The Effect of Panoramic Radiography Quality on the Agreement of Diagnosis of Apical Radiolucent Lesions in Maxillary Premolars

Panoramik Radyografi Kalitesinin Maksiller Premolar Dişlerdeki Apikal Radyolüsent Lezyonların Tanısal Doğruluğuna Etkisi

Sümeyye ÇELİK ÖZSOY[1] https://orcid.org/0000-0002-3294-2524
Samed ŞATIR[1] https://orcid.org/0000-0002-5863-5928
Usame Ömer OSMANOĞLU[2] https://orcid.org/0000-0002-1198-2447

[1]Karamanoglu Mehmetbey University, Faculty of Dentistry, Oral and Maxillofacial Radiology
[2]Karamanoglu Mehmetbey University, Faculty of Medicine, Biostatistics

Citation: Özsoy Çelik S, Şatır S, Osmanoğlu U Ö. The Effect of Panoramic Radiography Quality on the Agreement of Diagnosis of Apical Radiolucent Lesions in Maxillary Premolars. *Int Arc Dent Sci*. 2025;46(1):41-50.

ABSTRACT

INTRODUCTION: The aim was to evaluate the diagnostic agreement of apical radiolucent lesions in the maxillary premolars between panoramic radiography (PAN) and periapical radiography (PAR) and to examine the effect of PAN image quality on diagnostic agreement.

METHODS: Ninety patients who had PAN and PAR (XMind DC, Satelec Acteon, France; 70kVp, 8mA, 0.32 s) including all or part of the maxillary premolar teeth were included. The maxillary posterior crown region was masked on the panoramic radiography to avoid bias. 2 observers were asked to score the maxillary premolar teeth on radiographs as no lesion (0), lesion present (1), and no tooth (2). Observers were asked to evaluate PANs and classify them according to the diagnostic quality. Intra-observer and interobserver agreement were statistically evaluated using Cohen's kappa test.

RESULTS: The sensitivity, precision, and F1 score for interobserver agreement regarding the presence of a lesion (score 1) in the first premolar were lower in PAN compared to PAR. Additionally, the sensitivity, precision, and F1 score for lesions in all premolars were lower in the low PAN quality group compared to the high PAN quality group.

CONCLUSION: PAN quality can have an impact on the diagnostic accuracy of apical radiolucent lesions in the maxillary premolars.

Keywords: *Dental granuloma, Dental radiography, Panoramic radiography, Periapical granuloma, Premolar*

ÖZ

GİRİŞ ve AMAÇ: Amaç, maksiller premolarlardaki apikal radyolusent lezyonların panoramik radyografi (PAN) ile periapikal radyografi (PAR) arasındaki tanı uyumunu değerlendirmek ve PAN görüntü kalitesinin tanı uyumuna olan etkisini incelemektir.

YÖNTEM ve GEREÇLER: Çalışmaya, maksiller premolar dişlerinin tamamını veya bir kısmını içeren PAN (73 kVp, 10 mA, 13.5s tarama süresi; PCH 2500, Vatech, South Korea) ve PAR'a (XMind DC, Satelec Acteon, France; 70kVp, 8mA, 0.32 s) sahip 90 hasta dahil edildi. Önyargıdan kaçınmak için maksiller posterior kron bölgesi PAN'da sansürlendi. 2 gözlemciden radyografilerde maksiller premolar dişleri lezyon yok (0), lezyon mevcut (1) ve diş yok (2) olarak puanlamaları istendi. PAN kalitesi dört grupta sınıflandırıldı: ideal görüntü (seviye 1; 81-100), yeterli görüntü (seviye 2; 61-80), zayıf ancak tanısal görüntü (seviye 3; 41-60) ve tanı için çok zayıf ve tekrarlanması önerilen görüntü (seviye 4; 0-40). Gözlemcilerden PAN'ları değerlendirmeleri ve bunları tanı kalitesine göre sınıflandırmaları istendi. Gözlemci içi ve gözlemciler arası uyum, istatistiksel olarak Cohen'in kappa testi ile değerlendirildi.

BULGULAR: PAN'da birinci premolarda lezyonun (skor 1) varlığı için gözlemciler arası uyumun duyarlılığı, kesinliği ve F1 skoru PAR'dan daha düşüktü. Düşük PAN kalitesine sahip gruptaki tüm premolarlarda skor 1 için duyarlılık, kesinlik ve F1 skoru yüksek PAN kalitesine sahip gruptan daha düşüktü.

SONUÇ: PAN kalitesi, maksiller premolarlardaki apikal radyolusent lezyonların tanısal doğruluğunu etkileyebilir.

Anahtar Kelimeler: *Dental granülom, Dental radyografi, Panoramik radyografi, Periapikal granülom, Premolar*

Corresponding author: samedsatir@yahoo.com
Received Date: 20.09.2024
Accepted Date: 13.02.2025

INTRODUCTION

Intraoral radiographs (IO) are considered superior to panoramic radiography (PAN) in diagnosing minor pathologies such as approximal caries and apical lesions.[1-5] In order to increase the proximal caries diagnostic accuracy of PAN, bite-wing radiography feature is added to digital PAN device software or tomosynthesis is utilized.[2,6] The failure of PAN compared to periapical radiography (PAR) in the diagnosis of apical lesions, especially in the maxillary posterior teeth, is associated with the superposition of anatomical structures in the maxilla.[7] On the other hand, it is stated that PAN is more successful than PAR in imaging the apical position of maxillary second and third molars,[8] and there may not be a significant difference between PAN and PAR in the diagnosis of periapical pathologies,[9] depending on the observer. In addition, although it is superior to PAN, it is noted that IO is not sensitive enough to detect associated lesions at an early stage.[6,10]

