E- ISSN: 2148-9505

TURKISH JOURNAL OF ORTHODONTICS

TJO

The Official Journal of Turkish Orthodontic Society

galenos Publishing House

Volume 39
Issue 03
SEPTEMBER 2026

The Official Journal of Turkish Orthodontic Society

TURKISH JOURNAL OF ORTHODONTICS

TJO

galenos Publishing House

Publisher Contact
Address: Molla Gürani Mah. Kaçamak Sk. No: 21/1
34093 İstanbul, Türkiye
Phone: +90 (530) 177 30 97
E-mail: info@galenos.com.tr/yayin@galenos.com.tr
Web: www.galenos.com.tr
Publisher Certificate Number: 14521

Printing Date: September 2026
E-ISSN: 2148-9505
International scientific journal published quarterly.

The Official Journal of Turkish Orthodontic Society

TURKISH JOURNAL OF ORTHODONTICS

TJO

Editorial Board

Please refer to the journal's webpage (**https://turkjorthod.org/**) for "Ethical Policy", "Instructions to Authors" and "About Us".

The editorial and publication processes of the journal are shaped in accordance with the guidelines of the ICMJE, WAME, CSE,COPE, EASE, and NISO. Turkish Journal of Orthodontics is indexed in PubMed Central, Web of Science-Emerging Sources Citation Index, DOAJ, Scopus, CNKI, Gale, EBSCO and TUBITAK ULAKBİM TR Index.

The journal is published online.

Owner: Gökmen Kurt on behalf of the Turkish Orthodontic Society

Responsible Manager: Çağla Şar

Editor in Chief: Çağla Şar

Address: Sütlüce Mah. İmrahor Cad. No: 82 Beyoğlu, İstanbul/Türkiye

Phone: +90 (212) 416 61 13

E-mail: info@turkjorthod.org

The Official Journal of Turkish Orthodontic Society

TURKISH JOURNAL OF ORTHODONTICS

TJO

Contents

DOI: 10.4274/TurkJOrthod.2026.2025.104

TURKISH JOURNAL OF ORTHODONTICS

Original Article

Assessing the Reliability of Artificial Intelligence-Based Automatic Digital Model Analysis Compared with Manual Approaches

Zeynep Beyçe Karaman, Can Sever, Kaan Kahya, Can Arslan, Derya Çakan

Yeditepe University Faculty of Dentistry, Department of Orthodontics, İstanbul, Türkiye

Cite this article as: Karaman ZB, Sever C, Kahya K, Arslan C, Çakan D. Assessing the reliability of artificial intelligence-based automatic digital model analysis compared with manual approaches. *Turk J Orthod.* 2026; 39(3): 116-125

Main Points

- Artificial intelligence-based automatic digital models demonstrated reliability equivalent to that of manual digital methods, with no significant differences in arch length deficiency and Bolton discrepancy measurements; furthermore, both digital methods provided superior precision in mesiodistal tooth dimensions compared with traditional plaster models, except in the premolar region.
- Manual plaster models showed significantly higher measurement values than both digital methods, suggesting an increased probability of error due to conventional processing steps.
- Deep learning-enhanced automatic segmentation eliminates operator-dependent variability while maintaining diagnostic accuracy, offering substantial time savings in clinical workflow.
- The study addresses a critical gap in the orthodontic literature by directly comparing digital analysis methods without relying on traditional plaster model benchmarks.

ABSTRACT

Objective: This study was conducted to compare the accuracy and reliability of artificial intelligence (AI)-supported automatic digital (AD) model analysis systems with manual digital (MD) and manual plaster (MP) methods in orthodontic diagnosis.

Methods: Twenty-four individuals (8 male, 16 female; mean age 19.5±6.05 years) with permanent dentition and mild to moderate crowding were included. Linear measurements were performed using digital calipers on plaster models. Digital models were obtained through intraoral scanning and analysed using OrthoSystem Software (3Shape, Copenhagen, Denmark) with both manual and automatic segmentation. Mesiodistal tooth dimensions, arch length deficiency were evaluated across all three methods, and Bolton analysis. Arch width measurements were compared only between the digital and MP methods due to software limitations in the automatic detection of landmarks. Statistical analysis included repeated-measures analysis of variance, Friedman's test, and paired t-tests with significance set at $p<0.05$.

Results: All three methods demonstrated consistent results with low standard deviations and narrow confidence intervals. Significant differences were found in mesiodistal tooth measurements between digital systems and plaster models, excluding premolars ($p<0.05$). No significant differences were observed between AD and MD systems for arch length deficiency or Bolton discrepancy ($p>0.05$). The MP method yielded higher measurements than those from digital systems, suggesting a higher error rate attributable to conventional phase redundancy.

Conclusion: Artificial intelligence-based AD model analysis provides reliable, consistent measurements comparable to those of MD methods. The automatic segmentation approach, enhanced by deep learning algorithms, demonstrates clinical accuracy approaching that of MD analysis, offering a time-efficient alternative for orthodontic model assessment.

Keywords: Model analysis, digital orthodontics, intra-oral scanning, artificial intelligence, automatic segmentation

Corresponding author: Can Sever, **e-mail:** can.sever@std.yeditepe.edu.tr **ORCID:** orcid.org/0009-0006-6406-4865
Received: August 29, 2025 **Accepted:** June 04, 2026 **Publication Date:** September 25, 2026

INTRODUCTION

Model analysis plays a crucial role in evaluating the dental and skeletal structures of patients during orthodontic treatment planning. Traditionally, plaster models are created by taking alginate impressions of patients' teeth. Measurements such as tooth dimensions, arch length, intercanine distance, overjet, and overbite are performed on these models, and analyses, such as Hayce-Nance, are used to determine space deficiencies. Bolton analysis is used to evaluate the dimensional harmony of the teeth.[1,2] These measurements and analyses are essential for understanding malocclusion severity and for developing individualised treatment plans to achieve functional and aesthetic alignment. The accuracy of these measurements directly impacts treatment outcomes and patient satisfaction.[3]

Technological advancements have significantly influenced orthodontic model analysis.[4] Digital models provide enhanced dental arch visualisation, more accurate treatment outcome simulations, and more efficient communication among orthodontists, dental laboratories, and patients.[5] Additionally, the ability to store and modify digital records instantaneously has streamlined workflows. Within these digital systems, measurements can be obtained by manually segmenting standard tessellation language files uploaded to systems such as OrthoSystem.

Recent developments in AI-based orthodontic analysis have led to significant advances. Yu et al.[4] evaluated the reliability and time-efficiency of artificial intelligence (AI)-based automatic digital (AD) model analysis, reporting substantial time savings with maintained accuracy. Recent comprehensive reviews have further highlighted the expanding role of AI in orthodontic diagnostics. Kazimierczak et al.[6] conducted a systematic review demonstrating AI's high efficacy in dental diagnostics, cephalometric evaluation, and treatment planning; however, they noted the need for manual supervision in clinical settings. Similarly, Nordblom et al.[7] critically evaluated existing AI applications in orthodontics, emphasising the gap between research publications and regulatory approvals. Although various commercial platforms, including Invisalign's ClinCheck, SureSmile, and Orchestrate 3D, employ different AI-driven approaches for model analysis, the lack of recent head-to-head comparisons in the peer-reviewed literature represents a significant knowledge gap. These systems use deep learning algorithms and neural networks to perform automatic tooth segmentation and landmark identification, offering potential advantages in reduced operator dependency and improved workflow efficiency; however, questions remain regarding their diagnostic accuracy across diverse clinical scenarios.[8]

Automatic segmentation can now be performed using AI-based software, providing precise measurements of parameters for model analysis. AI-based AD model analysis represents a paradigm shift in orthodontics, leveraging advanced algorithms to analyse digital models with greater speed, efficiency, and consistency than traditional methods.[9] Although these systems promise transformative improvements, they also raise questions regarding their diagnostic accuracy and clinical utility. Evaluating the differences between digital and manual methods is essential to ensure the reliability and safety of patient care.[10]

This study was conducted to assess the accuracy and reliability of AI-based AD model analysis compared with manual digital (MD) measurements and manual plaster (MP) measurements in orthodontic diagnosis. Specifically, we hypothesised that AI-based AD model analysis would demonstrate measurement accuracy and reliability comparable to those of MD methods for tooth-specific parameters and Bolton analysis, and that both digital approaches would show systematic differences from traditional plaster models. Comparing manual and AD methods holds particular clinical significance as it directly addresses the practical question of whether AI-based systems can replace time-intensive manual segmentation without compromising diagnostic accuracy. This evaluation is essential for evidence-based clinical decision-making regarding the integration of AI tools into orthodontic workflows, as it determines whether practitioners can confidently adopt automated systems for treatment planning, space analysis, and Bolton discrepancy assessment. This pilot study addresses a significant gap in the orthodontic literature by directly comparing digital analysis methods in their own right, without solely relying on traditional plaster model benchmarks,[8,11] thereby providing preliminary evidence for the clinical applicability of AI-driven automatic segmentation in orthodontic practice.

METHODS

A total of 24 individuals (8 male and 16 female; mean age 19.5±6.05 years) undergoing orthodontic treatment at the Orthodontic Department of Yeditepe University Dental Hospital, were included in this study. The inclusion criteria were as follows: individuals presenting with permanent dentition, no congenitally missing teeth, mild to moderate dental crowding, and an absence of supernumerary teeth or other significant dental anomalies. Patients with extracted teeth (except third molars), severe crowding (>6 mm arch length deficiency), extensive restorations affecting proximal surfaces (which could interfere with accurate mesiodistal measurements), severe skeletal malocclusions (ANB >6° or <-2°), craniofacial syndromes, and poor-quality intraoral scans with artifacts or incomplete tooth surfaces were excluded. Only patients with mild to moderate crowding were included for two reasons: (1) to ensure adequate visualisation and segmentation of tooth boundaries, as severe crowding with significant overlapping may challenge both manual operators and AI algorithms in accurately identifying interproximal contact points, and (2) to represent the most common clinical scenario encountered in routine orthodontic practice, thereby maximising the clinical applicability of our findings. These criteria ensured that the study focused on a relatively uniform sample population suitable for both manual and automatic measurements. This pilot study was conducted to provide preliminary evidence for the reliability of AI-based automatic segmentation compared with traditional methods in orthodontic model analysis. The limited sample size (n=24) is acknowledged as a limitation that may affect generalisability, particularly to populations with severe malocclusions or diverse morphological variations. However, this sample size is consistent with similar methodological validation studies in orthodontic digital model analysis (Lippold et al.,[11] n=20; Reuschl et al.,[10] n=30; Yu et al.,[4] n=30) and was deemed adequate for this initial comparative assessment. Future larger-scale studies with more diverse patient populations will be necessary to establish broader clinical guidelines.

Study Design

This prospective cross-sectional comparative study was designed to evaluate three different orthodontic model analysis methods simultaneously. All measurements were performed on pretreatment models obtained from the same patients at a single time point, before any orthodontic intervention (bracket placement or aligner therapy), enabling a direct comparison between methods without temporal variables or treatment-related confounders. Patients were recruited consecutively as they presented for initial orthodontic consultation at the Orthodontic Department of Yeditepe University Dental Hospital.

This study was approved by the Yeditepe University Non-Interventional Clinical Research Ethical Committee (approval number: 20231040657, date: 11.10.2024). All methods were conducted in accordance with relevant guidelines and regulations, including the Declaration of Helsinki. Written informed consent was obtained from all participants before participation in the study. For participants under 18 years of age, written informed consent was obtained from their parents or legal guardians. Alginate impressions of the participants' dentition were obtained and used to create plaster models, which served as the basis for manual linear measurements. Measurements were taken directly on the plaster models using a high-precision digital caliper. Each measurement was performed three times by the same examiner to minimise human error and ensure intra-examiner reliability. The mean of these three measurements was recorded as the final value for each parameter. Representative images of the plaster model measurement process are presented in Figures 1 and 2. Intraoral scans were performed for each patient using a 3Shape intraoral scanner (3Shape, Copenhagen, Denmark) to obtain digital diagnostic models. The scanned data were exported and uploaded to the OrthoSystem Software (3Shape, Copenhagen, Denmark). Within the software, two different measurement methods were applied. First, manual tooth segmentation was performed by an experienced orthodontist, and the resulting measurements were recorded for further analysis. This step involved the careful delineation of tooth boundaries to ensure accurate representation and reliable measurement (Figure 3). Subsequently, the same digital model was uploaded as a separate file for automatic segmentation using the software's built-in AI algorithm. The AI segmentation algorithm employed by OrthoSystem utilises deep learning neural networks trained on thousands of annotated dental models. The workflow comprises the following: (1) automatic detection of tooth boundaries using convolutional neural networks that identify edges and surface discontinuities, (2) individual tooth segmentation based on morphological features including cuspal anatomy and interproximal contact areas, (3) landmark identification for measurements using pattern recognition algorithms to locate mesial and distal contact points, and (4) automatic calculation of predetermined parameters, including mesiodistal tooth dimensions and Bolton ratios. The entire automatic process takes approximately 2-3 minutes per model, compared with 10-15 minutes for manual segmentation.

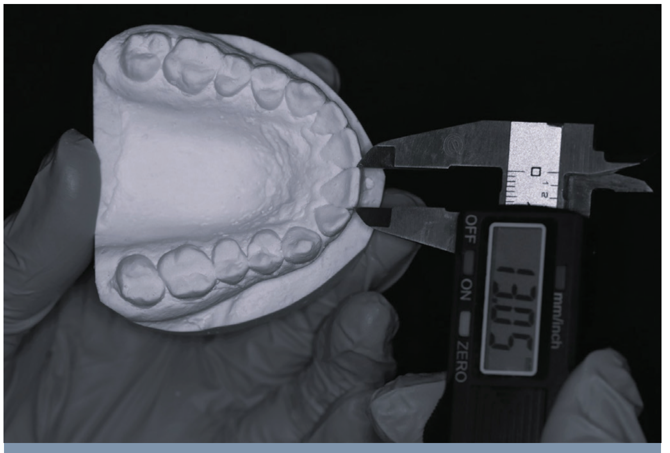


Figure 1. Measurement of mesiodistal tooth dimensions on plaster models using a high-precision digital calliper. The calliper was positioned at the widest mesiodistal diameter of the tooth. Each measurement was performed three times by the same examiner, and the mean value was recorded to ensure intra-examiner reliability and minimise measurement error.

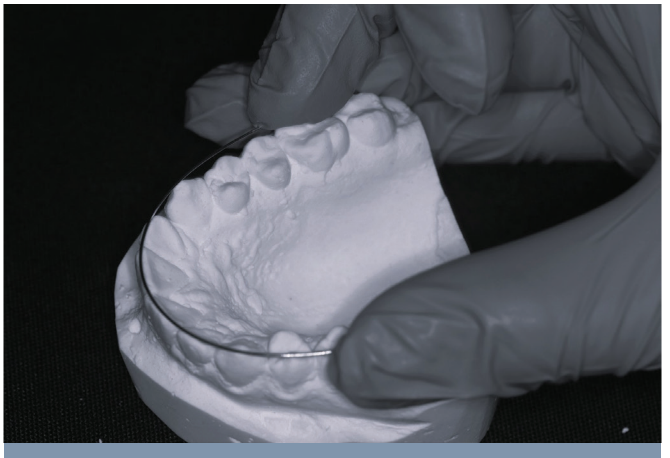


Figure 2. Measurement of arch dimensions on plaster models using digital callipers. Linear measurements, including intercanine width, intermolar width, and arch length, were obtained directly from these conventional models. The caliper beaks were positioned at the predefined anatomical landmarks to measure transverse arch dimensions in the posterior region.

The automatic segmentation feature analysed the model and generated corresponding measurements, which were recorded for comparison against the manually obtained results (Figure 4). In addition to individual tooth measurements, distance measurements were evaluated, including intercanine width, intermolar width, and arch length (Table 1). Due to software limitations requiring manual landmark placement, these measurements could be compared only between the MD and MP methods. The software's automatic segmentation feature does not currently support automated landmark identification for arch dimension measurements.

All measurements, both manual and automatic, were performed in a controlled environment to standardise conditions and minimise external variables.

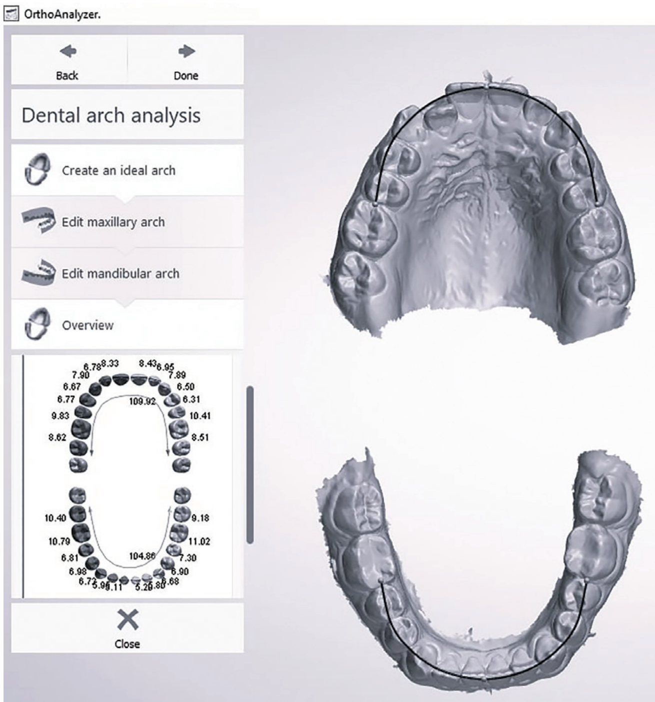


Figure 3. OrthoSystem digital model analysis interface (3Shape, Copenhagen, Denmark) showing maxillary and mandibular arch measurements. The left panel displays the dental arch analysis menu, with options for arch editing and measurements. The schematic diagram in the lower left shows individual tooth mesiodistal measurements in millimetres for both arches. The right panels show 3D digital models obtained through intraoral scanning in occlusal view. Manual analysis required operator-dependent tooth identification and typically took 10-15 minutes per patient.

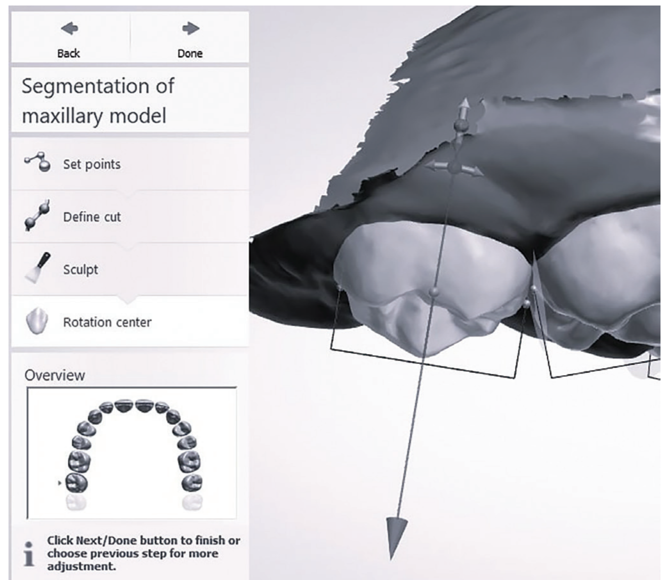


Figure 4. Manual segmentation interface in OrthoSystem software showing the tooth boundary delineation process. The left panel displays the segmentation tools, including set points for landmark placement, define cut for separating individual teeth, sculpt for refining tooth boundaries, and rotation center for orientation adjustments. The main view (right) presents a magnified buccal view of the maxillary posterior region, where the operator manually defines the separation lines between adjacent teeth (blue lines indicate the operator-defined cut lines). The overview panel (lower left) shows the complete dental arch with individually segmented teeth. This manual segmentation process required precise identification of the interproximal contact areas and typically took approximately 10-15 minutes per model, compared with 2-3 minutes for AI-based automatic segmentation.
AI, artificial intelligence.

Table 1. Definition of the variable measurements and model analysis

Variables measurements	Definition
Tooth widths	The widest mesiodistal width of the teeth in occlusal view
Sum of 10 teeth	Sum of tooth widths from the second premolar to the second premolar
Sum of 12 teeth	Sum of tooth widths from the first molar to the first molar
Sum of 14 teeth	Sum of tooth widths from the second molar to the second molar
Intercanine width	The distance between the cusp tips of the canines
Arch perimeter	Measure from the distal surface of one first molar to the distal surface of the opposite first molar along the occlusal curve
Analysis	
Bolton 6 (anterior ratio)	Ratio of sum of mesiodistal width of teeth 33-43 and sum of mesiodistal width of teeth 13-23
Bolton 12 (overall ratio	Ratio of sum of mesiodistal width of teeth 36-46 and sum of mesiodistal width of teeth 16-26
Arch length discrepancy	Difference of available arch perimeter and required arch perimeter

The dual approach, combining manual and automatic segmentation, enabled a comprehensive evaluation of the AI-based measurement system's reliability and accuracy relative to those of the traditional manual method. This protocol was designed to ensure methodological rigor and provide meaningful insights into AI performance in diagnostic orthodontic analysis.

Statistical Analysis

Statistical analyses were performed using SPSS, version 23.0 (IBM Corp., Armonk, NY, USA). The Shapiro-Wilk test was used to assess normality. Repeated-measures analysis of variance (ANOVA) was used to compare values across the three models that fit a normal distribution, and multiple comparisons were performed using the Bonferroni test.

Friedman's test was used to compare values across the three models that did not fit a normal distribution, and multiple comparisons were performed using Dunn's test. A paired two-sample t-test was used to compare values in two models that fit a normal distribution, and the Wilcoxon test was used to compare values in two models that did not fit a normal distribution. The analysis results are presented as the mean±standard deviation or median (minimum-maximum). The significance level was set at $p<0.05$. To assess reliability, digital and manual measurements were repeated 15 days after the initial measurements for 12 randomly selected individuals. Intra-observer correlation coefficients were found to be in the range of 0.91 to 0.96.

A post-hoc power analysis was conducted to evaluate the adequacy of the sample size. With 24 participants and three measurement methods (AD, MD, MP), using a repeated-measures ANOVA with an alpha level of 0.05, the study achieved statistical power of approximately 85% for detecting large effect sizes (f=0.40) and 60% for medium effect sizes (f=0.25). This indicates adequate power for detecting clinically meaningful differences between measurement methods. However, it should be acknowledged that the sample size was determined according to feasibility and consistency with similar pilot studies in the literature rather than a priori power calculation. Future studies with larger sample sizes (n>50) would provide greater power to detect smaller effect sizes and enable more robust subgroup analyses.

RESULTS

In the tables comparing three groups, pairwise comparisons were also performed and labelled a and b. Values are not labelled when there is no difference in pairwise comparisons between models, regardless of whether the p-value indicates statistical significance. When the data did not conform to normality, Friedman's test was applied, which is indicated by double asterisks. When comparing the mesiodistal dimension, the results showed no difference between the digital systems. However, there was a significant difference between the measurements made on the plaster model and on the digital systems ($p<0.05$; Tables 2 and 3). A comparison of digital measurements revealed internal consistency of the systems, with a significant disparity observed only in the measurements of the summed values for maxillary teeth 12 and 14.

Table 4 presents a direct comparison of the two digital measurement methods (AD vs. MD) to evaluate the consistency of digital systems independently of traditional plaster models.

The absence of significant differences between the AD and MD methods across most parameters is a key finding that directly addresses our study hypothesis. This consistency demonstrates that AI-based automatic segmentation achieves measurement reliability comparable to that of MD analysis, the current clinical standard for digital orthodontic model evaluation. The finding that automated systems can replicate MD measurements without significant deviation is clinically meaningful, as it validates the diagnostic accuracy of AI algorithms while offering substantial time savings (2-3 minutes vs. 10-15 minutes per model) and eliminating operator-dependent variability. Although both digital methods showed systematic differences from traditional plaster models, the equivalence between AD and MD methods supports the potential for AI-based systems to replace time-intensive manual segmentation in routine clinical practice without compromising measurement accuracy. This finding is particularly relevant given that the study's primary aim was to evaluate whether AI systems could achieve parity with MD methods rather than to demonstrate superiority. To address potential concerns about statistical power, a post-hoc power analysis was conducted, demonstrating that the study achieved 85% power to detect large effect sizes (f=0.40) with our sample of 24 participants. This finding indicates that the lack of significant differences between AD and MD methods reflects genuine measurement equivalence rather than insufficient statistical power to detect differences. The narrow confidence intervals and low standard deviations observed across both digital methods further support this interpretation, demonstrating consistent measurement precision independent of the segmentation approach (manual vs. automatic).

Significant differences were identified in mesiodistal tooth measurements when comparing the MP and AD methods (Table 5). However, no significant difference was found in Bolton analysis values.

Table 6 compares MD and MP to evaluate the accuracy of MD analysis relative to conventional methods. Distance measurements (intercanine width, intermolar width, and arch length) were evaluated between digital and MP methods.

Automatic segmentation could not be assessed for these parameters because the software requires manual landmark identification for distance measurements in both digital methods. This represents a significant limitation of existing AI algorithms, which cannot independently identify the anatomical landmarks required for arch dimension calculations. Consequently, transverse and sagittal measurements could be compared only between the MD and MP methods, precluding a comprehensive evaluation of AI-based automatic segmentation across all clinically relevant parameters. This technological constraint highlights an important area necessitating future AI development in orthodontic software. Significant differences were found between digital and MP measurements (p<0.05), with digital systems consistently showing higher values for transversal measurements.

DISCUSSION

The principal finding of this study is that AI-based AD-model analysis produced mesiodistal tooth measurements and Bolton ratios that were statistically equivalent to those obtained from MD analysis, supporting our primary hypothesis. This equivalence is clinically meaningful, suggesting that the 2-3-minute automatic segmentation workflow can replace the 10-15-minute manual process without sacrificing diagnostic accuracy in cases of mild-to-moderate crowding. These findings add to the growing body of evidence supporting digital orthodontic workflows[4,9] and are consistent with the findings of Yu et al.,[4] who similarly reported maintained accuracy alongside substantial time savings in an AI-based automatic model analysis system. Our study extends this evidence by directly benchmarking AD against MD methods, using the same intraoral scanning platform under controlled conditions, thereby isolating the effect of the segmentation method from the scanner variability.

The consistent finding that both digital methods yielded higher mesiodistal values than MP models aligns with prior literature. Cuperus et al.[12] and Kardach et al.[13] both reported

Table 2. Comparison of maxillary tooth size measurements between manual plaster, manual digital and automated digital groups

	Factor	MP	MD	AD	p
	11	8.47±0.51^{b}	8.8±0.41^{a}	8.8±0.42^{a}	**<0.001***
		8.6 (7.15-9.38)	8.82 (7.73-9.8)	8.82 (7.73-9.77)	
	12	6.57±0.71^{b}	6.87±0.57^{a}	6.88±0.57^{a}	**<0.001***
		6.52 (5-8.15)	6.91 (5.24-8.05)	6.93 (5.24-8.05)	
	13	7.92±0.51	8.07±0.68	8.11±0.69	**0.041****
		8.04 (6.8-8.56)	8.27 (6.2-9.12)	8.27 (62-9.12)	
M	14	7.1±0.52	7.23±0.6	7.03±0.59	**0.048****
		6.98 (6.39-8.31)	7.27 (5.19-8.43)	6.98 (5.19-8.43)	
A	15	6.85±0.4^{b}	7.03±0.4^{a}	6.91±0.35^{ab}	**0.017***
		6.84 (6.02-7.73)	7.04 (6.03-7.95)	6.85 (6.39-7.77)	
X	16	10.21±0.68^{b}	10.58±0.68^{a}	10.57±0.66^{ab}	**0.014***
		10.17 (8.9-12)	10.56 (9.29-11.92)	10.57 (9.29-11.7)	
I	17	9.82±0.88	9.91±0.76	9.78±0.8	0.596*
L		9.72 (8.13-12)	9.88 (8.5-11.27)	9.8 (8.5-11.27)	
L	21	8.55±0.42^{b}	8.75±0.42^{a}	8.75±0.42^{a}	**<0.001***
		8.62 (7.65-9.3)	8.75 (7.66-9.48)	8.75 (7.66-9.48)	
A	22	6.56±0.67	6.88±0.73	6.89±0.74	**<0.001****
		6.47 (4.49-7.68)^{b}	6.91 (4.45-7.93)^{a}	6.91 (4.45-7.93)^{a}	
	23	7.74±0.56^{b}	7.98±0.62^{a}	8.02±0.65^{a}	**0.010***
		7.8 (6.7-8.72)	8.04 (6.29-8.91)	8.04 (6.29-9.14)	
	24	7.02±0.44^{b}	7.22±0.38^{a}	6.99±0.49^{b}	**0.002***
		7 (6.19-8.08)	7.18 (6.68-8.35)	6.97 (6.16-8.2)	
	25	6.92±0.61	6.92±0.46	6.91±0.55	0.957*
		6.97 (5.7-8)	6.84 (6.22-7.87)	6.78 (6.07-8.17)	
	26	10.07±0.64^{b}	10.69±0.62^{a}	10.66±0.53^{a}	**<0.001***
		10.11 (8.6-11.56)	10.83 (9.55-11.78)	10.67 (9.55-11.63)	
	27	9.63±0.88^{b}	10.06±0.84^{a}	9.96±0.77^{a}	**0.003***
		9.7 (7.9-12)	9.95 (7.89-11.86)	9.95 (8.12-11.26)	

*Repeated measures analysis of variance; **Friedman test; mean ± standard deviation; median (minimum-maximum); $^{a-b}$No difference between models with the same letter. The reason why there is no lettering in the tables even though the p-value is significant is due to the fact that there is no difference in the results of the pairwise comparison of the models. Significance level was taken as p<0.05.
MP, manual plaster; MD, manual digital; AD, automated digital.

