies with panoramic radiographs should be mentioned in this context. With the exception of Kim et al. with more than 12,000 radiographs [37], almost all identified studies reported data sets with less than 2000 panoramic radiographs [18,30–36,38–40]. Moreover, our study allows the comparison of different CNNs for detecting PBL for each sextant. In addition, the data set included periapical radiographs with a broad spectrum of dental pathologies or restorations.

As a limiting factor of our study, the unbalanced image distribution across all sextants should be discussed. Although the number of radiographs from the maxilla was found to be similar to that of the mandible, less than half of the images were available from anterior teeth compared to posterior teeth (Table 2), which possibly indicates an imbalance in the

data set. The main reason leading to this unequal image distribution might be that under clinical conditions, the justification of an indication for radiography varies between the different sextants. In addition, moderate and severe PBL were also underrepresented. Such imbalances may negatively influence the diagnostic performance of CNNs. Therefore, it is crucial to safeguard a representative and well-balanced number of images for each sextant and severity score in order to improve the metrics of the models. Furthermore, this study utilized periapical radiographs only. However, both panoramic and periapical radiographs are considered relevant for PBL assessment. As for the aspect of comparing the diagnostic performance within different sextants, panoramic radiographs might be considered less applicable, since overlapping anatomical structures could potentially limit the diagnostic performance for the anterior region. Moreover, our data set was compiled from anonymized periapical X-rays; thus, no conclusions can be drawn about further, patient-specific diagnostic information. Additional diagnostic information, such as clinical attachment loss and pocket depths, would be particularly helpful for the initial diagnosis of periodontal disease, considering that the radiographic assessment of periodontal bone defects of low depth and buccolingual width might be restricted [52]. Here, the radiographic assessment of PBL becomes more relevant with further disease progression when the extent of osseous lesions can be visualized more accurately [53]. Another limitation to be mentioned is that we made a diagnosis for each image by distinguishing between a healthy periodontium and teeth affected with PBL (score 0 vs. 1–3). Considering that none of the five CNNs showed the hypothesized accuracy of 90%, this binary decision-making has to be understood as a limitation, which also negatively influenced the metrics of the models. It can be assumed that the precise annotation of PBL-related structures may increase the performance of the CNNs [23,29]. However, exact image labelling is time-consuming and requires extensive resources, especially with such large data sets. Nevertheless, it can be expected that precisely annotated radiographs representing a large and balanced data set would probably increase the precision of machine-based PBL detection.

5. Conclusions

In summary, the CNNs used showed nearly identical diagnostic performance in detecting PBL on periapical radiographs. However, different outcomes were documented among sextants, which can be primarily explained by the radiographic anatomy. With regard to comparable projects in the future, it is expected that the diagnostic performance can be further increased by precise annotations.

Author Contributions: Conceptualization, project administration, and supervision, J.K. and H.D.; study design, J.K., H.D. and O.M.; visualization, R.H. and V.G.; investigation, P.H., H.D., L.M., T.M., A.W., U.C.W. and J.K.; transformer network training and statistical analysis, O.M. and M.H.; writing—original draft preparation, P.H., J.K. and H.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was approved by the Ethics Committee of the Medical Faculty of the LMU Munich (project number 020-798, approved on 8 October 2020).

Informed Consent Statement: The procedures used in studies with human participants were all in accordance with the ethical standards of the institutional and/or national research committee and the 1964 Helsinki Declaration and its subsequent amendments or comparable ethical standards.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kassebaum, N.J.; Bernabe, E.; Dahiya, M.; Bhandari, B.; Murray, C.J.; Marcenes, W. Global burden of severe periodontitis in 1990–2010: A systematic review and meta-regression. *J. Dent. Res.* **2014**, *93*, 1045–1053. [CrossRef] [PubMed]
2. Frencken, J.E.; Sharma, P.; Stenhouse, L.; Green, D.; Laverty, D.; Dietrich, T. Global epidemiology of dental caries and severe periodontitis—A comprehensive review. *J. Clin. Periodontol.* **2017**, *44* (Suppl. S18), S94–S105. [CrossRef] [PubMed]
3. Nazir, M.A. Prevalence of periodontal disease, its association with systemic diseases and prevention. *Int. J. Health Sci.* **2017**, *11*, 72–80.
4. Damgaard, C.; Holmstrup, P.; Van Dyke, T.E.; Nielsen, C.H. The complement system and its role in the pathogenesis of periodontitis: Current concepts. *J. Periodontal. Res.* **2015**, *50*, 283–293. [CrossRef]
5. Abdulkareem, A.A.; Al-Taweel, F.B.; Al-Sharqi, A.J.B.; Gul, S.S.; Sha, A.; Chapple, I.L.C. Current concepts in the pathogenesis of periodontitis: From symbiosis to dysbiosis. *J. Oral Microbiol.* **2023**, *15*, 2197779. [CrossRef]
6. Könönen, E.; Gursoy, M.; Gursoy, U.K. Periodontitis: A Multifaceted Disease of Tooth-Supporting Tissues. *J. Clin. Med.* **2019**, *8*, 1135. [CrossRef]
7. Kandelman, D.; Petersen, P.E.; Ueda, H. Oral health, general health, and quality of life in older people. *Spec. Care Dent.* **2008**, *28*, 224–236. [CrossRef]
8. Tonetti, M.S.; Jepsen, S.; Jin, L.; Otomo-Corgel, J. Impact of the global burden of periodontal diseases on health, nutrition and wellbeing of mankind: A call for global action. *J. Clin. Periodontol.* **2017**, *44*, 456–462. [CrossRef]
9. Chistiakov, D.A.; Orekhov, A.N.; Bobryshev, Y.V. Links between atherosclerotic and periodontal disease. *Exp. Mol. Pathol.* **2016**, *100*, 220–235. [CrossRef]
10. Borgnakke, W.S.; Ylöstalo, P.V.; Taylor, G.W.; Genco, R.J. Effect of periodontal disease on diabetes: Systematic review of epidemiologic observational evidence. *J. Periodontol.* **2013**, *84* (Suppl. S4), S135–S152. [CrossRef]
11. Gomes-Filho, I.S.; Cruz, S.S.D.; Trindade, S.C.; Passos-Soares, J.S.; Carvalho-Filho, P.C.; Figueiredo, A.; Lyrio, A.O.; Hintz, A.M.; Pereira, M.G.; Scannapieco, F. Periodontitis and respiratory diseases: A systematic review with meta-analysis. *Oral Dis.* **2020**, *26*, 439–446. [CrossRef]
12. Preshaw, P.M. Detection and diagnosis of periodontal conditions amenable to prevention. *BMC Oral Health* **2015**, *15* (Suppl. S1), S5. [CrossRef]
13. Papapanou, P.N.; Sanz, M.; Buduneli, N.; Dietrich, T.; Feres, M.; Fine, D.H.; Flemmig, T.F.; Garcia, R.; Giannobile, W.V.; Graziani, F.; et al. Periodontitis: Consensus report of workgroup 2 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. *J. Periodontol.* **2018**, *89* (Suppl. S1), S173–S182. [CrossRef]
14. Tonetti, M.S.; Greenwell, H.; Kornman, K.S. Staging and grading of periodontitis: Framework and proposal of a new classification and case definition. *J. Periodontol.* **2018**, *89* (Suppl. S1), S159–S172. [CrossRef] [PubMed]
15. Garnick, J.J.; Silverstein, L. Periodontal probing: Probe tip diameter. *J. Periodontol.* **2000**, *71*, 96–103. [CrossRef] [PubMed]
16. Leroy, R.; Eaton, K.A.; Savage, A. Methodological issues in epidemiological studies of periodontitis—How can it be improved? *BMC Oral Health* **2010**, *10*, 8. [CrossRef]
17. Meusburger, T.; Wulk, A.; Kessler, A.; Heck, K.; Hickel, R.; Dujic, H.; Kuhnisch, J. The Detection of Dental Pathologies on Periapical Radiographs-Results from a Reliability Study. *J. Clin. Med.* **2023**, *12*, 2224. [CrossRef] [PubMed]
18. Kong, Z.; Ouyang, H.; Cao, Y.; Huang, T.; Ahn, E.; Zhang, M.; Liu, H. Automated periodontitis bone loss diagnosis in panoramic radiographs using a bespoke two-stage detector. *Comput. Biol. Med.* **2023**, *152*, 106374. [CrossRef]
19. Danks, R.P.; Bano, S.; Orishko, A.; Tan, H.J.; Moreno Sancho, F.; D'Aiuto, F.; Stoyanov, D. Automating Periodontal bone loss measurement via dental landmark localization. *Int. J. Comput. Assist. Radiol. Surg.* **2021**, *16*, 1189–1199. [CrossRef]
20. Alotaibi, G.; Awawdeh, M.; Farook, F.F.; Aljohani, M.; Aldhafiri, R.M.; Aldhoayan, M. Artificial intelligence (AI) diagnostic tools: Utilizing a convolutional neural network (CNN) to assess periodontal bone level radiographically a retrospective study. *BMC Oral Health* **2022**, *22*, 399. [CrossRef]
21. Chang, J.; Chang, M.F.; Angelov, N.; Hsu, C.Y.; Meng, H.W.; Sheng, S.; Glick, A.; Chang, K.; He, Y.R.; Lin, Y.B.; et al. Application of deep machine learning for the radiographic diagnosis of periodontitis. *Clin. Oral Investig.* **2022**, *26*, 6629–6637. [CrossRef] [PubMed]
22. Kabir, T.; Lee, C.T.; Chen, L.; Jiang, X.; Shams, S. A comprehensive artificial intelligence framework for dental diagnosis and charting. *BMC Oral Health* **2022**, *22*, 480. [CrossRef] [PubMed]
23. Lee, C.T.; Kabir, T.; Nelson, J.; Sheng, S.; Meng, H.W.; Van Dyke, T.E.; Walji, M.F.; Jiang, X.; Shams, S. Use of the deep learning approach to measure alveolar bone level. *J. Clin. Periodontol.* **2022**, *49*, 260–269. [CrossRef]
24. Tsoromokos, N.; Parinussa, S.; Claessen, F.; Moin, D.A.; Loos, B.G. Estimation of Alveolar Bone Loss in Periodontitis Using Machine Learning. *Int. Dent. J.* **2022**, *72*, 621–627. [CrossRef] [PubMed]
25. Chen, H.; Li, H.; Zhao, Y.; Zhao, J.; Wang, Y. Dental disease detection on periapical radiographs based on deep convolutional neural networks. *Int. J. Comput. Assist. Radiol. Surg.* **2021**, *16*, 649–661. [CrossRef] [PubMed]
26. Lee, J.-H.; Kim, D.-h.; Jeong, S.-N.; Choi, S.-H. Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *J. Periodontal Implant Sci.* **2018**, *48*. [CrossRef] [PubMed]
27. Lin, P.L.; Huang, P.Y.; Huang, P.W. Automatic methods for alveolar bone loss degree measurement in periodontitis periapical radiographs. *Comput. Methods Programs Biomed.* **2017**, *148*, 1–11. [CrossRef]
28. Lin, P.L.; Huang, P.W.; Huang, P.Y.; Hsu, H.C. Alveolar bone-loss area localization in periodontitis radiographs based on threshold segmentation with a hybrid feature fused of intensity and the H-value of fractional Brownian motion model. *Comput. Methods Programs Biomed.* **2015**, *121*, 117–126. [CrossRef]