PAN has the important advantage of showing both jaws, TMJ, maxillary sinus and all teeth in the dental arch with a single image and is therefore routinely used in today's dental practice.[3,5] PAN is used in the detection of many pathologies and anatomical landmarks especially in artificial intelligence (AI)-based diagnostic applications, which have recently become widespread, as it provides a very large area to be displayed for the maxillofacial region.[11-13]

In the last few years, it is seen that the diagnostic accuracy is at a high level in AI-based diagnostic applications with PAN. However, emphasis is placed on standardizing the data used to increase diagnostic accuracy.[11,14] A study showing that the diagnostic accuracy decreases when deep learning models obtained from dental radiological data of different institutions are tested with each other, supports the idea that the databases that form the basis of AI applications should be standardized.[15] Observer interpretation plays a crucial role in standardizing PAN data quality. In addition, it can be said that one of the first and most important steps in PAN standardization is PAN image quality.[2,11,16]

There are scales developed to assess whether PANs are adequate for diagnosis.[16] In this way, it is tried to be a guide in which PANs meet the standard conditions in terms of diagnostic image quality and when imaging should be repeated. The main reasons that reduce the diagnostic quality in PAN are motion artifact, presence of metallic foreign objects, patient positioning errors, certain landmarks not in the correct localization in PAN, and the focal trough not compatible with the dental arch.[16]

Effect of PAN image quality on the diagnostic accuracy in the detection of various pathologies has gained importance especially with the increase in AI applications. This retrospective study aimed to assess the diagnostic agreement of PAN and PAR for detecting apical radiolucent lesions in maxillary premolars and to evaluate how PAN image quality impacts diagnostic accuracy.

MATERIALS AND METHODS

Ethics committee approval was received for this study from Karamanoglu Mehmetbey University Faculty of Medicine Clinical Research Ethics Committee (03-2023/14). Written consent was obtained from the patients whose retrospective data were included in the study, that their data could be used on condition that their identities and personal data remain confidential.

Sample Size

The sample size calculation aimed to ensure adequate power to assess observer agreement. Based on McHugh's classification,[17] a sample size of 70 was required to achieve 80% power (K0 = 0.80, K1 = 0.60, $\alpha = 0.05$). With 90 patients included, the achieved power for the study was 87.49% (two-sided hypothesis).[18] Power analyses were performed using PASS 11 software (NCSS, LLC., USA).

Patient selection

Patients who applied to Karamanoglu Mehmetbey University Faculty of Dentistry between December 2021 and December 2022 over 18 years of age and have taken on the same day both PAN (90 patients; 46 female, 44 male; mean age: 42.2, age range: 18-75) and PAR (189 teeth; 91 first premolars, 98 second premolars; 10 patients 37 teeth bilateral, 80 patients 152 teeth unilateral) were included in the study. Those with orthodontic brackets, implants in the maxillary posterior region, impacted teeth, or those with a specific lesion/condition (such as odontoma, non-odontogenic cyst, sinus lifting graft) were not included in the study as it may cause bias in the observers. Since it is the main subject of the study, no elimination criteria were applied for situations that could negatively affect the PAN image quality.

PAN assessment

PANs (73 kVp, 10 mA, 13.5s scan time; *PCH 2500, Vatech*, South Korea) were evaluated in a standard observation room by observing 30 cm from the WLCD screen (*HP Pavilion 2211x Monitor*, 21.5-inch screen size and 1920 × 1080 resolution). To prevent cognitive bias, the crowns of maxillary posterior teeth were masked using Windows Paint (*Microsoft*, USA) before assessment (Figure 1a-2a-3a). In each PAN, a total of 4 maxillary premolar teeth were scored as no lesion (score 0), radiolucent lesion present (score 1), and no tooth (score 2).