Table 3. Comparison of mandibular tooth size measurements between manual plaster, manual digital and automated digital groups

	Factor	MP	MD	AD	p
	31	5.47±0.31	5.61±0.27	5.6±0.28	**0.019***
		5.49 (4.78-6.07)	5.59 (5.25-6.19)	5.58 (5.25-6.19)	
	32	6.03±0.36	6.13±0.33	6.14±0.33	**0.003****
		5.95 (5.5-6.8)^{b}	6.13 (5.7-6.74)^{a}	6.19 (5.7-6.74)^{a}	
	33	6.71±0.5^{b}	7.08±0.66^{a}	7.13 ±0.63^{a}	**0.002***
		6.77 (5.58-7.48)	7.04 (6.16-8.73)	7.04 (6.24-8.73)	
M	34	7.23±0.47	7.3±0.45	7.33±0.43	0.349*
		7.27 (6.26-8.3)	7.28 (6.59-8.2)	7.34 (6.55-8.2)	
A	35	7.21±0.57	7.32±0.49	7.33±0.55	0.783**
		7.2 (6.26-8.2)	7.27 (6.47-8.87)	7.29 (6.54-9.33)	
N	36	10.89±0.81^{b}	11.16±0.68^{a}	11.18±0.7^{a}	**0.011***
		10.96 (9.4-12.45)	11.21 (9.51-12.13)	11.16 (9.51-12.2)	
D	37	10.43±1.24	10.49±0.77	10.41±0.95	0.134**
		10.09 (9-15.34)	10.36 (9.16-12)	10.19 (8.19-12.08)	
I	41	5.5±0.32^{b}	5.83±0.53^{a}	5.84±0.52^{a}	**0.002***
		5.44 (4.94-6.06)	5.79 (4.92-6.99)	5.79 (4.92-6.99)	
B	42	5.86±0.41	6.21±0.35	6.21±0.35	**<0.001****
		5.85 (5.07-6.7)^{b}	6.08 (5.78-7.12)^{a}	6.08 (5.78-7.12)^{a}	
L	43	6.73±0.52	6.97±0.48	6.98±0.39	0.079**
		6.9 (5.85-7.5)	6.81 (5.74-7.67)	6.91 (5.99-7.64)	
E	44	7.12±0.63^{b}	7.39±0.51^{ab}	7.43±0.52^{a}	**0.012***
		7.12 (6.15-8.13)	7.25 (6.45-8.44)	7.26 (6.45-8.44)	
	45	7.07±0.66^{b}	7.35±0.55^{a}	7.3±0.56^{ab}	**0.005***
		7.21 (5.24-8.11)	7.32 (6.42-8.58)	7.34 (6.44-8.79)	
	46	10.74±1.05^{b}	11.13±0.86^{a}	11.14±0.8^{a}	**0.010***
		10.85 (8.53-12.65)	11.08 (9.28-12.62)	11.05 (9.52-12.62)	
	47	10.11±0.65	10.3±0.85	10.3±0.81	0.142*
		10.13 (8.78-10.98)	10.27 (7.88-11.85)	10.31 (7.88-11.85)	

*Repeated measures analysis of variance; **Friedman test; mean ± standard deviation; median (minimum-maximum); $^{a-b}$No difference between models with the same letter. The reason why there is no lettering in the tables even though the p-value is significant is due to the fact that there is no difference in the results of the pairwise comparison of the models. Significance level was taken as p<0.05.
MP, manual plaster; MD, manual digital; AD, automated digital.

Table 4. Comparison of linear measurements between manual digital and automatic digital groups

		MD	AD	Mean differences ± s.d	p
	Factor	Mean ± s.d	Mean ± s.d		
	Sum of 10 teeth	67.29±4.25	67.48±4.11	0.18±0.71	0.219*
	Sum of 12 teeth	89.39±4.83	89.57±4.61	0.18±0.81	0.295*
Mn	Sum of 14 teeth	110.18±5.88	110.17±5.73	-0.01±0.88	0.963*
	ALD	-4.93±4.29	-5.11±4.12	-0.18±0.71	0.219*
	Sum of 10 teeth	75.32±4.82	74.99±3.72	-0.33±2.67	0.546*
	Sum of 12 teeth	97.15±4.81	96.4±4.46	-0.75±1.57	**0.028***
Mx	Sum of 14 teeth	117.2±5.76	116.24±5.53	-0.96±1.8	**0.015***
	ALD	-1.03±5.28	-0.87±4.13	0.17±2.73	0.766*
	Bolton 12	0.72±3.71	1.31±3.56	0.33 (-1.36-7.68)	0.083**
	Bolton 6	1.38±1.35	1.45±1.36	0.01 (-0.7-0.8)	0.360**

*Paired two-sample t-test **Wilcoxon Test. Significance level was taken as p<0.05.
MD, manual digital; AD, automated digital; ALD, arch length discrepancy; s.d, standard deviation.

digital measurements exceeding plaster values, attributing this to the dimensional instability of gypsum and the superior proximal surface resolution of direct intraoral scanning. However, the absence of a significant difference in premolar measurements between the digital and MP methods diverges from the general pattern and is consistent with Kardach et al.'s[13] observation of region-specific variation. This finding likely reflects the more accessible morphology and favourable positioning of premolars in cases of mild crowding, where both digital and physical measurement tools provide comparable accuracy. In contrast, the anterior region is more susceptible to measurement discrepancies due to crowding-related overlap and to the challenge of precisely locating interproximal contact points. These findings suggest that AI segmentation performance is not uniform across the dental arch and may be influenced by regional morphological factors.

Bolton analysis showed no significant differences across all three methods, consistent with the findings of Mullen et al.[14] and Quimby et al.[15] This finding is likely attributable to the proportional nature of the Bolton index: systematic overestimation by digital methods affects the numerator and denominator similarly, thereby attenuating the net discrepancy. Furthermore, this finding supports the use of both digital methods for Bolton analysis in routine clinical practice, irrespective of plaster model benchmarks.

A key limitation that has direct clinical relevance is the AI system's current inability to automatically identify landmarks for arch-width and arch-length measurements. These parameters, including intercanine width, intermolar width, and total arch length are essential for space analysis and treatment planning; however, they require manual placement of landmarks in both digital methods. Fully automated orthodontic model analysis is not yet achievable with current technology, and the time-saving advantage of AI segmentation applies only to tooth-specific measurements. Kazimierczak et al.[6] and Nordblom et al.[7] similarly noted that current AI orthodontic

Table 5. Comparison of linear measurements between manual plaster and automatic digital groups

		MD	AD	Mean differences ± s.d	p
	Factor	Mean ± s.d	Mean ± s.d		
	Sum of 10 teeth	67.29±4.25	67.48±4.11	0.18±0.71	0.219*
	Sum of 12 teeth	89.39±4.83	89.57±4.61	0.18±0.81	0.295*
Mn	Sum of 14 teeth	110.18±5.88	110.17±5.73	-0.01±0.88	0.963*
	ALD	-4.93±4.29	-5.11±4.12	-0.18±0.71	0.219*
	Sum of 10 teeth	75.32±4.82	74.99±3.72	-0.33±2.67	0.546*
	Sum of 12 teeth	97.15±4.81	96.4±4.46	-0.75±1.57	**0.028***
Mx	Sum of 14 teeth	117.2±5.76	116.24±5.53	-0.96±1.8	**0.015***
	ALD	-1.03±5.28	-0.87±4.13	0.17±2.73	0.766*
	Bolton 12	0.72±3.71	1.31±3.56	0.33 (-1.36-7.68)	0.083**
	Bolton 6	1.38±1.35	1.45±1.36	0.01 (-0.7- 0.8)	0.360**

*Paired two-sample t-test **Wilcoxon test. Significance level was taken as p<0.05.
MD, manual digital; AD, automated digital; ALD, arch length discrepancy; s.d, standard deviation.

Table 6. Comparison of linear measurements between manual plaster and manuel digital groups

		MP	MD	Mean differences ± s.d	p
	Factor	Mean ± s.d	Mean ± s.d		
	Sum of 10 teeth	64.97±4.95	67.29±4.25	-2.33±3.46	**0.003***
	Sum of 12 teeth	86.38±5.61	89.39±4.83	-3.01±3.91	**0.001***
Mn	Sum of 14 teeth	106.87±6.73	110.18±5.88	-3.3±4.4	**0.001***
	ALD	1.81±5.83	-4.93±4.29	6.74±5.39	**<0.001***
	Sum of 10 teeth	73.23±3.8	-2.1±2.86	-2.1±2.86	**0.002***
	Sum of 12 teeth	93.78±4.14	-3.37±3.09	-3.37±3.09	**<0.001***
Mx	Sum of 14 teeth	113.25±5.08	-3.95±3.5	-3.95±3.5	**<0.001***
	ALD	3.92±5.21	4.96±5.94	4.96±5.94	**<0.001***
	Bolton 12	0.58±3.92	0.72±3.71	0.33 (-7.95-4.41)	0.742**
	Bolton 6	1±1.51	1.38±1.35	-0.71 (-2.58-2.28)	0.247**

*Paired two-sample t-test **Wilcoxon test. Significance level was taken as p<0.05.
MP, manual plaster; MD, manual digital; ALD, arch length discrepancy; s.d, standard deviation.

tools require continued manual supervision and that a gap remains between research outputs and regulatory-level clinical validation. Our findings reinforce this view and identify landmark detection as a specific technical bottleneck requiring targeted development.

These findings should be interpreted within the study's limitations, which are detailed in the Study Limitations section below.

Study Limitations

The sample size (n=24) and inclusion of only mild-to-moderate crowding cases limit generalisability to more severe or complex malocclusions. The sample size in this study is consistent with that in comparable methodological investigations (Lippold et al.,[11] n=20; Reuschl et al.,[10] n=30; Yu et al.,[4] n=30) and achieved adequate statistical power (85%) for large effect sizes; however, a priori power calculation was not performed. Inter-examiner reliability was not assessed, as all measurements were performed by a single orthodontist. However, intra-examiner reliability was high (intraclass correlation coefficient 0.91-0.96). Only one commercial AI platform (OrthoSystem, 3Shape) was evaluated, and performance may differ across systems with different training datasets and algorithms. The AI system's current inability to automatically identify landmarks for arch-width and arch-length measurements precluded a full comparative analysis across all clinically relevant parameters. This cross-sectional study assessed only measurement accuracy; the clinical impact on treatment planning and outcomes was not evaluated.

CONCLUSION

Within the limitations of this pilot study, AI-based AD model analysis demonstrated measurement reliability comparable to that of MDI methods for tooth-specific parameters and Bolton analysis, thereby largely supporting our study hypothesis. The absence of significant differences between automatic and MD methods confirms that AI-based automatic segmentation can achieve diagnostic reliability equivalent to that of MD analysis. However, the hypothesis is partially rejected regarding comprehensive parameter assessment, as AI systems could not evaluate arch-width measurements due to current limitations in automatic landmark detection.

These preliminary findings indicate potential for clinical application-particularly because of its improved time efficiency (2-3 minutes versus 10-15 minutes for manual segmentation). Nevertheless, larger studies with more diverse populations assessing multiple AI platforms are needed to establish definitive clinical guidelines. The current inability of AI systems to automatically identify landmarks for arch measurements, combined with the single-platform evaluation (3Shape OrthoSystem), represents significant limitations that require continued development and validation.

Both automatic and MD measurement methods showed systematic differences from conventional plaster models, with digital methods generally providing higher measurement values. This pilot study provides initial evidence supporting the integration of AI-based analysis in orthodontic practice. However, manual verification remains essential for complex cases. Future research should focus on expanding training datasets to include diverse morphological variations and developing standardised protocols for AI validation in orthodontic diagnostics.

Ethics

Ethics Committee Approval: This study was approved by the Yeditepe University Non-Interventional Clinical Research Ethical Committee (approval number: 20231040657, date: 11.10.2024).

Informed Consent: Written informed consent was obtained from all participants before participation in the study.

Footnotes

Author Contributions: Surgical and Medical Practices - Z.B.K., C.S., K.K., C.A.; Concept - Z.B.K.; Design - Z.B.K.; Data Collection and/or Processing - C.S., K.K., C.A.; Analysis and/or Interpretation - D.Ç.; Literature Search - Z.B.K., C.S., K.K., C.A.; Writing - Z.B.K.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. Rheude B, Sadowsky PL, Ferriera A, Jacobson A. An evaluation of the use of digital study models in orthodontic diagnosis and treatment planning. *Angle Orthod.* 2005;75(3):300-304. [CrossRef]
2. Yuen KK, So LL, Tang EL. Mesiodistal crown diameters of the primary and permanent teeth in Southern Chinese--a longitudinal study. *Eur J Orthod.* 1997;19(6):721-731. [CrossRef]
3. Kumar AA, Ananthakrishnan MG, Kumar S, Divakar G, Sekar S, Dharani S. Assessing the validity and reliability of tooth widths and bolton ratios obtained from digital models and plaster models. *J Pharm Bioallied Sci.* 2022;14(Suppl 1):S148-S151. [CrossRef]
4. Yu JH, Kim JH, Liu J, Mangal U, Ahn HK, Cha JY. Reliability and time-based efficiency of artificial intelligence-based automatic digital model analysis system. *Eur J Orthod.* 2023;45(6):712-721. [CrossRef]
5. Gül Amuk N, Karsli E, Kurt G. Comparison of dental measurements between conventional plaster models, digital models obtained by impression scanning and plaster model scanning. *Int Orthod.* 2019;17(1):151-158. [CrossRef]
6. Kazimierczak N, Kazimierczak W, Serafin Z, Nowicki P, Nożewski J, Janiszewska-Olszowska J. AI in orthodontics: revolutionizing diagnostics and treatment planning-a comprehensive review. *J Clin Med.* 2024;13(2):344. [CrossRef]
7. Nordblom NF, Büttner M, Schwendicke F. Artificial intelligence in orthodontics: critical review. *J Dent Res.* 2024;103(6):577-584. [CrossRef]
8. Alam MK, Alanazi DSA, Alruwaili SRF, Alderaan RAI. Assessment of AI models in predicting treatment outcomes in orthodontics. *J Pharm Bioallied Sci.* 2024;16(Suppl 1):S540-S542. [CrossRef]
9. Scott JD, English JD, Cozad BE, et al. Comparison of automated grading of digital orthodontic models and hand grading of

3-dimensionally printed models. *Am J Orthod Dentofacial Orthop.* 2019;155(6):886-890. [CrossRef]

10. Reuschl RP, Heuer W, Stiesch M, Wenzel D, Dittmer MP. Reliability and validity of measurements on digital study models and plaster models. *Eur J Orthod.* 2016;38(1):22-26. [CrossRef]
11. Lippold C, Kirschneck C, Schreiber K, et al. Methodological accuracy of digital and manual model analysis in orthodontics-a retrospective clinical study. *Comput Biol Med.* 2015;62:103-109. [CrossRef]
12. Cuperus AM, Harms MC, Rangel FA, Bronkhorst EM, Schols JG, Breuning KH. Dental models made with an intraoral scanner: a validation study. *Am J Orthod Dentofacial Orthop.* 2012;142(3):308-313. [CrossRef]
13. Kardach H, Szponar-Żurowska A, Biedziak B. A comparison of teeth measurements on plaster and digital models. *J Clin Med.* 2023;12(3):943. [CrossRef]
14. Mullen SR, Martin CA, Ngan P, Gladwin M. Accuracy of space analysis with emodels and plaster models. *Am J Orthod Dentofacial Orthop.* 2007;132(3):346-352. [CrossRef]
15. Quimby ML, Vig KW, Rashid RG, Firestone AR. The accuracy and reliability of measurements made on computer-based digital models. *Angle Orthod.* 2004;74(3):298-303. [CrossRef]

DOI: 10.4274/TurkJOrthod.2026.2026.60

TURKISH JOURNAL OF
ORTHODONTICS

Original Article

Diagnostic Concordance of Panoramic Radiographs and Cone-Beam Computed Tomography and their Influence on Treatment Decision-Making for Impacted Canines

Genta Agani Sabah[1], Müfide Bengü Erden[2], Mehmet Gümüş Kanmaz[3]

[1]İzmir Tınaztepe University Faculty of Dentistry, Department of Orthodontics, İzmir, Türkiye
[2]İzmir Tınaztepe University Faculty of Dentistry, Department of Oral and Maxillofacial Surgery, İzmir, Türkiye
[3]İzmir Tınaztepe University Faculty of Dentistry, Department of Periodontology, İzmir, Türkiye

Cite this article as: Agani Sabah G, Erden MB, Kanmaz MG. Diagnostic concordance of panoramic radiographs and cone-beam computed tomography and their influence on treatment decision-making for impacted canines. *Turk J Orthod.* 2026; 39(3): 126-135

Main Points

- Parameters derived from panoramic radiographs were significantly associated with the cone-beam computed tomography (CBCT)-derived KPG index, indicating diagnostic concordance between two-dimensional (2D) and three-dimensional (3D) imaging modalities for evaluating the severity of maxillary canine impaction.
- Higher severity values in both 2D and 3D imaging parameters were strongly associated with the clinical decision to surgically extract impacted maxillary canines rather than attempt orthodontic eruption.
- An α angle of ≥46.1° and a d distance of ≥12.3 mm were identified as 2D screening thresholds associated with the decision for surgical extraction of impacted maxillary canines.
- Given the comparable diagnostic performance of 2D and 3D parameters, panoramic radiographs may serve as effective first-line screening tools, with CBCT reserved for precise surgical planning in complex cases that exceed the identified 2D thresholds, thereby optimizing the balance between diagnostic benefit and patient safety.

ABSTRACT

Objective: To evaluate the diagnostic concordance between panoramic radiographs and cone-beam computed tomography (CBCT) and to determine their comparative diagnostic performance in relation to the treatment decision of impacted maxillary canines.

Methods: Radiographic records of 51 impacted maxillary canines in 44 patients (37 unilateral and 7 bilateral cases) were analyzed. Panoramic radiographic evaluations included sector classification, the α angle, and the d distance. The KPG index was used in CBCT assessments to determine impaction severity.

Results: All imaging parameters showed high intra- and interrater reliability (intraclass correlation coefficients ≥0.927; κ≥0.658; p<0.001). Significant associations were found between the panoramic radiographic parameters (sector, α angle, and d distance) and the CBCT-based KPG index, and between sector classification and buccopalatal position (p<0.05). Regarding treatment decisions, higher parameter values were significantly associated with the choice of surgical extraction (p<0.05). Receiver operating characteristic analysis demonstrated acceptable to excellent discriminatory ability for panoramic radiograph-based and CBCT-based parameters [area under the curve (AUC): 0.715 to 0.843], with no differences among the AUC values (p>0.05). A d distance ≥12.3 mm and an α angle ≥46.1° were identified as thresholds associated with the decision for surgical extraction.

Conclusion: Panoramic radiographic parameters demonstrated diagnostic concordance with CBCT, and their performance was comparable to CBCT-based assessment for treatment decisions. The identified 2D thresholds may serve as an effective first-line screening tool for identifying complex cases likely to require surgical extraction, signaling a level of severity at which the superior anatomical detail of CBCT is warranted for precise surgical planning.

Keywords: Cuspid, impacted tooth, panoramic radiography, cone-beam computed tomography

Corresponding author: Asst. Prof. Genta Agani Sabah, **e-mail:** genta.agani@gmail.com **ORCID:** orcid.org/0000-0002-7571-6922
Received: March 18, 2026 **Accepted:** June 17, 2026 **Epub:** August 07, 2026 **Publication Date:** September 25, 2026

INTRODUCTION

Maxillary canines play a vital role in establishing functional occlusion, supporting midfacial esthetics, and maintaining the continuity of the dental arch.[1] However, maxillary canines commonly experience disturbances during their eruption because they originate deep within the maxilla, lateral to the piriform fossa. They follow the longest and most complex path to reach their final position in the dental arch, increasing the risk of ectopic eruption and impaction.[2,3] The prevalence of impacted maxillary canines has been reported to range from 1% to 3% in the general population.[3,4] If left untreated, impacted canines may lead to various complications, including displacement or resorption of adjacent teeth, alveolar bone loss, cyst formation, infection, and gingival recession.[3] Therefore, early and accurate diagnosis is essential for effective intervention and optimal treatment outcomes.[5]

The management of impacted canines requires an interdisciplinary approach, with multiple factors influencing the final orthodontic and periodontal outcomes.[6] Pretreatment radiographic features observed on panoramic radiographs, such as the inclination of the impacted canine (α angle), vertical distance between the cusp tip of the impacted canine and the occlusal plane (d distance), and position of the cusp tip relative to adjacent teeth (sector classification),[7] have been studied as predictors of both the duration and outcomes of orthodontic and interceptive treatment for impacted canines.[8,9] However, the introduction of cone-beam computed tomography (CBCT) has greatly improved the diagnostic accuracy of impacted maxillary canines by providing three-dimensional (3D) visualization, that enables a more comprehensive assessment of impaction severity, buccopalatal position, depth, surrounding bone coverage, and root resorption of adjacent teeth.[1,10-12] The KPG index, a grid-based scoring system that assesses cusp and root tip positions along the x, y, and z axes, quantifies the difficulty of impaction, thereby supporting clinicians in formulating effective treatment strategies.[13]

Although CBCT has been shown to enhance diagnostic confidence[14] and influence treatment decisions regarding impacted canines,[15] other studies have reported no significant differences between two-dimensional (2D) and 3D modalities.[4] This conflicting evidence complicates the benefit-to-risk assessment during treatment planning, since concerns regarding higher radiation exposure due to the use of CBCT persist.[1]

Previous research on impacted maxillary canines has primarily focused on how imaging modality affects localization accuracy, diagnostic confidence, pre-surgical planning, and duration of treatment.[1,16-19] Although a recent study derived panoramic radiograph-based cut-off values for predicting orthodontic treatment success or failure after active traction had been attempted,[20] to the best of our knowledge, no previous study has evaluated panoramic radiograph parameters and the CBCT-derived KPG index within the same quantitative diagnostic-accuracy framework in relation to the treatment decision of orthodontic eruption versus surgical extraction. Consequently, evidence-based 2D screening thresholds specifically designed to predict the need for surgical extraction remain unavailable, leaving a critical gap in the ability to objectively justify CBCT use under the As Low As Reasonably Achievable (ALARA) principle. Furthermore, most previous studies have relied solely on orthodontic examiners,[16,21] potentially overlooking the interdisciplinary nature of these cases where oral surgeons play a key diagnostic role. Therefore, the aim of this study was to evaluate the diagnostic concordance between panoramic radiographs and the CBCT-derived parameters, assess inter-rater reliability between an orthodontist and an oral surgeon, and determine the comparative diagnostic performance of parameters derived from panoramic radiographs and CBCT for treatment decision-making, thereby providing quantitative evidence to guide the selective use of CBCT.

METHODS

This retrospective study was approved by the İzmir Tınaztepe University Non-Interventional Clinical Trials Ethics Committee (approval number: MUKCE2025/05, date: 20.02.2025) and was conducted in accordance with the ethical standards of the Declaration of Helsinki. Written informed consent was obtained from patients who agreed to the use of their radiographs in this study. Sample size calculation (version 3.1.9.4, G*Power; Kiel University, Kiel, Germany) was performed using a significance level of 5% and a statistical power of 80% to detect a Pearson correlation coefficient of r=0.4.[16] Based on this calculation, a minimum of 46 impacted maxillary canines was required. Radiographic records (panoramic radiographs and CBCT) of 44 patients with maxillary canine impactions (37 unilateral and 7 bilateral) were selected from the hospital archive, starting with the most recent and continuing backward until the required sample size was achieved.

All radiographic images were obtained from the Oral and Maxillofacial Radiology Department of İzmir Tınaztepe University. Panoramic radiographs were acquired using the MyRay Hyperion X5 system (Cefla, Imola, Italy) with exposure parameters of 76 kV, 7 mA, and an exposure time of 14.2 seconds. Cone-beam computed tomography images were obtained using the MyRay Hyperion X9 Pro 3D system (Cefla, Imola, Italy) at 90 kV, 8 mA, and an exposure time of 26.4 seconds. Inclusion was restricted to radiographs characterized by the clear delineation of key anatomical landmarks, including the cusp tip and apex of the impacted canine and the roots of adjacent teeth and by the absence of artifacts or positioning errors. Radiographs exhibiting anatomical superimposition or technical flaws that compromised landmark identification were excluded. Furthermore, patients presenting with cleft lip or palate, craniofacial syndromes, metabolic bone diseases, dental anomalies (e.g., agenesis, peg-shaped laterals, supernumerary teeth), or prior or ongoing orthodontic treatment were excluded.

The severity of maxillary canine impaction on panoramic radiographs was assessed using three parameters: the mesiodistal position of the cusp tip relative to adjacent teeth (sector classification), the angle between the long axis of the impacted canine and the midline constructed from the anterior nasal spine to the intermaxillary suture (α-angle), and the vertical distance from the cusp tip to the occlusal plane (d distance) (Figure 1).[7] Higher sector classifications, α angles, and d distances were associated with greater severity of impaction. Additionally, the complexity of the impacted canines was categorized as low or high based on the criteria described by Arriola-Guillén et al.[22] Canines located in sectors 1, 2, or 3 with an α angle ≤40° were considered low-complexity, whereas those in sector 3 with an α angle >40°, or in sectors 4 or 5, were considered high-complexity.

The KPG index was used to determine the severity of maxillary canine impaction on CBCT scans.[13] This method involves a grid-based scoring system applied to three axes (x, y, and z) to evaluate the severity of impaction. In each plane, the positions of the cusp tip and root tip of the impacted canine were assigned scores ranging from 0 to 5, and the total KPG score was calculated by summing the cusp and root tip scores across all three planes (Figure 2). Based on the final score, the level of treatment difficulty was categorized as easy (0-9), moderate (10-14), difficult (15-19), or extremely difficult (≥20). While the classification by Arriola-Guillén et al.[22] utilizes the term "complexity" and the KPG index[13] utilizes "treatment difficulty," both systems were interpreted in this study as identifying the severity of impaction and the associated orthodontic challenge. The buccopalatal position of the impacted canine within the dental arch (buccal, mid-alveolus, or palatal) was determined using the occlusal reference line along the z-axis of the KPG index.

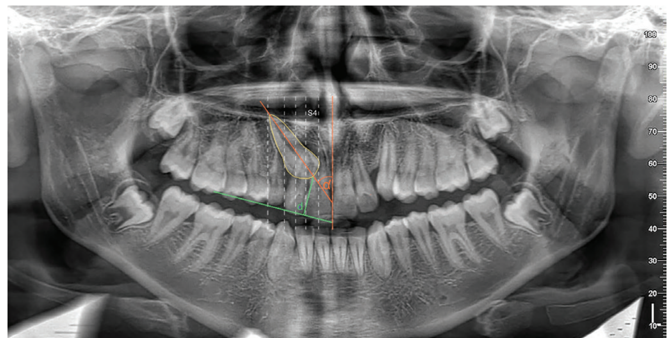


Figure 1. Parameters used during evaluation with panoramic radiographs. White lines: sector classification (S4) of the cusp tip of the impacted maxillary canine; orange lines: α angle between the long axis of the impacted canine and the midline; green lines: d distance from the cusp tip to the occlusal plane.

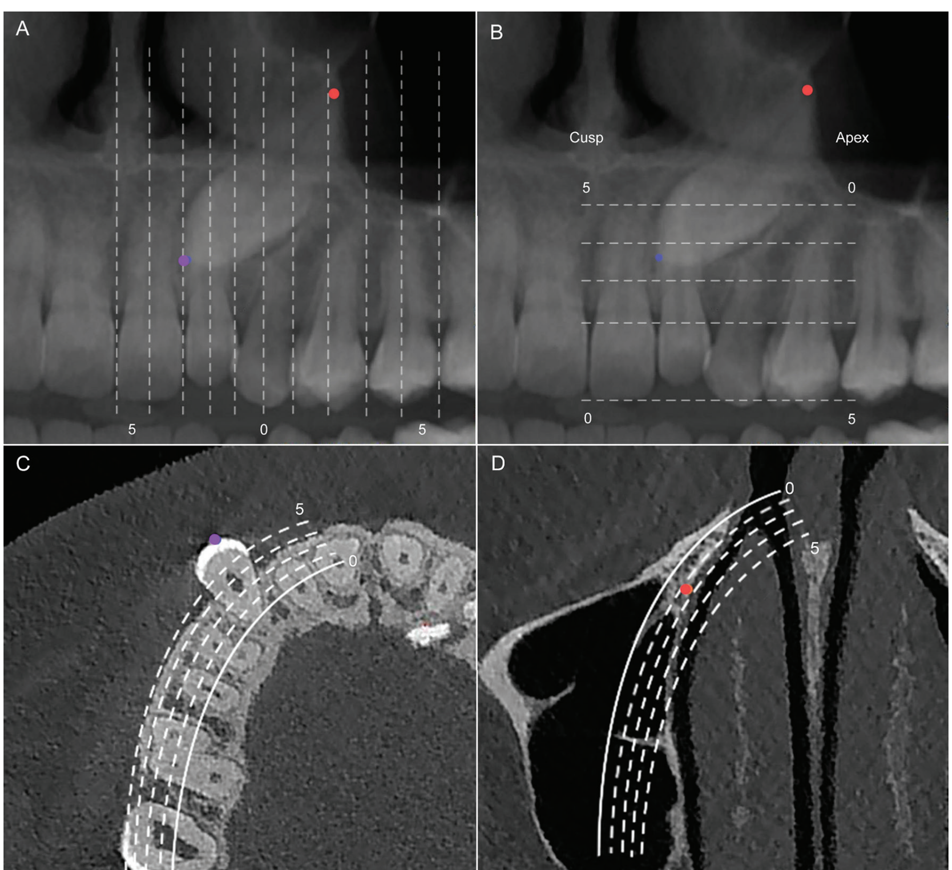


Figure 2. KPG index. The purple dot represents the cusp tip, and the orange dot represents the apex of the impacted maxillary canine. A) Mesiodistal dimension (x-axis); B) Vertical dimension (y-axis); C) Deviation of the cusp tip from the occlusal arch (z-axis); and D) Deviation of the apex from the occlusal arch (z-axis).