29. Chen, C.C.; Wu, Y.F.; Aung, L.M.; Lin, J.C.; Ngo, S.T.; Su, J.N.; Lin, Y.M.; Chang, W.J. Automatic recognition of teeth and periodontal bone loss measurement in digital radiographs using deep-learning artificial intelligence. *J. Dent. Sci.* **2023**, *18*, 1301–1309. [CrossRef]
30. Jiang, L.; Chen, D.; Cao, Z.; Wu, F.; Zhu, H.; Zhu, F. A two-stage deep learning architecture for radiographic staging of periodontal bone loss. *BMC Oral Health* **2022**, *22*, 106. [CrossRef]
31. Ertas, K.; Pence, I.; Cesmeli, M.S.; Ay, Z.Y. Determination of the stage and grade of periodontitis according to the current classification of periodontal and peri-implant diseases and conditions (2018) using machine learning algorithms. *J. Periodontal. Implant Sci.* **2023**, *53*, 38. [CrossRef] [PubMed]
32. Widyaningrum, R.; Candradewi, I.; Aji, N.; Aulianisa, R. Comparison of Multi-Label U-Net and Mask R-CNN for panoramic radiograph segmentation to detect periodontitis. *Imaging Sci. Dent.* **2022**, *52*, 383–391. [CrossRef] [PubMed]
33. Zadrozny, L.; Regulski, P.; Brus-Sawczuk, K.; Czajkowska, M.; Parkanyi, L.; Ganz, S.; Mijiritsky, E. Artificial Intelligence Application in Assessment of Panoramic Radiographs. *Diagnostics* **2022**, *12*, 224. [CrossRef] [PubMed]
34. Li, H.; Zhou, J.; Zhou, Y.; Chen, Q.; She, Y.; Gao, F.; Xu, Y.; Chen, J.; Gao, X. An Interpretable Computer-Aided Diagnosis Method for Periodontitis from Panoramic Radiographs. *Front. Physiol.* **2021**, *12*, 655556. [CrossRef]
35. Chang, H.J.; Lee, S.J.; Yong, T.H.; Shin, N.Y.; Jang, B.G.; Kim, J.E.; Huh, K.H.; Lee, S.S.; Heo, M.S.; Choi, S.C.; et al. Deep Learning Hybrid Method to Automatically Diagnose Periodontal Bone Loss and Stage Periodontitis. *Sci. Rep.* **2020**, *10*, 7531. [CrossRef]
36. Thanathornwong, B.; Suebnukarn, S. Automatic detection of periodontal compromised teeth in digital panoramic radiographs using faster regional convolutional neural networks. *Imaging Sci. Dent.* **2020**, *50*, 169–174. [CrossRef]
37. Kim, J.; Lee, H.S.; Song, I.S.; Jung, K.H. DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs. *Sci. Rep.* **2019**, *9*, 17615. [CrossRef]
38. Krois, J.; Ekert, T.; Meinhold, L.; Golla, T.; Kharbot, B.; Wittemeier, A.; Dorfer, C.; Schwendicke, F. Deep Learning for the Radiographic Detection of Periodontal Bone Loss. *Sci. Rep.* **2019**, *9*, 8495. [CrossRef]
39. Liu, Q.; Dai, F.; Zhu, H.; Yang, H.; Huang, Y.; Jiang, L.; Tang, X.; Deng, L.; Song, L. Deep learning for the early identification of periodontitis: A retrospective, multicentre study. *Clin. Radiol.* **2023**, *78*, e985–e992. [CrossRef]
40. Orhan, K.; Aktuna Belgin, C.; Manulis, D.; Golitsyna, M.; Bayrak, S.; Aksoy, S.; Sanders, A.; Onder, M.; Ezhov, M.; Shamshiev, M.; et al. Determining the reliability of diagnosis and treatment using artificial intelligence software with panoramic radiographs. *Imaging Sci. Dent.* **2023**, *53*, 199–208. [CrossRef]
41. Scott, J.; Biancardi, A.M.; Jones, O.; Andrew, D. Artificial Intelligence in Periodontology: A Scoping Review. *Dent. J.* **2023**, *11*, 43. [CrossRef] [PubMed]
42. Patil, S.; Joda, T.; Soffe, B.; Awan, K.H.; Fageeh, H.N.; Tovani-Palone, M.R.; Licari, F.W. Efficacy of artificial intelligence in the detection of periodontal bone loss and classification of periodontal diseases: A systematic review. *J. Am. Dent. Assoc.* **2023**, *154*, 795–804.e1. [CrossRef] [PubMed]
43. Turosz, N.; Checinska, K.; Checinski, M.; Brzozowska, A.; Nowak, Z.; Sikora, M. Applications of artificial intelligence in the analysis of dental panoramic radiographs: An overview of systematic reviews. *Dentomaxillofac. Radiol.* **2023**, *52*, 20230284. [CrossRef] [PubMed]
44. Bossuyt, P.M.; Reitsma, J.B.; Bruns, D.E.; Gatsonis, C.A.; Glasziou, P.P.; Irwig, L.; Lijmer, J.G.; Moher, D.; Rennie, D.; de Vet, H.C.; et al. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *BMJ* **2015**, *351*, h5527. [CrossRef]
45. Schwendicke, F.; Singh, T.; Lee, J.H.; Gaudin, R.; Chaurasia, A.; Wiegand, T.; Uribe, S.; Krois, J.; on behalf of the IADR E-oral Health Network and the ITU WHO Focus Group AI for Health. Artificial intelligence in dental research: Checklist for authors, reviewers, readers. *J. Dent.* **2021**, *107*, 103610. [CrossRef]
46. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
47. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *arXiv* **2019**, arXiv:1801.04381.
48. Liu, Z.; Mao, H.; Wu, C.-Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. *arXiv* **2022**, arXiv:2201.03545.
49. Matthews, D.E.; Farewell, V.T. *Using and Understanding Medical Statistics*; S.Karger AG: Basel, Switzerland, 2015.
50. Kim, J.E.; Nam, N.E.; Shim, J.S.; Jung, Y.H.; Cho, B.H.; Hwang, J.J. Transfer Learning via Deep Neural Networks for Implant Fixture System Classification Using Periapical Radiographs. *J. Clin. Med.* **2020**, *9*, 1117. [CrossRef]
51. Kearney, V.P.; Yansane, A.M.; Brandon, R.G.; Vaderhobli, R.; Lin, G.H.; Hekmatian, H.; Deng, W.; Joshi, N.; Bhandari, H.; Sadat, A.S.; et al. A generative adversarial inpainting network to enhance prediction of periodontal clinical attachment level. *J. Dent.* **2022**, *123*, 104211. [CrossRef]
52. Pepelassi, E.A.; Tsiklakis, K.; Diamanti-Kipioti, A. Radiographic detection and assessment of the periodontal endosseous defects. *J. Clin. Periodontol.* **2000**, *27*, 224–230. [CrossRef] [PubMed]
53. Fiorellini, J.P.; Sourvanos, D.; Sarimento, H.; Karimbux, N.; Luan, K.W. Periodontal and Implant Radiology. *Dent. Clin. N. Am.* **2021**, *65*, 447–473. [CrossRef] [PubMed]

10. Veröffentlichung II

Diagnostics
Volume 13, Issue 23, 3562
November 2023

Automatized Detection of Periodontal Bone Loss on Periapical Radiographs by Vision Transformer Networks

Helena Dujic, Ole Meyer, Patrick Hoss, Uta Christine Wölfle, Annika Wülk, Theresa Meusburger, Leon Meier, Volker Gruhn, Marc Hesenius, Reinhard Hickel, Jan Kühnisch

diagnostics

MDPI

Article

Automatized Detection of Periodontal Bone Loss on Periapical Radiographs by Vision Transformer Networks

Helena Dujic [1,*], Ole Meyer [2], Patrick Hoss [1], Uta Christine Wölfle [1], Annika Wülk [1], Theresa Meusburger [1], Leon Meier [1], Volker Gruhn [2], Marc Hesenius [2], Reinhard Hickel [1] and Jan Kühnisch [1,*]

[1] Department of Conservative Dentistry and Periodontology, LMU University Hospital, LMU Munich, 80336 Munich, Germany; patrick.hoss@t-online.de (P.H.); uwoelfle@dent.med.uni-muenchen.de (U.C.W.); annika.wuelk@gmx.de (A.W.); theresa.meusburger@hotmail.com (T.M.); meierl.leon@gmail.com (L.M.); hickel@dent.med.uni-muenchen.de (R.H.)