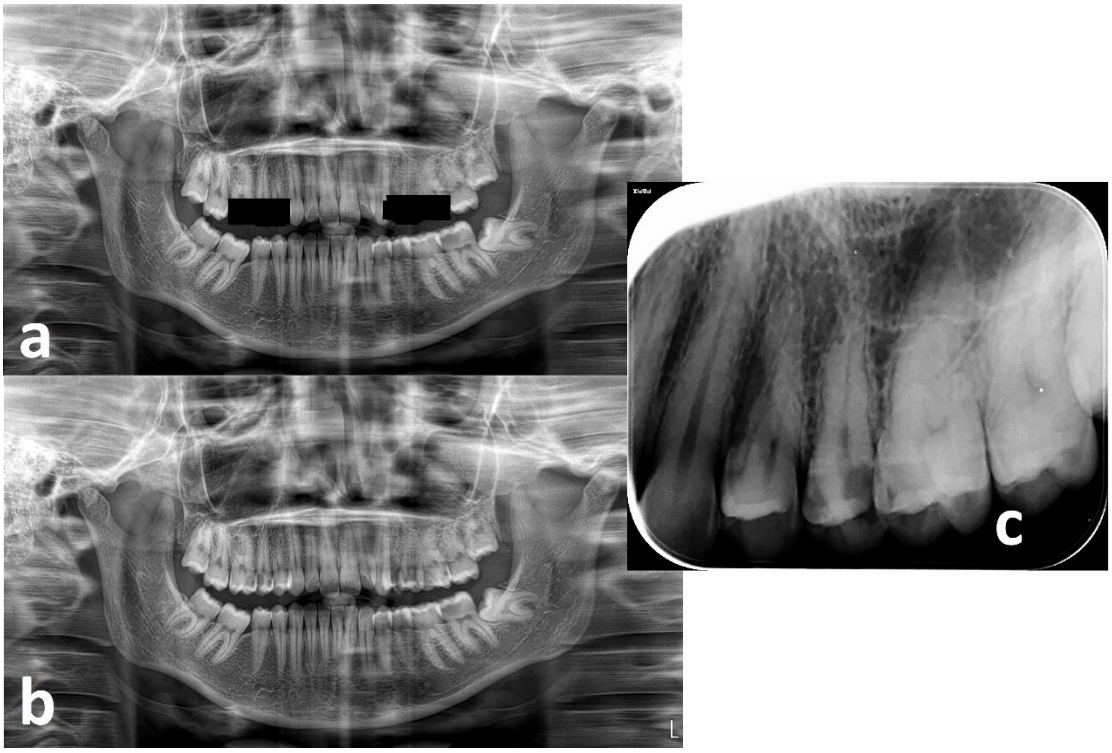

Figure 1. A 22-year-old female patient. Left maxillary first premolar scored 2 in one observation with PAN (a). For the same tooth, 0 was scored in all 4 observations with PAR. Note the radiolucent region on the mesial root surface in the PAR of this tooth (c). Observers included the unmasked version of this PAN in Level 3 in terms of image quality. Note the motion artifact on the left side (b).

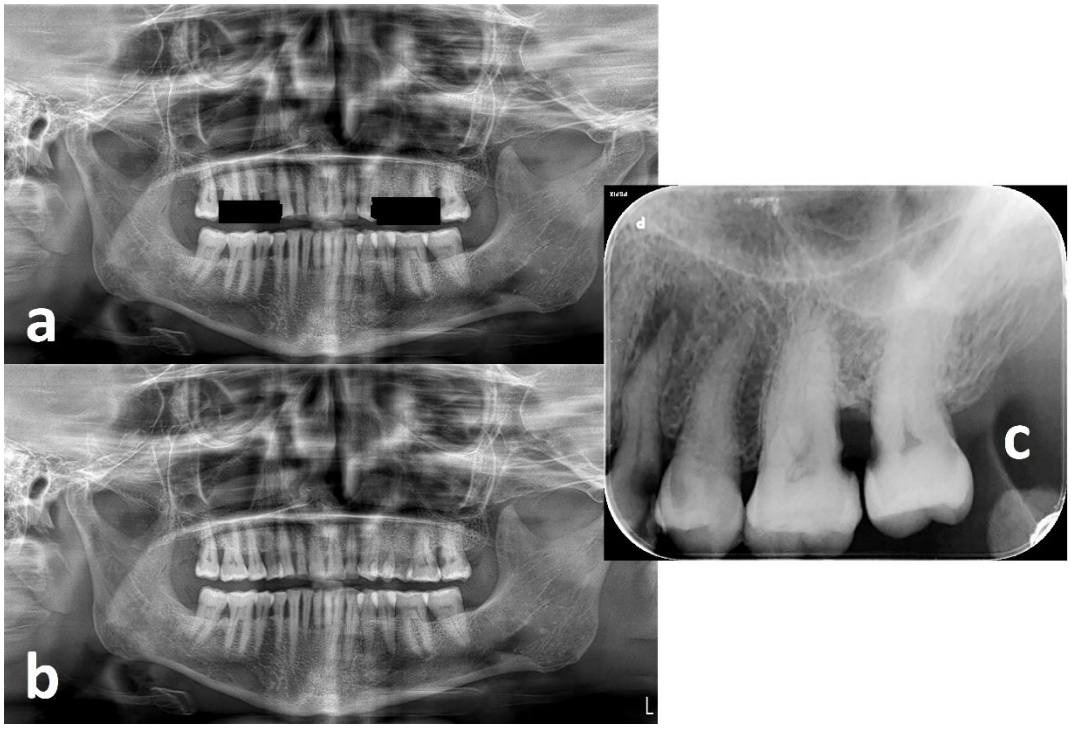

Figure 2. 56-year-old female patient. Left maxillary second premolar scored 2 in two observations with PAN (a). For the same tooth, 0 was scored in all 4 observations with PAR. While the left maxillary first premolar was scored as 0 in 4 observations with PAN, the score was 1 in 4 observations with PAR (c). Observers included the unmasked version of this PAN in Level 3 in terms of image quality. Note that the left ramus appears wider due to the midline positioning error (b).