In addition to the radiographic evaluation, patient records were retrospectively reviewed to classify cases into two groups based on the final treatment decision: orthodontic eruption or surgical extraction of the impacted canines. Receiver operating characteristic (ROC) curve analysis was subsequently performed to evaluate the discriminatory ability of parameters derived from panoramic radiographs and CBCT in differentiating between these treatment groups. The area under the curve (AUC) was calculated to assess overall discriminatory ability, and optimal cut-off points were determined by identifying the point on the ROC curve that maximized the sum of sensitivity and specificity.[23] Area under the curve values were interpreted as follows: ≤0.5 indicated no discrimination; 0.5-0.7, poor discrimination; 0.7-0.8, acceptable discrimination; 0.8-0.9, excellent discrimination; and >0.9, outstanding discrimination.[24]

All measurements on panoramic radiographs and relevant CBCT slices were performed using ImageJ software (version 1.54, Java 1.8.0_345, 64-bit; U.S. National Institutes of Health, Bethesda, MD, USA). Prior to the study, an orthodontist (G.A.S.) and an oral surgeon (M.B.E.) calibrated their measurement techniques using 10 randomly selected records. Both examiners then independently evaluated the full dataset under blinded conditions to determine interrater reliability. Intrarater reliability was assessed by the orthodontist, who repeated the measurements one month later. Only the orthodontist's initial measurements were used for subsequent statistical analyses.

Statistical Analysis

Descriptive statistics (mean, standard deviation, median, minimum, and maximum) were used to summarize all study variables. Intra- and interrater reliabilities were assessed using intraclass correlation coefficients (ICC) for continuous variables and weighted kappa or Cohen's kappa for categorical variables.

Because 7 of the 44 patients contributed bilateral impacted canines, generalized estimating equations (GEE) analyses were used to account for the clustering of canines within patients. Generalized estimating equations was applied to compare α-angle and d distance across KPG index categories and to identify associations between radiographic parameters (complexity group, α angle, d distance, KPG score) and treatment decision. Pairwise comparisons between KPG index categories were performed with Bonferroni adjustment for multiple comparisons. For the associations between complexity group and KPG index category and between buccopalatal position and sector classification, Fisher's exact tests with Monte Carlo simulation were used because GEE modelling for these outcomes was not feasible due to quasi-complete separation in the former (all easy cases were low-complexity and all extremely difficult cases were high-complexity) and sparse cell counts in the latter. Receiver operating characteristic curve analysis was performed to evaluate the discriminatory ability of each parameter to differentiate canines managed with orthodontic eruption from those managed with surgical extraction. Pairwise comparisons of AUC values were performed using the DeLong test.

All statistical tests were two-sided, and statistical significance was set at $p<0.05$. All analyses were performed using SPSS software (version 26.0; IBM Corp., Armonk, NY, USA), with the exception of the AUC confidence intervals and pairwise DeLong comparisons, which were calculated in R (version 4.6.0; R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Descriptive statistics for the panoramic radiographic and CBCT-derived parameters evaluated in this study, are presented in Table 1. Continuous parameters demonstrated almost perfect inter- and intrarater reliability (ICC ≥0.927), while categorical parameters exhibited substantial to almost perfect agreement (κ≥0.658) ($p<0.001$) (Table 2).

Table 1. Descriptive statistics of qualitative and quantitative variables

Qualitative variables	
Gender	
Female, n (%)	30 (68.2)
Male, n (%)	14 (31.8)
Sector	
S1, n (%)	2 (3.9)
S2, n (%)	9 (17.6)
S3, n (%)	8 (15.7)
S4, n (%)	18 (35.3)
S5, n (%)	14 (27.5)
Complexity group	
Low-complexity, n (%)	17 (33.3)
High-complexity, n (%)	34 (66.7)
Buccopalatal position	
Buccal, n (%)	13 (25.5)
Mid-alveolus, n (%)	10 (19.6)
Palatal, n (%)	28 (54.9)
Treatment decision	
Orthodontic eruption, n (%)	35 (68.6)
Surgical extraction, n (%)	16 (31.4)
Quantitative variables	
Age (years)	
Mean ± SD	20.13±10.17
Median	16.66
Min-max	10.51-59.62
α angle (°)	
Mean ± SD	39.15±18.11
Median	38.29
Min-max	2.85-83.61

Table 1. Continued

Qualitative variables	
d distance (mm)	
Mean ± SD	12.48±4.66
Median	11.95
Min-max	1.20-24.54
KPG score	
Mean ± SD	14.12±2.65
Median	14.00
Min-max	7-20

SD, standard deviation; Min-max, minimum-maximum.

The distributions of complexity group, α angle, and d distance across the four KPG index categories are presented in Table 3. A significant association was observed between the KPG index and the panoramic radiographic complexity group (p=0.026). All canines in the easy category were classified as low-complexity, whereas all canines in the extremely difficult category were classified as high-complexity. Generalized estimating equations analyses revealed significant differences in both the α angle and the d distance across the KPG index categories (p<0.001). The mean α angle increased progressively from 11.92±12.20° in the easy category to 67.41±10.59° in the extremely difficult category (p<0.001). Similarly, the mean d distance increased from 10.12±7.05 mm to 22.71±2.58 mm across the same range (p<0.001).

Table 2. Intrarater and interrater reliability of parameters derived from panoramic radiographs and CBCT

	Reliability coefficient	95% CI	p-value
		Lower; upper	
Panoramic radiograph parameters			
Sector			
Intrarater	0.985‡	0.955; 1.000	<0.001*
Interrater	0.940‡	0.881; 0.998	<0.001*
α angle			
Intrarater	0.995†	0.991; 0.997	<0.001*
Interrater	0.995†	0.991; 0.997	<0.001*
d distance			
Intrarater	0.927†	0.874; 0.958	<0.001*
Interrater	0.931†	0.881; 0.960	<0.001*
CBCT			
x-axis: cusp			
Intrarater	0.987‡	0.963; 1.000	<0.001*
Interrater	0.949‡	0.900; 0.999	<0.001*
x-axis: apex			
Intrarater	0.981‡	0.945; 1.000	<0.001*
Interrater	0.705‡	0.580; 0.830	<0.001*
y-axis: cusp			
Intrarater	1.000‡	1.000; 1.000	<0.001*
Interrater	0.935‡	0.863; 1.000	<0.001*
y-axis: apex			
Intrarater	1.000‡	1.000; 1.000	<0.001*
Interrater	0.658‡	0.634; 1.000	<0.001*
z-axis: cusp			
Intrarater	0.941‡	0.884; 0.999	<0.001*
Interrater	0.886‡	0.812; 0.961	<0.001*
z-axis: apex			
Intrarater	0.975‡	0.940; 1.000	<0.001*
Interrater	0.863‡	0.779; 0.947	<0.001*
KPG score			
Intrarater	0.911‡	0.834; 0.987	<0.001*
Interrater	0.773‡	0.693; 0.852	<0.001*

Table 2. Continued

	Reliability coefficient	95% CI Lower; upper	p-value
Buccopalatal position			
Intrarater	1.000[ǂ]	1.000; 1.000	<0.001*
Interrater	1.000[ǂ]	1.000; 1.000	<0.001*

*Significant at p<0.001.
[‡]Weighted Kappa: used for ordinal parameters (sector; x, y and z axes; KPG score); [†]ICC: used for continuous parameters (d distance and α angle); [ǂ]Cohen's Kappa: used for nominal parameters (buccopalatal position).
CI, confidence interval; CBCT, cone-beam computed tomography; ICC, intraclass correlation coefficient.

Table 3. Comparison of complexity group, α angle, and d distance according to the KPG index

Panoramic radiograph parameters	KPG index				p-value
	Easy (n=3)	Moderate (n=26)	Difficult (n=20)	Extremely difficult (n=2)	
Complexity group, n (%)[†]					0.026*
Low-complexity	3 (100.0)	10 (38.5)	4 (20.0)	0 (0.0)	
High-complexity	0 (0.0)	16 (61.5)	16 (80.0)	2 (100.0)	
α angle (°)[‡]					<0.001***
Mean ± SD	11.92±12.20	35.30±14.32	45.42±17.58	67.41±10.59	
Median	7.12	34.96	41.73	67.41	
Min-max	6.79-25.79	9.83-57.51	18.91-83.61	59.92-74.91	
d distance (mm)[‡]					<0.001***
Mean ± SD	10.12±7.05	11.09±4.02	13.63±3.80	22.71±2.58	
Median	8.46	10.26	13.32	22.71	
Min-max	4.04-17.86	6.78-19.44	7.04-24.42	20.88-24.54	

*Significant at p<0.05.***Significant at p<0.001.
[†]Fisher's exact test ($FET_{complexity}$=7.590); [‡]Generalized estimating equations analysis. p-values represent the overall test of differences across the four KPG index categories.
SD, standard deviation; Min-max, minimum-maximum.

Pairwise comparisons of the α angle and the d distance among KPG index categories are presented in Table 4. For the α angle, significant differences were observed in all comparisons involving the easy or extremely difficult categories. The easy category showed significantly lower α angle values than the moderate (mean difference =-23.76°, p=0.026), difficult (-33.79°, p<0.001), and extremely difficult (-55.45°, p<0.001) categories, while the extremely difficult category showed significantly higher values than the Moderate (31.69°, p<0.001) and difficult (21.66°, p=0.005) categories. The moderate and difficult categories did not differ significantly (mean difference =-10.03°, p>0.05). For the d distance, the extremely difficult category showed significantly greater values than all other KPG categories (p≤0.004), and the difficult category showed significantly greater values than the moderate category (mean difference =-3.08 mm, p=0.047). The easy category was not statistically distinguishable from the moderate or difficult categories (p>0.05), which likely reflects the limited sample size of the easy category (n=3).

A significant association was also found between sector classification and buccopalatal position (p<0.001). Buccal impactions were predominantly located in sectors 2 and 3; palatal impactions were predominantly located in sectors 4 and 5; mid-alveolus impactions were more widely distributed across the sectors.

Of the 51 canines, 35 (68.6%) were managed by orthodontic eruption and 16 (31.4%) underwent surgical extraction. All four radiographic parameters were found to be significantly associated with the treatment decision (p<0.05) (Table 5). High-complexity canines had greater odds of surgical extraction compared to low-complexity canines (OR=6.68; 95% CI, 1.10-40.55; p=0.039), although the wide confidence interval likely reflects the small number of extraction events in the low-complexity category (n=2 of 17) and should therefore be interpreted with caution. Higher α angles, greater d distances, and higher KPG scores were also significantly associated with increased odds of surgical extraction. Each 1° increase in the α angle was associated with a 9% increase in the odds of extraction [odds ratio (OR)=1.09, 95% CI 1.04-1.14, p=0.001], each 1-mm increase in the d distance was associated with a 43% increase (OR=1.43, 95% CI 1.18-1.73, p<0.001), and each 1-point increase in the KPG score was associated with a 71% increase (OR=1.71, 95% CI 1.21-2.41, p=0.002).

The ROC curve analysis demonstrated acceptable-to-excellent discrimination by all four parameters between canines

Table 4. Pairwise comparisons of α angle and d distance across KPG index categories

Pairwise comparison	α angle (°)			d distance (mm)		
	Mean difference	95% CI lower; upper	p-value	Mean difference	95% CI lower; upper	p-value
Easy vs. moderate	-23.76	-45.69; -1.83	0.026*	-1.37	-11.07; 8.32	1.000
Easy vs. difficult	-33.79	-55.41; -12.16	<0.001*	-4.45	-14.07; 5.16	1.000
Easy vs. extremely difficult	-55.45	-81.15; -29.75	<0.001*	-12.94	-22.94; -2.93	0.004*
Moderate vs. difficult	-10.03	-22.99; 2.94	0.248	-3.08	-6.14; -0.02	0.047*
Moderate vs. extremely difficult	-31.69	-47.34; -16.04	<0.001*	-11.57	-15.46; -7.67	<0.001*
Difficult vs. extremely difficult	-21.66	-38.89; -4.43	0.005*	-8.48	-12.65; -4.31	<0.001*

*Significant at Bonferroni-adjusted p<0.05. Generalized estimating equations analysis.
CI, confidence interval.

Table 5. Generalized estimating equations analysis of complexity group, α angle, d distance and KPG score in relation to treatment decision

	Orthodontic eruption	Surgical extraction	OR	95% CI lower; upper	p-value
Complexity group, n (%)†			6.68	1.10; 40.55	0.039*
Low-complexity	15 (88.2)	2 (11.8)			
High-complexity	20 (58.8)	14 (41.2)			
α angle (°)‡			1.09	1.04; 1.14	0.001***
Mean ± SD	32.96±15.94	52.70±15.26			
d distance (mm)‡			1.43	1.18; 1.73	<0.001***
Mean ± SD	10.78±3.85	16.21±4.15			
KPG score‡			1.71	1.21; 2.41	0.002**
Mean ± SD	13.29±2.51	16.00±2.12			

*Significant at p<0.05; **Significant at p<0.01; ***Significant at p≤0.001.
†For complexity group, low-complexity is the reference category.
‡For continuous predictors, the OR represents the change in odds of extraction per one-unit increase (per 1°, 1 mm, and 1 point, respectively).
OR, odds ratio; CI, confidence interval.

managed with orthodontic eruption and those managed with surgical extraction (Table 6, Figure 3). The d distance had the highest AUC (0.843; 95% CI: 0.721-0.939), followed by the α angle (0.813; 95% CI: 0.677-0.921), the KPG score (0.796; 95% CI: 0.661-0.910), and the sector classification (0.715; 95% CI: 0.559-0.847). Pairwise comparisons revealed no significant differences among the AUC values (p>0.05), indicating comparable discriminatory performance across the evaluated parameters. The optimal cut-off values were ≥4.5 for sector (sensitivity 50.0%, specificity 82.9%), ≥46.1° for α angle (sensitivity 68.8%, specificity 82.9%), ≥12.3 mm for d distance (sensitivity 87.5%, specificity 74.3%), and ≥15.5 for the KPG score (sensitivity 62.5%, specificity 85.7%).

DISCUSSION

To the best of our knowledge, this is the first study to directly compare panoramic radiograph parameters with the CBCT-based KPG index within a single diagnostic-accuracy framework regarding the clinical decision between orthodontic eruption and surgical extraction of impacted maxillary canines. The main finding was that panoramic radiographic parameters showed diagnostic concordance with the KPG index and provided comparable discriminatory ability in guiding this treatment decision. Consistent with previous studies,[7,13] our results suggest that radiographic indicators of impaction severity derived from both panoramic radiographs and CBCT may effectively inform treatment planning.

The management of impacted maxillary canines presents a complex diagnostic and therapeutic challenge for orthodontists. Although panoramic radiographs are the standard for initial radiographic examination, they are inherently limited by image distortion, magnification errors, structural superimposition, and lack of 3D information.[21] In contrast, CBCT offers superior anatomical detail and accuracy, significantly enhancing the localization of impacted canines and the evaluation of adjacent structures, thereby improving diagnostic reliability and treatment planning.[15] However, routine use of CBCT remains debated because its higher radiation dose and increased exposure raise concerns about long-term cancer risk, particularly in radiosensitive groups such as children and female patients.[25] Therefore, while CBCT provides significant clinical advantages, its application should be justified through evidence-based practice, and adherence to the ALARA principle is essential to ensure an optimal balance between diagnostic benefit and patient safety.[1,25]

Several previous studies have compared the diagnostic accuracy of panoramic radiography and CBCT to minimize radiation exposure. These studies suggest that CBCT consistently offers superior diagnostic accuracy for the localization of impacted maxillary canines[10,11,14,19] and highlight the limitations of 2D imaging, particularly in assessing the vertical position of the crown. Salari et al.[18] and Botticelli et al.[12] reported that conventional radiographs often overestimate the vertical position of the cusp tip, whereas CBCT provides a more accurate assessment. In the present study, although the d distance and KPG index represent fundamentally different measurement systems, the d distance increased progressively across the KPG index categories. This finding suggests that despite methodological differences, vertical measurements derived from 2D imaging may provide clinically meaningful insights into the severity of impaction.

The ROC curve analysis was used to evaluate the discriminatory ability of each radiographic parameter in the surgical extraction decision, given its recognition as the method of choice for quantifying the accuracy and discriminative performance of medical diagnostic tests.[26,27] All four evaluated parameters demonstrated good discriminatory ability, with panoramic radiographic and CBCT-based measurements providing comparable diagnostic information for the treatment decision. Among panoramic radiographic variables, an α angle exceeding 46.1° and a d distance of 12.3 mm or greater were identified as radiographic indicators associated with the clinical decision to extract impacted maxillary canines. The angulation threshold is consistent with the findings of Güllü and Çakmak Özlü,[20] who identified a cut-off value of 44.3° for treatment failure, suggesting that angulations above this threshold may significantly hinder eruption. In contrast, the lower threshold of 34.8° reported by Alqerban et al.[27] likely reflects the focus on early prediction of impaction rather than the surgical extraction decision evaluated in the present study. Regarding vertical position, our identified cut-off of 12.3 mm was notably lower than the 22.3 mm threshold reported by Güllü and Çakmak Özlü,[20] likely reflecting differences in clinical endpoints, as in their study the threshold was defined for the failure of active treatment, whereas the present study established a threshold for the initial decision to extract. For the KPG index, the ROC analysis identified a score of 16 or higher as the threshold associated with the decision to perform surgical extraction,

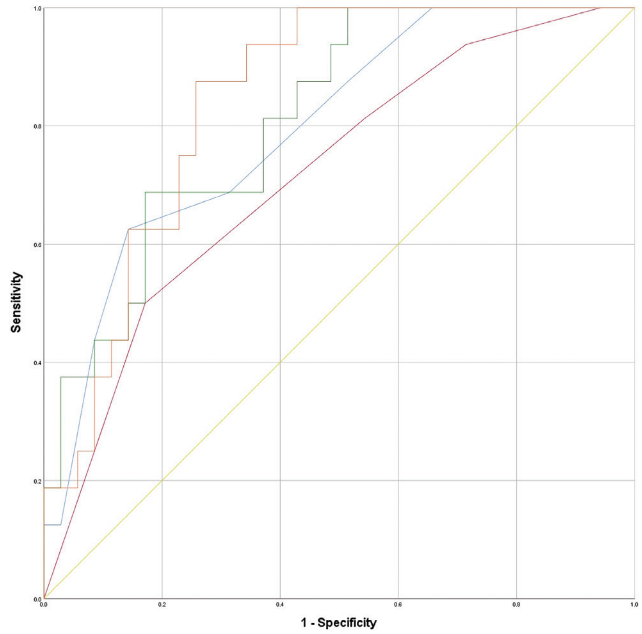


Figure 3. ROC curve of treatment decision. Blue line: KPG score; red line: sector; green line: α angle; orange line: d distance; yellow line: reference.

ROC, receiver operating characteristic.

which closely aligns with the transition from "moderate" (10-14) to "difficult" (15-19) complexity in the original classification. Taken together, the comparable diagnostic performance of panoramic radiographs and CBCT-based parameters, combined with the substantially lower radiation exposure associated with panoramic radiographs, suggests that 2D parameters available on initial panoramic radiographs may serve as a useful first-line screening tool to identify complex cases that require the superior anatomical detail of CBCT for precise surgical planning.

Consistent with previous research,[17,21] ta significant association was also observed between sector classification on panoramic radiographs and the buccopalatal position of impacted maxillary canines as determined by CBCT. Buccally impactions were predominantly observed in sectors 2 and 3, whereas palatal impactions were most frequently located in sectors 4 and 5. This distribution pattern suggests that sector classification on panoramic radiographs may offer predictive value for estimating

Table 6. AUC of parameters derived from panoramic radiographs and CBCT

Variable	AUC 95% CI (lower; upper)	Cut-off value	Prediction criteria	Sensitivity (%)	Specificity (%)
Sector	0.715 0.559; 0.847	4.5	Sector 5 indicates extraction	50.0	82.9
α angle (°)	0.813 0.677; 0.921	46.1	≥46.1° indicates extraction	68.8	82.9
d distance (mm)	0.843 0.721; 0.939	12.3	≥12.3 mm indicates extraction	87.5	74.3
KPG Score	0.796 0.661; 0.910	15.5	≥16 indicates extraction	62.5	85.7

Pairwise comparisons of AUCs were performed using the DeLong test. No statistically significant differences were found between AUC values (all p>0.05).
CBCT, cone-beam computed tomography; AUC, area under the curve; CI, confidence interval.

the buccopalatal orientation of impacted canines. However, previous studies have also reported only slight to fair agreement between 2D and 3D imaging in determining the buccopalatal position, with 2D modalities often suggesting a more palatal location; this may lead to inaccurate diagnoses, particularly in cases of mid-alveolar impactions.[12]

High reliability for parameters derived from both panoramic radiographs and CBCT was observed in this study and is consistent with previous research reporting intrarater agreement for key diagnostic features, such as sector location, buccopalatal position, and root resorption, across both imaging modalities.[17,21] Haney et al.[15] reported comparable intrarater agreement but observed slightly better consistency with 2D radiographs, although this difference was not significant. In contrast, Dalessandri et al.[16] found higher intrarater reliability using the KPG index compared to 2D methods, suggesting greater reproducibility with CBCT-based evaluations. Regarding interrater reliability, the present study is consistent with the findings of Ngo et al.[21] and Jung et al.,[17] who reported high agreement for various diagnostic features of impacted canines on both panoramic radiographs and CBCT images. In contrast, Alqerban et al.[5] and Dalessandri et al.[16] reported a higher interrater agreement with CBCT than with conventional 2D imaging. Notably, unlike studies that relied solely on orthodontists,[21] the inclusion of different specialists in the present study supports the findings of Wriedt et al.,[4] who found no diagnostic differences between dental specialties.

From a clinical perspective, the findings of this study support a more selective imaging strategy for impacted maxillary canines. While the superior anatomical detail of CBCT is undeniable, its routine use must be carefully balanced against the ALARA principle. This radiation burden is particularly critical in complex cases in which multiple 3D imaging sessions may be required during treatment. To guide selective CBCT referral, the present results establish specific panoramic thresholds to help identify high-complexity cases in which CBCT is most likely to provide critical added value for surgical planning. Finally, including both an orthodontist and an oral surgeon as examiners reflects the interdisciplinary nature of managing these cases and addresses a methodological limitation of previous comparative studies that relied solely on orthodontic raters.

Study Limitations

Several limitations of this study should be considered when interpreting the results. First, this was a single-center retrospective study with a relatively limited sample size, which may affect the generalizability of the results. The limited number of surgical extraction events (n=16) also precluded a multivariable assessment of whether combining radiographic parameters would improve discriminatory performance beyond that of individual parameters. Second, although the present study evaluated the association between radiographic parameters and the implemented treatment modality (surgical extraction versus orthodontic eruption), it did not assess long-term clinical success measures, such as total treatment duration and the incidence of eruption failure necessitating secondary extraction. Another important limitation is that the treatment decision, which is multifactorial and may also be influenced by clinician experience and patient preferences rather than anatomical difficulty alone, was used as an indicator of impaction complexity in this study. Beyond this clinical issue, a methodological limitation is the use of the actual treatment decision as the outcome variable in ROC analysis. Because the clinicians who made the treatment decisions had access to the same radiographic information evaluated in this study, some degree of circularity is unavoidable. Therefore, the discriminatory performance of the radiographic parameters may have been overestimated, and the identified cut-off values should be interpreted as thresholds associated with observed treatment decisions rather than as independently validated predictors of treatment success, eruption failure, or absolute need for extraction. Finally, important clinical variables, including occlusal characteristics, arch length discrepancy, and available space for the impacted canine were not evaluated and therefore could not be incorporated into the analyses, although they may also have influenced treatment decisions.

CONCLUSION

Within the limitations of this retrospective study, parameters derived from panoramic radiographs were significantly associated with the CBCT-derived KPG index and demonstrated diagnostic performance comparable to that of CBCT-based assessment for treatment decisions regarding impacted maxillary canines. Panoramic radiography may serve as an effective, low-radiation, first-line screening tool: specific indicators, such as the d distance or the α angle, can justify the use of CBCT for precise anatomical assessment. This stepwise imaging strategy supports the ALARA principle, ensuring that 3D imaging is reserved for cases where it provides critical added value for surgical planning. Prospective studies with larger sample sizes are needed to validate these thresholds and establish evidence-based imaging guidelines.

Ethics

Ethics Committee Approval: This study was approved by the İzmir Tınaztepe University Non-Interventional Clinical Trials Ethics Committee (approval number: MUKCE2025/05, date: 20.02.2025).

Informed Consent: Written informed consent was obtained from patients who agreed to the use of their radiographs in this study.