[2] Institute for Software Engineering, University of Duisburg-Essen, 45127 Essen, Germany; ole.meyer@uni-due.de (O.M.); volker.gruhn@paluno.uni-due.de (V.G.); marc.hesenius@uni-due.de (M.H.)

* Correspondence: h.dujic@med.uni-muenchen.de (H.D.); jkuehn@dent.med.uni-muenchen.de (J.K.); Tel.: +49-89-4400-59395 (H.D.)

Abstract: Several artificial intelligence-based models have been presented for the detection of periodontal bone loss (PBL), mostly using convolutional neural networks, which are the state of the art in deep learning. Given the emerging breakthrough of transformer networks in computer vision, we aimed to evaluate various models for automatized PBL detection. An image data set of 21,819 anonymized periapical radiographs from the upper/lower and anterior/posterior regions was assessed by calibrated dentists according to PBL. Five vision transformer networks (ViT-base/ViT-large from Google, BEiT-base/BEiT-large from Microsoft, DeiT-base from Facebook/Meta) were utilized and evaluated. Accuracy (ACC), sensitivity (SE), specificity (SP), positive/negative predictive value (PPV/NPV) and area under the ROC curve (AUC) were statistically determined. The overall diagnostic ACC and AUC values ranged from 83.4 to 85.2% and 0.899 to 0.918 for all evaluated transformer networks, respectively. Differences in diagnostic performance were evident for lower (ACC 94.1–96.7%; AUC 0.944–0.970) and upper anterior (86.7–90.2%; 0.948–0.958) and lower (85.6–87.2%; 0.913–0.937) and upper posterior teeth (78.1–81.0%; 0.851–0.875). In this study, only minor differences among the tested networks were detected for PBL detection. To increase the diagnostic performance and to support the clinical use of such networks, further optimisations with larger and manually annotated image data sets are needed.

Keywords: artificial intelligence; deep learning; machine learning; transformer; periapical radiographs; periodontitis; periodontal bone loss; diagnostics

Citation: Dujic, H.; Meyer, O.; Hoss, P.; Wölfle, U.C.; Wülk, A.; Meusburger, T.; Meier, L.; Gruhn, V.; Hesenius, M.; Hickel, R.; et al. Automatized Detection of Periodontal Bone Loss on Periapical Radiographs by Vision Transformer Networks. *Diagnostics* **2023**, *13*, 3562. https://doi.org/10.3390/diagnostics13233562

Academic Editor: Francesco Inchingolo

Received: 26 October 2023
Revised: 18 November 2023
Accepted: 27 November 2023
Published: 29 November 2023

1. Introduction

Periodontitis is a chronic inflammatory disease of the supporting dental tissues and affects a relevant proportion of the world's population [1–4]. Furthermore, periodontitis can also be associated with various risk factors such as smoking and stress, as well as systemic diseases such as diabetes mellitus or pulmonary diseases. Clinically, periodontitis is associated with periodontal bone loss (PBL), tooth loosening and tooth loss. All of these factors can further impair functionality, aesthetics and quality of life [5,6]. Considering the recommendations of the latest workshop on the classification of periodontal diseases [7,8], the initial diagnosis is primarily based on clinical assessment, bleeding on probing, repeated measurements of clinical attachment loss and probing pocket depth. The early manifestations of periodontitis are only clinically recognisable. Furthermore, staging based on the radiographic assessment of PBL is considered possible only with the progression of the disease. As a result, the importance of radiographs increases as the disease progresses,

since the extent of alveolar bone changes can be visualized more accurately [9,10]. However, a reliable assessment of PBL remains susceptible to diagnostic subjectivity among dentists [11,12]. Therefore, the use of image analysis tools based on artificial intelligence (AI) methods could possibly enable the automated assessment of PBL on radiographs and potentially improve diagnostic accuracy. Interestingly, several research groups have developed AI-based algorithms and published promising results on panoramic [11,13–21] and periapical radiographs [12,22–30]. Looking at the methodology of the studies published so far, almost all research groups have used an image set of a limited size to train different types of convolutional neural networks (CNNs). This has led to heterogeneous but promising results [31,32]. In particular, more than half of the studies published to date have reported a data set of less than 1000 X-ray images [12,14–19,21,26,29,30]. In addition, some studies used different exclusion criteria for their data set, meaning that radiographs with a specific tooth group or radiographs with caries or root canal treatment were excluded (e.g., [23,28]). In addition, variability in the architecture of the CNNs used can be observed, e.g., ResNet, U-Net and faster R-CNNs were trained for PBL detection [12,13,15,17–19,25]. Accurate manual annotation also contributed significantly to the reported results, as studies reporting the annotation of radiologic features of PBL described a better diagnostic performance, e.g., [13,25]. Moreover, none of the previously mentioned studies used recently introduced transformer networks for computer vision tasks, which are the most recent available technology for automatized image analysis and may possibly outperform current CNNs in the future [33]. On the one hand, CNNs have proven their value in tasks such as image classification and segmentation by efficiently processing large data sets. Among the most significant advantages is the ability of CNNs to recognize local patterns, such as edges or shapes. This proved to be particularly helpful for recognizing features in dental X-rays, such as tooth decay, different tooth shapes, etc. On the other hand, the vision transformer's attention mechanism allows the model to learn the correlation of parts of the image that may not be in direct proximity. In the case of PBL detection, these are primarily the cementoenamel junction, alveolar bone and apex, as well as other anatomical structures relevant for the evaluation. Notably, transformer networks usually require a larger amount of training data compared to CNNs. Following this, we aimed to compare the diagnostic performance of five different transformer networks for automatized PBL detection on periapical radiographs. Specifically, it was hypothesized that the diagnostic performance of the included transformer networks would be similar and that an overall diagnostic accuracy of 90% would be achievable.

2. Materials and Methods

2.1. Study Design

The Ethics Committee of the Medical Faculty of Ludwig Maximilian University (LMU) of Munich approved this study protocol (project number 020-798). The periapical radiographs used in this study were anonymized and obtained as part of previous clinical examinations. Consequently, we could not identify any of the patients and were therefore unable to obtain written informed consent. The reporting of this research followed the Standard for Reporting of Diagnostic Accuracy Studies (STARD) Steering Committee recommendations [34] as well as the recommendations for reporting AI studies in dentistry [35].

2.2. Periapical Radiographs

This study used anonymized periapical radiographs (Figure 1). All X-rays were taken at the Department of Conservative Dentistry and Periodontology (LMU University Hospital) and different dental practices. To ensure a high-quality image sample, exclusion criteria were previously defined. This involved excluding distorted radiographs, radiographs with overlapping teeth, radiographs with artifacts, and radiographs with incompletely imaged teeth for which an assessment of the periodontium was not possible. Furthermore, radiographs with implants, with endodontic treatments or photographed radiographs, were

also excluded. Further exclusion criteria were not defined. All periapical radiographs were stored in .jpg format and processed without downsizing the original resolution. Altogether, 21,819 periapical radiographs, divided into upper/lower anterior and posterior teeth, were selected for this study (Table 1). The majority of the radiographs were upper (N = 9461) and lower posterior teeth (N = 8425), outnumbering upper (N = 1944) and lower anterior teeth (N = 1989). Additionally, the radiographs were categorized according to PBL.

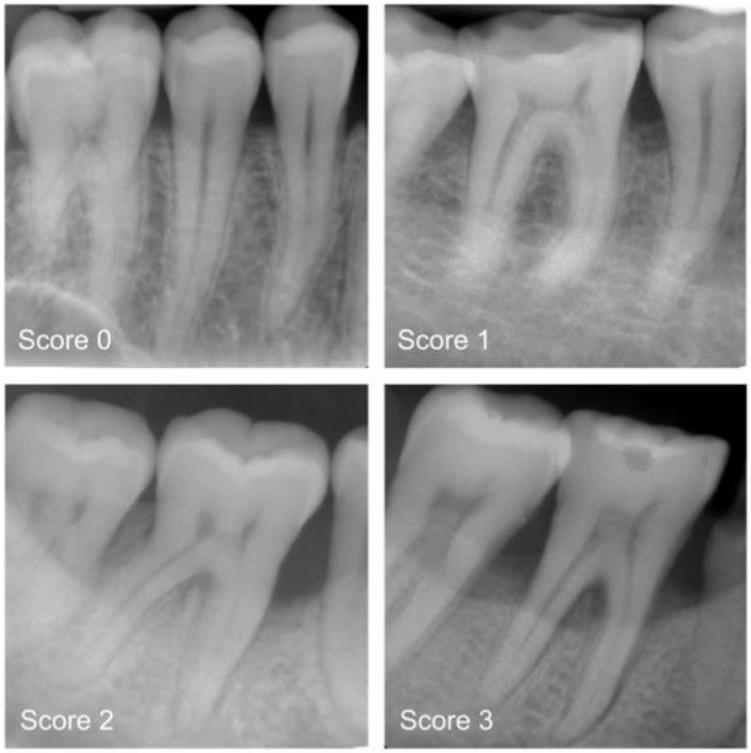

Figure 1. Examples of periapical radiographs for all categories: healthy periodontium (Score 0), mild radiographic periodontal bone loss (PBL) up to 15% of the root length (Score 1), moderate radiographic PBL between 15% and 33% of the root length (Score 2), and severe radiographic PBL extending to the mid–third of the root and beyond (Score 3).