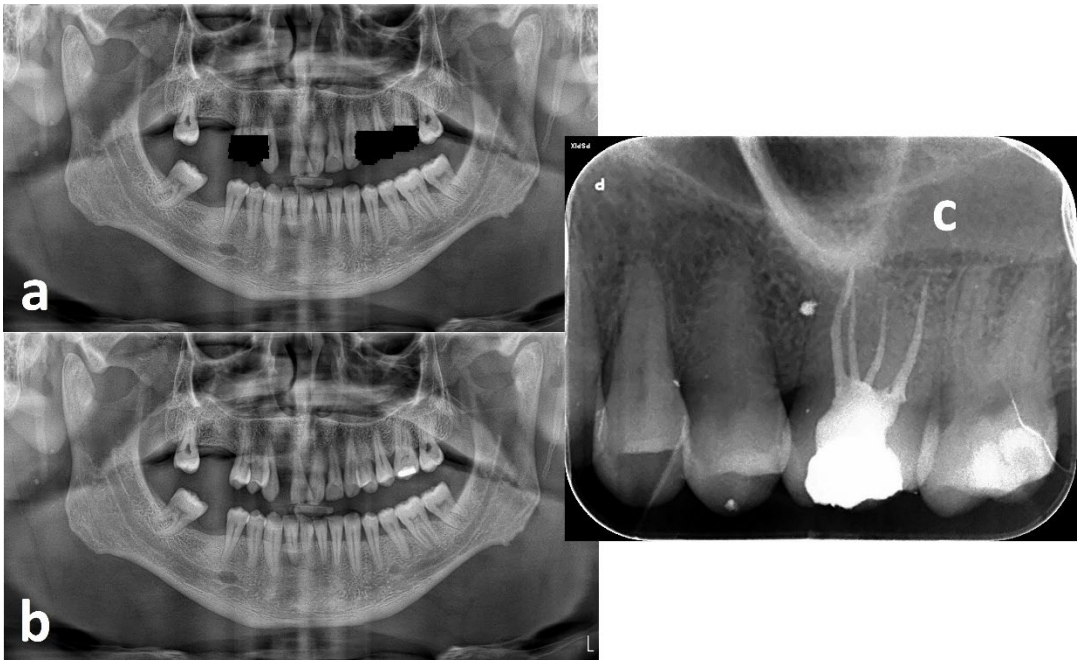

Figure 3. 41-year-old male patient. Left maxillary first premolar scored 2 in one observation with PAN (a). For the same tooth, 0 was scored in all 4 observations with PAR. PAR was taken due to root canal treatment of the left maxillary first molar. Both left premolars were only visible on the final radiography (c). Observers included the unmasked version of this PAN in Level 2 in terms of image quality (b).

PAR assessment

The PAR assessment was performed after the PAN assessment. PARs were taken by radiology technician (8 years' experience in oral radiology) with a wall-mounted periapical device (*XMind DC, Satelec Acteon*, France; 70kVp, 8mA, 0.32 s) and a phosphor plate (*Dürr Dental*, Germany) and visualized with a phosphor plate scanner (*VistaScan, Dürr Dental*, Germany). The same screen, the same distance and the same standard observation room were selected as in the PAN assessment. Each PAR was scored as no lesion (score 0), radiolucent lesion present (score 1), and no tooth (score 2), while premolars that did not imaged into the PAR region were not included in the study data. No classification for apical radiolucent lesion size was used. Observers were informed that the apical enlargement of the periodontal ligament should be evaluated as a score of 0.

PAN quality classification

The study of Choi et al. in 2012 was based on the PAN quality classification.[16] According to this classification, each PAN is scored between 0 and 100 (0 worst, 100 perfect). Of the total score, 8 are identification/information (gender, age, right/left marker, date), 6 are artifact/shadow (jewelry, dental prosthesis, unidentified foreign body), 8 are coverage (location and visibility of condyles, orbital inferior, and mandible inferior), 30 are patient position (midline location, occlusal plane location, artifact from patient movement, superposition of hyoid bone with mandible, anteroposterior location associated with bite plane), 38 are image properties (resolution, brightness, density and contrast) and 6 are overall image quality scoring (ideal, acceptable, poor but acceptable, very poor). According to the total score obtained, PAN quality is classified into four groups: ideal image (level 1; 81-100), adequate image (level 2; 61-80), poor but diagnostic image (level 3; 41-60), and very weak to diagnosis and recommended to be repeated image (level 4; 0-40).

Observers & the observation process

Two observers (SCO, 7 years' experience in oral radiology; SS, 6 years' experience in oral radiology) were designated for the PAN and PAR assessment and PAN quality classification. Both PAN and PAR assessments were made blindly and in an irregular order with observers independent of each other. Observers made their scoring twice, repeating all of them one month interval. One month after all scoring was completed, two observers evaluated the unmasked PANs together and determined the PAN image quality by consensus.

Statistical analysis

Statistical analyzes were evaluated using the IBM Statistical Package for Social Sciences 25.0 (*SPSS*, Chicago, IL) program. Descriptive statistics for continuous variables are presented as means (±standard deviation) or medians (Q1-Q3). Normality was assessed with the Shapiro-Wilk test, and parametric or non-parametric tests were applied accordingly. Interobserver

agreement was determined by Cohen's Kappa Test. McHugh's classification was used to determine the Cohen's Kappa Test compliance levels.[17] In addition, multiclass confusion matrix values (sensitivity, specificity, precision, F1 score, etc.) were used to detail the interobserver agreement. Relationships and comparisons between categorical variables were calculated by chi-square analysis. Due to the minimum expected frequency percentage in PAN quality grouping, appropriate group aggregation (level 1+2 and level 3+4) was made. Correlation heatmap plots were created using the Python 3.7.9 (Delaware, USA) software program. Statistical significance level was accepted as $p<0.05$.

RESULTS

Interobserver agreement of the first and second premolars separately is shown in Table 2. In the first premolars for PAN, the score 1 kappa coefficient was 0.59-0.60 for sensitivity, precision and F1 score, with poor-moderate agreement, 0.95-0.96 for specificity, with almost perfect agreement. For PAN, the score 1 kappa coefficient in second premolars was 0.64-0.69, 0.67-0.73 and 0.65-0.71 for sensitivity, precision and F1 score, respectively, with moderate agreement. All score kappa coefficients for the first premolars for PAR were 0.97-1.00, 0.73-1.00, 0.79-1.00 and 0.84-1.00 for sensitivity, precision, specificity and F1 score, respectively, with a strong-almost perfect agreement. In the second premolars for PAR, all score kappa coefficients for sensitivity, precision, specificity and F1 score were between 0.81-1.00, 0.60-1.00, 0.75-1.00 and 0.72-1.00, respectively, with strong agreement.

Intra-observer agreement for PAN and PAR is shown in Table 1. Kappa values between 0.84-1.00 and 0.76-1.00 were obtained for sensitivity and specificity, respectively. Overall accuracy was found to be between 93-95%. A strong level of agreement was found with the Cohen's Kappa coefficient between 0.83-0.86.

Interobserver agreement of all premolars is shown in Table 3. Overall accuracy for PAN is between 90-92% and for PAR it is 94-95%. Kappa coefficients were 0.76-0.80 (strong agreement) and 0.83-0.86 (strong agreement) for PAN and PAR, respectively.

Table 4 presents diagnostic accuracy comparisons based on PAN image quality. For cases with poor image quality, score 1 yielded a Kappa coefficient of 0.55 (sensitivity), 0.44 (precision), and 0.49 (F1 score), indicating poor agreement, with specificity at 0.97 (almost perfect agreement).