Footnotes

Author Contributions: Surgical and Medical Practices - G.A.S.; Concept - M.B.E., M.G.K.; Design - G.A.S.; Data Collection and/or Processing - G.A.S., M.B.E.; Analysis and/or Interpretation - G.A.S.; Literature Search - M.B.E., M.G.K.; Writing - M.B.E., M.G.K.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. Eslami E, Barkhordar H, Abramovitch K, Kim J, Masoud MI. Cone-beam computed tomography vs conventional radiography in visualization of maxillary impacted-canine localization: a systematic review of comparative studies. *Am J Orthod Dentofacial Orthop.* 2017;151(2):248-258. [CrossRef]
2. Bedoya MM, Park JH. A review of the diagnosis and management of impacted maxillary canines. *J Am Dent Assoc.* 2009;140(12):1485-1493. [CrossRef]
3. Bishara SE. Impacted maxillary canines: a review. *Am J Orthod Dentofacial Orthop.* 1992;101(2):159-171. [CrossRef]
4. Wriedt S, Jaklin J, Al-Nawas B, Wehrbein H. Impacted upper canines: examination and treatment proposal based on 3D versus 2D diagnosis. *J Orofac Orthop.* 2012;73(1):28-40. [CrossRef]
5. Alqerban A, Jacobs R, Fieuws S, Willems G. Comparison of two cone beam computed tomographic systems versus panoramic imaging for localization of impacted maxillary canines and detection of root resorption. *Eur J Orthod.* 2011;33(1):93-102. [CrossRef]
6. Nieri M, Crescini A, Rotundo R, Baccetti T, Cortellini P, Pini Prato GP. Factors affecting the clinical approach to impacted maxillary canines: a Bayesian network analysis. *Am J Orthod Dentofacial Orthop*. 2010;137(6):755-762. [CrossRef]
7. Ericson S, Kurol J. Early treatment of palatally erupting maxillary canines by extraction of the primary canines. *Eur J Orthod.* 1988;10(4):283-295. [CrossRef]
8. Crescini A, Nieri M, Buti J, Baccetti T, Pini Prato GP. Pre-treatment radiographic features for the periodontal prognosis of treated impacted canines. *J Clin Periodontol.* 2007;34(7):581-587. [CrossRef]
9. Leonardi M, Armi P, Franchi L, Baccetti T. Two interceptive approaches to palatally displaced canines: a prospective longitudinal study. *Angle Orthod.* 2004;74(5):581-586. [CrossRef]
10. Alfailany DT, Shaweesh AI, Hajeer MY, Brad B, Alhaffar JB. The diagnostic accuracy of cone-beam computed tomography and two-dimensional imaging methods in the 3D localization and assessment of maxillary impacted canines compared to the gold standard in-vivo readings: a cross-sectional study. *Int Orthod.* 2023;21(3):100780. [CrossRef]
11. Hajeer MY, Al-Homsi HK, Alfailany DT, Murad RMT. Evaluation of the diagnostic accuracy of CBCT-based interpretations of maxillary impacted canines compared to those of conventional radiography: an in vitro study. *Int Orthod.* 2022;20(2):100639. [CrossRef]
12. Botticelli S, Verna C, Cattaneo PM, Heidmann J, Melsen B. Two-versus three-dimensional imaging in subjects with unerupted maxillary canines. *Eur J Orthod.* 2011;33(4):344-349. [CrossRef]
13. Kau CH, Pan P, Gallerano RL, English JD. A novel 3D classification system for canine impactions--the KPG index. *Int J Med Robot.* 2009;5(3):291-296. [CrossRef]
14. Alqerban A, Hedesiu M, Baciut M, et al. Pre-surgical treatment planning of maxillary canine impactions using panoramic vs cone beam CT imaging. *Dentomaxillofac Radiol.* 2013;42(9):20130157. [CrossRef]
15. Haney E, Gansky SA, Lee JS, et al. Comparative analysis of traditional radiographs and cone-beam computed tomography volumetric images in the diagnosis and treatment planning of maxillary impacted canines. *Am J Orthod Dentofacial Orthop.* 2010;137(5):590-597. [CrossRef]
16. Dalessandri D, Migliorati M, Visconti L, Contardo L, Kau CH, Martin C. KPG index versus OPG measurements: a comparison between 3D and 2D methods in predicting treatment duration and difficulty level for patients with impacted maxillary canines. *Biomed Res Int.* 2014;2014:537620. [CrossRef]
17. Jung YH, Liang H, Benson BW, Flint DJ, Cho BH. The assessment of impacted maxillary canine position with panoramic radiography and cone beam CT. *Dentomaxillofac Radiol.* 2012;41(5):356-360. [CrossRef]
18. Salari B, Tofangchiha M, Padisar P, Reda R, Zanza A, Testarelli L. Diagnostic accuracy of conventional orthodontic radiographic modalities and cone-beam computed tomography for localization of impacted maxillary canine teeth. *Sci Prog.* 2024;107(1):368504241228077. [CrossRef]
19. Tsolakis AI, Kalavritinos M, Bitsanis E, et al. Reliability of different radiographic methods for the localization of displaced maxillary canines. *Am J Orthod Dentofacial Orthop.* 2018;153(2):308-314. [CrossRef]
20. Güllü YÖ, Çakmak Özlü F. Prediction of the success of orthodontic treatment of impacted maxillary canines using panoramic radiography parameters: a retrospective cross-sectional study. *BMC Oral Health.* 2024;24(1):1547. [CrossRef]
21. Ngo CTT, Fishman LS, Rossouw PE, Wang H, Said O. Correlation between panoramic radiography and cone-beam computed tomography in assessing maxillary impacted canines. *Angle Orthod.* 2018;88(4):384-389. [CrossRef]
22. Arriola-Guillén LE, Aliaga-Del Castillo A, Ruíz-Mora GA, Rodríguez-Cárdenas YA, Dias-Da Silveira HL. Influence of maxillary canine impaction characteristics and factors associated with orthodontic treatment on the duration of active orthodontic traction. *Am J Orthod Dentofacial Orthop.* 2019;156(3):391-400. [CrossRef]
23. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44(3):837-845. [CrossRef]
24. Hosmer DW, Lemeshow S. Applied logistic regression. 2nd edition. New York: John Wiley & Sons; 2000. [CrossRef]
25. Jha N, Kim YJ, Lee Y, Lee JY, Lee WJ, Sung SJ. Projected lifetime cancer risk from cone-beam computed tomography for orthodontic treatment. *Korean J Orthod.* 2021;51(3):189-198. [CrossRef]
26. Tseng YC, Pan CY, Chou ST, et al. Treatment of adult Class III malocclusions with orthodontic therapy or orthognathic surgery: receiver operating characteristic analysis. *Am J Orthod Dentofacial Orthop.* 2011;139(5):e485-e493. [CrossRef]
27. Alqerban A, Storms AS, Voet M, Fieuws S, Willems G. Early prediction of maxillary canine impaction. *Dentomaxillofac Radiol.* 2016;45(3):20150232. [CrossRef]

DOI: 10.4274/TurkJOrthod.2026.2025.75

TURKISH JOURNAL OF ORTHODONTICS

Original Article

Analysis of Orthodontic Treatment Requirements and Their Influence on Quality of Life and Self-Esteem Among Adults

Aslıhan Zeynep Öz[1], Abdullah Alper Öz[2], Özge Özaydın[3]

[1]Ondokuz Mayıs University Faculty of Dentistry, Department of Orthodontics, Samsun, Türkiye
[2]İstanbul Galata University Faculty of Dentistry, Department of Orthodontics, İstanbul, Türkiye
[3]Private Practice, Clinic of Orthodontics, İstanbul, Türkiye

Cite this article as: Öz AZ, Öz AA, Özaydın Ö. Analysis of orthodontic treatment requirements and their influence on quality of life and self-esteem among adults. *Turk J Orthod.* 2026; 39(3): 136-141

Main Points

- Greater orthodontic treatment need was weakly associated with poorer oral health-related quality of life.
- Self-esteem did not differ significantly according to orthodontic treatment need.
- Adults seek orthodontic treatment mainly for psychological and aesthetic reasons.

ABSTRACT

Objective: To evaluate the the association of orthodontic treatment need on oral health-related quality of life (OHRQoL) and self-esteem in adult patients seeking orthodontic care.

Methods: A total of 110 adults were recruited for the study. Orthodontic treatment need was evaluated using the index of complexity, outcome and need (ICON). Patients with a score of 43 and above were classified as Group 1; those with a score below 43 were classified as Group 2. Oral health-related quality of life was evaluated using the Oral Health Impact Profile-14 (OHIP-14), and self-esteem was assessed using the rosenberg self-esteem scale (RSES). Statistical evaluation involved Spearman's rank correlation, Mann-Whitney U tests, and independent samples t-tests.

Results: The mean OHIP-14 score was higher in Group 1 (12.63±7.46) than in Group 2 (10.61±7.46). However, this difference was not statistically significant (p=0.176). Mean RSES scores did not differ significantly between the groups (p=0.741). A slight but meaningful positive correlation emerged between ICON and OHIP-14 scores, indicating that as orthodontic treatment need increased, OHRQoL tended to decrease. No significant correlation was observed between ICON and RSES. Additionally, gender was not significantly associated with either OHIP-14 or RSES scores.

Conclusion: The analysis revealed that increased need for orthodontic treatment is associated with a slight worsening of OHRQoL but does not significantly affect self-esteem in adult subjects.

Keywords: Adult orthodontics, quality of life, self-esteem

INTRODUCTION

The most significant consequences of malocclusion are its psychosocial impact on individuals and deterioration of dentofacial aesthetics. The psychological and social consequences of poor dentofacial aesthetics can have a more serious impact on quality of life than physiological issues do.[1] The orofacial region plays a central role in interpersonal

Corresponding author: Assoc. Prof. Aslıhan Zeynep Öz, **e-mail:** aslihanzeynepoz@gmail.com **ORCID:** orcid.org/0000-0001-9168-5743
Received: July 07, 2025 **Accepted:** July 08, 2026 **Publication Date:** September 25, 2026

communication, serving as a key medium for vocal, physical, and emotional expression. Individuals seeking orthodontic treatment are typically motivated more by a desire to enhance their appearance and social acceptance than by functional or oral health concerns.

Over the past decade, orthodontics has seen growing attention paid to these factors owing to a shift from a purely biomedical focus to a broader biopsychosocial approach. Enhancing quality of life has become an important reason for seeking orthodontic treatment. The World Health Organization defines quality of life as an individual's perception of their position in life within the context of their culture and value systems and in relation to their goals, expectations, standards, and concerns.[2] It is defined as "the feeling of well-being arising from satisfaction or dissatisfaction with areas of life".[3] Oral health-related quality of life (OHRQoL), a more specific concept, is defined either as "the standard of oral and related tissue health that enables a person to eat, speak and socialize without active illness, discomfort or embarrassment" or as "the absence of negative effects of oral condition on social life and a positive sense of dentofacial self-confidence".[4]

It has long been recognised that individuals with malocclusion tend to be socially withdrawn and that their self-esteem is related to their facial and dental appearance. Therefore, orthodontic treatment is expected to improve self-esteem and reduce social anxiety. The complexity arises from the observation that certain adults are troubled by slight dental irregularities, while others with more severe malocclusions remain unconcerned. According to Birkeland et al.,[1] some individuals remained dissatisfied with their dental appearance both before and after treatment, whereas others expressed satisfaction at both time points. Previous research has also shown that individuals with malocclusion may feel embarrassed due to dental irregularities, and their self-confidence regarding facial appearance may be affected.[5,6] Among the expected psychosocial benefits of orthodontic treatment are increased self-esteem and reduced social anxiety.[1,5] Self-esteem is individuals' judgment about themselves and the attitudes they develop as a result of considering and evaluating the self. It is an individual's self-perception and self-assessment, as well as self-respect, self-confidence, self-acceptance, and self-approval. There is still a lack of definitive evidence that orthodontic treatment improves self-esteem.[7-9] However, there is evidence that personal self-esteem can be affected by situations or events such as illness.[10,11]

Oral health has long been recognized as a key determinant of quality of life, affecting individuals physically, socially, and psychologically.[12,13] Clinical assessment of malocclusion and orthodontic treatment need has traditionally relied on standardized occlusal indices. Instruments such as the index of orthodontic treatment need, dental aesthetic index, index of complexity, outcome and need (ICON), and Peer Assessment Rating provide standardized approaches for evaluating different aspects of malocclusion, treatment need, complexity, and treatment outcome.[14] In contrast, orthodontic treatment need indices have traditionally emphasized the presence of anatomical deviations rather than the broader consequences of such deviations on daily functioning and well-being. With the growing emphasis on patient-oriented perspectives, there is increasing interest in examining whether measures of malocclusion, as determined by occlusal indices, are associated with impaired OHRQoL.

Despite the widespread use of clinical indices to assess malocclusion severity, limited evidence exists regarding how that severity differentially affects OHRQoL and self-perception in adults. In particular, the relative influence of malocclusion severity on OHRQoL, compared with its influence on self-esteem, has not been clearly established.

Therefore, this study aimed to evaluate the associations of orthodontic treatment need, assessed using the ICON, with OHRQoL and self-esteem in adults seeking orthodontic treatment. The null hypothesis was that orthodontic treatment need would not be significantly associated with either OHRQoL or self-esteem.

METHODS

The Ondokuz Mayıs University Clinical Research Ethics Committee reviewed and approved this cross-sectional observational study (approval no: OMÜ KAEK 2025/05, date: 15.01.2025). Sample size analysis was performed using G*Power 3.1 software, based on a t-test to compare two independent group means. An effect size of d=0.50 was estimated from preliminary data, with a type I error rate of 5% (α=0.05), statistical power of 80%, and an allocation ratio (N2/N1) of 0.72. Based on these parameters, the minimum required sample size was calculated as approximately 104 participants.

A total of 110 patients aged 18 years and over who presented to the orthodontic clinic seeking treatment and did not require orthognathic surgery were included in the study. The mean patient age was 21.1 years. Written informed consent was obtained from all patients included in the study. The ICON index was used to assess the need for orthodontic treatment. In accordance with the original validation study by Daniels and Richmond,[15] a cut-off value of 43 was applied to distinguish cases requiring treatment. Patients with a score of 43 or above were classified as Group 1 (n=64); patients with a score below 43 were classified as Group 2 (n=46). All ICON assessments were performed by a single orthodontist with 15 years of clinical experience who was blinded to the participants' group allocation and questionnaire outcomes. To ensure reliability, 20 randomly selected cases were reassessed, and intra-examiner agreement was determined using the intraclass correlation coefficient (ICC), indicating high consistency (ICC=0.92).

The abbreviated version of the Oral Health Impact Profile-14 (OHIP-14) was used to assess OHRQoL. The total score ranges from 0 to 56, where higher scores indicate poorer oral health and diminished quality of life. The rosenberg self-esteem scale

Turk J Orthod 2026; 39(3): 136-141

(RSES) was used to assess self-esteem. The scale consists of 10 items assessing global self-esteem and can be interpreted as comprising two subdimensions: positive self-esteem (5 items) and negative self-esteem (5 items), with total scores ranging from 0 to 30. High scores indicate high self-esteem. The OHIP-14 and RSES questionnaires, originally developed by respective authors,[16,17] have previously been tested for reliability and validity in Turkish populations,[18,19] demonstrating high internal consistency and reproducibility. The OHIP-14 and RSES questionnaires were administered directly to all participants during their initial clinical visit. All participants completed the questionnaires at a single time point under standardized conditions in the clinical setting, without external influence.

Statistical Analysis

Statistical analyses were performed using the SPSS, version 22.0 (IBM Corp., Armonk, NY, USA). The normality of continuous variables was assessed with the Kolmogorov-Smirnov and Shapiro-Wilk tests. An independent samples t-test was conducted to compare the means of two groups for variables that were normally distributed. For non-normally distributed variables, the Mann-Whitney U test was used for group comparisons. Associations between continuous variables that did not meet the normality assumption was evaluated using the Spearman rank correlation test. Statistical significance was set at $p<0.05$.

RESULTS

Table 1 displays descriptive statistics (mean and standard deviation) of ICON values for Groups 1 and 2, stratified by gender, providing an overview of the distribution and variability of orthodontic treatment need across the two study groups.

Table 2 compares the OHIP-14 scores between the two groups of adult subjects based on their orthodontic treatment need, classified by ICON scores (Group 1: ICON ≥43, Group 2: ICON <43). The mean OHIP score was higher in Group 1 (12.63±7.46) than in Group 2 (10.61±7.46). However, this difference did not reach statistical significance (p=0.176).

Table 3 compares self-esteem scores between the two groups. The mean RSES scores were similar in both groups (Group 1: 13.60±1.99, Group 2: 13.63±2.11), and the difference was not statistically significant (p=0.741).

Table 4 presents the Spearman correlation coefficients for the associations among OHIP, RSES, ICON, and gender. A weak but statistically significant positive correlation was found between ICON and OHIP scores (r=0.231, p=0.015). No significant correlation was detected between ICON and RSES (r=-0.084, p=0.380). Additionally, gender showed no significant association with either OHIP (r=-0.017, p=0.862) or RSES (r=0.111, p=0.247) scores.

DISCUSSION

Traditionally, clinical assessments have been prioritized over patient-reported outcomes such as perceived functional capacity and psychological well-being. However, evidence consistently shows discrepancies between patients and clinicians in the evaluation of oral health and its consequences. Increasing recognition of these differences has led to a paradigm shift, emphasizing the integration of patients' perspectives-including expectations, satisfaction, and quality of life-into both research and clinical practice. The motivation of patients seeking orthodontic treatment is largely influenced

Table 1. Descriptive statistics of ICON values between groups and gender

	ICON			
	Male		Female	
	Mean (n)	SD	Mean (n)	SD
Group 1	56.6 (21)	10.87	56.11 (43)	8.39
Group 2	32.9 (12)	8.98	34.41 (34)	7.42

ICON, index of complexity, outcome and need; SD, standard deviation.

Table 2. Oral health-related quality of life measures of groups

OHIP	N	Mean	SD	Min	Max	p
Group 1 (ICON ≥43)	64	12.63	7.46	3	33	0.176
Group 2 (ICON <43)	46	10.61	7.46	1	47	

p<0.05.
ICON, index of complexity, outcome and need; SD, standard deviation; Min, minimum; Max, maximum.

Table 3. Rosenberg self-esteem measurements of groups (RSES)

RSES	N	Mean	SD	Min	Max	p
Group 1 (ICON ≥43)	64	13.60	1.99	9	18	0.741
Group 2 (ICON <43)	46	13.63	2.11	9	18	

p<0.05.
ICON, index of complexity, outcome and need; SD, standard deviation; Min, minimum; Max, maximum.

Table 4. Spearman correlations between OHIP and RSES with ICON and gender

		OHIP	RSES
ICON	Correlation coefficient	0.231*	-0.084
	p	0.015	0.380
	N	110	110
Gender	Correlation coefficient	-0.017	0.111
	p	0.862	0.247
	N	110	110

*Correlation is significant at the 0.05 level.
OHIP, Oral Health Impact Profile; ICON, index of complexity, outcome and need.

by aesthetic expectations rather than by objective clinical indications. This highlights the significant role of self-perceived dental appearance and social pressures in increasing demand for orthodontic procedures. The proportion of patients who applied for orthodontic treatment for aesthetic reasons was 80%, while the proportion who applied for functional reasons, such as chewing and speaking was as low as 20%.[20]

Numerous indices have been developed to assess orthodontic treatment need, each emphasizing different aspects of malocclusion, including dental aesthetics, occlusal characteristics, treatment need, and complexity. In the present study, ICON was selected because previous evidence demonstrates that ICON is more closely associated with variations in OHRQoL than other indices.[21] Specifically, ICON has been shown to better capture the relationship between occlusal traits and individuals' perceptions of the impact of oral health on quality of life, making it a more suitable measure for the objectives of this study.[21] The present study examined the association between orthodontic treatment need and OHRQoL while considering self-esteem, using the OHIP-14 and RSES. Furthermore, the use of the ICON index in an adult population allowed for a more nuanced assessment of psychosocial impacts, thereby contributing novel evidence to the existing literature.

Quality of life is a multidimensional concept encompassing physical, psychological, and social components, as well as overall well-being. Studies exploring malocclusion and its relationship to quality of life indicate that its effects vary across age groups, particularly among children, adolescents, and adults.[22] Previous studies have reported a close relationship between the need for orthodontic treatment and the emotional state of children and adolescents.[23,24] Liu et al.[21] demonstrated the lack of a relationship between physical discomfort, disability, and the need for orthodontic treatment in adult subjects, while Rusanen et al.,[25] stated that these problems are the most influential factors. Considering the studies conducted with similar age groups in the literature, the mean OHRQoL value in this study is similar to other studies,[26,27] but it is considerably lower than that of patients requiring orthognathic surgery with dentofacial deformity.[28,29]

Previous studies examining the relationship between malocclusion severity and OHRQoL have also produced conflicting findings. Some have reported significant associations between greater malocclusion severity and poorer OHRQoL,[30-32] consistent with the correlation observed in the present study, whereas others have found no significant relationship.[33-35] The inconsistency in the literature regarding the association between malocclusion severity and quality of life may be explained by several factors. First, cultural differences can influence how individuals perceive malocclusion, with some populations placing greater emphasis on esthetics while others prioritize functional aspects.[36] Second, the measurement instruments employed across studies vary considerably, which may account for divergent findings.[21,37] Third, sample characteristics, such as age distribution, may also play a role, as adolescents are more likely to be concerned with appearance-related issues, whereas adults may place greater weight on functional and psychosocial consequences.[38] Collectively, these factors may help explain why findings regarding the impact of malocclusion on quality of life remain mixed across studies.

The conceptual model proposed by Wilson and Cleary[39] emphasizes that individual and psychological characteristics can influence the relationship between clinical health status and perceived quality of life. For this reason, personal self-perception and self-assessment as well as quality of life were evaluated in this study. The need for orthodontic treatment among adult patients is associated with pain, self-consciousness, tension, embarrassment, and irritability.[40,41]

Research indicates that personal self-esteem is complex, influenced by factors beyond malocclusion, and can be enhanced by satisfaction with appearance.[31,32] In parallel with the findings of this study, Bahar et al.[42] reported no significant correlation between personal self-esteem and the malocclusion severity. The mean personal self-esteem score in our study population (13.62) was relatively low compared with previously reported values in similar cohorts. This discrepancy may be related to differences in demographic characteristics (including gender distribution) and sociocultural factors that influence perceptions of self-worth and dental appearance. Although previous studies[43,44] have reported a greater impact of malocclusion on OHRQoL among women, no significant sex-related differences were observed in the present study. These findings suggest that sex alone may not adequately explain individual variation in the psychosocial impact of malocclusion.

From a clinical perspective, these findings highlight the importance of considering psychosocial factors alongside clinical indices when evaluating orthodontic treatment needs in adult patients. Orthodontic interventions may contribute not only to the correction of malocclusion but also to improvements in OHRQoL, irrespective of malocclusion severity.

Study Limitations

Several limitations should be considered when interpreting the findings of this study. First, although the sample size met the requirements of the power analysis, the relatively modest sample may have limited the ability to detect small associations. Second, the unequal distribution of male and female participants may have affected sex-related comparisons. This imbalance could influence the results and limit the extent to which the findings can be generalized across genders. Future prospective studies should aim to recruit a larger and more gender-balanced sample to address these limitations.

CONCLUSION

Greater orthodontic treatment need was weakly associated with poorer OHRQoL in adults seeking orthodontic treatment. However, no significant association was observed between malocclusion severity and self-esteem. These findings indicate that while malocclusion severity may influence OHRQoL, it does not appear to significantly affect self-esteem in adults.

Ethics

Ethics Committee Approval: The Ondokuz Mayıs University Clinical Research Ethics Committee reviewed and approved this cross-sectional observational study (approval no: OMÜ KAEK 2025/05, date: 15.01.2025).

Informed Consent: Written informed consent was obtained from all patients included in the study.

Acknowledgment

This study was presented as an oral presentation at 15th International Symposium of the Turkish Orthodontic Society in 2017.

Footnotes

Author Contributions: Concept - A.Z.Ö., A.A.Ö.; Design - A.Z.Ö., A.A.Ö.; Data Collection and/or Processing - Ö.Ö.; Analysis and/or Interpretation - A.Z.Ö., A.A.Ö.; Literature Search - A.Z.Ö., A.A.Ö., Ö.Ö.; Writing - A.Z.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. Birkeland K, Bøe OE, Wisth PJ. Subjective assessment of dental and psychosocial effects of orthodontic treatment. *J Orofac Orthop.* 1997;58(1):44-61. [CrossRef]
2. The WHOQOL Group. The Development of the World Health Organization Quality of Life Assessment Instrument (the WHOQOL). In: Orley, J., Kuyken, W. (eds) Quality of Life Assessment: International Perspectives. Springer, Berlin, Heidelberg. 1994, pp. 41-57. [CrossRef]
3. Becker M, Diamond R, Sainfort F. A new patient focused index for measuring quality of life in persons with severe and persistent mental illness. *Qual Life Res.* 1993;2(4):239-251. [CrossRef]
4. Inglehart M. Oral health-related quality of life. *Quintessence Pub* 2002. [CrossRef]
5. Albino JE, Lawrence SD, Tedesco LA. Psychological and social effects of orthodontic treatment. *J Behav Med.* 1994;17(1):81-98. [CrossRef]
6. Shaw W. Factors influencing the desire for orthodontic treatment. *Eur J Orthod.* 1981;3(3):151-162. [CrossRef]
7. Kiyak HA. Does orthodontic treatment affect patients' quality of life? *J Dent Educ.* 2008;72(8):886-894. [CrossRef]
8. Shaw WC, Richmond S, Kenealy PM, Kingdon A, Worthington H. A 20-year cohort study of health gain from orthodontic treatment: psychological outcome. *Am J Orthod Dentofacial Orthop.* 2007;132(2):146-157. [CrossRef]
9. Tung AW, Kiyak HA. Psychological influences on the timing of orthodontic treatment. *Am J Orthod Dentofacial Orthop.* 1998;113(1):29-39. [CrossRef]
10. Crocker J, Major B. Social stigma and self-esteem: the self-protective properties of stigma. *Psychological Review.* 1989;96(4):608-630. [CrossRef]
11. Curbow B, Somerfield M, Legro M, Sonnega J. Self-concept and cancer in adults: theoretical and methodological issues. *Soc Sci Med.* 1990;31(2):115-128. [CrossRef]
12. Waldrop DP, Fabiano JA, Davis EL, Goldberg LJ, Nochajski T. Coexistent concerns: assessing the social and health needs of dental clinic patients. *Soc Work Health Care.* 2004;40(1):33-51. [CrossRef]
13. Gilbert GH, Duncan RP, Heft MW, Dolan TA, Vogel WB. Multidimensionality of oral health in dentate adults. *Med Care.* 1998;36(7):988-1001. [CrossRef]
14. Sun L, Wong HM, McGrath CPJ. The factors that influence oral health-related quality of life in young adults. *Health Qual Life Outcomes.* 2018;16:187. [CrossRef]
15. Daniels C, Richmond S. The development of the index of complexity, outcome and need (ICON). *J Orthod.* 2000;27(2):149-162. [CrossRef]
16. Slade GD, Spencer AJ. Development and evaluation of the Oral Health Impact Profile. *Community Dent Health.* 1994;11(1):3-11. [CrossRef]
17. Rosenberg M. Society and the adolescent self-image. Princeton university press Princeton, NJ; 1965. [CrossRef]
18. Başol ME, Karaağaçlıoğlu L, Yılmaz B. Developing a Turkish oral health impact profile-OHIP-14-TR. *Turkiye Klinikleri J Dental Sci.* 2014;20(2):85-92. [CrossRef]
19. Çuhadaroğlu F. Adolesanlarda benlik saygısı. *Uzmanlık Tezi, Hacettepe Üniversitesi Tıp Fakültesi Psikiyatri Anabilim Dalı, Ankara* 1986. [CrossRef]
20. Albino JE, Cunat JJ, Fox RN, Lewis EA, Slakter MJ, Tedesco LA. Variables discriminating individuals who seek orthodontic treatment. *J Dent Res.* 1981;60(9):1661-1667. [CrossRef]
21. Liu Z, McGrath C, Hägg U. Associations between orthodontic treatment need and oral health-related quality of life among young adults: does it depend on how you assess them? *Community Dent Oral Epidemiol.* 2011;39(2):137-144. [CrossRef]
22. De Baets E, Lambrechts H, Lemiere J, Diya L, Willems G. Impact of self-esteem on the relationship between orthodontic treatment

need and oral health-related quality of life in 11- to 16-year-old children. *Eur J Orthod.* 2012;34(6):731-737. [CrossRef]

23. Feu D, de Oliveira BH, de Oliveira Almeida MA, Kiyak HA, Miguel JA. Oral health-related quality of life and orthodontic treatment seeking. *Am J Orthod Dentofacial Orthop.* 2010;138(2):152-159. [CrossRef]
24. O'Brien C, Benson PE, Marshman Z. Evaluation of a quality of life measure for children with malocclusion. *J Orthod.* 2007;34(3):185-193; discussion 176. [CrossRef]
25. Rusanen J, Lahti S, Tolvanen M, Pirttiniemi P. Quality of life in patients with severe malocclusion before treatment. *Eur J Orthod.* 2010;32(1):43-48. [CrossRef]
26. Choi SH, Kim JS, Cha JY, Hwang CJ. Effect of malocclusion severity on oral health-related quality of life and food intake ability in a Korean population. *Am J Orthod Dentofacial Orthop.* 2016;149(3):384-390. [CrossRef]
27. Clijmans M, Lemiere J, Fieuws S, Willems G. Impact of self-esteem and personality traits on the association between orthodontic treatment need and oral health-related quality of life in adults seeking orthodontic treatment. *Eur J Orthod.* 2015;37(6):643-650. [CrossRef]
28. Khadka A, Liu Y, Li J, et al. Changes in quality of life after orthognathic surgery: a comparison based on the involvement of the occlusion. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology.* 2011;(112):719-725. [CrossRef]
29. Lee K-Y, Go E-J. Factors of influencing satisfaction with orthodontic treatment in orthodontic patients. *J Dent Hyg Sci.* 2010;(10):85. [CrossRef]
30. Ben Gassem AA, Aldweesh AH, Alsagob EI, et al. Psychosocial impact of malocclusion and self-perceived orthodontic treatment need among young adult dental patients. *Eur J Dent.* 2023;17(3):713-719. [CrossRef]
31. Bernabé E, Sheiham A, Tsakos G, Messias de Oliveira C. The impact of orthodontic treatment on the quality of life in adolescents: a case-control study. *Eur J Orthod.* 2008;30(5):515-520. [CrossRef]
32. Heravi F, Farzanegan F, Tabatabaee M, Sadeghi M. Do malocclusions affect the oral health-related quality of life? *Oral Health Prev Dent.* 2011;9(3):229-233. [CrossRef]
33. de Oliveira CM, Sheiham A, Tsakos G, O'Brien KD. Oral health-related quality of life and the IOTN index as predictors of children's perceived needs and acceptance for orthodontic treatment. *Br Dent J.* 2008;204(7):1-5; discussion 384-385. [CrossRef]
34. Mandall NA, Matthew S, Fox D, Wright J, Conboy FM, O'Brien KD. Prediction of compliance and completion of orthodontic treatment: are quality of life measures important? *Eur J Orthod.* 2008;30(1):40-45. [CrossRef]
35. Olkun HK, Sayar G. Impact of orthodontic treatment complexity on oral health-related quality of life in Turkish patients: a prospective clinical study. *Turk J Orthod.* 2019;32(3):125-131. [CrossRef]
36. Chen Y, Li T, Xu Y, et al. Oral health-related quality of life between Chinese and American orthodontic patients: a two-center cross-sectional study. *CCJPRS.* 2022;(4):171-177. [CrossRef]
37. Duisterwinkel F, Krijnen WP, Polder BJ, et al. Impact of malocclusion on oral health-related quality of life: insights from children with and without hypodontia. *J Orofac Orthop*. 2026;87(Suppl1):86-97. [CrossRef]
38. Pereira D, Machado V, Botelho J, et al. Impact of malocclusion, tooth loss and oral hygiene habits on quality of life in orthodontic patients: a cross-sectional study. *Int J Environ Res Public Health.* 2021;18(13):7145. [CrossRef]
39. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA.* 1995;273(1):59-65. [CrossRef]
40. Avontroodt S, Lemiere J, Cadenas de Llano-Pérula M, Verdonck A, Laenen A, Willems G. The evolution of self-esteem before, during and after orthodontic treatment in adolescents with dental malocclusion, a prospective cohort study. *Eur J Orthod.* 2020;42(3):257-262. [CrossRef]
41. Hassan AH, Amin Hel-S. Association of orthodontic treatment needs and oral health-related quality of life in young adults. *Am J Orthod Dentofacial Orthop.* 2010;137(1):42-47. [CrossRef]
42. Bahar AD, Sagi MS, Mohd Zuhairi FA, Wan Hassan WN. Dental aesthetics and self-esteem of patients seeking orthodontic treatment. *Healthcare (Basel).* 2024;12(16):1576. [CrossRef]
43. Rantavuori K, Silvola AS, Suominen A, Masood M, Suominen AL, Lahti S. Gender differences in the association between malocclusion traits and oral health-related quality of life in Finnish adults. *Eur J Oral Sci.* 2023;131(3):e12927. [CrossRef]
44. Närhi L, Tolvanen M, Pirttiniemi P, Silvola AS. Malocclusion severity and its associations with oral health-related quality of life in an adult population. *Eur J Orthod.* 2022;44(4):377-384. [CrossRef]

DOI: 10.4274/TurkJOrthod.2026.2025.190

TURKISH JOURNAL OF
ORTHODONTICS

Original Article

Assessing the Diagnostic Performance of ChatGPT-5.0 versus Machine Learning in Orthodontics: A Comparative Analysis for Extraction Treatment Planning

Artun Yangın, Hasan Camcı, Mehmet Soybelli

Afyonkarahisar Health Sciences University Faculty of Dentistry, Department of Orthodontics, Afyonkarahisar, Türkiye

Cite this article as: Yangın A, Camcı H, Soybelli M. Assessing the diagnostic performance of ChatGPT-5.0 versus machine learning in orthodontics: a comparative analysis for extraction treatment planning. *Turk J Orthod.* 2026; 39(3): 142-154

Main Points

- ChatGPT-5.0 achieved 75.77% accuracy in extraction treatment decisions.
- XGBoost achieved the highest overall accuracy (78.08%).
- ChatGPT-5.0 showed the highest sensitivity for extraction cases (76.68%).
- ChatGPT-5.0 outperformed random forest, support vector machine, logistic regression, and multi-layer perceptron.

