Aus der

Poliklinik für Zahnerhaltung und Parodontologie

Klinikum der Ludwig-Maximilians-Universität München

Direktor: Prof. Dr. med. dent. Falk Schwendicke

„Vision Transformer für die automatisierte Erkennung von parodontalem Knochenabbau auf apikalen Röntgenaufnahmen"

Dissertation

zum Erwerb des Doktorgrades der Zahnmedizin

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität München

vorgelegt von

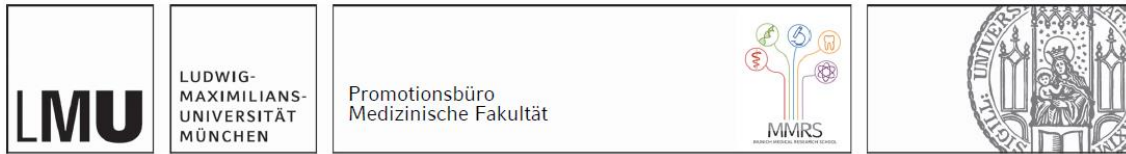Helena Dujić

aus

Nova Bila

Jahr

2024

Mit Genehmigung der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

Erstes Gutachten: Prof. Dr. med. dent. Jan Kühnisch

Zweites Gutachten: Prof. Dr. Roswitha Heinrich-Weltzien

Drittes Gutachten: Prof. Dr. Andrea Baur-Melnyk

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 28.05.2024

Affidavit

Eidesstattliche Versicherung

Dujić, Helena
_____
Name, Vorname

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

„Vision Transformer für die automatisierte Erkennung von parodontalem Knochenabbau auf apikalen Röntgenaufnahmen"

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, 29.05.2024   Helena Dujić
_____   _____
Ort, Datum   Unterschrift Doktorandin bzw. Doktorand

Inhaltsverzeichnis

Abkürzungsverzeichnis

ACC   Accuracy (dt. Genauigkeit)

AUC   Area under ROC curve (dt. Fläche unter der ROC-Kurve)

CNNs   Convolutional neural networks (dt. faltende neuronale Netze)

DL   Deep Learning

KI   Künstliche Intelligenz

ROC   Receiver operating characteristic curve

SE   Sensitivität

SP   Spezifität

ViT   Vision Transformer

Publikationsliste

Paper I (Erstautorschaft):

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

Paper II (Co-Autorschaft):

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). https://doi.org/10.3390/jcm12227189

1. Beitrag zu den Veröffentlichungen

Im Rahmen des Dissertationsprojekts und des Publikationsvorhabens waren folgende Arbeitsschritte erforderlich, die in der Arbeitsgruppe gemäß nachstehenden Tabellen durchgeführt wurden.

1.1 Beitrag zu Paper I

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

	Helena Dujić	Prof. Dr. Jan Kühnisch	Co-Autoren
Ethikantrag	-	100 %	-
Literaturrecherche	100 %	-	-
Projektidee/ Studiendesign	40 %	50 %	10 %
Röntgenbild-Auswahl	70 %	10 %	20 %
Befundung der Röntgenbilder	70 %	10 %	20 %
Modell-Training & Datenanalyse	-	-	100 %
Dateninterpretation und Auswahl veröffentlichungswürdiger Daten	80 %	10 %	10 %
Manuskript Management bis zur Veröffentlichung	90 %	10 %	-
Verfassen und Einreichen der Dissertation	100 %	-	-

1.2 Beitrag zu Paper II

Publikation "Detection of Periodontal Bone Loss on Periapical Radiographs –

A Diagnostic Study Using Different Convolutional Neural Networks"

	Helena Dujić	Prof. Dr. Jan Kühnisch	Co-Autoren
Ethikantrag	-	100 %	-
Literaturrecherche	-	-	100 % (Hoss P.)
Projektidee/ Studiendesign	40 %	50 %	10 %
Röntgenbild-Auswahl	20 %	10 %	70 % (Hoss P.)
Befundung der Röntgenbilder	20 %	10 %	70 % (Hoss P.)
Modell-Training & Datenanalyse	-	-	100 %
Dateninterpretation und Auswahl veröffentlichungswürdiger Daten	10 %	10 %	80 % (Hoss P.)
Manuskript Management bis zur Veröffentlichung	10 %	10 %	80 % (Hoss P.)

2. Einleitung

2.1 Hintergrund

2.1.1 Parodontitis – Definition, Ätiologie und Pathogenese

Parodontitis ist eine chronische, nicht übertragbare Erkrankung, die mit einer Destruktion des Zahnhalteapparats einhergeht. Weltweit ist ein hoher Bevölkerungsanteil an Parodontitis erkrankt [1, 2]. Zwar hat sich die Prävalenz der schweren Parodontitis nach der aktuellen Deutschen Mundgesundheitsstudie (DMS V) halbiert, dennoch ist in Deutschland jeder zweite junge Erwachsene und jüngere Senior von einer parodontalen Erkrankung betroffen [3]. In den vergangenen Jahren veränderte sich die Sichtweise auf die Pathogenese der Parodontitis. Inzwischen ist bekannt, dass die Erkrankung multifaktoriell bedingt ist und unter anderem von der Immunreaktion der Betroffenen abhängig ist [4-6]. Der Biofilm löst die primäre Entzündungsreaktion aus, die in eine Dysbiose der oralen bakteriellen Mikroflora übergehen kann. Klinisch macht sich zunächst eine anhaltende Blutungsneigung der Gingiva bemerkbar, die, wenn sie unbehandelt bleibt und der Biofilm persistiert, zu einer fortschreitenden Destruktion der parodontalen Fasern und schließlich zur Bildung von Parodontaltaschen führt. Parodontitis wird außerdem mit verschiedenen Risikofaktoren wie Rauchen und Stress sowie mit systemischen Erkrankungen wie Diabetes mellitus und kardiovaskulären Erkrankungen in Verbindung gebracht [7-10]. Das Initialstadium der Erkrankung ist primär durch einen messbaren klinischen Attachmentverlust (clinical attachment loss, CAL) gekennzeichnet. Unbehandelt führt die parodontale Erkrankung bei weiterem Fortschritt zu einer Resorption und einem pathologischen Abbau des Alveolarfortsatzes. Die Nichtbehandlung führt somit zu Zahnlockerungen und letztlich auch zum Zahnverlust. Für die Betroffenen bedeutet dies nicht nur eine Einschränkung der Kaufunktion und Funktionalität, sondern auch eine Beeinträchtigung der Ästhetik und Lebensqualität. Weiterhin erfordert der mögliche Verlust eines oder mehrerer Zähne einen prothetischen Zahnersatz, der mitunter mit hohen Kosten verbunden sein kann [6]. Die Parodontitis geht somit auch mit finanziellen Belastungen für die Betroffenen und das Gesundheitssystem einher.

Mit Blick auf die Verbreitung parodontaler Erkrankungen in der Bevölkerung kommt der frühzeitigen und validen Diagnostik eine hohe Bedeutung zu. Damit sollte es in einer Vielzahl von Fällen gelingen, weitere parodontale Destruktionen zumindest zu reduzieren bzw. optimalerweise zu verhindern.