Table 1. Interobserver agreement (first and second premolars separately) for PAN and PAR

First Premolar	PAN						PAR					
	Time 1			Time 2			Time 1			Time 2		
	0	1	2	0	1	2	0	1	2	0	1	2
TP	129	12	20	133	10	21	76	10	3	77	8	3
TN	32	152	157	31	156	157	13	79	88	11	80	88
FP	8	8	3	9	7	0	2	0	0	3	0	0
FN	11	8	0	7	7	2	0	2	0	0	3	0
Sensitivity	0,94	0,60	0,87	0,94	0,59	1,00	0,97	1,00	1,00	0,96	1,00	1,00
Precision	0,92	0,60	1,0	0,95	0,59	0,91	1,00	0,83	1,00	1,00	0,73	1,00
Specificity	0,80	0,95	0,98	0,78	0,96	1,00	0,87	1,00	1,00	0,79	1,00	1,00
F1 Score	0,93	0,60	0,93	0,94	0,59	0,95	0,99	0,91	1,00	0,98	0,84	1,00
Overall Accuracy	89,44%			91,11%			97,80%			96,70%		
Cohen's Kappa coefficient	0,723***			0,755***			0,918***			0,865***		
Second Premolar	**PAN**						**PAR**					
	Time 1			**Time 2**			**Time 1**			**Time 2**		
	0	**1**	**2**	**0**	**1**	**2**	**0**	**1**	**2**	**0**	**1**	**2**
TP	124	14	26	132	11	26	69	13	9	73	9	9
TN	40	151	153	37	160	152	22	78	89	18	82	89
FP	8	8	0	4	5	2	4	3	0	6	1	0
FN	8	7	1	7	4	0	3	4	0	1	6	0
Sensitivity	0,94	0,64	1,00	0,97	0,69	0,93	0,95	0,81	1,00	0,92	0,90	1,00
Precision	0,94	0,67	0,96	0,95	0,73	1,00	0,96	0,76	1,00	0,99	0,60	1,00
Specificity	0,83	0,94	1,00	0,90	0,97	0,99	0,85	0,96	1,00	0,75	0,99	1,00
F1 Score	0,94	0,65	0,98	0,96	0,71	0,96	0,95	0,79	1,00	0,95	0,72	1,00
Overall Accuracy	91,11%			93,89%			92,86%			92,86%		
Cohen's Kappa coefficient	0,791***			0,842***			0,828***			0,806***		

*TP: True Positive; TN: True Negative; FP: False Positive; FN: False Negative; Overall Accuracy; PAN: Panoramic Radiography; PAR: Periapical Radiography; Score 0: No lesion; Score 1: Radiolusent lesion; Score 2: No tooth; *** p<0,001 (Cohen's Kappa coefficient)*

Table 2. Intra-observer agreement (inter-time) for PAN and PAR

All Teeth	PAN						PAR					
	Observer SS			Observer SCO			Observer SS			Observer SCO		
	0	1	2	0	1	2	0	1	2	0	1	2
TP	264	27	46	263	30	46	145	23	12	150	17	12
TN	73	314	310	76	315	308	35	157	177	29	162	177
FP	15	5	3	15	3	3	6	3	0	9	1	0
FN	8	14	1	6	12	3	3	6	0	1	9	0
Sensitivity	0,95	0,84	0,94	0,95	0,91	0,94	0,96	0,88	1,00	0,94	0,94	1,00
Precision	0,97	0,66	0,98	0,98	0,71	0,94	0,98	0,79	1,00	0,99	0,65	1,00
Specificity	0,83	0,98	0,99	0,84	0,99	0,99	0,85	0,98	1,00	0,76	0,99	1,00
F1 Score	0,96	0,74	0,96	0,96	0,80	0,94	0,97	0,84	1,00	0,97	0,77	1,00
Overall Accuracy	93,61%			94,17%			95,24%			94,71%		
Cohen's Kappa coefficient	0,835***			0,852***			0,864***			0,830***		

*TP: True Positive; TN: True Negative; FP: False Positive; FN: False Negative; Overall Accuracy; PAN: Panoramic Radiography; PAR: Periapical Radiography; Score 0: No lesion; Score 1: Radiolusent lesion; Score 2: No tooth; *** p<0,001 (Cohen's Kappa coefficient)*

Table 3. Interobserver agreement (all teeth together) for PAN and PAR

All Teeth	PAN						PAR					
	Time 1			Time 2			Time 1			Time 2		
	0	1	2	0	1	2	0	1	2	0	1	2
TP	253	26	46	265	21	47	145	23	12	150	17	12
TN	72	303	310	68	316	309	35	157	177	29	162	177
FP	16	16	3	13	12	2	6	3	0	9	1	0
FN	19	15	1	14	11	2	3	6	0	1	9	0
Sensitivity	0,94	0,62	0,94	0,95	0,64	0,96	0,96	0,88	1,00	0,94	0,94	1,00
Precision	0,93	0,63	0,98	0,95	0,66	0,96	0,98	0,79	1,00	0,99	0,65	1,00
Specificity	0,82	0,95	0,99	0,84	0,96	0,99	0,85	0,98	1,00	0,76	0,99	1,00
F1 Score	0,94	0,63	0,96	0,95	0,65	0,96	0,97	0,84	1,00	0,97	0,77	1,00
Overall Accuracy	90,28%			92,5%			95,24%			94,71%		
Cohen's Kappa coefficient	0,760***			0,800***			0,864***			0,830***		

*TP: True Positive; TN: True Negative; FP: False Positive; FN: False Negative; Overall Accuracy; PAN: Panoramic Radiography; PAR: Periapical Radiography; Score 0: No lesion; Score 1: Radiolusent lesion; Score 2: No tooth; *** p<0,001 (Cohen's Kappa coefficient)*