ABSTRACT

Objective: To make accurate orthodontic extraction decisions, various clinical and cephalometric variables must be evaluated. This study aims to evaluate ChatGPT-5.0's performance in distinguishing orthodontic extraction decisions and to compare it with five supervised machine learning (ML) algorithms.

Methods: Of 550 retrospectively evaluated orthodontic records, 30 were reserved for calibration, leaving 520 for the main analysis. The reference standard was the consensus treatment decision of three expert orthodontists with more than 5 years of clinical experience. Overall, 23 variables were analyzed, including 13 clinical parameters, 7 cephalometric measurements, and photographs. ChatGPT-5.0's performance was evaluated using a 5-fold cross-validation design. It was compared with XGBoost, random forest, support vector machine (SVM), logistic regression, and multi-layer perceptron (MLP). Performance metrics included accuracy, sensitivity, specificity, precision, F1-score, and balanced accuracy, with 95% confidence intervals calculated. Statistical analyses utilized Cochran's Q test and the McNemar test with Holm-Bonferroni correction.

Results: Of the 520 main cases, 223 (42.88%) were extraction treatments and 297 (57.12%) were non-extraction treatments. XGBoost achieved the highest accuracy (78.08%), followed closely by ChatGPT-5.0 (75.77%). The overall performance difference among models was significant ($p \leq 0.001$). In pairwise comparisons, ChatGPT's accuracy was significantly higher than those of random forest, SVM, logistic regression, and MLP, but was found to be similar to XGBoost. ChatGPT-5.0 showed the highest sensitivity (76.68%), whereas XGBoost showed the highest specificity (82.15%).

Conclusion: ChatGPT-5.0 demonstrated performance comparable to, and in some cases superior to, traditional ML models for orthodontic extraction decisions. While XGBoost yielded the highest overall classification accuracy, ChatGPT-5.0's high sensitivity in detecting extraction cases was noteworthy.

Keywords: Artificial intelligence, ChatGPT, machine learning, orthodontic treatment planning, tooth extraction

Corresponding author: Res. Assist. Artun Yangın, **e-mail:** yangnartun@gmail.com **ORCID:** orcid.org/0009-0006-8229-6109
Received: January 05, 2026 **Accepted:** June 17, 2026 **Epub:** August 07, 2026 **Publication Date:** September 25, 2026

INTRODUCTION

Along with technological advancements, the disciplines of dentistry and orthodontics have undergone a comprehensive digital transformation.[1] Digital planning and workflows have become an integral part of routine practice.[2] In recent years, rapid progress in artificial intelligence (AI) has further accelerated this transformation.[3] Artificial intelligence applications in numerous areas such as cephalometric analysis,[4] patient cooperation prediction,[5] evaluation of cervical vertebral maturation stage,[6] unerupted tooth size prediction,[7] and treatment planning[8] have been widely documented in the literature. In this process, natural language processing systems that were initially text-based have evolved into more advanced, multimodal systems that include image-analysis capabilities, and have revealed remarkable potential in visually oriented areas such as the evaluation of orthodontic photographs and treatment planning.[9]

The use of large language models (LLMs) in orthodontics and dentistry has increased gradually in recent years.[1,10-12] However, there are significant differences, in terms of performance, consistency, and output structure, among various model families (such as ChatGPT, Gemini, DeepSeek, and Copilot) and among their versions.[13] In the literature, LLM applications are generally categorized into two main frameworks: text-based prompts and image-based or multimodal prompts.[12] While text-based applications mostly focus on patient communication, treatment planning, and clinical literature queries, performance on these tasks is evaluated using accuracy, readability, F1 score, and concordance metrics.[10] On the other hand, image-based applications requiring the interpretation of panoramic, cephalometric, or hand-wrist radiographs assess the visual processing capacity of multimodal LLMs and are generally evaluated using stricter quantitative metrics, such as mean absolute error, root mean square error, precision, and sensitivity.[14] Although previous studies have demonstrated that LLMs can demonstrate high reliability in text-based tasks, they also indicate that performance in image-based analyses varies significantly depending on the model and task, and in most cases such performance fails to reach the level of accuracy expected for clinical diagnostic use.[15-17]

In parallel with these developments, the systems used in orthodontics for diagnosis and determination of developmental stages are rapidly evolving from traditional deep learning (DL) approaches toward multimodal architectures.[18,19] Particularly for skeletal age determination, new-generation models that combine chaotic functional-connectivity matrices with metaheuristic methods, such as Puma Optimization, to analyze the nonlinear dynamics of cervical vertebrae have been reported to offer high accuracy and computational efficiency.[20] Similarly, task-specific DL architectures capable of pixel-level image recognition and feature extraction yield highly robust results in tasks such as cephalometric landmarking and radiological age determination.[21,22] Even when LLMs are guided by atlas-supported or example-enriched prompt strategies, they often lag behind task-specific machine learning (ML)/DL models in distinguishing fine visual details. Therefore, in light of current knowledge, LLMs are positioned as auxiliary clinical decision support tools rather than independent diagnostic systems.[22,23]

Orthodontic treatment planning is inherently a multidimensional process that requires the simultaneous evaluation of numerous variables, such as the patient's age, malocclusion type, cephalometric measurements, and dental relationships.[24-26] The decision between extraction and non-extraction treatment, one of the most critical stages of this process, can be particularly challenging for less-experienced clinicians, as it directly affects treatment mechanics, facial esthetics, and long-term stability.[25,27] For this reason, tools capable of processing multiple clinical inputs in a structured manner and providing support to the decision-making process are highly valuable in clinical practice.[28,29]

Although ML algorithms, which produce predictive results by identifying patterns in data, are increasingly used in orthodontic planning, studies directly comparing the performance of LLM-based approaches, particularly in multi-factor clinical decision tasks, with traditional ML/DL methods remain limited. Existing comparisons often remain superficial; the differences in data structures, training paradigms, evaluation strategies, and interpretability between these groups of models are not addressed in sufficient detail. This study aims to fill this gap by comparing ChatGPT's performance in distinguishing orthodontic extraction and non-extraction cases with that of traditional ML models on the same dataset.

The aim of the study was to compare the performance of ChatGPT with traditional ML models in predicting orthodontic extraction/non-extraction treatment decisions and to evaluate how the decision-making structures of these two approaches differ in terms of interpretability. A novel aspect of the study is that, on the one hand, it compares ChatGPT and traditional ML models on the same sample, and on the other hand, it demonstrates that the explanatory responses obtained from LLMs are not directly, methodologically, equivalent to quantitative feature importance metrics in ML models. The study presents a framework that jointly evaluates the potential and limitations of LLM use in orthodontics with respect to both performance and explainability. The null hypothesis (H_0) was that the LLM-based approach will not differ from ML algorithms in terms of its ability to distinguish cases; while the alternative hypothesis (H_1) predicts that there will be a statistically significant difference between the classification performances of these two approaches.

METHODS

This study was approved by the Institutional Afyonkarahisar Health Sciences University Non-Intervention Scientific Research Ethics Committee (approval no: 2024/8, date: 04.10.2024).

The principles of the Declaration of Helsinki were adhered to during the study process; written and verbal informed consent was obtained from all patients and their legal guardians.

Orthodontic records of patients who had completed treatment at our institution were retrospectively reviewed. Cases meeting the predefined inclusion and exclusion criteria were included in the study. The original plans of these patients had been determined and applied in accordance with a joint clinical decision by three expert orthodontists, each with at least five years of clinical experience, independent of the current study. Therefore, the treatment decisions made and applied in clinical practice were accepted as the reference standard in the study. To further validate the reference standard, the patients' post-treatment intraoral and extraoral records were also re-evaluated by three independent expert orthodontists. The evaluators were blinded to each other's evaluations, and the treatment outcomes were examined to determine whether they met the American Board of Orthodontics (ABO) criteria. Cases not meeting the ABO criteria were excluded from the study. Following this screening process, 550 cases fulfilled the study criteria. Of these, 520 cases comprised the primary analysis dataset, whereas the remaining 30 cases were used for prompt calibration.

The inclusion criteria for the study were: (1) fully erupted permanent dentition, (2) absence of dentofacial anomalies, (3) no indication for functional treatment, and (4) specific extraction patterns. In terms of treatment protocol, the study included non-extraction cases and cases in which the following premolars were extracted: upper and lower first premolars; upper first and lower second premolars; or only upper first premolars.

Exclusion criteria included impacted or supernumerary teeth, congenital tooth agenesis, retained primary teeth, diagnostic photographs of inadequate quality, changes in the treatment plan during treatment, and post-treatment records that failed to satisfy the ABO criteria.

The collected dataset consisted of intraoral and extraoral photographs, demographic information, diagnostic records, model analyses, and cephalometric measurements (Table 1). A total of 23 variables were used: 13 clinical parameters, 7 cephalometric measurements, 2 photograph

Table 1. Clinical, cephalometric, photographic, and demographic variables included in the analysis

Values	Parameter 1	Parameter 2	Parameter 3
Clinical values			
Age			
Gender	Female (0)	Male (1)	
Jaw position	Retrognathic (0)	Orthognathic (1)	Prognathic (2)
Molar relationship	Class I (0)	Class II (1)	Class III (2)
Overbite (mm)			
Overjet (mm)			
Vertical values	Decreased (0)	Normal (1)	Increased (2)
Upper midline	X mm to the right in the upper (+X)	X mm to the left in the upper (-X)	
Lower midline	X mm to the right in the lower (+X)	X mm to the left in the lower (-X)	
Upper crowding			
Lower crowding			
Upper diastema			
Lower diastema			
Cephalometric values			
SNA (°)			
SNB (°)			
ANB (°)			
FMA (°)			
U1-SN (°)			
IMPA (°)			
Nasolabial angle (°)			
Photos			
Intraoral			
Extraoral			
Treatment decision	Non-extraction (0)	Extraction (1)	

SNA, sella-nasion-A point angle; SNB, sella-nasion-B point angle; ANB, A point-nasion-B point angle; FMA, Frankfort-mandibular plane angle; U1-SN, upper incisor to sella-nasion plane angle; IMPA, incisor mandibular plane angle.

types, and 1 treatment decision. In the primary analysis, the treatment decision was operationalized as a binary outcome variable. Cases without tooth extraction were classified as "non-extraction", and treatment plans involving extraction of upper and/or lower premolars were classified as "extraction". Among the 520 cases in the main analysis dataset, 223 (42.88%) were in the extraction group and 297 (57.12%) were in the non-extraction treatment group (Table 2). Although the class distribution was not perfectly balanced, it was not significantly imbalanced; therefore, in addition to accuracy, balanced accuracy, sensitivity, specificity, precision, and F1-score were used in the performance evaluation to better reflect the potential impact of class imbalance.

All intraoral and extraoral photographs were obtained using a Canon EOS 60D digital camera (Tokyo, Japan), a Sigma 105-mm f/2.8 macro lens (Kawasaki, Japan), and a Godox ring flash (Shenzhen, China). Intraoral photographs (PNG format) were taken from five standard angles (frontal, right and left lateral, upper occlusal, and lower occlusal); extraoral photographs (PNG format) consisted of two images (frontal at rest and profile).

Evaluation of ChatGPT

All analyses were performed by a single researcher using a MacBook Air (MacBook Air, Apple Inc., Cupertino, California; Apple M1 chip, 8-core CPU, 8 GB RAM, 256 GB SSD). The analyses were conducted using ChatGPT-5.0 (OpenAI, San Francisco, CA, USA; October 2024 version), accessed via the web-based interface. To minimize information transfer between sessions, five separate sessions were opened, and the "memory" feature was disabled. Thus, each session was conducted independently, limiting contextual effects that could arise from previous conversations.

Thirty of the 550 cases in the initial dataset were set aside for prompt calibration purposes, and the remaining 520 cases formed the main analysis dataset. The 30 cases reserved for prompt calibration were not included in the final performance evaluation. The dataset split in this way and the schematic of the analysis workflow are shown in Figure 1.

The 520 cases constituting the main analysis dataset were divided into five equal subsets (fold 1 to fold 5), each consisting of 104 cases (Figure 1). A 5-fold cross-validation design was used.[20,30] In each iteration, one subset was used as the test set, while the remaining four subsets were presented to ChatGPT in the same session to provide contextual guidance. Accordingly, in each iteration, 416 cases were used for contextual prompting, and 104 cases were used for testing purposes. At the end of five iterations, the out-of-fold predictions from all folds were combined, and the final statistical analysis was performed on them. This process is summarized in Figure 1. In this study, the parameters of the ChatGPT-5.0 model were not updated, and no retraining or fine-tuning was applied.[31,32]

A standardized prompt template was created to test ChatGPT. To ensure standardization, all photographs were resized to 224×224 pixels. The final prompt used in the study was determined as follows:

"You are an orthodontic assistant. Based on the clinical data, cephalometric measurements, and photographs I provide, evaluate the patient's treatment plan. Decide whether the treatment should be non-extraction or extraction. Do not make any additional comments-only state the treatment decision".

Iterative prompt calibration (IPC) was performed to determine the final prompt. Seven prompts were tested in separate sessions using the 30 cases within a calibration subset which was not included in the overall performance analysis. The initial prompt was: "Based on this data, create the patient's treatment plan; decide whether it will be extraction or non-extraction". The final prompt presented above was selected after the calibration process. Figure 1 schematically shows the prompt calibration stage and its separation from the analysis dataset; it also presents the accuracy rates of the developed prompts. The 30 calibration cases used for prompt calibration were not included in the 5-fold cross-validation process. Therefore, the final performance metrics were calculated solely over the out-of-fold predictions obtained from the main analysis dataset.

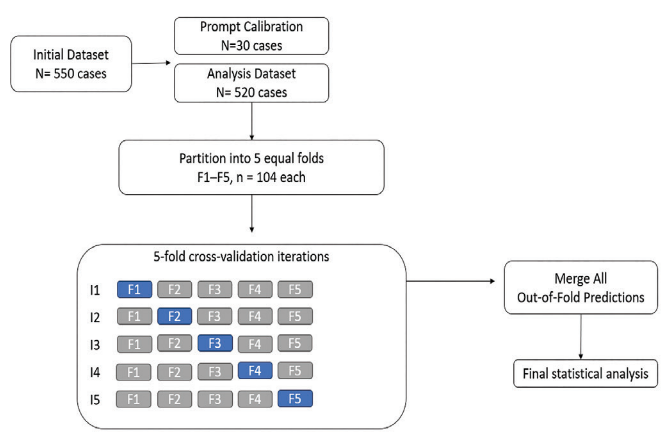


Figure 1. Flow diagram of the study.

Table 2. Demographic characteristics and treatment distribution of the study population

		Extraction type		
		Extraction	Non-extraction	Total
Sex	Female	151 (67.71%)	193 (64.98%)	344 (66.15%)
	Male	72 (32.29%)	104 (35.02%)	176 (33.85%)
Age		14.97	14.46	14.68
Total		223 (42.88%)	297 (57.12%)	520

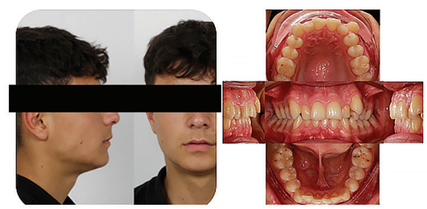
To qualitatively illustrate ChatGPT's decision process, a chain-of-thought (CoT) approach was additionally applied to a single sample case. However, this was not designed as a systematic analysis component and was not included in the overall performance analysis. The prompt given to ChatGPT and the sample reasoning output obtained are summarized in Table 3. The responses provided by ChatGPT for each case were recorded by the primary researcher and independently reviewed and verified by three expert orthodontists. A confusion matrix was created by comparing the actual results with the ChatGPT predictions. Furthermore, final performance metrics were calculated by combining the out-of-fold predictions obtained from each cross-validation iteration (Figure 2).

Evaluation of Machine Learning Algorithms

The dataset includes photographs, clinical data, cephalometric measurements, and treatment decisions of 520 patients (Table 1). Although ML algorithms can process numerical data, they lack the ability to directly analyze images. Therefore, a convolutional neural network (CNN) was used to convert visual data into a numerical format. To digitize visual data, the VGG16 CNN with pre-trained weights on the ImageNet dataset was used. A transfer learning approach was adopted in the training of the model. In this context, the final dense layers of the model responsible for classification were removed, and only the lower layers responsible for feature extraction were preserved. These lower layers were frozen and served as fixed feature extractors,

Table 3. Sample ChatGPT-5.0 output generated using a chain-of-thought (CoT)-optimized prompt

Prompt



You are my orthodontic assistant. Based on the clinical data, cephalometric measurements and photos I provide, evaluate the patient's treatment plan. Decide whether the treatment should be non-extraction or extraction. Do not make any other comments—only state the decision. Then, extract the chain of thought for this process.

CLINICAL VALUES

- Age: 17 years 8 months
- Gender: Male (1)
- Jaw position: Orthognathic (0)
- Molar relationship: Class II (1)
- Overbite (mm): 3
- Overjet (mm): 3.5
- Vertical values: Normal (1)
- Upper Midline: 3 mm to the left (–3)
- Lower Midline: Centered (0)
- Upper Crowding (mm): 10
- Lower Crowding (mm): 4.5

CEPHALOMETRIC VALUES

- SNA (°): 78.9
- SNB (°): 75.7
- ANB (°): 3.1
- FMA (°): 22.1
- U1-SN (°): 97.9
- IMPA (°): 94.7
- Nasolabial Angle (°): 114.5

Response

1. Crowding severity: The patient presents with significant crowding-10 mm in the upper arch and 4.5 mm in the lower-which cannot be resolved with non-extraction mechanics alone.
2. Skeletal Relationship: ANB is 3.1°, suggesting mild skeletal class II, but the jaw position is orthognathic, indicating that the discrepancy can be managed dentally.
3. Incisor inclinations: U1-SN is 97.9° and IMPA is 94.7°, both close to normal, indicating that there is no significant dental compensation; thus, space for alignment and retraction needs to be created.
4. Nasolabial angle: Measured at 114.5°, suggesting a favorable soft tissue profile that can tolerate retraction without causing aesthetic compromise.
5. Midline discrepancy: A 3 mm upper midline shift to the left is noted; bilateral extractions can aid in symmetry correction.
6. Overall alignment strategy: Extraction of all first premolars (14, 24, 34, 44) will provide adequate space for resolving crowding and allow correction of the molar and canine relationships, midline deviation, and incisor inclinations while maintaining soft tissue harmony.

producing 512-dimensional abstract feature vectors from the resized photographs (224×224 pixels). These 512-dimensional vectors obtained from each of the 7 photographs (5 intraoral, 2 extraoral) were concatenated, creating a single combined visual feature vector of 3584 dimensions (7×512) for each patient. The digitized image data were then combined with the patients' clinical and cephalometric data into a single CSV file, creating a numerical dataset suitable for testing ML algorithms. These numerical data were used to predict the binary outcome: non-extraction versus extraction treatment.

The ML algorithms used in this study are classified into different types according to their structural characteristics. As shown in Table 4, multi-layer perceptron (MLP) is a neural-network-based algorithm that processes data through forward and backward propagation. Random Forest is a tree-based method that performs classification by majority voting among numerous decision trees. XGBoost is also tree-based, but relies on a boosting approach and tries to correct previous errors by sequentially training decision trees. Support vector machines (SVM) and logistic regression are mathematical model-based methods. While SVM aims to determine the hyperplane that best separates the data, logistic regression calculates the probability of data points belonging to a specific class. Analyses were conducted on the Jupyter Notebook platform (Python 3.8.5, Anaconda, macOS). While the algorithms were trained with data obtained from 416 patients, 104 patients were reserved for testing. A strict evaluation strategy was followed

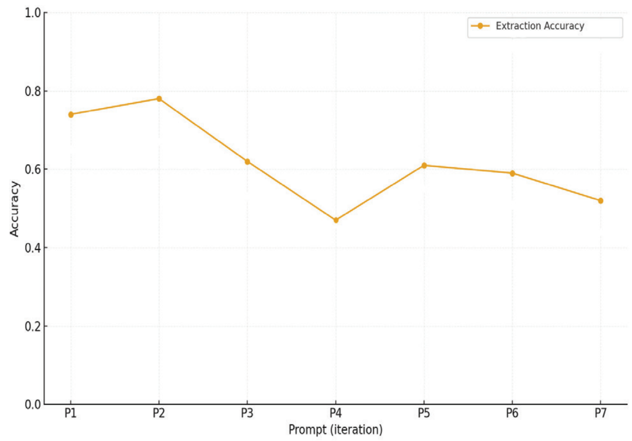


Figure 2. Iterative prompt calibration process and accuracy rates of 7 alternative prompts.

to minimize the risk of overfitting during model training and to increase model generalizability. To prevent the algorithms from memorizing the training data, 5-fold cross-validation was applied to the dataset of 520 cases. In each iteration, 80% of the data (n=416) were used for training, while the remaining 20% (n=104) that the model had never seen before were reserved as an independent test set. Furthermore, the hyperparameter optimization process (GridSearchCV) of the models was performed solely on the training sets (n=416). The test sets (n=104) were completely isolated from the optimization process, strictly preventing data leakage. Confusion matrices were generated using the combined classification results obtained for each case during the cross-validation process.

Statistical Analysis

All statistical analyses were performed using IBM SPSS Statistics for Windows (version 26.0; IBM Corp., Armonk, NY, USA). The performance of ChatGPT and ML algorithms in making extraction decisions was evaluated using accuracy, precision, sensitivity (recall), F1-score, balanced accuracy, and specificity (Table 5, Figure 3). 95% confidence intervals (CIs) for these metrics were obtained using the Wilson method.

Since the same cases were classified by all models, inter-model comparisons were made taking the paired data structure into account. Cochran's Q test was used to compare the overall accuracy of the six models. For pairwise comparisons between ChatGPT and each ML model, the McNemar test, which is appropriate for paired binary outcomes, was applied. A Holm-Bonferroni correction was performed to control type I error inflation due to multiple pairwise comparisons. Additionally, to strengthen clinical interpretation, differences between ChatGPT and other models for sensitivity among positive (extraction) cases and specificity among negative (non-extraction) cases were re-evaluated using the McNemar test, and the Holm-Bonferroni correction was applied. In all tests, the statistical significance level was accepted as two-tailed $p \leq 0.05$.

RESULTS

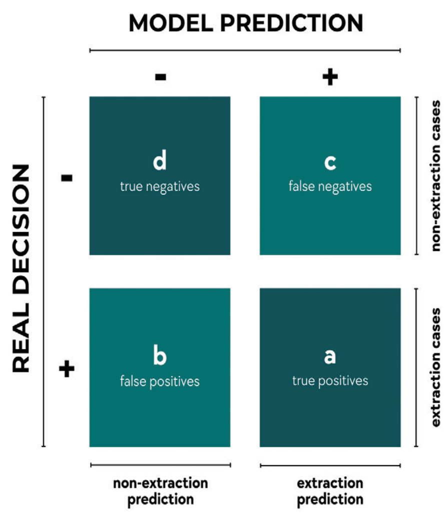
Data obtained from 520 patients (mean age: 14.68 years; standard deviation =2.75) who were included in the evaluation were analyzed (Table 2). Model performance is presented in Figure 4. Among the evaluated models, the highest accuracy was obtained by the XGBoost model (0.7808; 95% CI: 0.7432, 0.8142).

Table 4. The algorithms employed in the current study

Algorithm	Algorithm type	Working mechanism
MLP	Neural network-based	Processes data using artificial neurons; learns through forward and backward propagation.
Random forest	Tree-based	Builds multiple decision trees and classifies through majority voting.
XGBoost	Boosting (tree-based)	Trains decision trees sequentially, correcting previous errors during learning.
SVM	Mathematical model	Finds the hyperplane that best separates the data.
Logistic regression	Mathematical model	Calculates the probability of data points belonging to a class, separates them using the sigmoid function.

MLP, multi-layer perceptron; SVM, support vector machine.

Table 5. Metrics measuring model performance and their specific use or application		
Metric	**What it measures**	**When to use it**
Accuracy	The overall correctness of the model's predictions (ratio of correct predictions to total predictions).	When the dataset is balanced and all classes are equally important.
Precision	The proportion of true positives out of all predicted positives.	When false positives (e.g., false alarms) need to be minimized.
Sensitivity (recall)	The proportion of true positives out of all actual positives.	When false negatives (e.g., missed detections) need to be minimized.
F1 score	The harmonic mean of precision and recall, balancing the two metrics.	When both precision and recall are equally important, especially for imbalanced datasets.
Balanced accuracy	The average of sensitivity and specificity, useful for imbalanced datasets.	When the dataset is imbalanced and overall performance across classes is needed.
Specificity	The proportion of true negatives out of all actual negatives.	When false positives need to be minimized, especially in cases where detecting negatives correctly is crucial.



$$Accuracy = \frac{a+d}{a+b+c+d} \qquad Sensitivity = \frac{a}{a+b} \qquad Precision = \frac{a}{a+c}$$

$$Balanced\ Accuracy = \frac{\frac{a}{a+b} + \frac{d}{c+d}}{2} \qquad Specificity = \frac{d}{c+d} \qquad F1 - score = \frac{2a}{2a+b+c}$$

Figure 3. Confusion matrix structure and calculation of metrics for measuring model performance (positive values represent aggregated cases, while negative values represent non-aggregated cases).

ChatGPT ranked second with an accuracy of 0.7577 (95% CI: 0.7191, 0.7925) and outperformed random forest, SVM, logistic regression, and MLP. The lowest accuracy was observed for the MLP model (0.6288).

In terms of sensitivity, the highest value was determined for ChatGPT (0.7668; 95% CI: 0.7071, 0.8175). On the other hand, XGBoost showed the highest performance in terms of specificity, precision, F1-score, and balanced accuracy. When these findings are evaluated together, ChatGPT is seen to have a higher tendency to capture extraction cases, whereas XGBoost exhibits a stronger overall classification balance.

When the correct/incorrect classification patterns of the six models on the same cases were compared with Cochran's Q test, an overall statistically significant difference was found among the models (Q=47.432; p≤0.001) (Table 6). This result indicates that the evaluated models do not have equivalent performance.

Pairwise comparisons between ChatGPT and other models were performed using the McNemar test, and the Holm-Bonferroni correction was applied for multiple comparisons. Accordingly, ChatGPT showed a significantly higher accuracy compared to random forest (adjusted p=0.017), SVM (adjusted p=0.010), logistic regression (adjusted p≤0.001), and MLP (adjusted p<0.001) models. In contrast, no significant difference in accuracy was detected between ChatGPT and XGBoost (adjusted p=0.393) (Table 6).

In the analyses performed on extraction cases, the sensitivity of ChatGPT was found to be significantly higher than random forest, SVM, logistic regression, and MLP models (all adjusted p-values ≤0.001). No significant difference in sensitivity was detected between XGBoost and ChatGPT (adjusted p=0.368; Table 6). This finding supports the conclusion that ChatGPT shows a stronger tendency, especially in detecting extraction cases.

In the specificity analyses performed within non-extraction cases, although some differences were observed in crude comparisons, no statistically significant difference remained in terms of specificity between ChatGPT and other models after Holm-Bonferroni correction (all adjusted p≥0.05) (Table 6). This result indicates that, although XGBoost and some classical ML models have numerically higher specificity values, this superiority was not statistically confirmed after multiple comparisons.

Confusion Matrices

The confusion matrices presented in Figure 5 show the combined classification results obtained for all 520 cases. Examination of the confusion matrices from classification results for 520 cases showed that distributions closest to the ideal pattern were obtained for the XGBoost and ChatGPT models (Figure 5). All models classified non-extraction cases more accurately than they classified extraction cases. The model with the highest number of true positive classifications was ChatGPT (n=171), followed by XGBoost (n=162).