Table 1. Overview of the included periapical radiographs (N = 21,819) in relation to the corresponding regions and categories of periodontal bone loss.

Region of Periapical Radiograph	Healthy Periodontium (Score 0)	Mild PBL (Score 1)	Moderate PBL (Score 2)	Severe PBL (Score 3)	Total (N)
1st Quadrant	1701 (35.8%)	1826 (38.5%)	851 (18.0%)	367 (7.7%)	4745
2nd Quadrant	1231 (26.1%)	2080 (44.1%)	1093 (23.2%)	312 (6.6%)	4716
3rd Quadrant	1477 (34.7%)	2033 (47.7%)	593 (13.9%)	157 (3.7%)	4260
4th Quadrant	1282 (30.8%)	2027 (48.7%)	713 (17.1%)	143 (3.4%)	4165
Maxillary anteriors	653 (33.6%)	661 (34.0%)	433 (22.3%)	197 (10.1%)	1944
Mandibular anteriors	202 (10.2%)	676 (34.0%)	786 (39.5%)	325 (16.3%)	1989

2.3. Categorisation of Periodontal Bone Loss (Reference Standard)

All radiographs were precategorised by a group of graduate dentists (P.H., T.M., A.W. and L.M.) and later independently counterchecked by experienced examiners (H.D., U.W. and J.K.). For each of the periapical radiographs, a diagnosis was made by differentiating between healthy teeth and teeth affected by mild, moderate or severe PBL [7,8]. Clinical data were not available prior to decision making. In detail, the following diagnostic criteria were applied: 0—radiographic PBL not detectable; 1—mild radiographic PBL up to 15% of the root length; 2—moderate radiographic PBL between 15% and 33% of the root length; and 3—severe radiographic PBL extending to the mid-third of the root and beyond (Figure 1). In the case of divergent opinions, each radiograph was discussed until consensus was reached. Each dichotomized diagnostic decision (0 versus 1 to 3)—one per image—served as a reference standard for the cyclic training and repeated evaluation of the deep learning-based transformer network.

Before conducting this study, all participating dentists were trained during a 2-day workshop by the principal investigator (J.K.). Following this workshop, the effectiveness of training was determined during a calibration course. The inter- and intra-examiner reproducibility for PBL were assessed on 150 periapical radiographs. The corresponding Kappa values showed substantial reliability, ranging from 0.454 to 0.482 (inter-examiner). The intra-examiner reliability in terms of Cohen's Kappa amounted to 0.739 [36].

2.4. Training of the Deep Learning-Based Transformer Networks (Test Method)

A pipeline of well-established methods was used to train the transformer networks. In principle, the entire image set of 21,819 periapical radiographs was divided into a training set (N = 18,819) and a test set. The latter included 3000 randomly selected X-rays from the overall image set and served as an independent test set that was not included in the model training. Given the high number of periapical radiographs in our data set, image augmentation and preprocessing were not necessary. Furthermore, all X-rays had a standardized size.

The previously mentioned data set was used to train five different pre-trained transformer networks (Table 2) [33,37,38]. The learning performance was evaluated with the independent test set. The used transformer networks were trained by using backpropagation to determine the gradient for learning. Furthermore, the model training was accelerated by the use of Floating Point 16 and a university-based computer (i9 10850K 10 × 3.60 GHz, Intel Corp., Santa Clara, CA, USA) equipped with 64 GB RAM and a professional graphic card (RTX A6000 48 GB (Nvidia, Santa Clara, CA, USA). The batch size amounted to 16 randomly selected images. Each transformer was trained over 5 epochs with cross entropy loss as an error function and an application of the Adam optimizer (Betas 0.9 and 0.999, Epsilon × 10^{-8}).

Table 2. Model characteristics of the used transformer networks.

	ViT-Base (Google)	**ViT-Large (Google)**	**BEiT-Base (Microsoft)**	**BEiT-Large (Microsoft)**	**DeiT-Base (Facebook/Meta)**
Neural network	Vision transformer		Bidirectional encoder representation from image transformers		Data-efficient image transformer
Epochs	5	5	5	5	5
Learning rate	0.00005	0.00005	0.00005	0.00005	0.00005
FLOS	7.280×10^{15}	25.735×10^{15}	7.277×10^{15}	25.744×10^{15}	7.280×10^{15}
Samples per second	298.6	111.7	274.4	102.9	298.5
Parameter count	85.8×10^{6}	303.3×10^{6}	85.7×10^{6}	303.4×10^{6}	85.8×10^{6}

2.5. Statistical Analysis

The data were analysed using Python (version 3.8.5, http://www.python.org accessed on 28 November 2023). The diagnostic ACC was determined by calculating the number of

true negatives (TN), true positives (TP), false positives (FP) and false negatives (FN). In addition, the sensitivity (SE), specificity (SP), positive/negative predictive values (PPV/NPV) and area under the receiver operating characteristic (ROC) curve were calculated [39].

3. Results

In the present study, we calculated the diagnostic performance for automatized PBL detection on periapical radiographs for lower/upper and anterior/posterior teeth altogether (Table 3) and separately (Table 4) by using five different transformer networks. In general, when analysing the whole data set of periapical radiographs, the ACC ranged from 83.4% to 85.2%; the corresponding AUC values ranged from 0.899 to 0.918 (Figure 2). The detailed data analysis revealed generally better performance data for mandibular teeth than for maxillary teeth (Table 4). Here, the ACC ranged from 94.1% to 96.7% for mandibular anteriors and from 85.6% to 87.2% for mandibular posteriors. The corresponding data for maxillary anterior and posterior teeth varied between 86.7% and 90.2% as well as between 78.1% and 81.0%, respectively. Additionally, the AUC values tended to be similar or better for mandibular teeth (Table 4). Furthermore, the SE values were consistently higher than the SP values.

Table 3. Overview of the overall diagnostic performance of the five transformer neuronal networks where the independent test set (N = 3000 radiographs) was evaluated by the AI-based algorithm for the assessment of periodontal bone loss. Diagnostic accuracy (ACC), sensitivity (SE), specificity (SP), negative predictive value (NPV), positive predictive value (PPV) and area under the receiver operating characteristic curve (AUC) were calculated for all types of teeth.

All Apical Radiographs	True Positive (TP)		True Negative (TN)		False Positive (FP)		False Negative (FN)		Diagnostic Performance					
	N	%	*N*	%	*N*	%	*N*	%	ACC	SE	SP	NPV	PPV	AUC
ViT-base	1884	62.8	673	22.4	230	7.7	213	7.1	85.2	89.8	74.5	76.0	89.1	0.918
ViT-large	1831	61.0	671	22.4	232	7.7	266	8.9	83.4	87.3	74.3	71.6	88.8	0.899
BEiT-base	1885	62.8	649	21.6	254	8.5	212	7.1	84.5	89.9	71.9	75.4	88.1	0.914
BEiT-large	1914	63.8	631	21.0	272	9.1	183	6.1	84.8	91.3	69.9	77.5	87.6	0.907
DeiT-base	1879	62.6	646	21.5	257	8.6	218	7.3	84.2	89.6	71.5	74.8	88.0	0.908

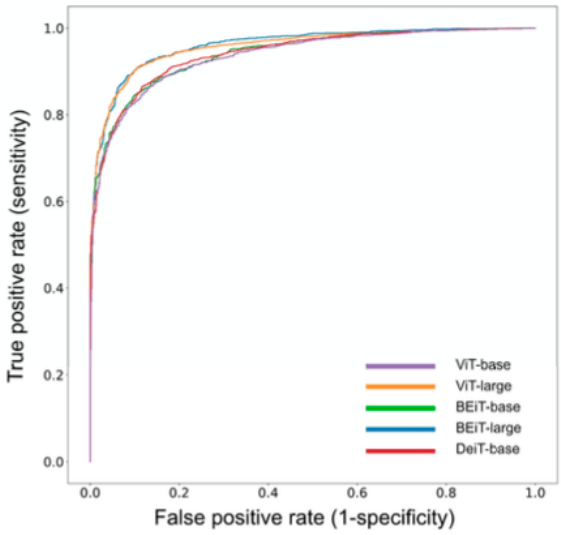

Figure 2. The receiver operating characteristic (ROC) curves illustrate the diagnostic performance of five different transformer networks for PBL detection.

Table 4. Overview of the diagnostic performance of the five transformer neuronal networks for mandibular and maxillary anterior and posterior teeth. Accuracy (ACC), sensitivity (SE), specificity (SP), negative predictive value (NPV), positive predictive value (PPV) and area under the receiver operating characteristic curve (AUC) were calculated.

		True Positive (TP)		True Negative (TN)		False Positive (FP)		False Negative (FN)		Diagnostic Performance					
		N	%	*N*	%	*N*	%	*N*	%	ACC	SE	SP	NPV	PPV	AUC
Mandibular anterior teeth	ViT-base	240	88.2	16	5.9	9	3.3	7	2.6	94.1	97.2	64.0	69.6	96.4	0.944
	ViT-large	241	88.6	18	6.6	7	2.6	6	2.2	95.2	97.6	72.0	75.0	97.2	0.960
	BEiT-base	242	89.0	21	7.7	4	1.5	5	1.8	96.7	98.0	84.0	80.8	98.4	0.963
	BEiT-large	245	90.1	18	6.6	7	2.6	2	0.7	96.7	99.2	72.0	90.0	97.2	0.952
	DeiT-base	242	89.0	15	5.5	10	3.7	5	1.8	94.5	98.0	60.0	75.0	96.0	0.970
Mandibular posterior teeth	ViT-base	700	61.6	287	25.3	78	6.9	70	6.2	87.0	90.9	78.6	80.4	90.0	0.937
	ViT-large	687	60.5	285	25.1	80	7.1	83	7.3	85.6	89.2	78.1	77.4	89.6	0.913
	BEiT-base	704	62.0	277	24.4	88	7.8	66	5.8	86.4	91.4	75.9	80.8	88.9	0.933
	BEiT-large	711	62.6	279	24.6	86	7.6	59	5.2	87.2	92.3	76.4	82.5	89.2	0.923
	DeiT-base	694	61.1	281	24.8	84	7.4	76	6.7	85.9	90.1	77.0	78.7	89.2	0.927
Maxillary anterior teeth	ViT-base	157	59.5	81	30.7	18	6.8	8	3.0	90.2	95.2	81.8	91.0	89.7	0.958
	ViT-large	156	59.1	77	29.2	22	8.3	9	3.4	88.3	94.5	77.8	89.5	87.6	0.948
	BEiT-base	158	59.8	73	27.7	26	9.8	7	2.7	87.5	95.8	73.7	91.3	85.9	0.954
	BEiT-large	157	59.5	73	27.7	26	9.8	8	3.0	87.1	95.2	73.7	90.1	85.8	0.954
	DeiT-base	154	58.3	75	28.4	24	9.1	11	4.2	86.7	93.3	75.8	87.2	86.5	0.954
Maxillary posterior teeth	ViT-base	787	59.2	289	21.8	125	9.4	128	9.6	81.0	86.0	69.8	69.3	86.3	0.875
	ViT-large	747	56.2	291	21.9	123	9.3	168	12.6	78.1	81.6	70.3	63.4	85.9	0.851
	BEiT-base	781	58.8	278	20.9	136	10.2	134	10.1	79.7	85.4	67.1	67.5	85.2	0.865
	BEiT-large	801	60.3	261	19.6	153	11.5	114	8.6	79.9	87.5	63.0	69.6	84.0	0.861
	DeiT-base	789	59.4	275	20.7	139	10.4	126	9.5	80.1	86.2	66.4	68.6	85.0	0.860

When comparing the metrics of the included transformer networks, only minor differences appeared in the results (Tables 3 and 4). However, the ACC and AUC values were found to be high in all scenarios, and SE was higher than SP.

4. Discussion

The present study aimed to compare the diagnostic performance of five different transformer networks for automatized PBL detection on periapical radiographs. Depending on the applied network, the overall diagnostic ACC and AUC values ranged from 83.4% to 85.2% and 0.899 to 0.918, respectively (Table 3, Figure 2). On the one hand, the ACC values must be evaluated as high; on the other hand, the hypothesized overall diagnostic ACC of 90% was not achieved. Therefore, the initially formulated hypothesis must be rejected.

When comparing the documented diagnostic performance data (Tables 3 and 4) with data from the literature, the following conclusion can be drawn. In general, the majority of comparable studies presented model performances in the same or lower order of magnitude [11–15,17,20,21,23,26,28,40], whereas only a few studies registered above-average values [25,41]. In detail, Lee et al. [25] reported an ACC for staging that ranged from 88% to 99%. They further stated that the ACC for periodontitis case classification was 85%. Specifically, 693 periapical radiographs were independently annotated by examiners prior to training the model, indicating regions of interest such as the alveolar bone, presence of teeth, cementoenamel junctions and presence of restorations. In addition, a further 644 periapical radiographs were used to assess the ACC of the model. In another study on staging, Widyaningrum et al. [41] stated that the detection rate was 95%, with the best performance shown for stage 4 periodontitis. Although the data set consisted of only 100 panoramic radiographs, two investigators annotated the previously mentioned radiographs before training the CNN. Accurate annotations were made by marking the alveolar ridge and the alveolar bone surrounding the teeth. In addition, the examiners added a number indicating the stage of periodontitis. Therefore, the few studies with better diagnostic performance seem remarkable compared to other studies with results of a lower magnitude. Here, other dental detection tasks should also be mentioned in comparison, where a higher ACC—typically approximately 90%—was usually registered with a similar methodology, e.g., in the detection of caries or periapical lesions on radiographs (e.g., [42–44]) and the detection of clinical pathologies or restorations on intraoral photographs (e.g., [45–49]). This may indicate that automatized PBL detection is more difficult to accomplish, which is supported by the fact that PBL characteristics are usually spread over the whole radiographic image and can have varying extents.

Our study revealed differences in the performance of the model in relation to the analysed group of teeth. In principle, automatized PBL detection performed better for mandibular teeth than for maxillary teeth, and better for anterior teeth compared to posterior teeth (Table 4). Only a few studies have considered this aspect thus far, e.g., by the exclusion of periapical radiographs with upper anterior and posterior teeth or by the inclusion of anterior teeth only [23,26]. To avoid the influence of data inconsistencies on the results of the trained CNN, Tsoromokos et al. [26] only considered periapical radiographs of the mandible and reported a data set with 446 radiographs. In addition, Alotaibi et al. [23] considered 1724 periapical radiographs of maxillary and mandibular anterior teeth only and excluded radiographs of teeth that had been restored with full crowns or root canal treatments, as well as radiographs of teeth that had undergone apical surgery with root resection. In this context, the study by Lee et al. [28] should also be mentioned, which included periapical radiographs of posterior teeth to identify periodontally compromised premolars and molars. Further exclusion criteria were root canal treatment and teeth with fully restorative crowns as well as moderate to severe caries and teeth with a shape deviating from the usual anatomical structure. When considering the data shown in Table 4, it must be concluded that the partial exclusion of periapical radiographs may bias the model's performance and limit the generalisability of the data shown. As is reasonable for this finding, the anatomical structures in the upper jaw in relation to the intraoral projec-

tion technique must be considered. Interestingly, this issue can be obviously downsized when using panoramic X-rays [15]. Nevertheless, a well-balanced inclusion of periapical radiographs from different groups of teeth may be relevant and should be implemented in future studies.

In this study, five well-established open-source transformer networks were trained: ViT-base and ViT-large from Google, BEiT-base and BEiT-large from Microsoft, and DeiT-base from Facebook/Meta [33,37,38]. The main differences between the transformer networks are in their size, training strategy and fine-tuning approach. "Base" and "large" models differ in size and computational complexity, whereby "large" models have more parameters. During training, ViTs process images as a sequence of patches and use an attention mechanism to learn the overall correlations within images. DeiT can achieve a high performance even with limited training data. Here, a smaller model learns to imitate a larger, already pre-trained model and benefits from a large data set without directly using it. In contrast, BEiT is trained in a two-stage process: pre-training on a large data set to capture general visual features, followed by fine-tuning for specific tasks. Transformer networks have rarely been applied for computer vision tasks in dentistry and not specifically for the detection of PBL. So far, only three studies using transformer networks were published; however, none of them focused on PBL assessment in periapical radiographs [50–52]. Nevertheless, there have been studies in which CNNs were used for PBL detection on periapical and panoramic radiographs (e.g., [11,14,15,17,21–26,40]). Here, the majority of investigations used only a low to moderate number of radiographs for model development, and most studies on periapical radiographs included a maximum of a few thousand images [13,22,23,25,27,28]. In contrast, Kim et al. [20] annotated the PBL in an extensive set of 12,179 panoramic radiographs, which may have potentially enhanced the internal study strength. The reported model-dependent AUC values ranged from 0.92 to 0.95 [20], which were slightly higher than the results from our study setup (Table 2). Therefore, it can be argued that the chosen study setup produced comparable data in the moment, which in part might be attributed to the use of transformer networks. Interestingly, we observed similar performance data with each of the included transformer networks. There was a minor tendency for less-complex transformer networks, e.g., Google's vision transformer/base, to perform better than their more complex counterparts (Tables 2 and 3). However, further improvements might be possible, especially by employing exact annotations in a large image set. Such features could enable precise object segmentation [20].

This study has several strengths and limitations. From a methodological point of view, this study used a large and well-balanced set of periapical radiographs (N = 21,819) in which all X-rays were diagnosed by dental professionals following the latest recommendations for PBL assessment [7,8]. Another unique feature seems to be the comparison of five transformer networks for the detection of PBL on periapical radiographs, as no other studies with the same methodology could be identified. In addition, the following limitations must be taken into account. In this study, we used categorial diagnostic scoring per image only. In detail, this means that the exact areas of PBL on periapical radiographs remained unmarked, which can be interpreted as a limitation. The exact annotation must be understood as a crucial feature to localize PBL precisely on X-rays. The exact annotation of the pathological structures would require the detection, classification and segmentation of PBL on each radiographic image. In particular, the marking of pathological segments must be understood as a time-consuming procedure that needs to be addressed in future projects. Another limitation is that only periapical radiographs were examined in this study and that panoramic radiographs have not been considered so far. However, in view of the fact that both types of radiographs are commonly used to assess PBL, but the format, size and radiographic anatomy differ, a separate analysis was justified. In addition, no clinical information was available for the anonymized radiographs in this study. Another limitation might be that we did not include any other transformer networks or CNNs in this study.

5. Conclusions

From the results of this study, it can be concluded that it was possible to achieve good diagnostic performance for automatized PBL detection when using a large set of periapical radiographs and several transformer networks. However, it can be hypothesized that the model performance can be improved by using exact annotations.

Author Contributions: Conceptualisation, project administration and supervision J.K.; study design, H.D., J.K., O.M., M.H., V.G. and R.H.; investigation, H.D., P.H., L.M., T.M., A.W., U.C.W. and J.K.; transformer network training and statistical analysis, O.M. and M.H.; writing—original draft preparation, H.D., J.K. and P.H. All authors contributed equally to the interpretation of data and reviewed, edited and approved the final manuscript version. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was approved by the Ethics Committee of the Medical Faculty of the LMU Munich (project number 020-798, approved on 8 October 2020).

Informed Consent Statement: Procedures used in studies with human participants were all in accordance with the ethical standards of the institutional and/or national research committee and the 1964 Helsinki Declaration and its subsequent amendments or comparable ethical standards.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no potential conflicts of interest with respect to the authorship and publication of this article.

References

1. Nazir, M.; Al-Ansari, A.; Al-Khalifa, K.; Alhareky, M.; Gaffar, B.; Almas, K. Global Prevalence of Periodontal Disease and Lack of Its Surveillance. *Sci. World J.* **2020**, *2020*, 2146160. [CrossRef] [PubMed]
2. Frencken, J.E.; Sharma, P.; Stenhouse, L.; Green, D.; Laverty, D.; Dietrich, T. Global epidemiology of dental caries and severe periodontitis—A comprehensive review. *J. Clin. Periodontol.* **2017**, *44* (Suppl. 18), S94–S105. [CrossRef]
3. Tonetti, M.S.; Jepsen, S.; Jin, L.; Otomo-Corgel, J. Impact of the global burden of periodontal diseases on health, nutrition and wellbeing of mankind: A call for global action. *J. Clin. Periodontol.* **2017**, *44*, 456–462. [CrossRef]
4. Kassebaum, N.J.; Bernabe, E.; Dahiya, M.; Bhandari, B.; Murray, C.J.; Marcenes, W. Global burden of severe periodontitis in 1990–2010: A systematic review and meta-regression. *J. Dent. Res.* **2014**, *93*, 1045–1053. [CrossRef] [PubMed]
5. Papapanou, P.N.; Susin, C. Periodontitis epidemiology: Is periodontitis under-recognized, over-diagnosed, or both? *Periodontol. 2000* **2017**, *75*, 45–51. [CrossRef] [PubMed]
6. Petersen, P.E.; Ogawa, H. The global burden of periodontal disease: Towards integration with chronic disease prevention and control. *Periodontol. 2000* **2012**, *60*, 15–39. [CrossRef]
7. Papapanou, P.N.; Sanz, M.; Buduneli, N.; Dietrich, T.; Feres, M.; Fine, D.H.; Flemmig, T.F.; Garcia, R.; Giannobile, W.V.; Graziani, F.; et al. Periodontitis: Consensus report of workgroup 2 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. *J. Periodontol.* **2018**, *89* (Suppl. 1), S173–S182. [CrossRef]
8. Tonetti, M.S.; Greenwell, H.; Kornman, K.S. Staging and grading of periodontitis: Framework and proposal of a new classification and case definition. *J. Periodontol.* **2018**, *89* (Suppl. 1), S159–S172. [CrossRef]
9. Tonetti, M.S.; Sanz, M. Implementation of the new classification of periodontal diseases: Decision-making algorithms for clinical practice and education. *J. Clin. Periodontol.* **2019**, *46*, 398–405. [CrossRef] [PubMed]
10. Fiorellini, J.P.; Sourvanos, D.; Sarimento, H.; Karimbux, N.; Luan, K.W. Periodontal and Implant Radiology. *Dent. Clin. N. Am.* **2021**, *65*, 447–473. [CrossRef]
11. Kong, Z.; Ouyang, H.; Cao, Y.; Huang, T.; Ahn, E.; Zhang, M.; Liu, H. Automated periodontitis bone loss diagnosis in panoramic radiographs using a bespoke two-stage detector. *Comput. Biol. Med.* **2023**, *152*, 106374. [CrossRef] [PubMed]
12. Danks, R.P.; Bano, S.; Orishko, A.; Tan, H.J.; Moreno Sancho, F.; D'Aiuto, F.; Stoyanov, D. Automating Periodontal bone loss measurement via dental landmark localisation. *Int. J. Comput. Assist. Radiol. Surg.* **2021**, *16*, 1189–1199. [CrossRef] [PubMed]
13. Kabir, T.; Lee, C.T.; Chen, L.; Jiang, X.; Shams, S. A comprehensive artificial intelligence framework for dental diagnosis and charting. *BMC Oral Health* **2022**, *22*, 480. [CrossRef] [PubMed]
14. Ertas, K.; Pence, I.; Cesmeli, M.S.; Ay, Z.Y. Determination of the stage and grade of periodontitis according to the current classification of periodontal and peri-implant diseases and conditions (2018) using machine learning algorithms. *J. Periodontal. Implant Sci.* **2022**, *53*, 38. [CrossRef]

15. Jiang, L.; Chen, D.; Cao, Z.; Wu, F.; Zhu, H.; Zhu, F. A two-stage deep learning architecture for radiographic staging of periodontal bone loss. *BMC Oral Health* **2022**, *22*, 106. [CrossRef] [PubMed]
16. Zadrozny, L.; Regulski, P.; Brus-Sawczuk, K.; Czajkowska, M.; Parkanyi, L.; Ganz, S.; Mijiritsky, E. Artificial Intelligence Application in Assessment of Panoramic Radiographs. *Diagnostics* **2022**, *12*, 224. [CrossRef] [PubMed]
17. Li, H.; Zhou, J.; Zhou, Y.; Chen, Q.; She, Y.; Gao, F.; Xu, Y.; Chen, J.; Gao, X. An Interpretable Computer-Aided Diagnosis Method for Periodontitis From Panoramic Radiographs. *Front. Physiol.* **2021**, *12*, 655556. [CrossRef]
18. Thanathornwong, B.; Suebnukarn, S. Automatic detection of periodontal compromised teeth in digital panoramic radiographs using faster regional convolutional neural networks. *Imaging Sci. Dent.* **2020**, *50*, 169–174. [CrossRef]
19. Chang, H.J.; Lee, S.J.; Yong, T.H.; Shin, N.Y.; Jang, B.G.; Kim, J.E.; Huh, K.H.; Lee, S.S.; Heo, M.S.; Choi, S.C.; et al. Deep Learning Hybrid Method to Automatically Diagnose Periodontal Bone Loss and Stage Periodontitis. *Sci. Rep.* **2020**, *10*, 7531. [CrossRef]
20. Kim, J.; Lee, H.S.; Song, I.S.; Jung, K.H. DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs. *Sci. Rep.* **2019**, *9*, 17615. [CrossRef]
21. Krois, J.; Ekert, T.; Meinhold, L.; Golla, T.; Kharbot, B.; Wittemeier, A.; Dorfer, C.; Schwendicke, F. Deep Learning for the Radiographic Detection of Periodontal Bone Loss. *Sci. Rep.* **2019**, *9*, 8495. [CrossRef] [PubMed]
22. Chen, C.C.; Wu, Y.F.; Aung, L.M.; Lin, J.C.; Ngo, S.T.; Su, J.N.; Lin, Y.M.; Chang, W.J. Automatic recognition of teeth and periodontal bone loss measurement in digital radiographs using deep-learning artificial intelligence. *J. Dent. Sci.* **2023**, *18*, 1301–1309. [CrossRef]
23. Alotaibi, G.; Awawdeh, M.; Farook, F.F.; Aljohani, M.; Aldhafiri, R.M.; Aldhoayan, M. Artificial intelligence (AI) diagnostic tools: Utilizing a convolutional neural network (CNN) to assess periodontal bone level radiographically—A retrospective study. *BMC Oral Health* **2022**, *22*, 399. [CrossRef]
24. Chang, J.; Chang, M.F.; Angelov, N.; Hsu, C.Y.; Meng, H.W.; Sheng, S.; Glick, A.; Chang, K.; He, Y.R.; Lin, Y.B.; et al. Application of deep machine learning for the radiographic diagnosis of periodontitis. *Clin. Oral Investig.* **2022**, *26*, 6629–6637. [CrossRef] [PubMed]
25. Lee, C.T.; Kabir, T.; Nelson, J.; Sheng, S.; Meng, H.W.; Van Dyke, T.E.; Walji, M.F.; Jiang, X.; Shams, S. Use of the deep learning approach to measure alveolar bone level. *J. Clin. Periodontol.* **2022**, *49*, 260–269. [CrossRef]
26. Tsoromokos, N.; Parinussa, S.; Claessen, F.; Moin, D.A.; Loos, B.G. Estimation of Alveolar Bone Loss in Periodontitis Using Machine Learning. *Int. Dent. J.* **2022**, *72*, 621–627. [CrossRef] [PubMed]
27. Chen, H.; Li, H.; Zhao, Y.; Zhao, J.; Wang, Y. Dental disease detection on periapical radiographs based on deep convolutional neural networks. *Int. J. Comput. Assist. Radiol. Surg.* **2021**, *16*, 649–661. [CrossRef]
28. Lee, J.-H.; Kim, D.-h.; Jeong, S.-N.; Choi, S.-H. Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *J. Periodontal Implant. Sci.* **2018**, *48*, 114–123. [CrossRef]
29. Lin, P.L.; Huang, P.Y.; Huang, P.W. Automatic methods for alveolar bone loss degree measurement in periodontitis periapical radiographs. *Comput. Methods Programs Biomed.* **2017**, *148*, 1–11. [CrossRef]
30. Lin, P.L.; Huang, P.W.; Huang, P.Y.; Hsu, H.C. Alveolar bone-loss area localization in periodontitis radiographs based on threshold segmentation with a hybrid feature fused of intensity and the H-value of fractional Brownian motion model. *Comput. Methods Programs Biomed.* **2015**, *121*, 117–126. [CrossRef]
31. Patil, S.; Joda, T.; Soffe, B.; Awan, K.H.; Fageeh, H.N.; Tovani-Palone, M.R.; Licari, F.W. Efficacy of artificial intelligence in the detection of periodontal bone loss and classification of periodontal diseases: A systematic review. *J. Am. Dent. Assoc.* **2023**, *154*, 795–804.e791. [CrossRef] [PubMed]
32. Scott, J.; Biancardi, A.M.; Jones, O.; Andrew, D. Artificial Intelligence in Periodontology: A Scoping Review. *Dent. J.* **2023**, *11*, 43. [CrossRef] [PubMed]
33. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. Available online: https://arxiv.org/abs/2010.11929 (accessed on 23 October 2023).
34. Bossuyt, P.M.; Reitsma, J.B.; Bruns, D.E.; Gatsonis, C.A.; Glasziou, P.P.; Irwig, L.; Lijmer, J.G.; Moher, D.; Rennie, D.; de Vet, H.C.; et al. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *BMJ* **2015**, *351*, h5527. [CrossRef]
35. Schwendicke, F.; Singh, T.; Lee, J.H.; Gaudin, R.; Chaurasia, A.; Wiegand, T.; Uribe, S.; Krois, J. Artificial intelligence in dental research: Checklist for authors, reviewers, readers. *J. Dent.* **2021**, *107*, 103610. [CrossRef]
36. Meusburger, T.; Wulk, A.; Kessler, A.; Heck, K.; Hickel, R.; Dujic, H.; Kühnisch, J. The Detection of Dental Pathologies on Periapical Radiographs-Results from a Reliability Study. *J. Clin. Med.* **2023**, *12*, 2224. [CrossRef]
37. Bao, H.; Dong, L.; Piao, S.; Wei, F. BEiT: BERT Pre-Training of Image Transformers. *arXiv* **2022**, arXiv:2106.08254v2. Available online: https://arxiv.org/abs/2106.08254 (accessed on 23 October 2023).
38. Touvron, H.; Cord, M.; Douze, M.; Massa, F.; Sablayrolles, A.; Jégou, H. Training Data-Efficient Image Transformers & Distillation through Attention. *arXiv* **2021**, arXiv:2012.12877v2. Available online: https://arxiv.org/abs/2012.12877 (accessed on 23 October 2023).
39. Matthews, D.E.; Farewell, V.T. *Using and Understanding Medical Statistics*; S.Karger AG: Basel, Switzerland, 2015.
40. Kurt, S.; Celik, O.; Bayrakdar, I.S.; Orhan, K.; Bilgir, E.; Odabas, A.; Aslan, A.F. Success of artificial intelligence system in determining alveolar bone loss from dental panoramic radiography images. *Cumhuriyet. Dent. J.* **2020**, *23*, 318–324. [CrossRef]

41. Widyaningrum, R.; Candradewi, I.; Aji, N.; Aulianisa, R. Comparison of Multi-Label U-Net and Mask R-CNN for panoramic radiograph segmentation to detect periodontitis. *Imaging Sci. Dent.* **2022**, *52*, 383–391. [CrossRef]
42. Lian, L.; Zhu, T.; Zhu, F.; Zhu, H. Deep Learning for Caries Detection and Classification. *Diagnostics* **2021**, *11*, 1672. [CrossRef]
43. Moidu, N.P.; Sharma, S.; Chawla, A.; Kumar, V.; Logani, A. Deep learning for categorization of endodontic lesion based on radiographic periapical index scoring system. *Clin. Oral Investig.* **2022**, *26*, 651–658. [CrossRef]
44. Lee, J.H.; Kim, D.H.; Jeong, S.N.; Choi, S.H. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J. Dent.* **2018**, *77*, 106–111. [CrossRef] [PubMed]
45. Schonewolf, J.; Meyer, O.; Engels, P.; Schlickenrieder, A.; Hickel, R.; Gruhn, V.; Hesenius, M.; Kühnisch, J. Artificial intelligence-based diagnostics of molar-incisor-hypomineralization (MIH) on intraoral photographs. *Clin. Oral Investig.* **2022**, *26*, 5923–5930. [CrossRef] [PubMed]
46. Engels, P.; Meyer, O.; Schonewolf, J.; Schlickenrieder, A.; Hickel, R.; Hesenius, M.; Gruhn, V.; Kühnisch, J. Automated detection of posterior restorations in permanent teeth using artificial intelligence on intraoral photographs. *J. Dent.* **2022**, *121*, 104124. [CrossRef] [PubMed]
47. Kühnisch, J.; Meyer, O.; Hesenius, M.; Hickel, R.; Gruhn, V. Caries Detection on Intraoral Images Using Artificial Intelligence. *J. Dent. Res.* **2022**, *101*, 158–165. [CrossRef] [PubMed]
48. Zhang, X.; Liang, Y.; Li, W.; Liu, C.; Gu, D.; Sun, W.; Miao, L. Development and evaluation of deep learning for screening dental caries from oral photographs. *Oral Dis.* **2022**, *28*, 173–181. [CrossRef]
49. Schlickenrieder, A.; Meyer, O.; Schoenewolf, J.; Engels, P.; Hickel, R.; Gruhn, V.; Hesenius, M.; Kühnisch, J. Automatized Detection and Categorization of Fissure Sealants from Intraoral Digital Photographs Using Artificial Intelligence. *Diagnostics* **2021**, *11*, 1608. [CrossRef]
50. Zhou, X.; Yu, G.; Yin, Q.; Yang, J.; Sun, J.; Lv, S.; Shi, Q. Tooth Type Enhanced Transformer for Children Caries Diagnosis on Dental Panoramic Radiographs. *Diagnostics* **2023**, *13*, 689. [CrossRef]
51. Gao, S.; Li, X.; Li, X.; Li, Z.; Deng, Y. Transformer based tooth classification from cone-beam computed tomography for dental charting. *Comput. Biol. Med.* **2022**, *148*, 105880. [CrossRef]
52. Ying, S.; Wang, B.; Zhu, H.; Liu, W.; Huang, F. Caries segmentation on tooth X-ray images with a deep network. *J. Dent.* **2022**, *119*, 104076. [CrossRef]

11. Literaturverzeichnis

Abdulkareem, A. A., Al-Taweel, F. B., Al-Sharqi, A. J. B., Gul, S. S., Sha, A., Chapple, I. L. C. (2023). Current concepts in the pathogenesis of periodontitis: From symbiosis to dysbiosis. *Journal of oral microbiology*, *15*(1), 2197779. https://doi.org/10.1080/20002297.2023.2197779

Alotaibi, G., Awawdeh, M., Farook, F. F., Aljohani, M., Aldhafiri, R. M., Aldhoayan, M. (2022). Artificial intelligence (AI) diagnostic tools: utilizing a convolutional neural network (CNN) to assess periodontal bone level radiographically-a retrospective study. *BMC oral health*, *22*(1), 399. https://doi.org/10.1186/s12903-022-02436-3

Chen, X., Guo, J., Ye, J., Zhang, M., Liang, Y. (2022). Detection of Proximal Caries Lesions on Bitewing Radiographs Using Deep Learning Method. *Caries research*, *56*(5-6), 455-463. https://doi.org/10.1159/000527418

Chen, C. C., Wu, Y. F., Aung, L. M., Lin, J. C., Ngo, S. T., Su, J. N., Lin, Y. M., Chang, W. J. (2023). Automatic recognition of teeth and periodontal bone loss measurement in digital radiographs using deep-learning artificial intelligence. *Journal of dental sciences*, *18*(3), 1301-1309. https://doi.org/10.1016/j.jds.2023.03.020

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers For Image Recognition at Scale. Online verfügbar: https://arxiv.org/abs/2010.11929 (Zugang am 18. Januar 2024).

Ekert, T., Krois, J., Meinhold, L., Elhennawy, K., Emara, R., Golla, T., Schwendicke, F. (2019). Deep learning for the Radiographic Detection of Apical Lesions. *Journal of endodontics*, *45*(7), 917-922.e5. https://doi.org/10.1016/j.joen.2019.03.016

Felsch, M., Meyer, O., Schlickenrieder, A., Engels, P., Schönewolf, J., Zöllner, F., Heinrich-Weltzien, R., Hesenius, M., Hickel, R., Gruhn, V., Kühnisch, J. (2023). Detection and localization of caries and hypomineralization on dental photographs with a vision transformer model. *NPJ digital medicine*, *6*(1), 198. https://doi.org/10.1038/s41746-023-00944-2

Fiorellini, J. P., Sourvanos, D., Sarimento, H., Karimbux, N., Luan, K. W. (2021). Periodontal and Implant Radiology. *Dental clinics of North America*, *65*(3), 447-473. https://doi.org/10.1016/j.cden.2021.02.003

Gao, S., Li, X., Li, X., Li, Z., Deng, Y. (2022). Transformer based tooth classification from cone-beam computed tomography for dental charting. *Computers in biology and medicine*, *148*, 105880. https://doi.org/10.1016/j.compbiomed.2022.105880

Greenhill, A. T., Edmunds, B. R. (2020). A primer of artificial intelligence in medicine. *Techniques and Innovations in Gastrointestinal Endoscopy*, *22*(2), 85-89. https://doi.org/10.1016/j.tgie.2019.150642

Hwang, J. J., Jung, Y. H., Cho, B. H., Heo, M. S. (2019). An overview of deep learning in the field of dentistry. *Imaging science in dentistry*, *49*(1), 1-7. https://doi.org/10.5624/isd.2019.49.1.1

Jiang, L., Chen, D., Cao, Z., Wu, F., Zhu, H., Zhu, F. (2022). A two-stage deep learning architecture for radiographic staging of periodontal bone loss. *BMC oral health*, *22*(1), 106. https://doi.org/10.1186/s12903-022-02119-z

Jordan, R. und Micheelis, W. (2016). Fünfte Deutsche Mundgesundheitsstudie (DMS V). Institut der Deutschen Zahnärzte (Materialienreihe Band 35). Köln, Deutscher Zahnärzte Verlag DÄV.

Kim, J., Lee, H. S., Song, I. S., Jung, K. H. (2019). DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs. *Scientific reports*, *9*(1), 17615. https://doi.org/10.1038/s41598-019-53758-2

Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems 25*, 1097-1105.

Krois, J., Ekert, T., Meinhold, L., Golla, T., Kharbot, B., Wittemeier, A., Dörfer, C., Schwendicke, F. (2019). Deep Learning for the Radiographic Detection of Periodontal Bone Loss. *Scientific reports*, *9*(1), 8495. https://doi.org/10.1038/s41598-019-44839-3

Kühnisch, J., Meyer, O., Hesenius, M., Hickel, R., Gruhn, V. (2022). Caries Detection on Intraoral Images Using Artificial Intelligence. *Journal of dental research*, *101*(2), 158-165. https://doi.org/10.1177/00220345211032524

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

Lee, J. H., Kim, D. H., Jeong, S. N., Choi, S. H. (2018b). Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *Journal of periodontal & implant science*, *48*(2), 114-123. https://doi.org/10.5051/jpis.2018.48.2.114

Lee, C. T., Kabir, T., Nelson, J., Sheng, S., Meng, H. W., Van Dyke, T. E., Walji, M. F., Jiang, X., Shams, S. (2022). Use of the deep learning approach to measure alveolar bone level. *Journal of clinical periodontology*, *49*(3), 260-269. https://doi.org/10.1111/jcpe.13574

Li, C. W., Lin, S. Y., Chou, H. S., Chen, T. Y., Chen, Y. A., Liu, S. Y., Liu, Y. L., Chen, C. A., Huang, Y. C., Chen, S. L., Mao, Y. C., Abu, P. A. R., Chiang, W. Y., Lo, W. S. (2021a). Detection of Dental Apical Lesions Using CNNs on Periapical Radiograph. *Sensors (Basel), 21*(21), 7049. https://doi.org/10.3390/s21217049

Li, H., Zhou, J., Zhou, Y., Chen, Q., She, Y., Gao, F., Xu, Y., Chen, J., Gao, X. (2021b). An Interpretable Computer-Aided Diagnosis Method for Periodontitis From Panoramic Radiographs. *Frontiers in physiology*, *12*, 655556. https://doi.org/10.3389/fphys.2021.655556

Lian, L., Zhu, T., Zhu, F., Zhu, H. (2021). Deep Learning for Caries Detection and Classification. *Diagnostics*, *11*(9), 1672. https://doi.org/10.3390/diagnostics11091672

Meusburger, T., Wülk, A., Kessler, A., Heck, K., Hickel, R., Dujic, H., Kühnisch, J. (2023). The Detection of Dental Pathologies on Periapical Radiographs–Results from a Reliability Study. *Journal of clinical medicine*, *12*(6), 2224. https://doi.org/10.3390/jcm12062224

Moidu, N. P., Sharma, S., Chawla, A., Kumar, V., Logani, A. (2022). Deep learning for categorization of endodontic lesion based on radiographic periapical index scoring system. *Clinical oral investigations*, *26*(1), 651-658. https://doi.org/10.1007/s00784-021-04043-y

Papapanou, P. N., Sanz, M., Buduneli, N., Dietrich, T., Feres, M., Fine, D. H., Flemmig, T. F., Garcia, R., Giannobile, W. V., Graziani, F., Greenwell, H., Herrera, D., Kao, R. T., Kebschull, M., Kinane, D. F., Kirkwood, K. L., Kocher, T., Kornman, K. S., Kumar, P. S., Loos, B. G., Machtei, E., Meng, H., Mombelli, A., Needleman, I., Offenbacher, S., Seymour, G. J., Teles, R., Tonetti, M. S. (2018). Periodontitis: Consensus report of workgroup 2 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. *Journal of periodontology*, *89 Suppl. 1*, S173-S182. https://doi.org/10.1002/JPER.17-0721

Pethani, F. (2021). Promises and perils of artificial intelligence in dentistry. *Australian dental journal*, *66*(2), 124-135. https://doi.org/10.1111/adj.12812

Pepelassi, E. A., Tsiklakis, K., Diamanti-Kipioti, A. (2000). Radiographic detection and assessment of the periodontal endosseous defects. *Journal of clinical*

periodontology, *27*(4), 224-230. https://doi.org/10.1034/j.1600-051x.2000.027004224.x

Roshanzamir, A., Aghajan, H., Soleymani Baghshah, M. (2021). Transformer-based deep neural network language models for Alzheimer´s disease risk assessment from targeted speech. *BMC medical informatics and decision making*, *21*(1), 92. https://doi.org/10.1186/s12911-021-01456-3

Schönewolf, J., Meyer, O., Engels, P., Schlickenrieder, A., Hickel, R., Gruhn, V., Hesenius, M., Kühnisch, J. (2022). Artificial intelligence-based diagnostics of molar-incisor-hypomineralization (MIH) on intraoral photographs. *Clinical oral investigations*, *26*(9), 5923-5930. https://doi.org/10.1007/s00784-022-04552-4

Schwendicke, F., Samek, W., Krois, J. (2020). Artificial Intelligence in Dentistry: Chances and Challenges. *Journal of dental research*, *99*(7), 769-774. https://doi.org/10.1177/0022034520915714

Schwendicke, F., Singh, T., Lee, J. H., Gaudin, R., Chaurasia, A., Wiegand, T., Uribe, S., Krois, J. (2021). Artificial intelligence in dental research: Checklist for authors, reviewers, readers. *Journal of dentistry*, *107*, 103610. https://doi.org/10.1016/j.jdent.2021.103610

Scott, J., Biancardi, A. M., Jones, O., Andrew, D. (2023). Artificial Intelligence in Periodontology: A Scoping Review. *Dentistry journal*, *11*(2), 43. https://doi.org/10.3390/dj11020043

Sheng, C., Wang, L., Huang, Z., Wang, T., Guo, Y., Hou, W., Xu, L., Wang, J., Yan, X. (2022). Transformer-Based Deep Learning Network for Tooth Segmentation on Panoramic Radiographs. *Journal of systems science and complexity*, *36*(1), 257-272. https://doi.org/10.1007/s11424-022-2057-9

Tonetti, M. S., Greenwell, H., Kornman, K. S. (2018). Staging and grading of periodontitis: Framework and proposal of a new classification and case definition. *Journal of periodontology*, *89 Suppl 1*, S159-S172. https://doi.org/10.1002/JPER.18-0006

Tsoromokos, N., Parinussa, S., Claessen, F., Moin, D. A., Loos, B. G. (2022). Estimation of Alveolar Bone Loss in Periodontitis Using Machine Learning. *International dental journal*, *72*(5), 621-627. https://doi.org/10.1016/j.identj.2022.02.009

Widyaningrum, R., Candradewi, I., Aji, N. R. A., Aulianisa, R. (2022). Comparison of Multi-Label U-Net and Mask R-CNN for panoramic radiograph segmentation to detect periodontitis. *Imaging science in dentistry*, *52*(4), 383-391. https://doi.org/10.5624/isd.20220105

Ying, S., Wang, B., Zhu, H., Liu, W., Huang, F. (2022). Caries segmentation on tooth X-ray images with a deep network. *Journal of dentistry*, *119*, 104076. https://doi.org/10.1016/j.jdent.2022.104076

Zhou, X., Yu, G., Yin, Q., Yang, J., Sun, J., Lv, S., Shi, Q. (2023). Tooth Type Enhanced Transformer for Children Caries Diagnosis on Dental Panoramic Radiographs. *Diagnostics*, *13*(4), 689. https://doi.org/10.3390/diagnostics13040689

Danksagung

Ich bedanke mich herzlichst bei allen, die mich während der Anfertigung meiner Promotionsschrift unterstützt haben und ohne deren Mithilfe diese so nicht möglich gewesen wäre.

Für die ausgezeichnete Betreuung während der Promotionsarbeit möchte ich allen voran meinem Doktorvater Herrn. Prof. Dr. Jan Kühnisch besonderen Dank entgegenbringen, der mir jederzeit mit fachlichem Wissen und enormem Einsatz zur Seite stand.

Darüber hinaus bedanke ich mich vielmals bei meinen Co-Autoren für die angenehme Zusammenarbeit und den kollegialen Austausch. Hierbei möchte ich insbesondere Helena Dujic großen Dank für die durchweg unermüdliche Unterstützung und Hilfsbereitschaft aussprechen.

Außerdem gilt mein herzlichster Dank natürlich auch meiner Familie und Partnerin für die unentbehrliche, moralische Unterstützung und Stärkung sowie ihr stets offenes Ohr.

.