DISCUSSION

The low interobserver agreement and low sensitivity for PAN in the presence of lesions (score 1) indicates that PAN is less successful in diagnosing apical radiolucent lesions in the maxilla premolars than PAR (Table 1 and 3). In the present study, it was found that sensitivity was lower than specificity in the diagnosis of apical lesions with both PAN and PAR, which is consistent with the literature.[7,8,10,19,20] In the diagnosis of radiolucent lesion with PAR, the sensitivity for the second premolar was lower than for the first premolar (Table 1). This may be because the second premolars are more likely to be in superposition with anatomical landmarks and in close proximity to the maxillary sinus.[7,8] In a diagnostic study with PAN with eye tracking, it was stated that lesions localized in the maxillary sinus were more frequently missed.[21] It is also stated that the diagnosis with PAR may be restricted when the lesion size is small.[10] In this retrospective study, a classification related to lesion size was not used for PAR and the lesion sizes of the first and second premolars were not standardized. Therefore, in the scenario where the lesions on the first premolars were larger in size, they might have been detected by the observers with higher agreement. Although it was not statistically significant in the diagnosis of radiolucent lesion with PAN, the sensitivity of the first premolars was lower than the second premolars. This may be since the first premolars are more closely related to the fossa canina than the second premolars and are more likely to be out of the focal trough.

It is seen that in most of the studies in which PAN is used for AI applications, no information is given about the image quality of the PAN included in the study for modeling,[7,22,23] or that insufficient and non-standardized criteria (such as good visualization, clear image) are used.[5,13,24,25] This situation suggests that although the necessity of PAN standardization for deep learning models has been emphasized in some studies,[26,27] the issue of PAN image quality, which should be standardized first, has been ignored. As a first step to achieve PAN standardization, it can be planned to classify PAN image quality and to include PANs that can meet a certain quality level in AI models. The results of in the present study showed that the statistical diagnostic accuracy of maxillary premolar apical radiolucent lesions in PANs with high image quality is slightly higher than in PANs with low image quality. In addition, it was determined that sensitivity, precision and F1 scores were

lower in PANs with low image quality compared to PANs with high image quality, especially in score 1 (Table 4). There was not statistically and clinically significant difference in diagnostic agreement between groups separated by PAN image quality.

Table 4. Interobserver agreement in the effect of PAR image quality on diagnostic accuracy

Time 1	**Panoramic Radiography Image Quality**					
	High			**Low**		
	0	**1**	**2**	**0**	**1**	**2**
TP	159	20	18	94	6	28
TN	38	180	195	34	123	115
FP	7	9	3	9	7	0
FN	12	7	0	7	8	1
Sensitivity	0,96	0,69	0,86	0,91	0,46	1,00
Precision	0,93	0,74	1,00	0,93	0,43	0,97
Specificity	0,84	0,95	0,98	0,79	0,95	1,00
F1 Score	0,94	0,71	0,92	0,92	0,44	0,98
Overall Accuracy		91,20%			88,89%	
Cohen's Kappa coefficient		0,760***			0,753***	
Time 2	**Panoramic Radiography Image Quality**					
	High			**Low**		
	0	**1**	**2**	**0**	**1**	**2**
TP	165	16	19	100	5	28
TN	35	185	196	33	131	113
FP	6	10	0	7	2	2
FN	10	5	1	4	6	1
Sensitivity	0,96	0,62	1,00	0,93	0,71	0,93
Precision	0,94	0,76	0,95	0,96	0,45	0,97
Specificity	0,85	0,95	1,00	0,83	0,98	0,98
F1 Score	0,95	0,68	0,97	0,95	0,56	0,95
Overall Accuracy		92,59%			92,36%	
Cohen's Kappa coefficient		0,781***			0,817***	
All Teeth	**Panoramic Radiography Image Quality**					
	High			**Low**		
	0	**1**	**2**	**0**	**1**	**2**
TP	324	36	37	194	11	56
TN	73	365	391	67	254	228
FP	13	19	3	16	9	2
FN	22	12	1	11	14	2
Sensitivity	0,96	0,65	0,93	0,92	0,55	0,97
Precision	0,94	0,75	0,97	0,95	0,44	0,97
Specificity	0,85	0,95	0,99	0,81	0,97	0,99
F1 Score	0,95	0,70	0,95	0,93	0,49	0,97
Overall Accuracy		91,90%			90,63%	
Cohen's Kappa coefficient		0,770***			0,784***	

*TP: True Positive; TN: True Negative; FP: False Positive; FN: False Negative; Overall Accuracy; Score 0: No lesion; Score 1: Radiolusent lesion; Score 2: No tooth; *** p<0,001 (Cohen's Kappa coefficient)*

In studies comparing the diagnostic accuracy of radiographic techniques, the number and competence (such as being an oral radiologist and experience) of observers is an important indicator for the reliability of the findings. The use of one,[6,28] two,[7,29] three[2,26] or more[20,30] observers to provide a criterion close to the gold standard in diagnostic accuracy studies like present study shows that there is no standard in the number of observers and that the researchers determine the number of observers in accordance with their own conditions. The results of present study are compatible with the literature. Therefore, it can be said that the fact that only two observers were used in present study did not adversely affect the reliability of the study data. Despite her/his expertise in oral radiology and many years of experience, it should not be denied that there are many factors that influence an observer's diagnostic decisions. Therefore, it can be predicted that in the future, new criteria aiming to optimize visual perception (such as eye rest time and mood markers) may be added to the observer competences (such as number, expertise, and experience) determined in studies planned to increase the diagnostic accuracy. Therefore, in addition to numerical criteria such as years of experience and number of observers,

qualitative criteria such as instantaneous performance measurements also affect the results, and qualitative performance criteria need to be standardized.

The most important limitation of present study is that histopathology was not used as the gold standard. This limitation can be excused due to the nature of the cases where the biological objects used in retrospective studies are not converted into ex-vivo material.[14,31] Another limitation of the inability to use histopathology as the gold standard was that only premolar teeth were included in the study. Molar teeth may be in close proximity to radiopaque anatomical landmarks such as the zygomatic arch and the floor of the maxillary sinus. Only premolars were included in this study since the superposition of their roots in the buccolingual direction is more confusing and the study[8] in the literature claiming that PAN may be superior to PAR in the diagnosis accuracy of apical radiolucent lesions in maxillary second and third molars. Since it is a retrospective study, the fact that the observers did not have information about the clinical history of the patients whose radiographs were used can be shown as a limitation. However, this turned into an advantage in present study as it prevented the bias of the observers. Because the aim of this study was to observe how observers make decisions when comparing different radiography techniques by simply making observations. It can be said that observer agreement is used as the gold standard, as in similar studies in the literature.[2,20,26] The compatibility of our findings with the literature in line with the expectations suggests that the subjective decisions of the observers can be trusted. Nevertheless, as the gold standard, histopathology is a more realistic and appropriate choice than observers regardless of number.

The classification model used for image quality in present study presents four different subgroups for PAN.[16] Since the data included in the study were concentrated under two subgroups (scores 2 and 3) according to the PAN quality classification, they had to be grouped under two groups as high and low classification. Further studies can be planned to include more PANs to reveal how diagnostic accuracy is affected for each sub-heading of the present classification. It can be predicted that the findings obtained from the two main subgroups formed in present study on diagnostic accuracy will contribute to the literature and similar studies to be planned in the future. Finally, the effect of only maxillary premolar apical radiolucent lesions on diagnostic accuracy was evaluated using the PAN image quality classification. In a study on the effect of patient positioning on the visualization of landmarks in cone-beam computed tomography (CBCT), it was mentioned that positioning by focusing on the cause of CBCT may adversely affect the imaging of other anatomical and pathological conditions.[32] The PAN image quality classification we used in present study[16] is not standardized for any lesion or anatomical landmark and aims to optimize the image quality of the entire region that can be imaged with PAN. It is possible that studies that can be created by different PAN quality classification methods and testing different pathological/anatomical structures will contribute to both the literature.

CONCLUSION

The statistical diagnostic accuracy of maxillary premolar apical radiolucent lesions in PANs with high image quality is slightly higher than in PANs with low image quality. It should not be overlooked that it would be a good idea to evaluate with PAR to detect possible residual roots in the maxillary premolar region, which is thought to be edentulous after imaging with PAN. These findings suggest the importance of considering both imaging techniques for comprehensive evaluation and emphasize the need for standardizing image quality to optimize diagnostic outcomes in clinical practice.

REFERENCES

1. Akarslan ZZ, Akdevelioğlu M, Güngör K, Erten H. A comparison of the diagnostic accuracy of bitewing, periapical, unfiltered and filtered digital panoramic images for approximal caries detection in posterior teeth. *Dentomaxillofac Radiol.* 2008;37(8):458-463. doi:10.1259/dmfr/84698143
2. Jeon KJ, Han SS, Lee C, Choi YJ, Jung HI, Kim YH. Application of panoramic radiography with a multilayer imaging program for detecting proximal caries: a preliminary clinical study. *Dentomaxillofac Radiol.* 2020;49(8):20190467. doi:10.1259/dmfr.20190467
3. Arslan ZB, Demir H, Berker Yıldız D, Yaşar F. Diagnostic accuracy of panoramic radiography and ultrasonography in detecting periapical lesions using periapical radiography as a gold standard. *Dentomaxillofac Radiol.* 2020;49(6):20190290. doi:10.1259/dmfr.20190290
4. Wenzel A. Radiographic modalities for diagnosis of caries in a historical perspective: from film to machine-intelligence supported systems. *Dentomaxillofac Radiol.* 2021;50(5):20210010. doi:10.1259/dmfr.20210010
5. Kılıc MC, Bayrakdar IS, Çelik Ö, et al. Artificial intelligence system for automatic deciduous tooth detection and numbering in panoramic radiographs. *Dentomaxillofac Radiol.* 2021;50(6):20200172. doi:10.1259/dmfr.20200172