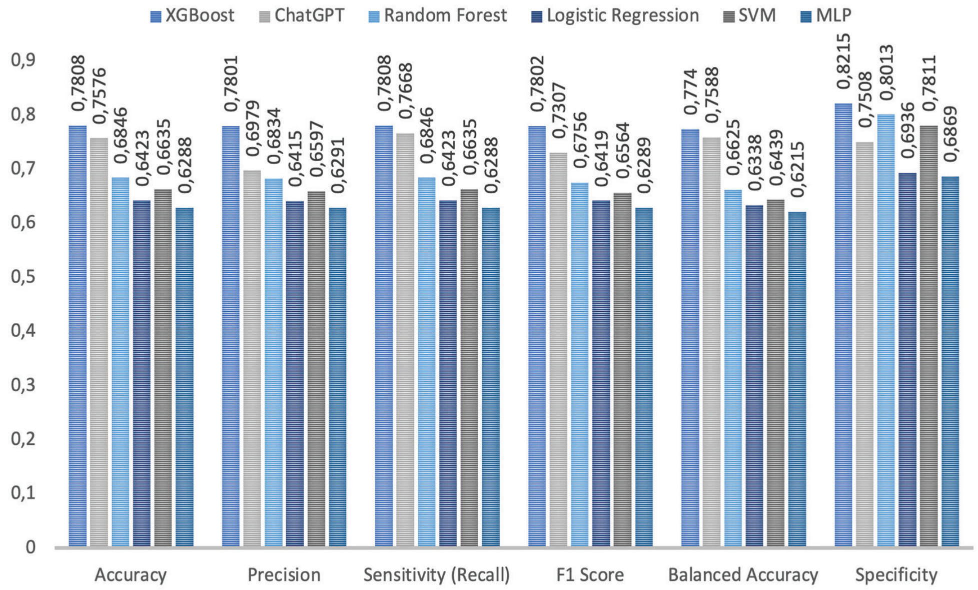


Figure 4. Comparison of metrics used to evaluate model performance in treatment decisions, including extraction and non-extraction outcomes. SVM, support vector machine; MLP, multi-layer perceptron.

Table 6. Pairwise comparison of ChatGPT and machine learning models

Pairwise comparison of ChatGPT and machine learning models in terms of accuracy					
Comparison	**ChatGPT accuracy**	**Comparator model accuracy**	**Unadjusted p-value**	**Holm-adjusted p-value**	**Interpretation**
Cochran's Q test (6 models)	-	-	<0.001	-	A statistically significant overall difference was observed among the models
ChatGPT vs. XGBoost	0.7577	0.7808	0.393	0.393	Not statistically significant
ChatGPT vs. random forest	0.7577	0.6846	0.009	0.017	Statistically significant
ChatGPT vs. SVM	0.7577	0.6731	0.003	0.010	Statistically significant
ChatGPT vs. logistic regression	0.7577	0.6500	<0.001	<0.001	Statistically significant
ChatGPT vs. MLP	0.7577	0.6288	<0.001	<0.001	Statistically significant
Pairwise comparisons of ChatGPT and machine learning models in terms of sensitivity and specificity					
Metric	**Comparison**	**ChatGPT**	**Other model**	**Raw p**	**Holm-adjusted p**
Sensitivity	ChatGPT vs. XGBoost	0.7668	0.7265	0.368	0.368
Sensitivity	ChatGPT vs. random forest	0.7668	0.5067	<0.001	<0.001
Sensitivity	ChatGPT vs. SVM	0.7668	0.5067	<0.001	<0.001
Sensitivity	ChatGPT vs. logistic regression	0.7668	0.5740	<0.001	<0.001
Sensitivity	ChatGPT vs. MLP	0.7668	0.6009	<0.001	<0.001
Specificity	ChatGPT vs. XGBoost	0.7508	0.8215	0.032	0.128
Specificity	ChatGPT vs. random forest	0.7508	0.8182	0.043	0.128
Specificity	ChatGPT vs. SVM	0.7508	0.7980	0.198	0.396
Specificity	ChatGPT vs. logistic regression	0.7508	0.7071	0.250	0.396
Specificity	ChatGPT vs. MLP	0.7508	0.6498	0.012	0.061

p<0.05 was considered statistically significant.
SVM, support vector machine; MLP, multi-layer perceptron.

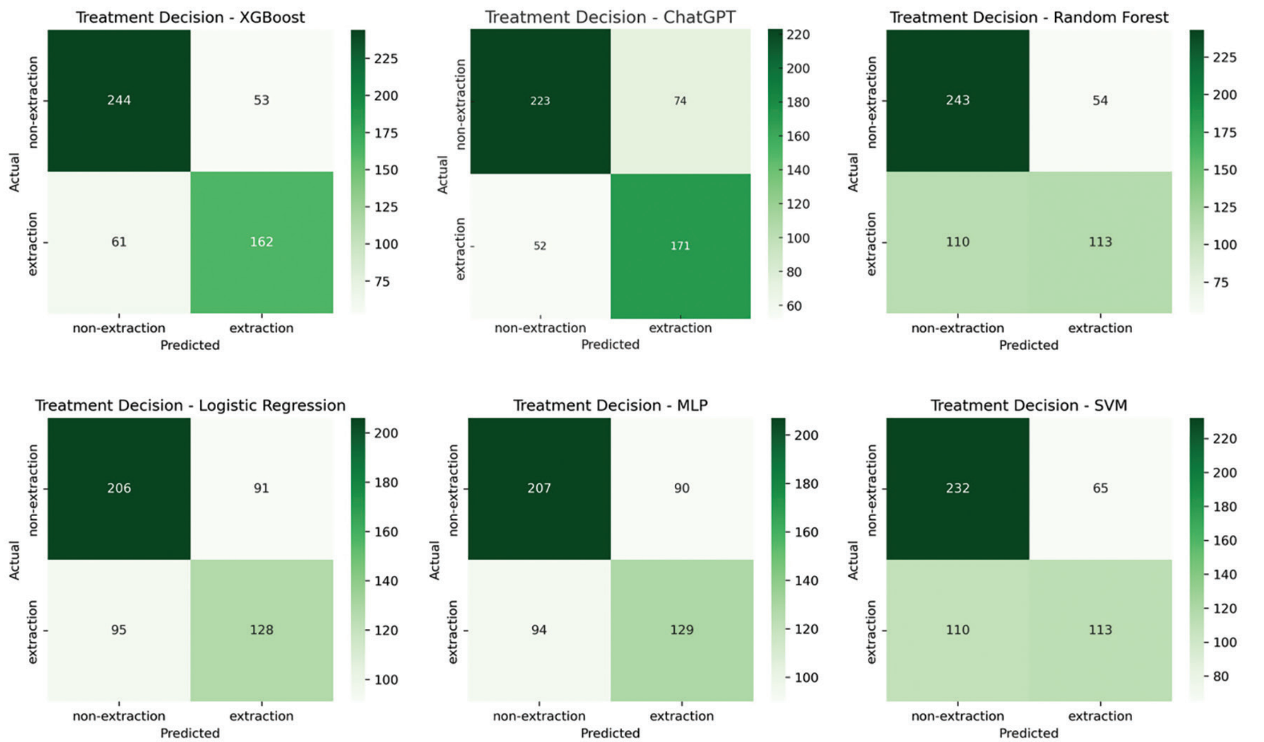


Figure 5. Visualization of model performance in the treatment decision presented as a confusion matrix.
MLP, multi-layer perceptron; SVM, support vector machine.

This finding indicates that, although ChatGPT lagged behind XGBoost in overall performance metrics, it exhibited higher sensitivity in identifying extraction cases. In contrast, the number of false-negative cases was highest in the random forest and SVM models; in each model, 110 cases that were actual extractions were classified as non-extraction. In terms of the number of false positive cases, that is, cases that were actually non-extraction but were classified as extraction, the lowest values were observed for the XGBoost (n=53) and random forest (n=54) models.

Feature Importance

Feature importance was evaluated using clinical and cephalometric variables (Figure 6). Across XGBoost, random forest, and logistic regression models, variable contributions were examined using quantitative feature-importance metrics. For ChatGPT, a quantitative feature importance calculation in the classical sense was not performed. Instead, the prominent variables in the explanatory responses accompanying the model's classification decision were evaluated qualitatively. Support vector machine and MLP models were not included in the relevant analysis because they did not provide directly interpretable variable-contribution outputs with the method used in this study.

In ML models, maxillary crowding was identified as one of the most influential variables and was frequently emphasized in ChatGPT's explanatory responses (Figure 6). ChatGPT referred to the nasolabial angle more prominently in its decision justifications, whereas the relative contribution of this parameter was lower in the XGBoost model. The logistic regression model assigned higher weight to jaw position and ANB angle than other ML models. While the contribution of the diastema variable remained low in the random forest model, it was more pronounced in the XGBoost and logistic regression models. In ChatGPT's explanations, maxillary arch diastemas were identified as particularly noteworthy parameters in the decision-making process.

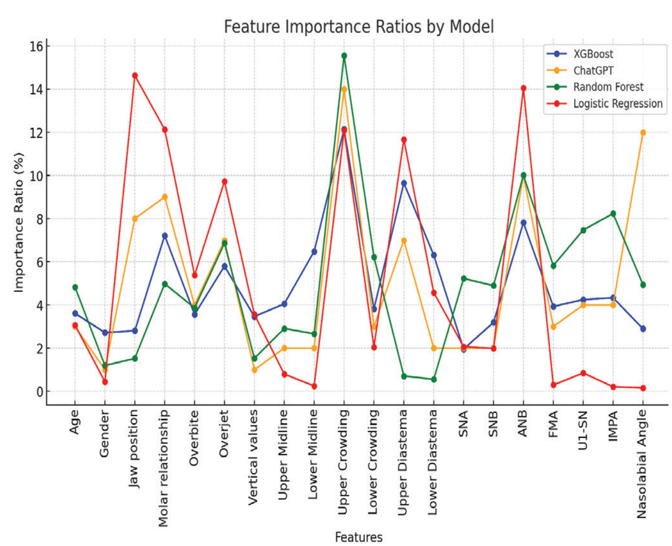


Figure 6. Feature importance percentages for the artificial intelligence models.

DISCUSSION

The use of AI in orthodontics has expanded rapidly in recent years. In particular, the accuracy of different algorithms for diagnosis, treatment planning, and cephalometric analysis has been investigated.[11,21,33] While early studies mostly focused on traditional ML models trained for specific tasks,[33] the potential of LLMs such as ChatGPT, Microsoft Copilot, and Gemini in clinical decision processes has recently started to draw attention.[10,13] In this context, the present study aimed to directly compare the performance of ChatGPT-5.0 in distinguishing between extraction and non-extraction treatment decisions with that of traditional ML algorithms. To the best of our knowledge, this study is the first to perform the mentioned comparison on the same sample.

Traditional ML algorithms are designed and optimized for specific tasks.[34] In contrast, LLMs like ChatGPT-5.0 cannot be retrained by the user. Their performance can be guided only by prompt engineering.[31] For this reason, the two approaches handle the clinical problem in fundamentally different ways. General-purpose models such as ChatGPT-5.0, thanks to their multimodal architectures, can simultaneously evaluate different data types, including age, cephalometric measurements, crowding severity, and extraoral photographs, both visually and textually within a clinical context.[35] However, since they are not specialized for learning from numerical data, traditional algorithms are expected to perform better on pure classification tasks. In the study, the parameters of the ChatGPT-5.0 model were not updated, and no retraining or fine-tuning was applied.[31,32] Therefore, the approach used should not be considered model training in the classical sense. Instead, the study should be defined as a large-language-model evaluation design supported by a 5-fold cross-validation performed under many-shot in-context guidance using numerous examples, conducted after prompt calibration.[36] Consequently, this method is not part of the standard zero-shot or few-shot evaluation paradigm but rather of a contextually guided, cross-validated evaluation approach.

The findings of our study support this framework. While XGBoost achieved the highest performance with 78.08% accuracy, ChatGPT-5.0 ranked second with 75.77% accuracy. The performance ranking being XGBoost ≥ ChatGPT > Random Forest > SVM > Logistic Regression > MLP is consistent with the large-scale study conducted by Prasad et al.[28] and the models reaching the highest accuracy rates being XGBoost, Random Forest, and Decision Tree-based algorithms. Similarly, in the study of Köktürk et al.,[37] the most successful predictions in the orthodontic extraction decision were obtained with Gradient Boosted Trees (the algorithm family forming the basis of XGBoost), SVM, and random forest models. Considering the complex structure of clinical, cephalometric, and demographic variables in our dataset, it is an expected result and is consistent with the literature that algorithms that optimize decision trees, such as XGBoost, show superior performance compared with neural networks, such as the traditional (MLP; 62.88% accuracy). However, ChatGPT-5.0 demonstrated performance comparable to traditional models. In particular, the model, which exhibits higher sensitivity than many algorithms, suggests that it has a strong tendency to detect extraction cases. That ChatGPT-5.0 obtained its highest score in sensitivity (0.7668) and its lowest score in precision (0.6979) indicates that the model is more aggressive in detecting extraction cases but has a higher margin of error. Although ChatGPT-5.0 appears promising as a tool to support clinical decision-making, ask-specific models trained with expert consensus are more reliable for precise treatment decisions.

The current findings should be interpreted in the context of previous studies on the use of AI in orthodontics. For example, Kunz et al. [38] reported an accuracy of over 90% using CNNs in cephalometric landmark detection. Similarly, studies evaluating extraction/non-extraction treatment planning with CNN architectures such as ResNet and VGG have reported success rates varying between 70% and 90%.[33] When evaluated from this perspective, the performance demonstrated by ChatGPT-5.0 in this study falls within the reported success range of traditional algorithms and is at an acceptable level.

That orthodontic planning is inherently subjective raises the question of how the “ground truth” will be defined in such studies. While approaches based on a single expert opinion were used by Jung and Kim,[27] designs that brought together the opinions of multiple experts were reported by Prasad et al.[28] and Mason et al.[39] While the single expert approach increases the risk of individual bias, the inclusion of multiple experts can decrease data consistency. The consensus method used in this study by three expert orthodontists offered a more balanced solution between the two extreme approaches.

The data distribution in our study appears consistent with clinical reality. Of the 520 patients, 223 (42.88%) were in the extraction treatment group and 297 (57.12%) were in the non-extraction treatment group. Although examples balancing the case distribution as 50% to 50% have been reported by Köktürk et al.[37] and Mason et al.,[39] there are also studies preserving natural clinical distributions where non-extraction cases reach 75%. The distribution in our study reflects the slight class imbalance encountered in clinical practice. Indeed, Etemad et al.[40] emphasized that this imbalance is not a data flaw, but a natural consequence of the conservative approach adopted by clinicians in borderline cases.

When the performance profiles of the models were examined, the vast majority achieved their highest scores for specificity. This finding is consistent with the low sensitivity (0.29, 0.53) and high specificity (0.94, 0.96) results reported by Etemad et al.[40] In other words, the models appear to be more successful in identifying non-extraction cases.

With respect to the variables affecting the extraction decision, our findings are consistent with the literature.[37,39,40] Mason et al.[39] reported that crowding, facial profile, and lower incisor inclination are among the most influential parameters in the extraction/non-extraction treatment decision. In our study, maxillary crowding also emerged as one of the most important variables in traditional ML models in which quantitative feature importance was calculated. This result is consistent with the literature, which indicates that the discrepancy between available and required arch lengths is one of the main determinants in extraction decision-making. In contrast, the greater prominence of variables, such as the nasolabial angle and upper-arch diastemas, in ChatGPT-5.0's explanatory responses suggests that the model processes clinical information on a different representational plane.

The difference in explainability between traditional ML models and LLMs should not be ignored when interpreting these findings. In models such as XGBoost, feature importance is a quantitative metric that indicates each feature's contribution to model predictions. Conversely, LLMs, such as ChatGPT, lack a similarly standardized method for calculating statistical feature importance. By their nature, these models are black-box systems and their decision mechanisms are not fully transparent.[41] In the literature, to partially mitigate this opacity, prompting strategies such as CoT are used, in which the model is asked to generate step-by-step explanations.[42,43] However, these explanations are not quantitative evidence that directly reflects the internal classification mechanism of the model; rather, they are textual justifications generated post-decision. Therefore, it is not correct to evaluate the deterministic feature importance values obtained from ML models and ChatGPT's more frequent references to certain parameters on the same methodological grounds. ChatGPT's frequent reference to certain parameters should be interpreted not as a finding equivalent to quantitative feature importance but as an post-hoc explanatory output generated after the decision. For this reason, the importance of correct prompt selection to ensure that generated responses meet the same standard has been reported in the literature, and the IPC method has been proposed.[44,45] Iterative prompt calibration can guide models that tend to generate open-ended, variable responses to produce more structured and consistent outputs.[15,43,45] This approach is particularly important for reducing the risk of hallucination.[46] However, it is necessary to conduct calibration processes on small test subsets that are independent of the main dataset to prevent data leakage and to maintain scientific validity.[45] In our study, an additional dataset of 30 patients was used for IPC, and the most appropriate prompt was obtained using the aforementioned method.

The findings of the present study provide insights into how LLMs such as ChatGPT-5.0 and ML algorithms such as XGBoost can be integrated into routine orthodontic decision-making. These models can be particularly useful in two clinical scenarios. First, these algorithms can be used as a second opinion and verification tool for orthodontic students and physicians with limited clinical experience. In situations where complex, multifactorial data (age, degree of crowding, cephalometric angles) need to be evaluated simultaneously, AI systems can provide a standardized decision-making process by systematically filtering clinical parameters that clinicians might otherwise overlook. Second, especially in borderline cases when the physician is undecided between extraction and non-extraction treatment, the predictions offered by these models can serve as a clinical prompt, triggering are-evaluation of treatment risks. In our study, ChatGPT-5.0's relatively higher sensitivity in detecting extraction cases suggests that the model might have a more aggressive tendency when making such risky decisions and might help physicians critically question their own decisions. Consequently, the aim of these systems is not to replace the physician but to create an evidence-based, reproducible hybrid decision-support system that minimizes clinical oversights as patient loads increase.

Study Limitations

Certain limitations should be considered when evaluating the findings obtained from this study. Since this is, to the best of our knowledge, one of the first studies to evaluate ChatGPT-5.0's performance in orthodontic extraction versus non-extraction treatment decisions, the direct comparability of the findings with previous literature is limited. Data obtained from a single center may limit the generalizability of the treatment protocols. Although the local, isolated system structure used in the study was chosen for data security, it does not completely eliminate the risk of unauthorized access or other systemic vulnerabilities. For ChatGPT-5.0, intra-model consistency between sessions or across time was not directly evaluated. To reduce variability, all evaluations were conducted on the same day by a single researcher using a standardized prompt structure; a new session was started after each case. However, this approach is a methodological standardization measure, not a direct consistency analysis, and it does not completely eliminate the risk of user-dependent bias. Furthermore, ChatGPT-5.0's image-processing capacity is still limited compared to that of traditional ML algorithms, and the performance of NLP-based models is sensitive to prompt quality; therefore, the obtained results reflect ChatGPT-5.0's capabilities at the time the research was conducted. Despite this, by comparing ChatGPT-5.0 with traditional ML models using the same sample, this study makes an original contribution to the existing literature. Additionally, it provides a methodological framework by demonstrating that ChatGPT-5.0's explanatory outputs are not directly equivalent to quantitative feature importance metrics in traditional models. Future studies should focus on improving LLMs with multimodal training approaches, evaluating clinician-LLM hybrid decision support systems, and testing these models with longitudinal clinical validation designs. In this respect, the study constitutes a starting point for future research involving misclassification patterns, subgroup performance, and more detailed explanatory analyses.

CONCLUSION

This study compared the performance of ChatGPT-5.0's with that of traditional ML algorithms in distinguishing between orthodontic extraction and non-extraction treatment decisions. The findings showed a statistically significant difference in performance when all models were evaluated together. Therefore, the null hypothesis (H0), which predicted no difference between the LLM-based approach and ML algorithms in classification performance, was rejected, and the alternative hypothesis (H1), which predicted a statistically significant difference between the two approaches, was accepted.

However, the fact that ChatGPT-5.0 did not show a statistically significant difference in accuracy compared to XGBoost in pairwise comparisons, while it exhibited significantly higher performance than random forest, SVM, logistic regression, and MLP models, indicates that this result is not consistent across models. Although ChatGPT-5.0 performed better than or differently from some traditional ML models, it yielded results similar to those of the highest-performing model, XGBoost.

These findings indicate that ChatGPT-5.0 has notable potential as a clinical decision support tool in orthodontics; however, some task-oriented optimized ML algorithms can maintain their superiority, especially in terms of overall classification balance and specificity.

Ethics

Ethics Committee Approval: This study was approved by the Institutional Afyonkarahisar Health Sciences University Non-Intervention Scientific Research Ethics Committee (approval no: 2024/8, date: 04.10.2024).

Informed Consent: Written and verbal informed consent was obtained from all patients and their legal guardians.

Footnotes

Author Contributions: Concept - A.Y.; Design - A.Y., H.C.; Data Collection and/or Processing - M.S.; Analysis and/or Interpretation - H.C., M.S.; Literature Search - A.Y., M.S., H.C.; Writing - A.Y., H.C., M.S.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. Camcı H, Salmanpour F. Comparing the esthetic impact of virtual mandibular advancement, bichectomy, jawline, and their combination. *Am J Orthod Dentofacial Orthop.* 2023;163(6):756-765. [CrossRef]
2. Camcı H, Salmanpour F. Impact of intraoral scanning conditions on the accuracy virtual aligners (VA). *Australasian Orthodontic Journal.* 2022;38(1):102-110. [CrossRef]
3. Salmanpour F, Camcı H. Prediction of patient cooperation before orthodontic treatment: handwriting and artificial intelligence. *J World Fed Orthod.* 2024;13(6):303-309. [CrossRef]
4. Ahmed N, Abbasi MS, Zuberi F, et al. Artificial intelligence techniques: analysis, application, and outcome in dentistry-a systematic review. *Biomed Res Int.* 2021;2021:9751564. [CrossRef]
5. Salmanpour F, Camci H. Artificial intelligence for predicting orthodontic patient cooperation: voice records versus frontal photographs. *APOS Trends Orthod.* 2024;14:255-263. [CrossRef]
6. Radwan MT, Sin Ç, Akkaya N, Vahdettin L. Artificial intelligence-based algorithm for cervical vertebrae maturation stage assessment. *Orthod Craniofac Res.* 2023;26(3):349-355. [CrossRef]
7. Camcı H, Salmanpour F. Estimating the size of unerupted teeth: moyers vs deep learning. *Am J Orthod Dentofacial Orthop.* 2022;161(3):451-456. [CrossRef]
8. Fawaz P, Sayegh PE, Vannet BV. What is the current state of artificial intelligence applications in dentistry and orthodontics? *J Stomatol Oral Maxillofac Surg.* 2023;124(5):101524. [CrossRef]
9. Dipalma G, Inchingolo AD, Inchingolo AM, et al. Artificial intelligence and its clinical applications in orthodontics: a systematic review. *Diagnostics (Basel).* 2023;13(24):3677. [CrossRef]
10. Salmanpour F, Camcı H, Geniş Ö. Comparative analysis of AI chatbot (ChatGPT-4.0 and microsoft copilot) and expert responses to common orthodontic questions: patient and orthodontist evaluations. *BMC Oral Health.* 2025;25(1):896. [CrossRef]
11. Hakami Z, Saheb SAK, Bawazeer OA. Orthodontic knowledge assessment: a comparison of five AI chatbots. *Saudi Dent J.* 2026;38(3):20. [CrossRef]
12. Zheng J, Ding X, Pu JJ, et al. Unlocking the potentials of large language models in orthodontics: a scoping review. *Bioengineering.* 2024;11(11):1145. [CrossRef]
13. Çelik İH, Camcı H, Salmanpour F. Bridging the information gap in pediatric dentistry: a comparison of ChatGPT-4o, Google Gemini Advanced, and expert responses based on evaluations by parents and pediatric dentists. Journal of Clinical Pediatric Dentistry. 2026. 50(1);147-155. [CrossRef]
14. Ozkan E, Koyun M. Atlas-assisted bone age estimation from hand-wrist radiographs using multimodal large language models: a comparative study. *Diagnostics.* 2026;16(3):487. [CrossRef]
15. Akpınar M, Salmanpour F. Chat generative pretrained transformer-4.0's accuracy in assessing cervical vertebrae and hand-wrist maturation stages: a retrospective study. *Am J Orthod Dentofacial Orthop.* 2025;168(6):753-763. [CrossRef]
16. Giannakopoulos K, Kavadella A, Aaqel Salim A, Stamatopoulos V, Kaklamanos EG. Evaluation of the performance of generative AI large language models ChatGPT, Google Bard, and Microsoft Bing Chat in supporting evidence-based dentistry: comparative mixed methods study. *J Med Internet Res.* 2023;25:e51580. [CrossRef]
17. Morishita M, Fukuda H, Muraoka K, et al. Evaluating GPT-4V's performance in the Japanese national dental examination: a challenge explored. *J Dent Sci.* 2024;19(3):1595-1600. [CrossRef]
18. Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. *Nature.* 2023;616(7956):259-265. [CrossRef]
19. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature.* 2023;620(7972):172-180. Erratum in: *Nature.* 2023;620(7973):E19. [CrossRef]
20. Cicek O, Özçelik YB, Altan A. A new approach based on metaheuristic optimization using chaotic functional connectivity matrices and fractal dimension analysis for AI-driven detection of orthodontic growth and development stage. *Fractal and Fractional.* 2025;9(3):148. [CrossRef]
21. Subramanian AK, Chen Y, Almalki A, Sivamurthy G, Kafle D. Cephalometric analysis in orthodontics using artificial intelligence-a comprehensive review. *Biomed Res Int.* 2022;2022:1880113. [CrossRef]

22. Yıldırım A, Cicek O, Genç YS. Can AI-based chatgpt models accurately analyze hand-wrist radiographs? A comparative study. *Diagnostics (Basel)*. 2025;15(12):1513. [CrossRef]
23. Farzanegan P, Zarabadi M, Najary S, Tahmasbi MA, Behnaz M. The role of artificial intelligence in orthodontics for determining skeletal age based on cervical vertebra maturation degree: a comprehensive review. *Health Sci Rep*. 2025;8(11):e71487. [CrossRef]
24. Suhail Y, Upadhyay M, Chhibber A, Kshitiz. Machine learning for the diagnosis of orthodontic extractions: a computational analysis using ensemble learning. *Bioengineering (Basel)*. 2020;7(2):55. [CrossRef]
25. Choi HI, Jung SK, Baek SH, et al. Artificial intelligent model with neural network machine learning for the diagnosis of orthognathic surgery. *J Craniofac Surg*. 2019;30(7):1986-1989. Erratum in: J Craniofac Surg. 2020;31(4):1156. [CrossRef]
26. Ganguly R, Suri L, Patel F. A literature review of t extraction decision and outcomes in orthodontic treatment. *J Mass Dent Soc*. 2016;65(2):28-31. [CrossRef]
27. Jung SK, Kim TW. New approach for the diagnosis of extractions with neural network machine learning. *Am J Orthod Dentofacial Orthop*. 2016;149(1):127-133. [CrossRef]
28. Prasad J, Mallikarjunaiah DR, Shetty A, Gandedkar N, Chikkamuniswamy AB, Shivashankar PC. Machine learning predictive model as clinical decision support system in orthodontic treatment planning. *Dent J (Basel)*. 2022;11(1):1. [CrossRef]
29. Kılınç DD, Mansız D. Examination of the reliability and readability of chatbot generative pretrained transformer's (ChatGPT) responses to questions about orthodontics and the evolution of these responses in an updated version. *Am J Orthod Dentofacial Orthop*. 2024;165(5):546-555. [CrossRef]
30. Kök H, Acilar AM, İzgi MS. Usage and comparison of artificial intelligence algorithms for determination of growth and development by cervical vertebrae stages in orthodontics. *Prog Orthod*. 2019;20(1):41. [CrossRef]
31. Abuabara A, do Nascimento TVPM, Trentini SM, et al. Evaluating the accuracy of generative artificial intelligence models in dental age estimation based on the Demirjian's method. *Front Dent Med*. 2025;6:1634006. [CrossRef]
32. Çelik İH, Salmanpour F. Can ChatGPT-5 estimate dental age? A comparative study with Demirjian and Willems methods in paediatric dentistry. *BMC Oral Health*. 2026;26(1):849. [CrossRef]
33. Mohammad-Rahimi H, Nadimi M, Rohban MH, Shamsoddin E, Lee VY, Motamedian SR. Machine learning and orthodontics, current trends and the future opportunities: a scoping review. *Am J Orthod Dentofacial Orthop*. 2021;160(2):170-192.e4. [CrossRef]
34. Khanagar SB, Al-Ehaideb A, Vishwanathaiah S, et al. Scope and performance of artificial intelligence technology in orthodontic diagnosis, treatment planning, and clinical decision-making - a systematic review. *J Dent Sci*. 2021;16(1):482-492. [CrossRef]
35. OpenAI. Karşınızda GPT-5. [CrossRef]
36. Vassis S, Powell H, Petersen E, et al. Large-language models in orthodontics: assessing reliability and validity of ChatGPT in pretreatment patient education. *Cureus*. 2024;16(8):e68085. [CrossRef]
37. Köktürk B, Pamukçu H, Gözüaçık Ö. Evaluation of different machine learning algorithms for extraction decision in orthodontic treatment. *Orthod Craniofac Res*. 2024;27 Suppl 2(Suppl 2):13-24. [CrossRef]
38. Kunz F, Stellzig-Eisenhauer A, Zeman F, et al. Artificial intelligence in orthodontics. *J Orofac Orthop*. 2010;81:52-68. [CrossRef]
39. Mason T, Kelly KM, Eckert G, Dean JA, Dundar MM, Turkkahraman H. A machine learning model for orthodontic extraction/non-extraction decision in a racially and ethnically diverse patient population. *Int Orthod*. 2023;21(3):100759. [CrossRef]
40. Etemad LE, Heiner JP, Amin AA, et al. Effectiveness of machine learning in predicting orthodontic tooth extractions: a multi-institutional study. *Bioengineering (Basel)*. 2024;11(9):888. [CrossRef]
41. Nordblom NF, Büttner M, Schwendicke F. Artificial intelligence in orthodontics: critical review. *J Dent Res*. 2024;103(6):577-584. [CrossRef]
42. Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models. *Adv. Neural Inf. Process. Syst.* 2022;35. [CrossRef]
43. Salmanpour F, Akpınar M. Performance of chat generative pretrained transformer-4.0 in determining labiolingual localization of maxillary impacted canine and presence of resorption in incisors through panoramic radiographs: a retrospective study. *American Journal of Orthodontics and Dentofacial Orthopedics*. 2025;168(2):220-231. [CrossRef]
44. Jha S, Jha SK, Lincoln P, Bastian ND, Velasquez A, Neema S. dehallucinating large language models using formal methods guided iterative prompting. Proceedings - 2023 IEEE International Conference on Assured Autonomy ICAA. 2023:149-152. [CrossRef]
45. Wang J, Sun Y, Liang Y, Li X, Gong B. Iteratively calibrating prompts for unsupervised diverse opinion summarization. *Frontiers in Artificial Intelligence and Applications*. 2024;392:3939-3946. [CrossRef]
46. Zhao Z, Wang B, Ouyang L, Dong X, Wang J, He C. Beyond hallucinations: enhancing LVLMs through hallucination-aware direct preference optimization. *arXiv*. 2023. [CrossRef]

DOI: 10.4274/TurkJOrthod.2026.2026.17

TURKISH JOURNAL OF
ORTHODONTICS

Original Article

Evaluation of Oxidizable Substance Interaction Associated with Orthodontic Clear Aligners and an Acrylic Material: A Preliminary *in vitro* Study

Merve Kurnaz[1], Bora İsmail Biren[1], Ataberk Kayhan[1], Gökhan Biçim[2], Sibel Biren[1], Yegane Güven[2]

[1]İstanbul Kent University Faculty of Dentistry, Department of Orthodontics, İstanbul, Türkiye
[2]İstanbul Kent University Faculty of Dentistry, Department of Basic Medical Sciences Biochemistry, İstanbul, Türkiye

Cite this article as: Kurnaz M, Biren Bİ, Kayhan A, Biçim G, Biren S, Güven Y. Evaluation of oxidizable substance interaction associated with orthodontic clear aligners and an acrylic material: a preliminary *in vitro* study. *Turk J Orthod.* 2026; 39(3): 155-160

Main Points

- All tested materials showed measurable interaction within 24 hours, although the extent varied among materials.
- The greatest interaction of oxidizable substances was observed in ClearQuartz, whereas SmartTrack exhibited the lowest levels.
- Potassium permanganate may be useful as a simple screening approach, but it does not identify or quantify specific compounds.