2.1.2 Parodontitis – Diagnostik und Klassifikation

Für die Diagnostik von parodontalen Erkrankungen wurde gemäß der neuen Klassifikation für parodontale und periimplantäre Erkrankungen [11, 12] ein klinischer Behandlungspfad definiert [13, 14]. Dieser umfasst die Identifikation von Betroffenen und die Bestätigung der Verdachtsdiagnose, die durch Staging und Grading ergänzt werden sollen. Staging bezieht sich auf die Einstufung der Parodontitis, die den Schweregrad der Erkrankung wiedergibt. Das Grading bezieht sich auf das Risikoprofil und die individuellen Patientenfaktoren, die den Krankheitsverlauf beeinflussen können. Daraus leitet sich auch die spezifische Bewertung für die Komplexität des Behandlungsverlaufs ab [14]. Für die Bestätigung der Verdachtsdiagnose sind klinisch in den Anfangsstadien der Erkrankung primär das Bluten auf Sondieren (engl. bleeding on probing, BOP) und der klinische Attachmentverlust von Relevanz. Mit weiterem Fortschreiten und bei unzureichender Behandlung werden auch erste radiologische Anzeichen für den parodontalen Knochenabbau sichtbar [15]. Der prozentuale Anteil des parodontalen Knochenabbaus wird anhand anatomischer Referenzpunkte – Schmelz-Zement-Grenze, Apex, tiefster Punkt des marginalen Knochenabbaus – im Verhältnis zur gesamten Wurzellänge bestimmt. Praktisch gesehen unterliegt die Beurteilung des parodontalen Knochenabbaus der Subjektivität der Behandler. Aufgrund ihrer klinischen Erfahrung können mehrere Behandler den parodontalen Knochenabbau unterschiedlich einstufen [16]. Die Beleuchtungsstärke und der Ort, an dem der radiologische Knochenabbau beurteilt wurde, können ebenfalls eine Rolle spielen. Dies kann Auswirkungen auf das Staging und Grading der Parodontitis haben, insbesondere bei einem Behandlerwechsel. Um die Qualität der Evaluation zu verbessern, wäre es möglich, den gesamten Bewertungsprozess für den parodontalen Knochenabbau automatisiert durchzuführen. Dies würde nicht nur die zeit- und arbeitsintensive Befundung von Röntgenbildern vereinfachen, sondern auch eine lückenlose Verlaufskontrolle der Patienten ermöglichen. Eine derartige

Automatisierung lässt sich mit künstlicher Intelligenz (KI) realisieren. In den vergangenen Jahren wurden bereits Veröffentlichungen zu diesem Thema publiziert, welche sich mit apikalen Einzelzahnaufnahmen und Panoramaschichtaufnahmen befassten, z. B. [17-20]. Unter anderem wurden Studien veröffentlicht, die sich mit Staging und Grading auf der Basis von klinischen Informationen befassen [21].

2.1.3 Künstliche Intelligenz in der parodontalen Diagnostik

Die KI könnte den Klinikalltag erleichtern, indem durch die automatisierte Befundung Zeit eingespart und durch standardisierte Prozesse die individuelle Subjektivität bei der Befundung ausgeschlossen werden kann [22, 23]. Zudem könnten auch Patienten ihre eigenen Gesundheitsdaten in verschiedenen Anwendungsprogrammen nutzen, um Auskunft über ihre Zahngesundheit zu erhalten. Da bei der klinischen Diagnostik täglich große Datenmengen generiert werden – Röntgen-, DVT-, CT- und MRT-Aufnahmen – finden KI-Modelle vorwiegend bei Bilderkennungsaufgaben Anwendung. Die Spezialgebiete der KI, Deep Learning (DL) und Transfer Learning, sind für die Entwicklung von KI-Modellen relevant, die komplexere Zusammenhänge zwischen den abgebildeten Strukturen erkennen und Normabweichungen erfassen können. Insbesondere in der zahnmedizinischen Diagnostik wurden bisher Convolutional neural networks (CNNs, dt. faltende neuronale Netze) eingesetzt, um Karies oder apikale Parodontitiden zu erkennen, z. B. [24-28]. Ferner finden die seit einigen Jahren in der Spracherkennung etablierten Transformer-Netze auch für Bilderkennungsaufgaben als Vision Transformer (ViT) Verwendung [29]. Im Vergleich zu CNNs sind ViT durch ihren Aufmerksamkeitsmechanismus in der Lage, die Zusammenhänge zwischen nicht unmittelbar aneinander angrenzenden Bildbereichen zu interpretieren. Für die radiologische Beurteilung des parodontalen Knochenabbaus könnte dies von Vorteil sein. Bereits veröffentlichte Daten haben gezeigt, dass verschiedene CNN-Modelle in der Lage sind, parodontalen Knochenabbau sowohl auf apikalen Röntgenbildern, z. B. [30-33], als auch auf Panoramaschichtaufnahmen zu erkennen, z. B. [34-37]. Weiterhin konnte gezeigt werden, dass selbst erfahrene Parodontologen eine vergleichbare diagnostische Leistung mit dem KI-Modell haben, welches jedoch die Beurteilung des

parodontalen Knochenverlustes deutlich schneller durchführen kann [38]. Die meisten der bisher veröffentlichten Arbeiten verwenden unterschiedliche CNN-Modelle und unterschiedliche statistische Variablen, die die diagnostische Leistung des KI-Modells beschreiben. Ein direkter Vergleich der diagnostischen Leistung ist aufgrund der Heterogenität der bisherigen Literatur allerdings nicht immer möglich. Die ViT wurden zwar für die Entwicklung von DL-Modellen für die Kariesdiagnostik verwendet [28, 39, 40], in der Literatur gibt es jedoch keine Daten über ihre Verwendung für die Bewertung des parodontalen Knochenabbaus.

2.2   Fragestellung und Hypothese

Ziel dieser Studie war es, verschiedene ViT für die automatisierte Erkennung von parodontalem Knochenabbau zu testen. Zur Analyse wurden fünf frei verfügbare ViT (ViT-base und ViT-large von Google, BEiT-base und BEiT-large von Microsoft, und DeiT-base von Facebook/Meta) herangezogen, die bereits mit einem größeren Datensatz vortrainiert wurden [41]. Die fünf ViT wurden in einer Klassifikationsaufgabe zur Erkennung von parodontalem Knochenabbau an einzelnen apikalen Röntgenbildern trainiert und feinjustiert. Die Studie sollte ermöglichen, verschiedene ViT-Modelle miteinander zu vergleichen. Es stellte sich die Frage, ob die diagnostische Leistungsfähigkeit der einzelnen ViT-Modelle variiert und ob sie im Vergleich zu den CNNs überlegen wäre. Außerdem wurde mit Blick auf die unterschiedlichen Zahngruppen die Frage gestellt, ob es einen Unterschied in der diagnostischen Leistung geben würde. Obwohl keine Studien mit ViT für ein ähnliches Vorhaben vorliegen, wurde in Anlehnung an die auf CNN-Modellen basierenden Veröffentlichungen die Hypothese aufgestellt, dass eine diagnostische Gesamtgenauigkeit von mindestens 90 % erreichbar ist. Außerdem wurde angenommen, dass sich die diagnostische Leistung zwischen den einzelnen ViT nicht signifikant unterscheiden würde.

3. Material und Methode

3.1 Studiendesign

Im Rahmen dieses Forschungsvorhabens wurde ein Ethikantrag mit der Projektnummer 020-798 eingereicht und von der Ethikkommission der Medizinischen Fakultät der Ludwig-Maximilians-Universität (LMU) genehmigt. Durch die Anonymisierung der Röntgenaufnahmen war eine Einverständniserklärung der Patienten nicht erforderlich. Um eine korrekte Darstellung der Studie und der erzielten Ergebnisse zu gewährleisten, wurden die Empfehlungen der STARD-Checkliste [42] sowie die Empfehlungen für die Darstellung von KI-Studien in der Zahnmedizin befolgt. [43].

Weiterhin haben alle am Projekt beteiligten Untersucher an einem zweitägigen Workshop teilgenommen, in dem die Unterweisung vom Studienleiter Prof. Dr. Jan Kühnisch erfolgte. Anschließend wurde die Zuverlässigkeit dieser Unterweisung im Rahmen eines Kalibrierungskurses überprüft. Anhand von 150 apikalen Röntgenaufnahmen wurde die Reproduzierbarkeit der Beurteilung des parodontalen Knochenabbaus bewertet. Die berechneten Kappa-Werte zeigten eine moderate Interrater- Reliabilität (0,454 – 0,482) und eine zufriedenstellende Intrarater-Reliabilität (0,739) [16].

3.2 Datensatz (Referenzstandard)

Um sicherzustellen, dass der Datensatz so repräsentativ wie möglich ist, wurden bestimmte Ausschlusskriterien für die Auswahl der Röntgenaufnahmen festgelegt. So wurden etwa Röntgenaufnahmen mit Implantaten oder Wurzelkanalbehandlungen ausgeschlossen. Des Weiteren wurden verzerrte Röntgenaufnahmen, Aufnahmen mit unvollständig abgebildeten Zähnen und Aufnahmen mit Artefakten nicht einbezogen. Im Ergebnis wurden insgesamt 21.819 anonymisierte Röntgenaufnahmen von Front- und Seitenzähnen aus dem Ober- und Unterkiefer ausgewählt.

Zusätzlich wurde die 2018 veröffentlichte Klassifikation parodontaler und periimplantärer Erkrankungen [11] berücksichtigt, um diagnostische Kriterien für die Röntgenbefunde zu bestimmen. Dabei wurden die Scores wie folgt definiert:

- Score 0 – parodontaler Knochenabbau nicht erkennbar
- Score 1 – parodontaler Knochenabbau <15 % des koronalen Wurzeldrittels
- Score 2 – parodontaler Knochenabbau 15 – 33 % des koronalen Wurzeldrittels
- Score 3 – parodontaler Knochenabbau bis zum mittleren Wurzeldrittel und darüber hinaus

Mit den diagnostischen Scores erfolgte die Befundung von 21.819 Röntgenaufnahmen in zwei Schritten (Tabelle 1). Diese wurden zunächst von den teilnehmenden Zahnärzten mit weniger als einem Jahr klinischer Erfahrung entsprechend den Scores kategorisiert, woraufhin diese Befunde durch Zahnärzte mit mehr klinischer Erfahrung gegengeprüft wurden. Der Datensatz diente als Referenz für die KI-Modelle und wurde dann in einen Trainingsdatensatz (N=18.819) und einen Testdatensatz (N=3.000) unterteilt. Der Trainingsdatensatz wurde zum Trainieren der KI-Modelle verwendet, während der Testdatensatz dazu diente, den Trainingserfolg zu überprüfen.

Parodontaler Knochenabbau	Score 0 (N)	Score 1 (N)	Score 2 (N)	Score 3 (N)	Summe (N)
Oberkieferseitenzähne					
1. Quadrant	1.701	1.826	851	367	4.745
2. Quadrant	1.231	2.080	1.093	312	4.716
Unterkieferseitenzähne					
3. Quadrant	1.477	2.033	593	157	4.260
4. Quadrant	1.282	2.027	713	143	4.165
Frontzähne					
Oberkiefer	653	661	433	197	1.944
Unterkiefer	202	676	786	325	1.989
Summe (N)	6.546	9.303	4.469	1.501	21.819

Tabelle 1   Übersicht und Anzahl der apikalen Röntgenaufnahmen in den einzelnen Zahngruppen und die dazugehörigen diagnostischen Kriterien.

3.3   Training von KI-basierten Modellen

Die fünf ViT (ViT-base und ViT-large von Google, BEiT-base und BEiT-large von Microsoft und DeiT-base von Facebook/Meta) wurden mit Backpropagation trainiert. Das Training der Modelle wurde durch den Einsatz von Floating Point 16 an Universitätsrechnern (i9 10850K 10 × 360 GHz, Intel Corp., Santa Clara CA, USA) mit 64 GB RAM und professionellen Grafikkarten (RTX A6000 48 GB, Nvidia, Santa Clara CA, USA) beschleunigt. Die *Batch*-Größe belief sich auf 16 zufällig ausgewählte Röntgenbilder, und die Trainingsdauer wurde auf fünf Epochen festgelegt. Bei den genannten KI-Modellen handelt es sich um bereits auf ImageNet vortrainierte Netze [41]. Die Lernrate für die Feinabstimmung bei der Aufgabe zur Erkennung von parodontalem Knochenabbau lag bei 5×10^{-5}. Mit dem Testdatensatz (N=3.000) wurde die Leistung der ViT anschließend evaluiert.

3.4   Statistische Analyse

Die Auswertung der Daten erfolgte mit Python (Version 3.8.5, http://www.python.org). Die statistische Analyse umfasste die diagnostische Genauigkeit (ACC, engl. accuracy) berechnet durch die Anzahl der echten Negativen (TN, engl. true negatives), der echten Positiven (TP, engl. true positives), der falschen Positiven (FP, engl. false positives) und der falschen Negativen (FN, engl. false negatives). Des Weiteren wurden Sensitivität (SE), Spezifität (SP), positive/negative prädiktive Werte (PPV/NPV, eng. positive/negative predictive values), die Fläche unter der Kurve (AUC, engl. area under the curve) und die Receiver Operating Characteristic (ROC) Kurve berechnet.

4.   Ergebnisse

Die diagnostische Leistung der ViT für die automatisierte Erkennung von parodontalem Knochenabbau wurde sowohl für alle Röntgenbilder zusammen als auch für die einzelnen Zahngruppen bestimmt. Aus dem gesamten Datensatz von 21.819 Einzelzahnaufnahmen stammte die Mehrheit der Bilder aus dem Seitenzahnbereich (81,9 %), während die Frontzähne 18,1 % des gesamten Datensatzes ausmachten. Die detaillierte Anzahl der Röntgenbilder in den einzelnen Zahngruppen und die entsprechenden diagnostischen Scores sind in Tabelle 1 aufgeführt. Die ViT wiesen für alle Zahngruppen folgende Werte auf: ACC 83,4 – 85,2 %, SE 87,3 – 91,3 %, SP 69,9 – 74,5 % und AUC 0,899 – 0,918. Eine detaillierte Darstellung der Werte für die einzelnen ViT-Modelle bietet die Tabelle 2, der Genauigkeitsverlauf während der Auswertung mit dem Testdatensatz ist in Abbildung 1 dargestellt. Betrachtet man die Zahngruppen und die zugehörigen Ergebnisse getrennt, so zeigt sich für die ViT, dass die höchste diagnostische Leistung für die Unterkieferfrontzähne erzielt wurde (ACC 94,1 – 96,7 %, SE 97,2 – 99,2 %,   SP 60,0 – 84,0 %,   AUC 0,944 – 0,970) (Abbildung 2). Die niedrigsten Werte wurden für das Oberkieferseitenzahngebiet ausgerechnet (ACC 78,1 – 81,0 %, SE 81,6 – 87,5 %, SP 63,0 – 70,3 %, AUC 0,851 – 0,875) (Abbildung 3).

ViT	True Positives		True Negatives		False Positives		False Negatives		Diagnostische Leistung					
	N	%	N	%	N	%	N	%	ACC	SE	SP	NPV	PPV	AUC
ViT-base	1.884	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	1.831	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	1.885	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	1.914	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	1.879	62,6	646	21,5	257	8,6	218	7,3	84,2	89,6	71,5	74,8	88,0	0,908

Tabelle 2   Übersicht über die gesamte diagnostische Leistung von ViT für die automatisierte Erkennung von parodontalem Knochenabbau. Dargestellt sind Angaben zu echten Positiven und Negativen (TP, TN) sowie falsch Positiven und Negativen (FP, FN). Genauigkeit (ACC in %), Sensitivität (SE in %), Spezifität (SP in %), negativer prädiktiver Wert (NPV in %), positiver prädiktiver Wert (PPV in %) und die Fläche unter der Kurve (AUC) wurden berechnet.
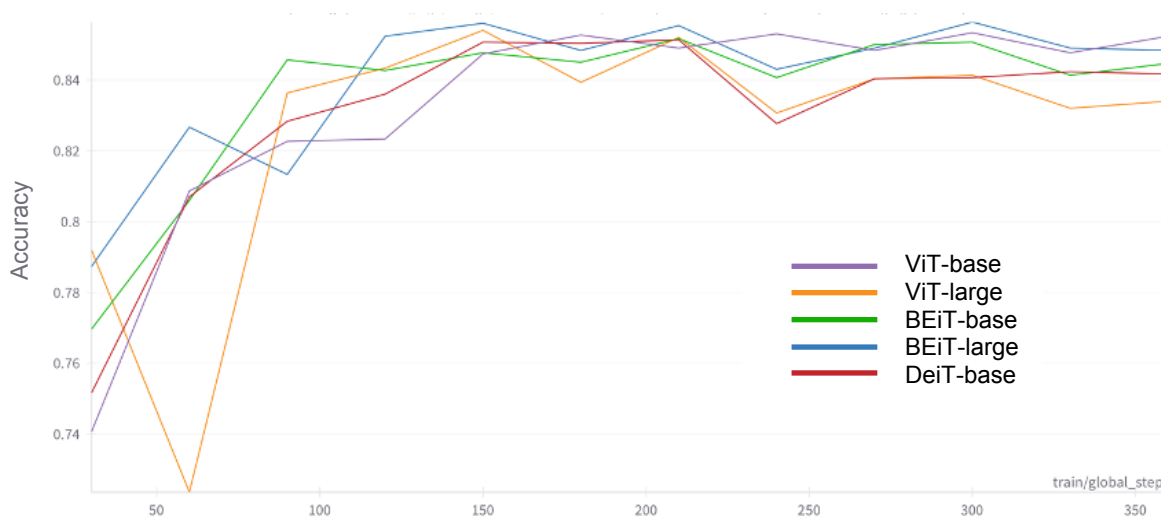

Abbildung 1   Genauigkeitsverlauf (Accuracy) aus dem Evaluationsprozess mit dem Testdatensatz (N=3.000) für die fünf ViT: ViT-base und ViT-large von Google, BEiT-base und BEiT-large von Microsoft und DeiT-base von Facebook/Meta.
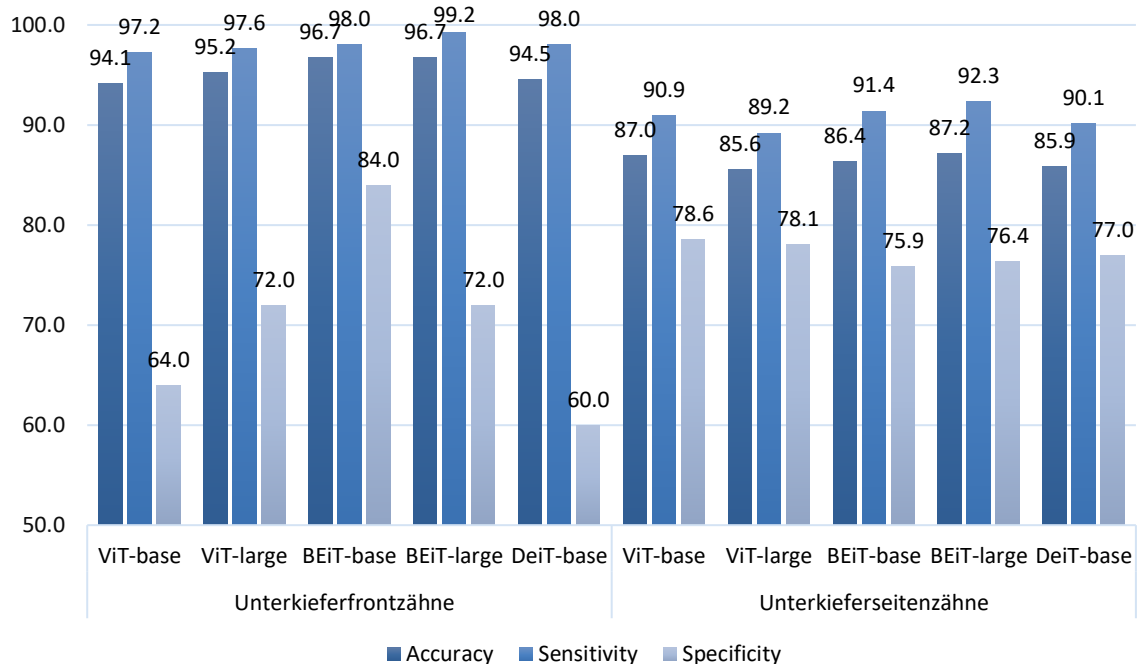
Abbildung 2   Übersicht und Vergleich der diagnostischen Leistung von ViT für die automatisierte Erkennung von parodontalem Knochenabbau an Front- und Seitenzähnen im Unterkiefer. Folgende Werte sind grafisch in Prozent (%) dargestellt: Genauigkeit (Accuracy), Sensitivität (Sensitivity) und Spezifität (Specificity).
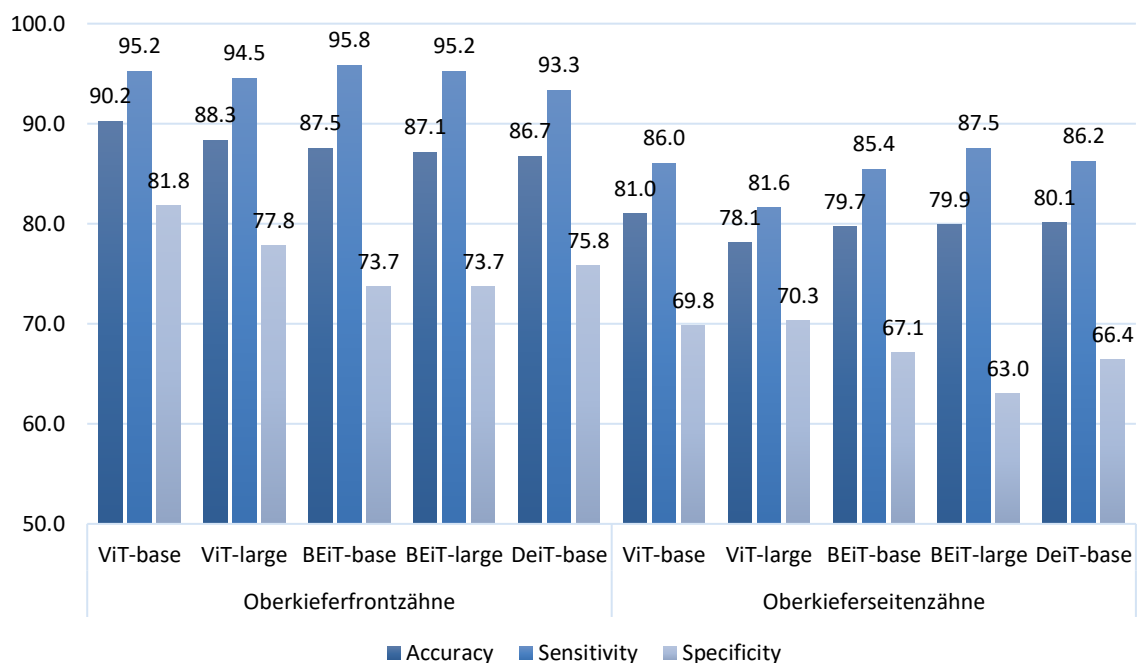

Abbildung 3   Übersicht und Vergleich der diagnostischen Leistung von ViT für die automatisierte Erkennung von parodontalem Knochenabbau an Front- und Seitenzähnen im Oberkiefer. Folgende Werte sind in Prozent (%) grafisch dargestellt: Genauigkeit (Accuracy), Sensitivität (Sensitivity) und Spezifität (Specificity).

5.   Diskussion

Mit diesem Projekt wurde das Ziel verfolgt, verschiedene ViT-Modelle für die automatisierte Erkennung von parodontalem Knochenabbau zu trainieren und sie miteinander zu vergleichen. Obwohl die erzielten Ergebnisse (Tabelle 2, Abbildung 1 – Abbildung 3) mit CNNs vergleichbar oder sogar besser sind, erreichte keines der getesteten ViT-Modelle eine Gesamtgenauigkeit von mehr als 90 %. Die Ausgangshypothese musste daher verworfen werden. Allerdings zeigten die verschiedenen ViT nur geringe Unterschiede in ihrer diagnostischen Leistung, was unsere Annahme bestätigte.

In den einzelnen Zahngruppen wurden Unterschiede in der diagnostischen Leistung festgestellt. Besonders hervorzuheben sind hier die Unterkieferfrontzähne, bei denen die ViT hohe Ergebnisse zeigten. Die Genauigkeitswerte lagen hier zwischen 94,1 und 96,7 % (Abbildung 2). Demgegenüber wurden die niedrigsten Werte für die Oberkieferseitenzähne berechnet (Genauigkeit 78,1 – 81,0 %; Abbildung 3). Die Sensitivität und Spezifität folgten der zuvor genannten Hierarchie, bei der sich die Sensitivität mit Werten zwischen 81,6 und 99,2 % als durchweg hoch einstufen lässt. Da die ViT ähnliche Werte zeigten, scheinen die Differenzen in den einzelnen Zahngruppen in der Auswahl der Röntgenaufnahmen begründet zu sein. In diesem Zusammenhang sei auf die Anatomie der abgebildeten Region, z. B. die Kieferhöhle und das Jochbein im Oberkieferseitenzahngebiet, und die damit einhergehenden Summationseffekte hingewiesen [44]. Ferner hätte die Wahl der intraoralen Aufnahmetechnik die Ergebnisse beeinflussen können, z. B. die Halbwinkeltechnik anstelle der Rechtwinkeltechnik im Seitenzahnbereich. Die fehlende Literatur mit einem ähnlichen Studiendesign lässt zwar keinen direkten Vergleich zu, dennoch wurden in Studien mit CNNs einige dieser Gründe berücksichtigt. So wurden beispielsweise nur apikale Röntgenaufnahmen von bestimmten Zahngruppen einbezogen. Infolgedessen wurde die diagnostische Leistung auf der Grundlage von Einzelzahnaufnahmen ausschließlich aus dem Frontzahngebiet [17] oder lediglich aus dem Unterkiefer [20] bestimmt. An dieser Stelle sei betont, dass ein ausgeglichener Datensatz eine wichtige Voraussetzung für die Vergleichbarkeit der Ergebnisse ist.

Um die anvisierte diagnostische Leistung mit KI-Modellen zu erreichen, sollten einige Faktoren in Betracht gezogen werden. An erster Stelle ist der Datensatz bzw. die Datensatzgröße zu nennen. Die gewählte KI-Architektur sollte mit so vielen repräsentativen Bildern wie möglich trainiert und fein abgestimmt werden. Dies wurde in den vergangenen fünf Jahren nur in wenigen Studien berücksichtigt [35, 45-47], wobei die meisten mit einem Datensatz von weniger als 2000 Röntgenbildern arbeiteten [19, 21, 36, 48, 49]. Die Leistung der KI-Modelle kann jedoch auch bei kleineren Datensätzen mithilfe von manuellen Annotationen gesteigert werden. Diese sind zwar äußerst aufwendig und zeitintensiv, ermöglichen aber auch das Erlernen wichtiger anatomischer Punkte, die zur Beurteilung von Röntgenaufnahmen und parodontalem Knochenabbau von Bedeutung sind. So konnten Lee et al. [21] mit 693 annotierten Röntgenbildern ein KI-Modell für die Segmentierung von Knochen, Zähnen und Schmelz-Zement-Grenzen entwickeln, welches die Stadien 1 – 3 mit einer Genauigkeit von bis zu 90 % bestimmen konnte. Je nach gewähltem Annotationsstil konnten im Vergleich dazu aber auch schlechtere Ergebnisse verzeichnet werden [32]. Daraus lässt sich schließen, dass manuelle Annotationen zwar die Leistung der KI erhöhen können, aber von Zahnärzten, Parodontologen oder Radiologen mit möglichst hoher Pixelgenauigkeit durchgeführt werden sollten. Obwohl die 21.819 Röntgenaufnahmen in unserem Projekt nicht manuell annotiert wurden, konnten die ausgewählten ViT-Modelle eine mit der Literatur vergleichbare Leistung bieten. Allerdings wurden diese nur in einer Klassifikationsaufgabe trainiert und getestet, daher ist es denkbar, dass die Ergebnisse mit manueller Annotation zusätzlich optimiert werden könnten.

Bei den gewählten KI-Architekturen sollte primär der Unterschied zwischen den einzelnen ViT-Modellen hervorgehoben werden. Transformer-Netze wurden 2017 erstmals für die Verarbeitung natürlicher Sprache vorgestellt [50]. Mit den ViT konnten diese Netze auch für Bilderkennungsaufgaben eingesetzt werden [29]. Dabei werden die einzelnen Bildbereiche wie eine Aufeinanderfolge betrachtet, deren Zusammenhang die ViT mit ihrem Aufmerksamkeitsmechanismus erkennen können. Wenn dieser Aufmerksamkeitsmechanismus speziell auf dentale Röntgenaufnahmen angewendet wird, können Strukturen wie die Wurzelspitze, die Schmelz-Zement-Grenze und der Alveolarknochen, die zwar nicht miteinander verbunden sind, zur Beurteilung des parodontalen Knochenabbaus herangezogen

werden. Bei den für diese Studie ausgewählten ViT handelt es sich um frei verfügbare und vortrainierte Netze. Dabei unterscheiden sich das ViT-base und ViT-large von Google in erster Linie durch ihre Größe, wobei das größere Modell eine bessere Leistung erbringt und mehr Rechenkapazität erfordert [29]. DeiT-base von Facebook/Meta, ähnlich aufgebaut wie ViT-base, zeichnet sich durch dateneffizientes Training aus und liefert vergleichbare Ergebnisse mit weniger Trainingsdaten [51]. BEiT-base und BEiT-large von Microsoft unterscheiden sich ebenfalls in ihrer Größe. Das Hauptmerkmal dieser Modelle ist ein Vortraining auf einem größeren Datensatz, der mitunter maskierte Bilder enthält, auf denen die fehlenden Bildbereiche durch das Modell vorhergesagt werden können [52]. Da in der Literatur keine weiteren Studien identifiziert werden konnten, in denen ViT zur Erkennung von parodontalem Knochenabbau eingesetzt wurden, wurde der Literaturvergleich mit Studien durchgeführt, die CNNs für ähnliche oder gleiche Aufgaben eingesetzt haben. CNNs stellen dabei auch DL-Modelle dar, die sich jedoch in ihrer Struktur und ihrem Mechanismus von den ViT wesentlich unterscheiden. Sie sind besonders nützlich für Bilderkennungsaufgaben, bei denen der Fokus beispielsweise auf einzelnen Zähnen oder konkreten zahnbezogenen pathologischen Befunden (z. B. Karies) liegt [26, 49].

Das beschriebene Projekt weist Stärken und Schwächen auf, die nachfolgend erläutert werden. Als Stärke ist hier zunächst der Vergleich mehrerer ViT-Modelle für dieselbe Klassifikationsaufgabe zu nennen. Bisher veröffentlichte Studien haben den direkten Vergleich aufgrund selbst entwickelter CNN-Modelle und unterschiedlicher Parameterwahl bei der statistischen Analyse erschwert [53]. In dieser Studie konnte gezeigt werden, welche Leistung fünf verschiedene, frei verfügbare KI-Modelle bei gleichem Studiendesign erreichen können. Zu den Stärken zählt auch der Datensatz mit 21.819 apikalen Röntgenaufnahmen, die vor dem KI-Training von Zahnärzten entsprechend dem parodontalen Knochenabbau kategorisiert wurden [11]. Der Datensatz umfasste sowohl Oberkiefer- als auch Unterkieferzähne und hebt diese Studie als eine der wenigen Studien hervor, die über einen derart großen Datensatz verfügte [35, 46]. Als Schwachpunkt ist die Tatsache zu nennen, dass in dieser Studie nur apikale Röntgenaufnahmen verwendet wurden. Klinisch werden zur Beurteilung des parodontalen Knochenabbaus neben apikalen Röntgenaufnahmen auch

Panoramaschichtaufnahmen verwendet. Die bisher veröffentlichte Literatur enthält auch einige Studien mit KI-Modellen, die mit Panoramaschichtaufnahmen trainiert wurden, z. B. [18, 19, 35, 37, 48, 54]. Folglich wäre diese Studie auch mit Panoramaschichtaufnahmen möglich. Obwohl für den ausgewählten Datensatz das Vorhandensein von parodontalem Knochenabbau festgestellt wurde, wurde das Staging nach der neuen Klassifikation nicht mit den KI-Modellen als Teil der Klassifikationsaufgabe bewertet. Dies kann ebenfalls als Schwachpunkt eingeordnet werden. Zugleich bedarf es aber eines anderen Studiendesigns oder manueller Annotationen von anatomischen Merkmalen auf den Röntgenbildern, die für die Berechnung des Knochenabbaus relevant sind. Wie bereits erwähnt, ist die manuelle Annotation solcher großen Datensätze zeitaufwendig und erfordert mehrere kalibrierte Zahnärzte oder Parodontologen. Dennoch könnten diese die Leistung des Modells optimieren.

Zusammenfassend hat die beschriebene Arbeit bestätigt, dass ViT den parodontalen Knochenabbau auf apikalen Röntgenbildern mit akzeptabler Genauigkeit erkennen können. Es sind jedoch weitere Studien erforderlich, einschließlich größerer und genau annotierter Datensätze, die zudem einen Vergleich mit erfahrenen Spezialisten ermöglichen.

6. Zusammenfassung

Die radiologische Beurteilung des parodontalen Knochenabbaus unterliegt häufig der Subjektivität des Behandlers. Eine Optimierung der Zuverlässigkeit radiologischer Bewertungen könnte durch die Automatisierung mittels künstlicher Intelligenz realisiert werden. In zahnmedizinischen Diagnostikaufgaben werden zunehmend neuronale Netze eingesetzt, wobei in den letzten Jahren auch Vision Transformer (ViT) für Bilderkennungsaufgaben vorgestellt wurden. Ziel der vorliegenden Studie war es fünf frei verfügbare ViT (ViT-base/ViT-large von Google, BEiT-base/BEiT-large von Microsoft, DeiT-base von Facebook/Meta) für die automatisierte Erkennung von parodontalem Knochenabbau auf apikalen Röntgenaufnahmen zu testen. Dabei wurde hypothetisch angenommen, dass die erzielte Gesamtgenauigkeit >90 % betragen würde und die Leistungsunterschiede zwischen den einzelnen ViT gering ausfallen würden.

Ein anonymisierter Datensatz von 21.819 Einzelzahnaufnahmen (Trainingssatz N=18.819/ Testsatz N=3.000) wurde von kalibrierten Zahnärzten mit vordefinierten diagnostischen Kriterien kategorisiert, wobei zwischen gesundem Parodontium und parodontalem Knochenabbau unterschieden wurde (Score 0: gesund, Score 1: Knochenabbau <15 % des koronalen Wurzeldrittels, Score 2: Knochenabbau 15 – 33 % des koronalen Wurzeldrittels, Score 3: Knochenabbau bis zum mittleren Wurzeldrittel und darüber hinaus). Die ViT wurden über fünf Epochen trainiert.

Die Gesamtgenauigkeit lag zwischen 83,4 – 85,2 % für alle ViT. Unterschiede in der diagnostischen Leistung zeigten sich für Unterkiefer- (94,1 – 96,7 %) und Oberkieferfrontzähne (86,7-90,2%), sowie für Unterkiefer- (85,6 – 87,2 %) und Oberkieferseitenzähne (78,1 – 81,0 %).

Obwohl die Hypothese mit der erzielten Gesamtgenauigkeit nicht bestätigt wurde, zeigte die beschriebene Studie, dass ViT den parodontalen Knochenabbau auf apikalen Röntgenaufnahmen vergleichbar gut erkennen können. Dies deutet auf die Notwendigkeit weiterer Studien mit größeren und annotierten Datensätzen hin, welche zukünftig einen Vergleich mit erfahrenen Spezialisten ermöglichen würden.

7.   Abstract (English)

The radiographic evaluation of periodontal bone loss (PBL) is often influenced by the dentist's subjectivity. Artificial intelligence could optimize the reliability of radiographic assessment. Neural networks are increasingly used in dental diagnostics, and in recent years Vision Transformers (ViT) have been introduced for computer vision tasks. The aim of the present study was to test five open-source ViT (ViT-base/ViT-large from Google, BEiT-base/BEiT-large from Microsoft, DeiT-base from Facebook/Meta) for the automatized detection of PBL on periapical radiographs. A hypothesis was made that the overall accuracy would be >90% and that the performance differences among the ViT would be minor.

A data set of 21,819 anonymized periapical radiographs (training set N=18,819/ test set N=3,000) was categorized by calibrated dentists using predefined diagnostic criteria, distinguishing between a healthy periodontium and PBL (score 0: healthy, score 1: PBL <15% of the coronal third of the root, score 2: PBL 15 – 33% of the coronal third of the root, score 3: PBL up to the middle third of the root and beyond). The ViT were trained over five epochs.

The overall accuracy (ACC) ranged from 83.4 to 85.2%. Differences in diagnostic performance were seen for mandibular (ACC 94.1 – 96.7%) and maxillary anterior teeth (86.7 – 90.2%), as well as mandibular (85.6 – 87.2%) and maxillary posterior teeth (78.1 – 81.0%).

Although the overall accuracy did not confirm the hypothesis, this study showed that ViT can detect PBL on periapical radiographs comparably well. This indicates the need for further studies with larger and annotated data sets, which would allow a comparison with experienced specialists in the future.

8.   Paper I

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

Dujic, H., Meyer, O., Hoss, P., Wölfle, U. C., Wülk, A., Meusburger,T., Meier, L., Gruhn, V., Hesenius, M., Hickel, R., & Kühnisch, J.

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

Helena Dujic [1,*][iD], 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,*][iD]

[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

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.
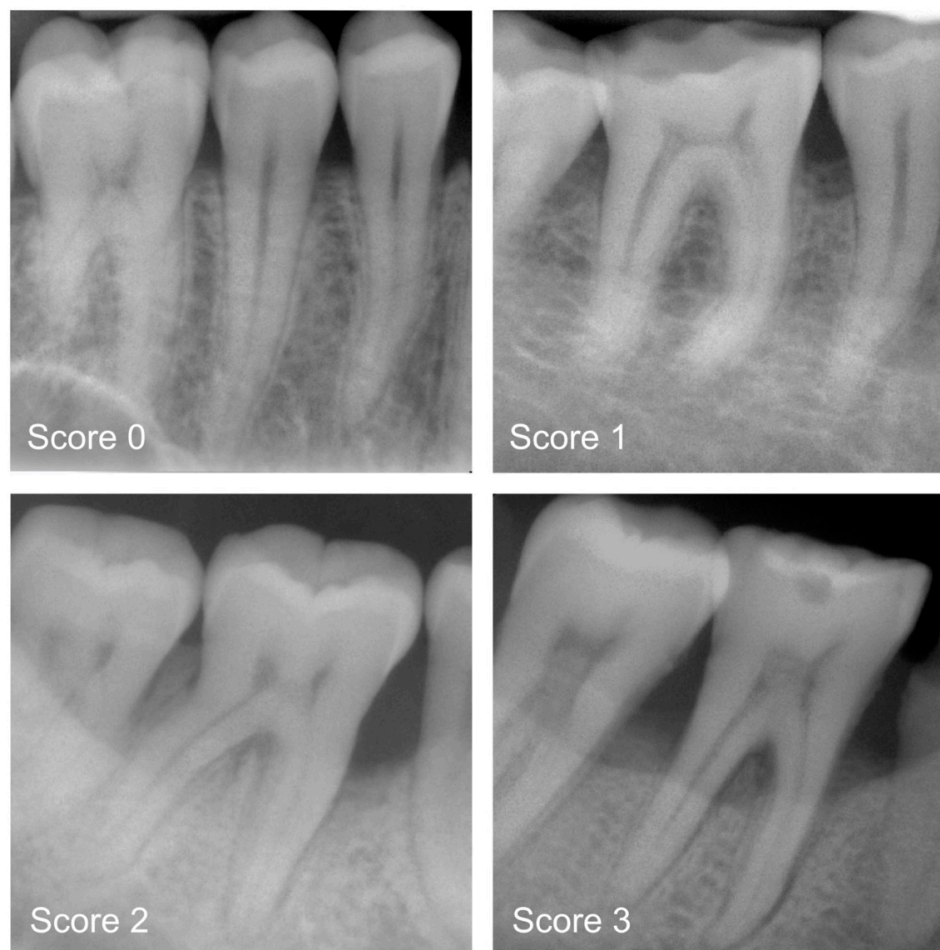
Figure 1. Examples of periapical radiographs for all categories: healthy periodontium (Score 0), mild radiographic periodontal bone loss (PBL) up to 15% of the root length (Score 1), moderate radiographic PBL between 15% and 33% of the root length (Score 2), and severe radiographic PBL extending to the 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 $\times 10^{-8}$).

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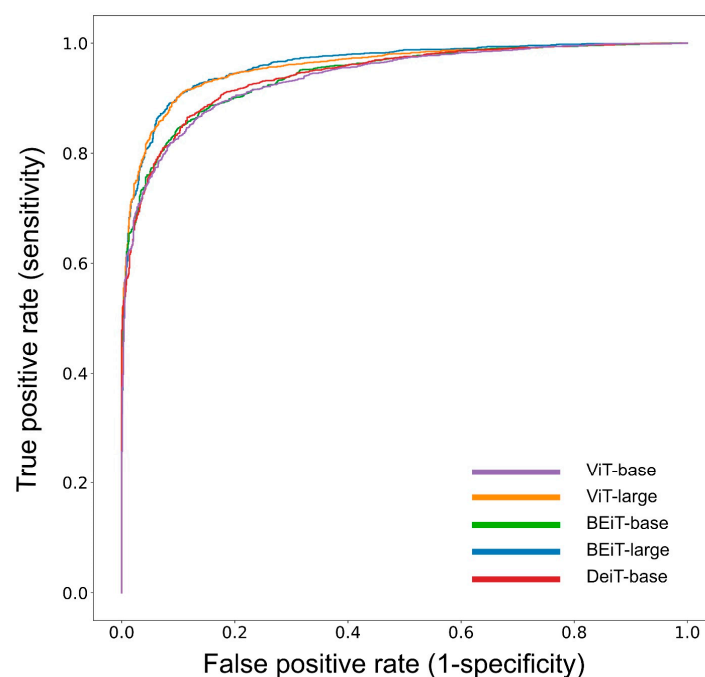
true negatives (TN), true positives (TP), false positives (FP) and false negatives (FN). In addition, the sensitivity (SE), specificity (SP), positive/negative predictive values (PPV/NPV) and area under the receiver operating characteristic (ROC) curve were calculated [39].

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.

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]

9. Paper II

Detection of Periodontal Bone Loss on Periapical Radiographs –

A Diagnostic Study Using Different Convolutional Neural Networks

Hoss, P., Meyer, O., Wölfle, U. C., Wülk, A., Meusburger, T.,
Meier, L., Hickel, R., Gruhn, V., Hesenius, M., Kühnisch, J., &
Dujic, H.

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

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.
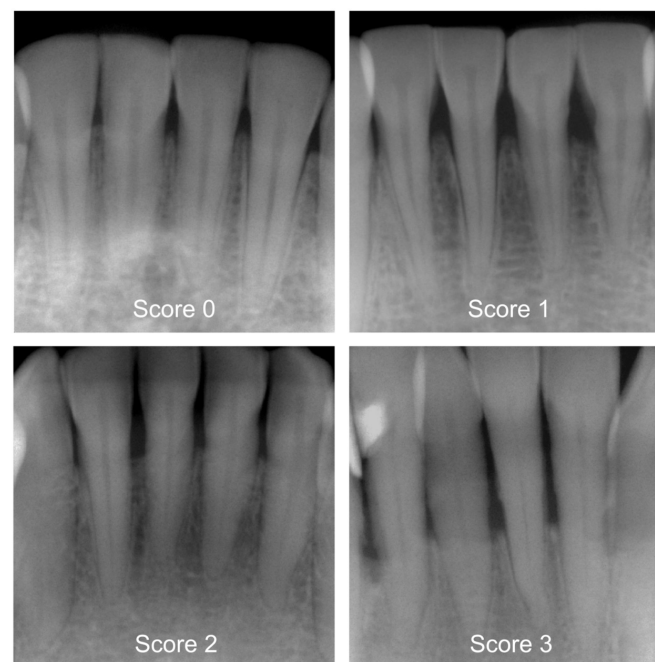
2.3. Categorization of Periodontal Bone Loss (Reference Standard)

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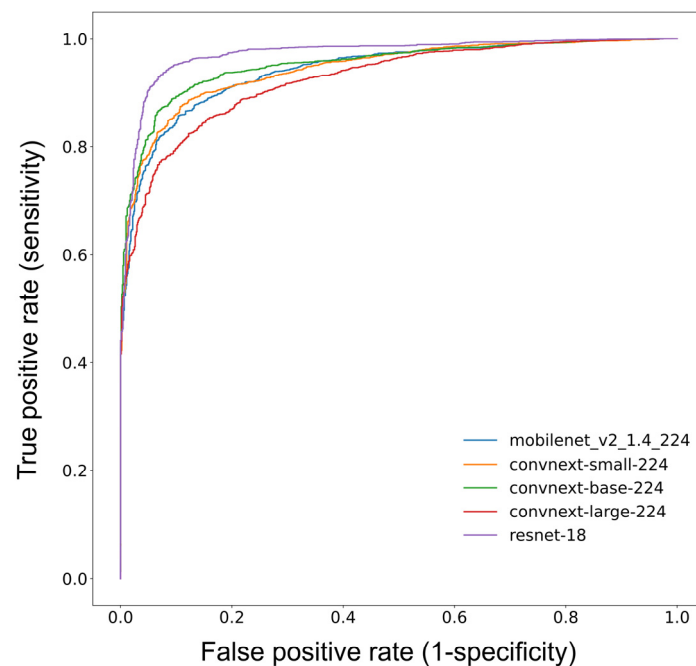
Figure 2. The receiver operating characteristic (ROC) curves graphically visualize the diagnostic performance of the developed convolutional neural networks (CNNs) in detecting PBL.

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.

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. Literaturverzeichnis

1.   Kassebaum NJ, Bernabe E, Dahiya M, Bhandari B, Murray CJ, Marcenes W. Global Burden of Severe Periodontitis in 1990-2010: A Systematic Review and Meta-Regression. J Dent Res. 2014;93(11):1045-53.

2.   Collaborators GBDOD, Bernabe E, Marcenes W, Hernandez CR, Bailey J, Abreu LG, et al. Global, Regional, and National Levels and Trends in Burden of Oral Conditions from 1990 to 2017: A Systematic Analysis for the Global Burden of Disease 2017 Study. J Dent Res. 2020;99(4):362-73.

3.   Jordan AR, Hoffmann T, Kocher T, Micheelis W. Update Zu Parodontalerkrankungen in Deutschland 2014. Zentrale Ergebnisse Der Fünften Deutschen Mundgesundheitsstudie. Parodontologie. 2016;27(4):403-12.

4.   Hajishengallis G, Chavakis T. Local and Systemic Mechanisms Linking Periodontal Disease and Inflammatory Comorbidities. Nat Rev Immunol. 2021;21(7):426-40.

5.   Meyle J, Chapple I. Molecular Aspects of the Pathogenesis of Periodontitis. Periodontol 2000. 2015;69(1):7-17.

6.   Sanz M, Herrera D, Kebschull M, Chapple I, Jepsen S, Beglundh T, et al. Treatment of Stage I-Iii Periodontitis-the Efp S3 Level Clinical Practice Guideline. J Clin Periodontol. 2020;47 Suppl 22(Suppl 22):4-60.

7.   Coelho JMF, Miranda SS, da Cruz SS, Trindade SC, Passos-Soares JS, Cerqueira EMM, et al. Is There Association between Stress and Periodontitis? Clin Oral Investig. 2020;24(7):2285-94.

8.   Nociti FH, Jr., Casati MZ, Duarte PM. Current Perspective of the Impact of Smoking on the Progression and Treatment of Periodontitis. Periodontol 2000. 2015;67(1):187-210.

9.   Sanz M, Ceriello A, Buysschaert M, Chapple I, Demmer RT, Graziani F, et al. Scientific Evidence on the Links between Periodontal Diseases and Diabetes: Consensus Report and Guidelines of the Joint Workshop on Periodontal Diseases and Diabetes by the International Diabetes Federation and the European Federation of Periodontology. J Clin Periodontol. 2018;45(2):138-49.

10. Sanz M, Marco Del Castillo A, Jepsen S, Gonzalez-Juanatey JR, D'Aiuto F, Bouchard P, et al. Periodontitis and Cardiovascular Diseases: Consensus Report. J Clin Periodontol. 2020;47(3):268-88.

11. Papapanou PN, Sanz M, Buduneli N, Dietrich T, Feres M, Fine DH, 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-S82.

12. Tonetti MS, Greenwell H, Kornman KS. Staging and Grading of Periodontitis: Framework and Proposal of a New Classification and Case Definition. J Periodontol. 2018;89 Suppl 1:S159-S72.

13. Deutsche Gesellschaft für Parodontologie e.V. (DG PARO), Deutsche Gesellschaft für Zahn-, Mund- und Kieferheilkunde e.V. (DGZMK). S3-Leitlinie Die Behandlung von Parodontitis Stadium I bis III - die Deutsche Implementierung der S3-Leitlinie „Treatment of Stage I – III Periodontitis" der European Federation of Periodontology (EFP) Version 1.2. 2020.

14. Tonetti MS, Sanz M. Implementation of the New Classification of Periodontal Diseases: Decision-Making Algorithms for Clinical Practice and Education. J Clin Periodontol. 2019;46(4):398-405.

15. Fiorellini JP, Sourvanos D, Sarimento H, Karimbux N, Luan KW. Periodontal and Implant Radiology. Dent Clin North Am. 2021;65(3):447-73.

16. 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(6).

17. Alotaibi G, Awawdeh M, Farook FF, Aljohani M, Aldhafiri RM, 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(1):399.

18. Chang HJ, Lee SJ, Yong TH, Shin NY, Jang BG, Kim JE, et al. Deep Learning Hybrid Method to Automatically Diagnose Periodontal Bone Loss and Stage Periodontitis. Sci Rep. 2020;10(1):7531.

19. Krois J, Ekert T, Meinhold L, Golla T, Kharbot B, Wittemeier A, et al. Deep Learning for the Radiographic Detection of Periodontal Bone Loss. Sci Rep. 2019;9(1):8495.

20. Tsoromokos N, Parinussa S, Claessen F, Moin DA, Loos BG. Estimation of Alveolar Bone Loss in Periodontitis Using Machine Learning. Int Dent J. 2022;72(5):621-7.

21. Lee CT, Kabir T, Nelson J, Sheng S, Meng HW, Van Dyke TE, et al. Use of the Deep Learning Approach to Measure Alveolar Bone Level. J Clin Periodontol. 2022;49(3):260-9.

22. Schwendicke F, Samek W, Krois J. Artificial Intelligence in Dentistry: Chances and Challenges. J Dent Res. 2020;99(7):769-74.

23. Topol EJ. High-Performance Medicine: The Convergence of Human and Artificial Intelligence. Nat Med. 2019;25(1):44-56.

24. Ekert T, Krois J, Meinhold L, Elhennawy K, Emara R, Golla T, Schwendicke F. Deep Learning for the Radiographic Detection of Apical Lesions. J Endod. 2019;45(7):917-22.e5.

25. Lee JH, Kim DH, Jeong SN, Choi SH. Detection and Diagnosis of Dental Caries Using a Deep Learning-Based Convolutional Neural Network Algorithm. J Dent. 2018;77:106-11.

26. Lian L, Zhu T, Zhu F, Zhu H. Deep Learning for Caries Detection and Classification. Diagnostics (Basel). 2021;11(9).

27. Song IS, Shin HK, Kang JH, Kim JE, Huh KH, Yi WJ, et al. Deep Learning-Based Apical Lesion Segmentation from Panoramic Radiographs. Imaging Sci Dent. 2022;52(4):351-7.

28. 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.

29. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale2021. Available from: https://doi.org/10.48550/arXiv.2010.11929 (aufgerufen am 24.01.2024).

30. Chang J, Chang MF, Angelov N, Hsu CY, Meng HW, Sheng S, et al. Application of Deep Machine Learning for the Radiographic Diagnosis of Periodontitis. Clin Oral Investig. 2022;26(11):6629-37.

31. Chen CC, Wu YF, Aung LM, Lin JC, Ngo ST, Su JN, et al. Automatic Recognition of Teeth and Periodontal Bone Loss Measurement in Digital Radiographs Using Deep-Learning Artificial Intelligence. J Dent Sci. 2023;18(3):1301-9.

32. Danks RP, Bano S, Orishko A, Tan HJ, 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(7):1189-99.

33. 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. Journal of Periodontal & Implant Science. 2018;48(2).

34. Ertas K, Pence I, Cesmeli MS, Ay ZY. 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.

35. Kim J, Lee HS, Song IS, Jung KH. Dentnet: Deep Neural Transfer Network for the Detection of Periodontal Bone Loss Using Panoramic Dental Radiographs. Sci Rep. 2019;9(1):17615.

36. 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.

37. 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(4):383-91.

38. Liu Q, Dai F, Zhu H, Yang H, Huang Y, Jiang L, et al. Deep Learning for the Early Identification of Periodontitis: A Retrospective, Multicentre Study. Clinical Radiology. 2023;78(12):E985-E92.

39. 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 (Basel). 2023;13(4).

40. Felsch M, Meyer O, Schlickenrieder A, Engels P, Schonewolf J, Zollner F, et al. Detection and Localization of Caries and Hypomineralization on Dental Photographs with a Vision Transformer Model. NPJ Digit Med. 2023;6(1):198.

41. Deng J, Dong W, Socher R, Li LJ, Kai L, Li F-F, editors. Imagenet: A Large-Scale Hierarchical Image Database. 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009 20-25 June 2009.

42. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. Stard 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. BMJ. 2015;351:h5527.

43. Schwendicke F, Singh T, Lee JH, Gaudin R, Chaurasia A, Wiegand T, et al. Artificial Intelligence in Dental Research: Checklist for Authors, Reviewers, Readers. J Dent. 2021;107:103610.

44. Pasler FA. Zahnärztliche Radiologie. Stuttgart: Georg Thieme Verlag KG; 2021.

45. 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(4):649-61.

46. Kearney VP, Yansane AM, Brandon RG, Vaderhobli R, Lin GH, Hekmatian H, et al. A Generative Adversarial Inpainting Network to Enhance Prediction of Periodontal Clinical Attachment Level. J Dent. 2022;123:104211.

47. Kurt S, Celik O, Bayrakdar IS, Orhan K, Bilgir E, Odabas A, Aslan AF. Success of Artificial Intelligence System in Determining Alveolar Bone Loss from Dental Panoramic Radiography Images. Cumhuriyet Dent J. 2020;23:318-24.

48. 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(1):106.

49. Kabir T, Lee CT, Chen L, Jiang X, Shams S. A Comprehensive Artificial Intelligence Framework for Dental Diagnosis and Charting. BMC Oral Health. 2022;22(1):480.

50. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need2017. Available from: https://doi.org/10.48550/arXiv.1706.03762 (aufgerufen am 24.01.2024).

51. Touvron H, Cord M, Douze M, Massa F, Sablayrolles A, Jégou H. Training Data-Efficient Image Transformers & Distillation through Attention2021. Available from: https://doi.org/10.48550/arXiv.2012.12877 (aufgerufen am 24.1.2024).

52. Bao H, Dong L, Piao S, Wei F. Beit: Bert Pre-Training of Image Transformers.2022. Available from: https://doi.org/10.48550/arXiv.2106.08254 (aufgerufen am 24.01.2024).

53. Scott J, Biancardi AM, Jones O, Andrew D. Artificial Intelligence in Periodontology: A Scoping Review. Dent J (Basel). 2023;11(2).

54. 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(2):169-74.

Danksagung

Ich möchte mich an dieser Stelle bei allen bedanken, die mich während meiner Promotion unterstützt haben und ohne deren Mithilfe diese Dissertation nicht möglich gewesen wäre.

Mein besonderer Dank gilt meinem Doktorvater, Herrn Prof. Dr. Jan Kühnisch, für die Bereitstellung dieses interessanten Themas. Die fachliche Betreuung, sein außerordentliches Engagement und die durchweg positive Zusammenarbeit haben mich immer wieder motiviert, an dem Thema weiterzuarbeiten.

Bedanken möchte ich mich auch bei meinem Co-Autor, Herrn Patrick Hoss und allen anderen Co-Autoren, sowie der gesamten Arbeitsgruppe für die angenehme Arbeitsatmosphäre und den fachlichen Austausch.

Meinem Partner Domagoj möchte ich für die unermüdliche Hilfsbereitschaft danken. Seine geduldige und liebevolle Unterstützung war und ist meine Stütze, die alle Hindernisse überwindbar macht. Danke an meine Freunde, die immer ein offenes Ohr für mich hatten.

Nicht zuletzt gilt mein größter Dank meinen liebevollen Eltern, die mir das Studium und die Promotion ermöglicht und mich immer ermutigt haben. Hvala!