6. Mauriello SM, Broome AM, Platin E, et al. The role of stationary intraoral tomosynthesis in reducing proximal overlap in bitewing radiography. *Dentomaxillofac Radiol.* 2020;49(8):20190504. doi:10.1259/dmfr.20190504
7. Kwon O, Yong TH, Kang SR, et al. Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network. *Dentomaxillofac Radiol.* 2020;49(8):20200185. doi:10.1259/dmfr.20200185
8. Ridao-Sacie C, Segura-Egea JJ, Fernández-Palacín A, Bullón-Fernández P, Ríos-Santos JV. Radiological assessment of periapical status using the periapical index: comparison of periapical radiography and digital panoramic radiography. *Int Endod J.* 2007;40(6):433-440. doi:10.1111/j.1365-2591.2007.01233.x
9. Rohlin M, Kullendorff B, Ahlqwist M, Stenström B. Observer performance in the assessment of periapical pathology: a comparison of panoramic with periapical radiography. *Dentomaxillofac Radiol.* 1991;20(3):127-131. doi:10.1259/dmfr.20.3.1807995
10. Brasil DM, Yamasaki MC, Santaella GM, Guido MCZ, Freitas DQ, Haiter-Neto F. Influence of VistaScan image enhancement filters on diagnosis of simulated periapical lesions on intraoral radiographs. *Dentomaxillofac Radiol.* 2019;48(2):20180146. doi:10.1259/dmfr.20180146
11. Hung K, Montalvao C, Tanaka R, Kawai T, Bornstein MM. The use and performance of artificial intelligence applications in dental and maxillofacial radiology: A systematic review. *Dentomaxillofac Radiol.* 2020;49(1):20190107. doi:10.1259/dmfr.20190107
12. Heo MS, Kim JE, Hwang JJ, et al. Artificial intelligence in oral and maxillofacial radiology: what is currently possible?. *Dentomaxillofac Radiol.* 2021;50(3):20200375. doi:10.1259/dmfr.20200375
13. Song IS, Shin HK, Kang JH, et al. Deep learning-based apical lesion segmentation from panoramic radiographs. *Imaging Sci Dent.* 2022;52(4):351-357. doi:10.5624/isd.20220078
14. Silva VKS, Vieira WA, Bernardino ÍM, et al. Accuracy of computer-assisted image analysis in the diagnosis of maxillofacial radiolucent lesions: A systematic review and meta-analysis. *Dentomaxillofac Radiol.* 2020;49(3):20190204. doi:10.1259/dmfr.20190204
15. Nishiyama M, Ishibashi K, Ariji Y, et al. Performance of deep learning models constructed using panoramic radiographs from two hospitals to diagnose fractures of the mandibular condyle. *Dentomaxillofac Radiol.* 2021;50(7):20200611. doi:10.1259/dmfr.20200611
16. Choi BR, Choi DH, Huh KH, et al. Clinical image quality evaluation for panoramic radiography in Korean dental clinics. *Imaging Sci Dent.* 2012;42(3):183-190. doi:10.5624/isd.2012.42.3.183
17. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb).* 2012;22(3):276-282.
18. Flack VF, Afifi AA, Lachenbruch PA, Schouten HJA. Sample Size Determinations for the Two Rater Kappa Statistic. *Psychometrika.* 1988;53(3):321-325. doi:10.1007/BF02294215
19. Heck K, Litzenburger F, Ullmann V, Hoffmann L, Kunzelmann KH. In vitro comparison of two types of digital X-ray sensors for proximal caries detection validated by micro-computed tomography. *Dentomaxillofac Radiol.* 2021;50(3):20200338. doi:10.1259/dmfr.20200338
20. Takahashi N, Lee C, Da Silva JD, et al. A comparison of diagnosis of early stage interproximal caries with bitewing radiographs and periapical images using consensus reference. *Dentomaxillofac Radiol.* 2019;48(2):20170450. doi:10.1259/dmfr.20170450
21. Vogel D, Schulze R. Viewing patterns regarding panoramic radiographs with different pathological lesions: an eye-tracking study. *Dentomaxillofac Radiol.* 2021;50(8):20210019. doi:10.1259/dmfr.20210019
22. Hiraiwa T, Ariji Y, Fukuda M, et al. A deep-learning artificial intelligence system for assessment of root morphology of the mandibular first molar on panoramic radiography. *Dentomaxillofac Radiol.* 2019;48(3):20180218. doi:10.1259/dmfr.20180218
23. Jeon SJ, Yun JP, Yeom HG, et al. Deep-learning for predicting C-shaped canals in mandibular second molars on panoramic radiographs. *Dentomaxillofac Radiol.* 2021;50(5):20200513. doi:10.1259/dmfr.20200513
24. Amasya H, Yildirim D, Aydogan T, Kemaloglu N, Orhan K. Cervical vertebral maturation assessment on lateral cephalometric radiographs using artificial intelligence: comparison of machine learning classifier models. *Dentomaxillofac Radiol.* 2020;49(5):20190441. doi:10.1259/dmfr.20190441
25. Lee JS, Adhikari S, Liu L, Jeong HG, Kim H, Yoon SJ. Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: a preliminary study. *Dentomaxillofac Radiol.* 2019;48(1):20170344. doi:10.1259/dmfr.20170344
26. Nakamoto T, Hatsuta S, Yagi S, Verdonschot RG, Taguchi A, Kakimoto N. Computer-aided diagnosis system for osteoporosis based on quantitative evaluation of mandibular lower border porosity using panoramic radiographs. *Dentomaxillofac Radiol.* 2020;49(4):20190481. doi:10.1259/dmfr.20190481
27. Kuwana R, Ariji Y, Fukuda M, et al. Performance of deep learning object detection technology in the detection and diagnosis of maxillary sinus lesions on

panoramic radiographs. *Dentomaxillofac Radiol.* 2021;50(1):20200171. doi:10.1259/dmfr.20200171

28. Svystun O, Wenzel A, Schropp L, Spin-Neto R. Image-stitching artefacts and distortion in CCD-based cephalograms and their association with sensor type and head movement: ex vivo study. *Dentomaxillofac Radiol.* 2020;49(3):20190315. doi:10.1259/dmfr.20190315
29. Sato H, Da Silva JD, Lee C, et al. Effects of healthcare policy and education on reading accuracy of bitewing radiographs for interproximal caries. *Dentomaxillofac Radiol.* 2021;50(2):20200153. doi:10.1259/dmfr.20200153
30. Kuramoto T, Takarabe S, Okamura K, et al. Effect of differences in pixel size on image characteristics of digital intraoral radiographic systems: a physical and visual evaluation. *Dentomaxillofac Radiol.* 2020;49(6):20190378. doi:10.1259/dmfr.20190378
31. Akkaya N, Kansu O, Kansu H, Cagirankaya LB, Arslan U. Comparing the accuracy of panoramic and intraoral radiography in the diagnosis of proximal caries. *Dentomaxillofac Radiol.* 2006;35(3):170-174. doi:10.1259/dmfr/26750940
32. Miller CA, Lee Y, Avey GD, Vorperian HK. Head position classification of medical imaging studies: an assessment and development of a protocol. *Dentomaxillofac Radiol.* 2020;49(4):20190220. doi:10.1259/dmfr.20190220