ABSTRACT

Objective: This study aimed to comparatively evaluate the interactions of oxidizable substances associated with different orthodontic aligner materials using potassium permanganate ($KMnO_4$).

Methods: In this study, materials from three different clear-aligner manufacturers (SmartTrack, ClearQuartz, and Scheu Dental CA Pro+) and one acrylic block were used. The materials were mechanically fragmented and exposed to $KMnO_4$, a strong oxidizing agent. Potassium permanganate solutions were prepared at six concentrations (1.24 mM, 0.62 mM, 0.31 mM, 0.16 mM, 0.08 mM, 0.04 mM) to establish a calibration curve (R^2=0.9999). Based on this calibration, 0.2 mM and 1 mM solutions were selected for the experimental phase. The color change was evaluated spectrophotometrically at 1, 4, and 24 hours to assess interactions with oxidizable substances.

Results: The amount of oxidizable substances interacting with $KMnO_4$ was estimated in mM. Oxidizable substances were detected in all samples (Acrylic Block 0.0723 mM, ClearQuartz 0.0826 mM, Scheu Dental CA Pro+ 0.0584 mM, SmartTrack 0.0271 mM); however, the estimated levels of oxidizable-substance interaction varied among samples. The highest level was observed in ClearQuartz at 24 hours.

Conclusion: Orthodontic aligner materials exhibit varying levels of interaction with oxidizable substances, particularly during the first 24 hours. The method may serve as a simple preliminary screening tool. Future studies should investigate the composition of the released substances using more sensitive analytical methods.

Keywords: Clear aligners, oxidizable substances, potassium permanganate, polymer degradation

INTRODUCTION

In recent years, clear aligners have been widely used in orthodontics as an alternative to traditional fixed appliances.[1] These clear aligners should be used for at least 22 hours per day to deliver sufficient force to achieve optimal orthodontic treatment outcome. The number of companies producing clear aligners is increasing. As the number of companies increases, clear aligner treatments and the materials used are also evolving.[2,3]

Corresponding author: Bora İsmail Biren, **e-mail:** borabiren96@gmail.com **ORCID:** orcid.org/0009-0003-2938-5717
Received: January 29, 2026 **Accepted:** June 22, 2026 **Epub:** July 7, 2026 **Publication Date:** September 25, 2026

Bisphenol A (BPA) is an industrial chemical used in the production of polycarbonate plastics and epoxy resins.[4] Bisphenol A is a well-known endocrine disruptor with hormone-like effects. Due to its estrogen-like structure, BPA may cause hormonal imbalance.[4] In addition, BPA can cause oxidative stress, deoxyribonucleic acid damage, apoptosis, and immune system suppression.[5-7] Recent studies have also examined possible relationships between BPA exposure and some types of cancer.[5-7]

Clear aligners used in orthodontics are generally manufactured from thermoplastic materials based on polyurethane, such as ClearQuartz (Straumann Group, Switzerland) and SmartTrack (Align Technology, USA), or polyethylene terephthalate glycol, such as CA Pro+ (Scheu Dental, Iserlohn, Germany).[8,9]

Despite growing evidence of chemical leaching from aligner materials, most analytical techniques for identifying specific compounds require advanced instrumentation and are not always practical for routine screening.[10] Potassium permanganate ($KMnO_4$), a strong but non-specific oxidizing agent, has been proposed as a simple method for estimating total oxidizable substance interaction associated with thermoplastic polymers by measuring discoloration resulting from oxidation-reduction reactions.[10]

In a previous study, BPA release from three commercially available orthodontic aligner brands was evaluated in artificial saliva, artificial gastric fluid, and ethanol.[11] Samples were analyzed using high-performance liquid chromatography (HPLC)-tandem mass spectrometry, and all brands released BPA, with great variability.[11] In another study, four different clear aligners were sectioned and immersed in glass vials containing ethanol at different concentrations, and the leached substances were detected by gas chromatography-mass spectrometry (GC-MS).[12]

However, there is ongoing debate about whether other components of clear aligners degrade due to oral temperature, salivary enzymes, and mechanical abrasion, releasing of a variety of oxidizable products over time.

In light of the literature, the present study aimed to evaluate the interaction of oxidizable substances released from different brands of clear aligners and a conventionally used acrylic material, using a simple biochemical assessment method that can be performed under basic laboratory conditions. By comparing the interaction profiles of these orthodontic materials using a practical, accessible approach, this study sought to provide relevant insights into their potential chemical interactions and biocompatibility. The null hypothesis was that there would be no significant difference in the interaction of oxidizable substances among the tested materials.

METHODS

Orthodontic Material Selection

In this study, three thermoformed clear-aligner materials from different brands and one acrylic material commonly used in removable orthodontic appliances were evaluated. The clear aligner materials included ClearQuartz (Straumann Group, Switzerland), SmartTrack (Align Technology, USA), and CA Pro+ (Scheu Dental, Iserlohn, Germany). The aligners were obtained directly from the manufacturers. All samples were unused and had not been exposed to clinical conditions, and were removed from their original packaging immediately before testing without further processing. The acrylic material (Integra Orthodontic Acrylic, BGD, Türkiye) was fabricated in block form by an orthodontic laboratory in İstanbul, Türkiye.

Sample Solution Selection

Potassium permanganate was used to evaluate the interaction of oxidizable substances because of its simplicity, applicability as a preliminary screening method, and its well-established role as a strong but non-specific oxidizing agent.[10]

Oxidizable Substance Interaction Experiment

Preparation of Orthodontic Materials

The materials were mechanically fragmented with a slow-speed handpiece (KAVO, Biberach, Germany) fitted with a 12-blade tungsten carbide bur (Meisinger, Neuss, Germany) at 25,000 rpm (Figure 1). Fragmentation was performed to standardize the surface area of the specimens and enhance their interaction with the $KMnO_4$ solution in this preliminary screening model. This approach may increase the surface area and potentially overestimate interactions with oxidizable substances compared with those under clinical conditions. To prevent contamination, unused, a new sterilized bur was used for each material. A standardized sample weight of 1 g was prepared for each specimen.

Figure 1. Representative image of the fragmented ClearQuartz sample prepared for the potassium permanganate ($KMnO_4$) assay.

Potassium Permanganate Sample Preparation

Potassium permanganate solutions were prepared using distilled water (IsoLab, extra pure, 960.172). For calibration, six different concentrations (1.24 mM, 0.62 mM, 0.31 mM, 0.16 mM, 0.08 mM, and 0.04 mM) were prepared by serial dilution. Spectrophotometric measurements were performed at a wavelength of 525 nm. A calibration curve was constructed using these standard solutions, and linear regression analysis demonstrated an excellent linear relationship between concentration and absorbance (absorbance =2.578× Concentration + 0.0717; R^2=0.9999). The absorbance values obtained from the experimental samples were converted into Potassium permanganate concentrations using the rearranged calibration equation:

Concentration (mM) = (Absorbance - 0.0717)/2.578.

Based on the validated linear calibration range, 0.2 mM and 1 mM $KMnO_4$ solutions were selected for the experimental phase, as both concentrations fell within the calibration range and were suitable for the experimental conditions.[10]

Assessment of Discoloration

All experiments were conducted under controlled laboratory conditions at 37 °C. A fixed sample-to-solution ratio of 1 g of material per 100 mL of $KMnO_4$ solution was used for all groups to ensure consistency. All solutions were prepared in distilled water at approximately neutral pH (pH ≈ 7). A negative control consisting of $KMnO_4$ solution without any test material was included to determine baseline absorbance values.

Color change was evaluated spectrophotometrically at 1, 4, and 24 hours to assess interaction with oxidizable substances.

Ethical Approval and Consent

Ethical approval for this study was obtained from the İstanbul Kent University Ethics Committee (approval no: 2025-10, date: 10.12.2025). All procedures were conducted in accordance with relevant ethical guidelines and institutional standards. As no biological or patient-derived materials were used and the study was performed *in vitro*, informed consent was not required.

Statistical Analysis

This study was designed as a preliminary investigation to observe interactions between oxidizable substances in different orthodontic materials under standardized laboratory conditions. Each material was tested in 0.2 mM and 1 mM $KMnO_4$ solutions, and measurements were recorded after 1 hour, 4 hours, and 24 hours. Since the same experimental setup was monitored over time, the measurements reflect changes observed within each material group rather than the results of independent replicate experiments.

Changes in $KMnO_4$ concentration during the observation period were used to compare the behavior of the tested materials. Lower $KMnO_4$ values were associated with greater interaction with oxidizable substances. The purpose of this preliminary study was to identify general patterns and compare the materials under the same conditions.

Because independent replicate experiments were not conducted, statistical comparisons were not performed. The findings are therefore presented descriptively.

RESULTS

The fragmented material samples were exposed to 0.2 mM and 1 mM $KMnO_4$ solutions. Color change was measured spectrophotometrically (PG Instruments T80+, UK) at 1, 4, and 24 hours.

Oxidizable substances were detected indirectly in all samples, with variations across materials and time intervals. The amount of oxidizable substance interacting with $KMnO_4$ was estimated in mM from absorbance values using the calibration curve. Progressive discoloration over time was observed, corresponding to a decrease in $KMnO_4$ concentration and indicating increased interaction with oxidizable substances. The greatest change was observed in the ClearQuartz group at 24 hours, whereas the smallest change was recorded in the SmartTrack group. The quantitative results are presented in Table 1.

Table 1. Measured oxidizable substance levels (mM) calculated based on $KMnO_4$ reduction

Materials	1 hour concentration (mM)	4 hour concentration (mM)	24 hour concentration (mM)	Concentration level
Blank	1.00	1.00	**1.00**	1 mM
Acrylic block	1.00	0.99	**0.94**	1 mM
ClearQuartz	0.96	0.89	**0.71**	1 mM
Scheu Dental	0.98	0.96	**0.89**	1 mM
SmartTrack	1.00	1.00	**0.98**	1 mM
Blank	0.20	0.20	**0.20**	0.2 mM
Acrylic block	0.20	0.18	**0.12**	0.2 mM
ClearQuartz	0.17	0.16	**0.09**	0.2 mM
Scheu Dental	0.18	0.14	**0.12**	0.2 mM
SmartTrack	0.19	0.18	**0.17**	0.2 mM

Lower $KMnO_4$ concentration values indicate greater interaction of oxidizable substances.
$KMnO_4$, potassium permanganate.

In the 0.2 mM $KMnO_4$ solution, all materials showed progressive discoloration, indicating a reduction in $KMnO_4$ concentration (Figure 2).

In the 1 mM $KMnO_4$ solution, a similar trend was observed, with increasing discoloration over time. The greatest change occurred in the ClearQuartz group at 24 hours, whereas the SmartTrack group showed only minimal variation (Figure 3).

DISCUSSION

The use of clear aligners has increased substantially in recent years, and this trend has been accompanied by continuous developments in material composition. Clear aligners are thermoplastic materials thermoformed to fit three-dimensional dental models; their composition varies among manufacturers. As a result, their potential biological effects remain a subject of ongoing investigation.[8]

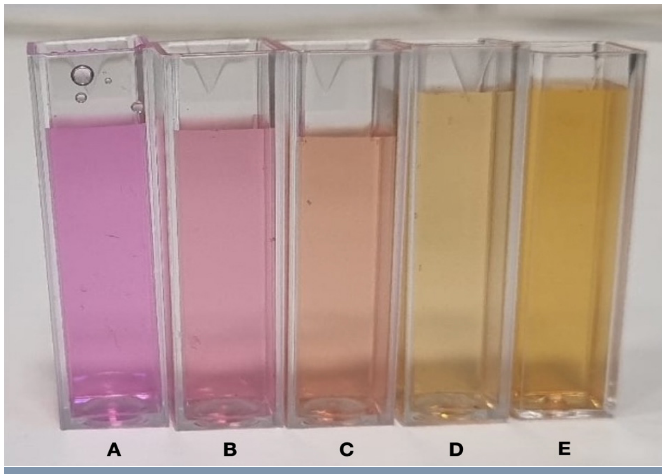


Figure 2. Representative appearance of the samples immersed in 0.2 mM potassium permanganate ($KMnO_4$) solution after 24 hours of incubation at 37 °C. From left to right: (A) Control, (B) SmartTrack, (C) CA Pro+ (Scheu Dental), (D) ClearQuartz, and (E) Acrylic block.

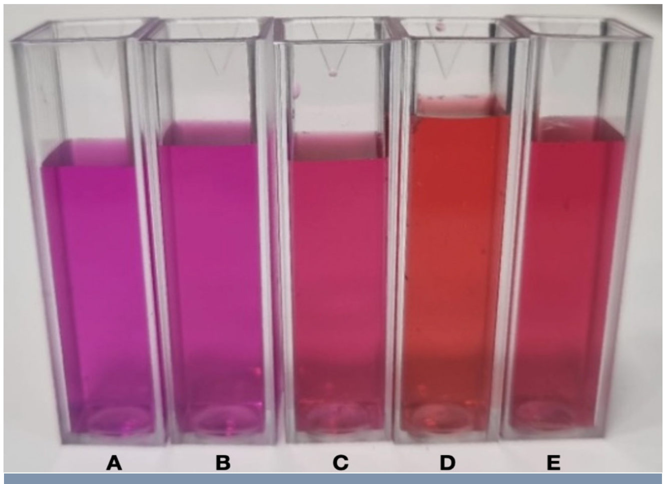


Figure 3. Representative appearance of the samples immersed in 1.0 mM potassium permanganate ($KMnO_4$) solution after 24 hours of incubation at 37 °C. From left to right: (A) Control, (B) SmartTrack, (C) CA Pro+ (Scheu Dental), (D) ClearQuartz, and (E) Acrylic block.

The present *in vitro* study demonstrated that all tested orthodontic materials reduced $KMnO_4$ within the first 24 hours, indicating the presence of oxidizable substances capable of reacting with the oxidizing agent under the experimental conditions. These findings highlight the importance of considering the early chemical behavior of orthodontic materials, as even initial interactions may contribute to cumulative biological exposure during treatment.

Furthermore, the results emphasize the potential value of simple biochemical evaluation methods for providing clinically relevant information about the chemical stability and biocompatibility of materials routinely used in orthodontic practice. The increase in discoloration over time reflected the reduction of $KMnO_4$. Since $KMnO_4$ is a strong but non-specific oxidizing agent, the observed discoloration should be considered an indicator of overall oxidizable load rather than a direct measurement of specific monomers or degradation products.[10,12]

Although $KMnO_4$ is not present in the oral environment, it was selected as a strong, non-specific oxidizing agent to estimate the overall oxidizable load associated with orthodontic materials. Similar approaches have been used in polymer research as preliminary screening tools to assess chemical reactivity. This approach differs from conventional methods used in previous studies, in which materials are incubated in artificial saliva, gastric fluid, or organic solvents such as ethanol, followed by compound-specific analyses using techniques such as HPLC or GC-MS.[11,12]

Compared with these established methodologies, the $KMnO_4$ method offers a simple and inexpensive screening approach that does not require sophisticated instrumentation. This may enhance its applicability as a preliminary assessment tool in material comparison studies.[10] However, unlike immersion models that allow identification and quantification of specific leached compounds, such as BPA, the $KMnO_4$ method provides only indirect and non-specific information on overall interactions with oxidizable substances.[11,12]

In a previous study, three commercially available clear aligner systems (SmileDirect, Invisalign, and Essix Ace) were immersed in artificial saliva, artificial gastric fluid, and ethanol, and BPA release was quantified using HPLC-tandem mass spectrometry. Samples were analyzed using HPLC-tandem mass spectrometry, and BPA release was detected in all materials.[11] These findings are consistent with the present results and suggest that clear aligner materials may exhibit oxidative reactivity under experimental conditions.

In another study, materials from four clear aligner systems (Invisalign, Eon, SureSmile, and Clarity) were evaluated in ethanol at different concentrations, and leaching was detected by GC-MS. A total of eleven different chemical compounds were identified.[12] These findings support the presence of multiple leachable substances in clear aligner materials and are consistent with the present results. Although the current method does not allow identification of specific compounds, the observed permanganate reduction may reflect interactions with the overall oxidizable substance.[13,14]

Azhagudurai et al.[13] reported measurable amounts BPA release from four different clear aligner materials in an *ex vivo* setting, with the highest release occurring within the first 24 hours. These findings indicate that the initial period of aligner wear is critical for chemical leaching; this is consistent with the present results, which show greater permanganate consumption at earlier time points.

In a similar study, BPA leaching from both thermoformed clear retainer materials and acrylic resin was investigated. Detectable BPA release was observed in both groups, particularly under thermal and mechanical stress. These findings are comparable with the present results and further support the oxidative reactivity observed in orthodontic materials.[14]

The method used in the present study is intended as a comparative screening approach rather than a definitive analytical technique.

Clinical Relevance

From a clinical perspective, immersion models using artificial saliva may better simulate the oral environment by incorporating factors such as of saliva and temperature on material degradation.[11,13] In contrast, the $KMnO_4$ method does not directly simulate the oral environment, but it may provide a simple way to compare the oxidative behavior of different materials.

It should also be considered that some immersion methods, particularly those using ethanol, may increase substance release because of their strong extraction effect.[12] Unlike compound-specific analytical methods, the $KMnO_4$ approach provides indirect information about the overall interaction between the tested material and the oxidizing solution.

Clear aligners and acrylic orthodontic appliances remain in prolonged contact with the oral environment during clinical use. The present findings indicate that orthodontic materials may exhibit oxidative behavior, particularly during the initial phase of use. Although the biological effects of these substances were not directly evaluated, the material-dependent differences observed in leaching behavior may still provide preliminary insight into materials' behavior. However, the present findings cannot be directly translated to clinical conditions. These findings also highlight the importance of continued research and material development to improve the chemical stability and biocompatibility of orthodontic appliances.

Study Limitations

The present study has some limitations that should be considered when interpreting the findings. First, $KMnO_4$ is a non-specific oxidizing agent; therefore, the observed reduction reflects the total oxidizable content rather than the release of specific monomers or degradation products. In addition, the method may overestimate substance release by including reactions with residual particles or polymer surface oxidation.

Another limitation is that the materials were mechanically fragmented into small particles before testing to standardize the surface area and enhance interaction with the $KMnO_4$ solution. While this approach allowed a controlled comparison between materials, it may have increased the effective surface area and led to an overestimation of oxidizable substance interactions compared with intact aligners used under clinical conditions.

Additionally, the experiments were conducted *in vitro* using distilled water, which does not adequately simulate the complex oral environment, including salivary composition, temperature variations, and mechanical forces. Therefore, the findings should be interpreted as preliminary screening results rather than a direct representation of clinical behavior.

Although immersion techniques remain the gold standard for identifying and quantifying specific leachables, the $KMnO_4$ method may still be useful as a practical, complementary tool for preliminary screening and comparative evaluation of material behavior. Future studies combining both approaches under clinically relevant conditions could help clarify material biocompatibility and degradation processes.

Despite these limitations, the $KMnO_4$ method remains a simple and practical approach for the initial evaluation of the interaction of oxidizable substances in orthodontic materials. Further studies under clinically relevant conditions, supported by more advanced analytical techniques, are needed to identify the specific compounds involved and better understand their possible biological effects.

CONCLUSION

Within the limitations of this preliminary *in vitro* study, all tested orthodontic materials demonstrated measurable interactions with $KMnO_4$, indicating the presence of oxidizable substances. The extent of interaction varied among the investigated materials, with ClearQuartz exhibiting the highest interaction with oxidizable substances and SmartTrack showing the lowest interaction levels under the tested conditions.

These findings suggest that clear aligners and acrylic orthodontic materials may differ in their chemical reactivity and potential degradation behavior. Although the $KMnO_4$ method does not identify or quantify specific compounds, it may serve as a simple, practical, and accessible preliminary screening approach for the comparative evaluation of orthodontic materials. Future studies should investigate the composition of the released substances using more sensitive analytical methods.

Ethics

Ethics Committee Approval: Ethical approval for this study was obtained from the İstanbul Kent University Ethics Committee (approval no: 2025-10, date: 10.12.2025).

Informed Consent: As no biological or patient-derived materials were used and the study was performed *in vitro*, informed consent was not required.

Footnotes

Author Contributions: Concept - M.K., B.İ.B., A.K., G.B., S.B., Y.G.; Design - M.K., B.İ.B., A.K., G.B., S.B., Y.G.; Data Collection and/or Processing - B.İ.B., A.K., G.B.; Analysis and/or Interpretation - B.İ.B., A.K., G.B.; Literature Search - M.K., B.İ.B., A.K., G.B., S.B., Y.G.; Writing - M.K., B.İ.B., A.K., G.B., S.B., Y.G.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. Zheng M, Liu R, Ni Z, Yu Z. Efficiency, effectiveness and treatment stability of clear aligners: a systematic review and meta-analysis. *Orthod Craniofac Res.* 2017;20(3):127-133. [CrossRef]
2. Timm LH, Farrag G, Baxmann M, Schwendicke F. Factors influencing patient compliance during clear aligner therapy: a retrospective cohort study. *J Clin Med.* 2021;10(14):3103. [CrossRef]
3. AlMogbel A. Clear aligner therapy: up to date review article. *J Orthod Sci.* 2023;12:37. [CrossRef]
4. Frankowski R, Zgoła-Grześkowiak A, Grześkowiak T, Sójka K. The presence of bisphenol A in the thermal paper in the face of changing European regulations - a comparative global research. *Environ Pollut.* 2020;265(Pt A):114879. [CrossRef]
5. Lee SG, Kim JY, Chung JY, et al. Bisphenol A exposure during adulthood causes augmentation of follicular atresia and luteal regression by decreasing 17β-estradiol synthesis via downregulation of aromatase in rat ovary. *Environ Health Perspect.* 2013;121(6):663-669. [CrossRef]
6. Hassan ZK, Elobeid MA, Virk P, et al. Bisphenol A induces hepatotoxicity through oxidative stress in rat model. *Oxid Med Cell Longev.* 2012;2012:194829. [CrossRef]
7. Rochester JR. Bisphenol A and human health: a review of the literature. *Reprod Toxicol.* 2013;42:132-155. [CrossRef]
8. Sharma R, Rajput A, Gupta KK, Sharma H. Clear aligner: invisalign: a review. *IP Indian Journal of Orthodontics and Dentofacial Research.* 2018;4(4):173-175. [CrossRef]
9. Madhyan D, Khatri PT, Kolge NE, Rodrigues LW, Lodd MM, Chaudhari R. Comparative assesment of optical properties of PETF, PU & PETC based orthodontic clear aligners exposed to a common staining agent: an- invitro study. *JCHR*. 2024;14(4):899-910. [CrossRef]
10. Ansari SA, Chaudhari AR, Patil C, & Pullori SR. Simple method of testing polymer leaching from thermoplastic sheets used for clear aligner. *APOS Trends Orthod.* 2014;4(2):66-69. [CrossRef]
11. Katras S, Ma D, Dayeh AA, Tipton D. Bisphenol a release from orthodontic clear aligners: an in-vitro study. *Recent Progress in Materials.* 2021;3(3):034. [CrossRef]
12. Alhendi A, Khounganian R, Almudhi A, Ahamad SR. Leaching of different clear aligner systems: an in vitro study. *Dent J (Basel).* 2022;10(2):27. [CrossRef]
13. Azhagudurai N, Rajendran R, Aishwarya K, Rajendrababu S, Kumar S, Reddy M. Detecting bisphenol a leaching from four different commercially available clear aligner sheets: an ex vivo study. *J Contemp Dent Pract*. 2024;25(6):535-539. [CrossRef]
14. Kotyk MW, Wiltshire WA. An investigation into bisphenol-a leaching from orthodontic materials. *Angle Orthod.* 2014;84(3):516-520. [CrossRef]

DOI: 10.4274/TurkJOrthod.2026.2026.8

TURKISH JOURNAL OF
ORTHODONTICS

Original Article

In vivo Evaluation of Gingival Thickness and Gingival Recession after Fixed Orthodontic Treatment in Patients with Normally Positioned Mandibular Incisors

Ahmet Can Yılmaz[1], Özer Alkan[2]

[1]Kütahya Health Sciences University Faculty of Dentistry, Department of Orthodontics, Kütahya, Türkiye
[2]Ankara Yıldırım Beyazıt University Faculty of Dentistry, Department of Orthodontics, Ankara, Türkiye

Cite this article as: Yılmaz AC, Alkan Ö. *In vivo* evaluation of gingival thickness and gingival recession after fixed orthodontic treatment in patients with normally positioned mandibular incisors. *Turk J Orthod.* 2026; 39(3): 161-168

Main Points

- No clinical gingival recession was observed in patients with normally positioned mandibular incisors during the observation period.
- Tooth- and level-specific changes in gingival thickness were detected after fixed orthodontic treatment.
- Labial gingival thickness should be considered during orthodontic treatment planning and periodontal monitoring.

ABSTRACT

Objective: This *in vivo* study evaluated clinical gingival recession and labial gingival thickness following fixed orthodontic treatment in patients with normally positioned mandibular incisors.

Methods: Twenty-six patients (18 females and 8 males; aged 11-25 years) with normally positioned mandibular incisors, defined by an initial 1-NB distance of 3-5 mm, were included. All patients received non-extraction fixed orthodontic treatment with the straight-wire technique. Gingival recession was clinically recorded at treatment onset (T0) and 20 days after appliance removal (T1). Gingival thickness was measured by transgingival probing at two labial sites on each mandibular anterior tooth: one apical to the free gingival groove and one coronal to the mucogingival junction. The T0-T1 changes were analyzed using paired samples t-tests; intergroup gingival thickness comparisons were analyzed using independent samples t-tests or Mann-Whitney U tests, with a Bonferroni correction where appropriate.

Results: The mean treatment duration was 28.26 months. The 1-NB distance increased significantly from 4.05±0.94 mm to 4.67±1.51 mm ($p=0.003$), whereas the 1-NB angle and the incisor mandibular plane angle changes were not significant. No clinical gingival recession was detected at T1. Significant tooth- and level-specific changes in gingival thickness were observed.

Conclusion: Within this sample and observation period, no clinical gingival recession was observed after treatment. However, changes in gingival thickness indicate that labial gingival thickness should be considered during orthodontic treatment planning and periodontal monitoring.

Keywords: Orthodontic appliances-fixed, incisors, gingival recession, gingival thickness

INTRODUCTION

Gingival recession is defined as the apical displacement of the gingival margin beyond the cementoenamel junction, resulting in exposure of the root surface to the oral environment.[1,2] Its clinical relevance includes dentin hypersensitivity, root caries, aesthetic concerns, plaque accumulation, and potential loss of periodontal support in advanced cases.[2-4] The etiology of gingival recession is multifactorial and involves periodontal inflammation, traumatic tooth brushing, aging,

Corresponding author: Asst. Prof. Ahmet Can Yılmaz, **e-mail:** dtahmetcanyilmaz@gmail.com **ORCID:** orcid.org/0000-0003-2738-5337
Received: February 06, 2026 **Accepted:** July 08, 2026 **Publication Date:** September 25, 2026

smoking, abnormal frenulum attachment, reduced width of keratinized tissue, a thin periodontal phenotype, alveolar bone dehiscence or fenestration, malocclusion, and orthodontic tooth movement.[1,4,5] The mandibular anterior region is particularly relevant in patients undergoing orthodontic treatment because gingival recession is frequently observed around the mandibular incisors, which are often positioned within a relatively narrow alveolar envelope.[2,3,6]

From an orthodontic perspective, labial displacement or proclination of mandibular incisors toward the limits of the alveolar bone may reduce labial bone support and increase the susceptibility of the marginal soft tissues to recession.[7-9] Cone-beam computed tomography studies have suggested that increases in mandibular incisor advancement and inclination may be associated with reductions in alveolar bone height and width, particularly when tooth movement exceeds periodontal limits.[8] Nevertheless, the relationship between orthodontic treatment, mandibular incisor inclination, and gingival recession remains controversial because previous studies have reported inconsistent findings.[3,7,10] Some studies have found associations between incisor proclination and gingival recession or reduced periodontal support, whereas others have reported no clinically relevant association between orthodontic tooth movement and gingival recession when periodontal conditions are properly controlled.[3,8-11] A recent prospective controlled study reported a higher incidence rate of gingival recession after fixed orthodontic treatment than in untreated controls, although most recession defects were limited in severity.[12]

Periodontal phenotype, particularly gingival thickness, is an important modifier of periodontal tissue response during orthodontic treatment.[1,5,13] A thin periodontal phenotype has been associated with greater marginal tissue instability and increased susceptibility to gingival recession, especially in the presence of inflammation, mechanical trauma, or tooth movement toward the alveolar limits.[1,2,13] Recent evidence has also shown that bleeding on probing, plaque accumulation, age, and gingival phenotype may influence gingival recession during the post-orthodontic or retention period.[5,14] In addition to soft-tissue parameters, recent radiographic studies have indicated that orthodontic treatment and gingival recession may be associated with changes in alveolar trabecular architecture, supporting the need to consider both periodontal soft tissues and alveolar bone structures in orthodontic–periodontal assessment.[14,15]

Despite this growing body of evidence, many previous studies have focused mainly on the presence or prevalence of gingival recession rather than prospectively evaluating site-specific changes in gingival thickness in the mandibular incisor region.[2,7,10] Moreover, evidence remains limited regarding patients whose mandibular incisors are within normal cephalometric limits at T0; this subgroup is clinically relevant for evaluating periodontal soft-tissue responses to routine orthodontic tooth movement.[2,3,9]

Therefore, the prospective clinical study aims to evaluate changes in gingival recession and labial gingival thickness in patients with normally positioned mandibular incisors prior to fixed orthodontic treatment and to determine whether anterior-posterior positional changes of the mandibular incisors are associated with these periodontal soft-tissue outcomes. The null hypotheses of the present study were as follows: $H0_1$: Positional changes in the mandibular incisors during fixed orthodontic treatment are not associated with development of clinical gingival recession. $H0_2$: Positional changes of the mandibular incisors during fixed orthodontic treatment are not associated with changes in labial gingival thickness measured at the apical level of the free gingival groove and at the coronal level of the mucogingival junction.

METHODS

This prospective, single-center clinical study was conducted at the Department of Orthodontics, Van Yüzüncü Yıl University Faculty of Medicine. A total of 26 patients (18 females and 8 males, aged 11-25 years) who were scheduled to receive fixed orthodontic treatment were enrolled. Prior to inclusion, all participants and the parents or legal guardians of participants aged <18 years were informed in detail about the study procedures and objectives, and written informed consent was obtained. The study was conducted in accordance with the Declaration of Helsinki and approved by the Van Yüzüncü Yıl University Faculty of Medicine Clinical Research Ethics Committee (approval no: 07, date: 23.05.2017).

Based on Cohen's conventional effect-size framework, the available sample of 26 patients was sufficient to detect moderate-to-large within-subject effects at a two-sided significance level of 0.05 with 80% statistical power. This sample size was considered adequate to detect associations of approximately r≥0.50. Accordingly, the study was primarily powered to detect moderate-to-large longitudinal changes; small effects and subgroup comparisons were interpreted with caution.[16]

Eligibility was assessed at T0. Before orthodontic treatment, all patients received standardized oral hygiene instructions and had periodontal health clinically confirmed. Individuals were included if they had clinically healthy periodontal tissues and mandibular incisors normally positioned according to Steiner's cephalometric analysis, with an initial 1-NB distance of 3-5 mm as the operational inclusion criterion.[17] Patients were also required to have complete pre- and post-treatment clinical and radiographic records, and only those whose orthodontic treatment was completed according to the American Board of Orthodontics[18] criteria were included in the final analysis. The exclusion criteria were as follows: pregnancy or lactation; smoking; periodontal pockets >4 mm; anterior crossbite or clinically evident traumatic anterior occlusion at baseline; structural defects of the teeth, crowns or large restorations that could interfere with clinical measurements; congenital anomalies, such as cleft lip and palate; systemic diseases; and the use of medications known to affect periodontal tissues, including immunosuppressive drugs, calcium channel blockers, and anticonvulsants.

Orthodontic Treatment Protocol

All patients were treated without extractions, using fixed orthodontic appliances with 0.018-inch-slot Roth-prescription brackets (Gemini Roth System, 3M Unitek, CA, USA) bonded to both maxillary and mandibular teeth. Leveling and alignment were performed sequentially with 0.012-, 0.014-, and 0.016-inch nickel-titanium archwires. Subsequently, 0.016×0.016-inch and 0.016×0.022-inch nickel-titanium archwires were used to express the prescribed bracket tip and torque values. Maxillary and mandibular elastics were used when clinically indicated. In the finishing phase, 0.016×0.016-inch and 0.016×0.022-inch stainless steel archwires were used, and vertical elastics were applied in the anterior or posterior regions when necessary.

Evaluation of Gingival Recession

Gingival recession was clinically evaluated at the beginning of orthodontic treatment (T0) and 20 days after the completion of active fixed orthodontic treatment and the removal of the fixed appliances (T1). All clinical evaluations were performed by the same examiner. Gingival recession was defined as the apical displacement of the gingival margin relative to the cementoenamel junction. Cases in which the cementoenamel junction was clinically visible due to apical migration of the gingival margin were recorded as having gingival recession. Sites without clinical exposure of the cementoenamel junction were recorded as 0.

Measurement of Gingival Thickness

Gingival thickness in the mandibular anterior region was evaluated by a single examiner at T0 and T1. Measurements were performed in the mandibular canine-to-canine region at two standardized labial points for each tooth: the apical level of the free gingival groove and the coronal level of the mucogingival junction.

After marking the measurement points with a marker pen, a topical xylocaine spray (Vemcain 10% Lidocaine, Vem İlaç Sanayi ve Ticaret Ltd Şti, Türkiye) was applied to improve patient comfort. When local anesthesia was required, a local anesthetic solution (Maxicaine, lidocaine hydrochloride) was injected slowly (Maxicaine, lidocaine hydrochloride, Vem İlaç Sanayi ve Ticaret Ltd Şti, Türkiye) in a limited volume of 0.1 mL to minimize anesthesia-related changes in mucosal volume. In these cases, gingival thickness measurements were performed 10-20 minutes after injection.

Measurements were performed using a size 10 endodontic spreader (G-Star Medical Co., Ltd., Guangdong, China) with a silicone stopper (Figure 1). The spreader was inserted perpendicular to the gingival surface at the marked points until a light contact with the alveolar bone was felt. Care was taken to apply only light pressure, because excessive force could cause the spreader to pass beyond the soft tissue and penetrate into the alveolar bone. Penetration depth was recorded as the distance between the tip of the spreader and the silicone stopper, measured using a digital caliper with a resolution of 0.01 mm (Figure 2).

To assess intra-examiner reliability, the same examiner repeated gingival thickness measurements in five patients during the same anesthetic period after a 10-minute interval. This interval was chosen to avoid additional patient discomfort and to maintain comparable local clinical conditions. Intraclass correlation coefficients were calculated to assess measurement reproducibility.

Cephalometric Evaluation of Mandibular Incisor Position

Lateral cephalometric radiographs obtained at T0 and T1 were used to evaluate the anteroposterior positional changes of the mandibular incisors. The radiographs were imported into the cephalometric analysis software (Nemoceph® NX 2005, Nemotec, Madrid, Spain), and image calibration was performed before tracing and measurement. All cephalometric measurements were performed by the same investigator using the same software and measurement protocol.

The most anterior mandibular incisor at baseline was used as the reference tooth for sagittal and angular measurements. Positional and angular changes in the mandibular incisors were calculated by comparing measurements obtained at T0 and T1. The cephalometric parameters used for this evaluation were the 1-NB distance (mm) and the 1-NB angle, both according to Steiner's analysis, and the incisor mandibular plane angle (IMPA) according to Tweed's analysis.

Statistical Analysis

Statistical analyses were performed using SPSS version 23 (IBM Corp., Armonk, NY, USA). Descriptive statistics were presented as mean, standard deviation, median, minimum, and maximum values. Intra-examiner reliability of gingival thickness measurements was assessed using intraclass correlation coefficients based on repeated measurements obtained from five patients. The normality of continuous variables was assessed using the Shapiro-Wilk test and visual inspection of Q-Q plots. Within-subject comparisons between T0 and T1 were performed using the paired samples t-test. Comparisons between independent groups were performed using the independent samples t-test or the Mann-Whitney U test, according to the distributional characteristics of the variables. For multiple site-specific comparisons of gingival thickness, the Bonferroni correction was applied where appropriate. Because gingival recession values were recorded as 0 at both T0 and T1, no inferential statistical comparison was performed for this variable. Statistical significance was set at $p<0.05$.

RESULTS

The intra-examiner reliability of gingival thickness measurements was excellent. The intraclass correlation coefficient calculated from repeated measurements in five patients was 0.994, with a 95% confidence interval of 0.987-0.997.

A total of 26 patients were included in the study: 18 women and 8 men. The mean age at T0 was 16.32±3.67 years, and the mean age at the end of treatment was 18.68±3.71 years. The mean duration of orthodontic treatment was 28.26 months (Table 1).

Descriptive cephalometric measurements obtained at T0 and T1 are presented in Table 2. The mean 1-NB distance increased from 4.05±0.94 mm at T0 to 4.67±1.51 mm at T1. The mean 1-NB angle remained almost unchanged, changing from 24.90°±3.84° at T0 to 24.92°±5.04° at T1. The mean IMPA angle decreased slightly from 92.77°±5.66° at T0 to 92.37°±7.13° at T1. Because changes were calculated as T0 minus T1, the mean changes were -0.62 mm for 1-NB distance, -0.03° for 1-NB angle, and 0.39° for IMPA.

Clinical gingival recession was not detected in any mandibular anterior tooth at either T0 or T1. Therefore, all gingival recession values were recorded as 0 at both time points, and no inferential statistical comparison was performed for this variable.

Gingival thickness values by tooth region, measurement level, and direction of mandibular incisor positional change are presented in Table 3. Patients were categorized as having proclined (n=14) or retroclined (n=12) mandibular incisors, based on the direction of positional change during treatment. At the apical level of the free gingival groove, the retroclined group showed statistically significant changes from T0 to T1 for teeth 33, 31, and 41. At this level, an intergroup difference at T1 was observed for tooth 32, with lower gingival thickness in the proclined group than in the retroclined group.

At the coronal level of the mucogingival junction, gingival thickness for tooth 32 in the proclined group decreased significantly from T0 to T1. In the retroclined group, significant decreases were observed for teeth 32 and 42. In addition, baseline gingival thickness at the mucogingival junction differed significantly between the proclined and retroclined groups for tooth 32. No other site-specific comparisons showed statistically significant differences after the applied statistical testing.

DISCUSSION

This prospective clinical study evaluated changes in clinical gingival recession and labial gingival thickness in patients with normally positioned mandibular incisors prior to fixed orthodontic treatment. The main findings were that no clinical gingival recession was detected in the mandibular anterior region at the end of active treatment, whereas site-specific changes in gingival thickness were observed at both the apical level of the free gingival groove and the coronal level of the mucogingival junction. In addition, the mean 1-NB distance increased significantly during treatment, whereas changes in the 1-NB angle and IMPA were limited. Based on these findings, the first null hypothesis was accepted, whereas the second was partially rejected.

The absence of clinical gingival recession in the mandibular anterior region at the end of active fixed orthodontic treatment indicates that no apical displacement of the gingival margin

Table 1. Demographic characteristics and treatment duration of the study sample

		Mean	SD	Min	Max
Female (n=18)	Age at T0 (y)	17.01	4.04	10.92	23.00
	Age at T1 (y)	19.39	4.08	13.11	26.06
	Treatment duration (mo)	28.46	6.17	15.18	37.72
Male (n=8)	Age at T0 (y)	14.77	2.14	12.42	18.25
	Age at T1 (y)	17.09	2.11	14.63	20.16
	Treatment duration (mo)	27.81	3.17	22.90	33.25
Total (n=26)	Age at T0 (y)	16.32	3.67	10.92	23.00
	Age at T1(y)	18.68	3.71	13.11	26.06
	Treatment duration (mo)	28.26	5.65	15.18	37.72

T0, beginning of orthodontic treatment; T1, end of active fixed orthodontic treatment; y, years; mo, months; SD, standard deviation; Min, minimum; Max, maximum.

Table 2. Cephalometric measurements of mandibular incisor position at T0 and T1

Variable	T0 Mean ± SD	T1 Mean ± SD	Change Mean ± SD	t	p
1-NB distance, mm	4.05±0.94	4.67±1.51	-0.62±0.95	-3.33	**0.003**
1-NB angle, °	24.90±3.84	24.92±5.04	-0.03±4.45	-0.03	0.973
IMPA, °	92.77±5.66	92.37±7.13	0.39±4.28	0.46	0.646

All p-values were calculated using the paired samples t-test. Change was calculated as T0-T1; therefore, negative values indicate an increase from T0 to T1. Statistical significance was set at (p<0.05).
T0, beginning of orthodontic treatment; T1, end of active fixed orthodontic treatment; SD, standard deviation; IMPA, incisor mandibular plane angle.

Table 3. Gingival thickness changes according to mandibular incisor movement direction

Measurement level	Tooth	Proclined group T0	Proclined group T1	Retroclined group T0	Retroclined group T1	Significant comparison
Apical to the free gingival groove	33	0.62±0.21 (0.36-1.00)	0.66±0.18 (0.50-1.15)	0.59±0.15 (0.40-0.94)	0.78±0.19 (0.48-1.04)	†
	32	0.69±0.16 (0.40-1.00)	0.68±0.16 (0.43-0.89)	0.70±0.20 (0.46-1.08)	0.81±0.17 (0.43-0.99)	§
	31	0.74±0.16 (0.50-0.99)	0.70±0.15 (0.50-1.00)	0.67±0.13 (0.46-0.94)	0.77±0.19 (0.38-1.12)	†
	41	0.73±0.23 (0.37-1.07)	0.74±0.21 (0.46-1.12)	0.72±0.16 (0.40-0.92)	0.83±0.18 (0.63-1.14)	†
	42	0.79±0.21 (0.49-1.04)	0.71±0.13 (0.46-0.92)	0.77±0.21 (0.48-1.07)	0.84±0.17 (0.45-1.05)	-
	43	0.63±0.22 (0.33-1.10)	0.67±0.18 (0.46-1.05)	0.65±0.25 (0.30-1.35)	0.71±0.20 (0.35-1.05)	-
Coronal to the mucogingival junction	33	0.66±0.27 (0.35-1.30)	0.61±0.16 (0.44-0.96)	0.63±0.20 (0.35-1.01)	0.67±0.22 (0.37-1.04)	-
	32	0.67±0.19 (0.45–1.04)	0.55±0.09 (0.45-0.78)	0.86±0.20 (0.60-1.26)	0.61±0.25 (0.38-1.35)	*†‡
	31	0.69±0.23 (0.27-1.04)	0.68±0.12 (0.42-0.84)	0.78±0.18 (0.44-1.07)	0.64±0.20 (0.38-1.14)	-
	41	0.71±0.24 (0.31-1.04)	0.68±0.17 (0.25-0.88)	0.81±0.20 (0.48-1.08)	0.70±0.31 (0.37-1.47)	-
	42	0.75±0.29 (0.25-1.20)	0.68±0.26 (0.35-1.11)	0.89±0.23 (0.52-1.30)	0.66±0.23 (0.28-1.21)	†
	43	0.58±0.20 (0.30-1.04)	0.63±0.17 (0.26-0.91)	0.67±0.30 (0.33-1.51)	0.67±0.19 (0.41-1.15)	-

Values are presented as mean ± standard deviation (minimum-maximum).
*A significant difference was observed between T0 and T1 in the proclined group.
†Significant difference between T0 and T1 in the retroclined group.
‡Significant difference between the proclined and retroclined groups at T0.
§Significant difference between the proclined and retroclined groups at T1.
Statistical significance was set at $p<0.05$ after the Bonferroni correction, where appropriate.
T0, beginning of orthodontic treatment; T1, end of active fixed orthodontic treatment.

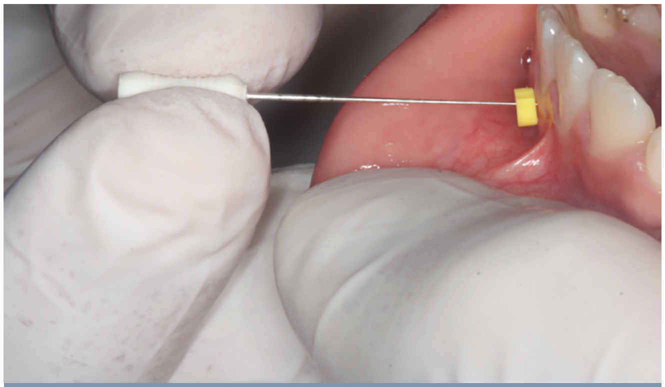


Figure 1. Transgingival probing using a 10-gauge endodontic spreader.

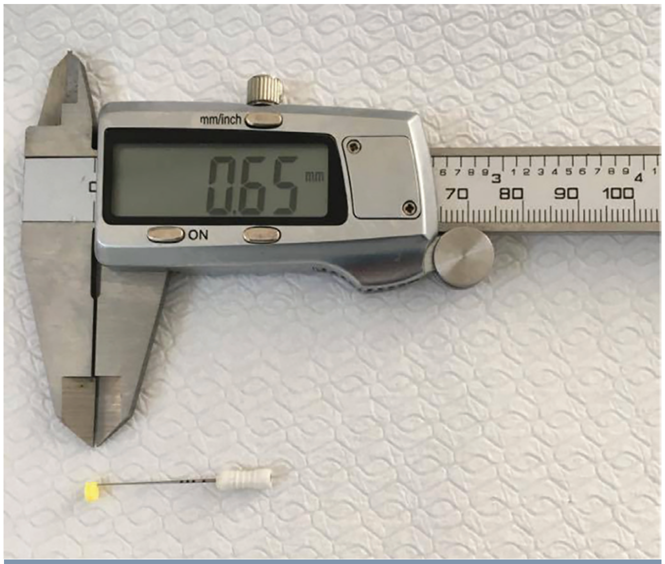


Figure 2. Digital caliper with 0.01 mm precision and endodontic spreader of 10- gauge.

was clinically detectable in this specific group of patients who had healthy periodontal tissues and normally positioned mandibular incisors at baseline. However, this finding should not be interpreted as evidence that fixed orthodontic treatment has no effect on gingival recession risk in general. Gingival recession is a multifactorial condition influenced by periodontal phenotype, baseline periodontal status, oral hygiene, inflammation, age, traumatic tooth brushing, alveolar bone support, and the length of the observation period.[2,5] Recent systematic reviews and controlled studies have shown that the relationship between orthodontic treatment and gingival recession remains controversial, with some studies reporting an increased risk after treatment and others finding no clinically relevant association, particularly when periodontal conditions are controlled.[3,10,12] Therefore, the absence of gingival recession in the present study should be interpreted within the limitations of the sample size, the baseline periodontal health of the patients, the normally positioned mandibular incisors at treatment onset, and the fact that periodontal evaluation was performed at the end of active treatment rather than after long-term retention.

One of the clinically relevant findings of the present study was that site-specific changes in gingival thickness were observed in the mandibular anterior region despite the absence of clinical gingival recession. This finding suggests that evaluating periodontal soft-tissue response only by the presence or absence of gingival recession may not fully reflect the tissue alterations occurring during fixed orthodontic treatment. Gingival thickness is an important component of the periodontal phenotype; a thin phenotype has been associated with increased susceptibility to marginal tissue instability, particularly in the presence of inflammation, mechanical trauma, or tooth movement toward the limits of the alveolar bone.[1,2,13] Therefore, reductions in gingival thickness may not necessarily result in short-term clinical gingival recession; however, they should be considered during periodontal risk assessment, especially when labial movement of the mandibular incisors is planned. The site-specific changes observed in the present study indicate that the periodontal soft-tissue response to mandibular incisor movement may vary according to tooth region and measurement level.

Although the 1-NB distance increased significantly during treatment, the limited changes observed in the 1-NB angle and IMPA suggest that mandibular incisor movement was characterized mainly by mild sagittal positional change rather than by marked proclination. Therefore, the present findings should be interpreted primarily in patients whose mandibular incisors were normally positioned at baseline and who did not undergo substantial protrusion or retrusion during treatment. This does not reduce the clinical relevance of the study, but it limits the direct generalizability of the findings to cases in which the periodontal boundaries are challenged by marked labial displacement of the mandibular incisors. Previous studies have suggested that mandibular incisor advancement or proclination toward the anatomical limits of the alveolar bone may be associated with reduced alveolar bone height and width, dehiscence formation, or increased risk of gingival recession in susceptible patients Matsumoto et al.;[8] Lee et al.;[11] Kalina et al.[9] Accordingly, the absence of gingival recession in the present sample should not eliminate the need for careful evaluation of periodontal phenotype and alveolar bone limits, particularly when greater mandibular incisor protrusion is planned.

The site-specific changes in gingival thickness observed in the present study are partly consistent with previous evidence that gingival thickness may change after fixed orthodontic treatment. Alkan et al.[19] reported decreases in gingival thickness in maxillary and mandibular anterior teeth after fixed orthodontic treatment but found no consistent significant association between sagittal tooth movement and gingival thickness changes. This finding suggests that changes in gingival thickness may not be explained solely by the amount or direction of sagittal incisor movement. Similarly, in the present study, changes in gingival thickness were not uniformly distributed across all mandibular anterior teeth but were more evident at specific teeth and measurement levels. This site-specific pattern may reflect the influence of baseline gingival thickness, local alveolar bone morphology, tooth position, and regional soft-tissue characteristics on the periodontal response to orthodontic tooth movement.

That the observed changes in gingival thickness did not progress to clinical gingival recession may suggest a degree of short-term adaptation of the periodontal soft tissues during active orthodontic treatment. Nevertheless, in individuals with a thin periodontal phenotype, limited labial bone support may reduce the stability of the marginal tissues, particularly when additional factors such as inflammation, plaque accumulation, or traumatic tooth brushing are present.[1,5,13] Therefore, when labial movement of mandibular incisors is planned, treatment planning should not rely solely on cephalometric incisor position but should also include assessment of gingival thickness, keratinized tissue dimensions, plaque control, bleeding on probing, and the anatomical limits of the alveolar bone.[2,12,20] The present findings suggest that clinical gingival recession may not occur by the end of active treatment in patients with normally positioned mandibular incisors and healthy periodontal tissues at baseline; however, site-specific reductions in gingival thickness should still be considered relevant in periodontal risk assessment.

A strength of the present study is its prospective clinical design, which allowed periodontal soft-tissue changes to be evaluated before and after fixed orthodontic treatment in the same patients. The inclusion of patients with normally positioned mandibular incisors at baseline provided a well-defined clinical sample that allowed assessment of the periodontal response to limited positional changes of the mandibular incisors. Another methodological strength was that gingival thickness was quantitatively assessed by transgingival probing rather than by visual inspection or probe transparency alone, since direct measurement methods provide a more objective evaluation of soft-tissue thickness.[19,21] In addition, measurements were performed at two standardized levels in the mandibular canine-to-canine region, allowing changes in gingival thickness to be evaluated by both tooth region and measurement level. The excellent intra-examiner reliability observed in the present study further supports the reproducibility of the gingival thickness measurements.

Study Limitations

Several limitations should be acknowledged. First, the sample size was relatively small, and no untreated control group was included. Therefore, the observed soft-tissue changes cannot be attributed solely to orthodontic treatment or distinguished from growth-related changes, age-related tissue alterations, or individual variability in periodontal response. Second, periodontal evaluation was performed at the end of active orthodontic treatment, and the long-term retention period was not included in the evaluation. Although patients with anterior crossbite or clinically evident traumatic anterior occlusion at baseline were excluded, detailed quantitative measurements of overjet and overbite were not systematically analyzed. Therefore, the possible influence of anterior occlusal relationships on periodontal outcomes should be considered when interpreting the results. Because gingival recession may develop or progress over time, longer follow-up studies are required to clarify the post-treatment periodontal response Morris et al.;[3] Gebistorf et al.;[4] Pernet et al.[6] Third, although all patients received standardized oral hygiene instructions and periodontal health was clinically confirmed before treatment, detailed periodontal variables, such as plaque index, bleeding on probing, clinical attachment level, and toothbrushing habits, were not systematically analyzed. In addition, frenulum tension in the mandibular anterior region and Pg-NB measurements were not evaluated. Mandibular incisor position was assessed on lateral cephalometric radiographs using the most anterior mandibular incisor as the reference tooth, which precluded analysis of individual three-dimensional positional changes of all mandibular incisors. Accordingly, the present findings should be interpreted primarily for periodontally healthy patients with normally positioned mandibular incisors at baseline.

Future studies should include larger samples, untreated or matched control groups, and long-term follow-up throughout the retention period to clarify the relationship between mandibular incisor movement and periodontal soft-tissue changes. In addition, systematic assessment of periodontal variables such as plaque index, bleeding on probing, clinical attachment level, keratinized tissue width, toothbrushing habits, and frenulum tension would help identify patient- and site-related factors associated with gingival recession risk.[2,5,12] Three-dimensional evaluation of alveolar bone boundaries may also strengthen periodontal risk assessment, particularly in patients for whom mandibular incisor protrusion is planned.[8,11] Overall, the present findings suggest that clinical gingival recession may not be observed at the end of fixed orthodontic treatment in periodontally healthy patients with normally positioned mandibular incisors; however, site-specific changes in gingival thickness may still occur and should be considered during orthodontic treatment planning and periodontal monitoring.

CONCLUSION

Within the limitations of this prospective clinical study, no clinical gingival recession was observed in the mandibular anterior region at the end of active fixed orthodontic treatment in patients whose mandibular incisors were normally positioned at baseline. However, site-specific changes in gingival thickness were detected at different teeth and at different measurement levels, suggesting that periodontal soft-tissue alterations may occur even in the absence of clinically detectable gingival recession. These findings indicate that labial gingival thickness should be considered during orthodontic treatment planning and periodontal monitoring, particularly when mandibular incisor positional changes are expected. Further controlled studies with larger sample sizes and long-term follow-up of retention period are needed to determine the clinical relevance of changes in gingival thickness and their potential association with future gingival recession.

Ethics

Ethics Committee Approval: The study procedures were approved by the Van Yüzüncü Yıl University Faculty of Medicine Clinical Research Ethics Committee (approval no: 07, date: 23.05.2017).

Informed Consent: Signed informed consent forms were obtained.

Footnotes

Author Contributions: Concept - A.C.Y., Ö.A.; Design - Ö.A.; Data Collection and/or Processing - A.C.Y.; Analysis and/or Interpretation - A.C.Y., Ö.A.; Literature Search - A.C.Y.; Writing - A.C.Y., Ö.A.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. Rasperini G, Acunzo R, Cannalire P, Farronato G. Influence of periodontal biotype on root surface exposure during orthodontic treatment: a preliminary study. *Int J Periodontics Restorative Dent.* 2015;35(5):665-675. [CrossRef]
2. Cadenas de Llano-Pérula M, