





Article

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

Patrick Hoss ¹, Ole Meyer ², Uta Christine Wölfle ¹, Annika Wülk ¹, Theresa Meusbürger 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.meusbuerger@hotmail.com (T.M.); meier.leon@gmail.com (L.M.); hickel@dent.med.uni-muenchen.de (R.H.); h.dujic@med.uni-muenchen.de (H.D.)

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

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

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

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



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

Academic Editor: Agostino Guida

Received: 3 November 2023

Revised: 14 November 2023

Accepted: 18 November 2023

Published: 20 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

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

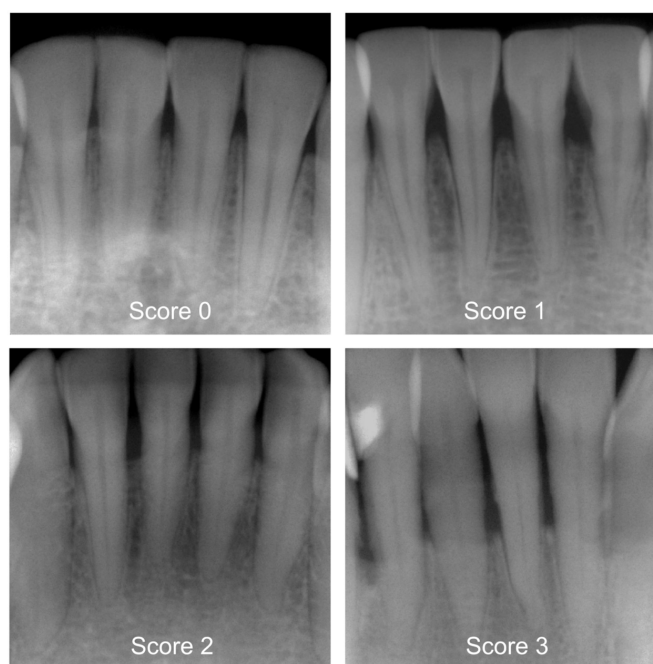


Figure 1. Examples of periapical radiographs for all categories: healthy periodontium, periodontal bone loss (PBL) not detectable (Score 0), mild radiographic PBL up to 15% in the coronal third of the tooth (Score 1), moderate radiographic PBL between 15% and 33% of the root length (Score 2), and severe 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)	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
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)	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Radiographs with maxillary anterior teeth								
ResNet-18	155	58.7	72	27.3	27	10.2	10	3.8
MobileNetV2	154	58.3	79	29.9	20	7.6	11	4.2
ConvNeXT/s ¹	155	58.7	79	29.9	20	7.6	10	3.8
ConvNeXT/b ²	157	59.5	77	29.2	22	8.3	8	3.0
ConvNeXT/l ³	158	59.8	74	28.0	25	9.5	7	2.7
Radiographs with maxillary posterior teeth								
ResNet-18	786	59.1	263	19.8	151	11.4	129	9.7
MobileNetV2	798	60.0	239	18.0	175	13.2	117	8.8
ConvNeXT/s ¹	783	58.9	275	20.7	139	10.5	132	9.9
ConvNeXT/b ²	794	59.7	278	20.9	136	10.2	121	9.1
ConvNeXT/l ³	794	59.8	266	20.0	148	11.1	121	9.1

Table 5. Cont.

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

¹ small, ² base, ³ large.

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

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

¹ small, ² base, ³ large.

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

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

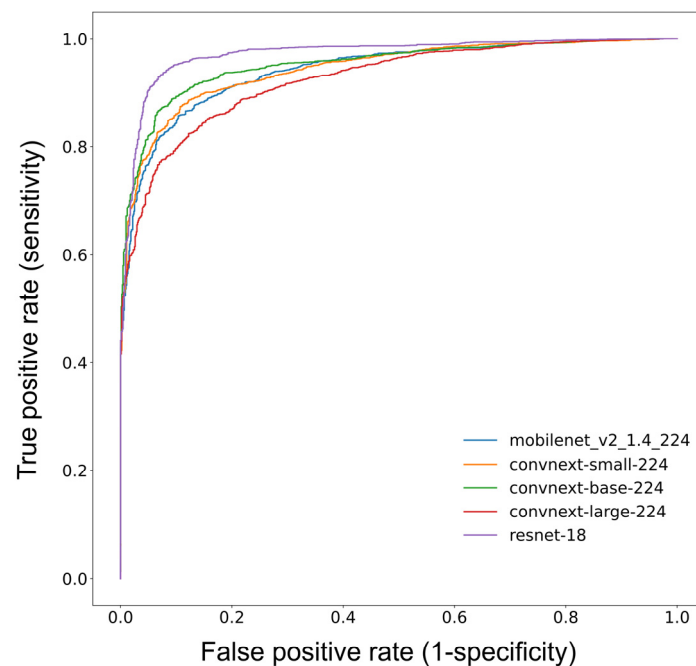


Figure 2. The receiver operating characteristic (ROC) curves graphically visualize the diagnostic performance of the developed convolutional neural networks (CNNs) in detecting PBL.

4. Discussion

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

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

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

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

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

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

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

5. Conclusions

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

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

Funding: This research received no external funding.

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

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

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

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

References

1. Kassebaum, N.J.; Bernabe, E.; Dahiya, M.; Bhandari, B.; Murray, C.J.; Marcenes, W. Global burden of severe periodontitis in 1990–2010: A systematic review and meta-regression. *J. Dent. Res.* **2014**, *93*, 1045–1053. [[CrossRef](#)] [[PubMed](#)]
2. Frencken, J.E.; Sharma, P.; Stenhouse, L.; Green, D.; Laverty, D.; Dietrich, T. Global epidemiology of dental caries and severe periodontitis—A comprehensive review. *J. Clin. Periodontol.* **2017**, *44* (Suppl. S18), S94–S105. [[CrossRef](#)] [[PubMed](#)]
3. Nazir, M.A. Prevalence of periodontal disease, its association with systemic diseases and prevention. *Int. J. Health Sci.* **2017**, *11*, 72–80.
4. Damgaard, C.; Holmstrup, P.; Van Dyke, T.E.; Nielsen, C.H. The complement system and its role in the pathogenesis of periodontitis: Current concepts. *J. Periodontol. 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. Kónönen, 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. Papananou, 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. Meusbarger, 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. Tsromokos, 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-Kipiotti, 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](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.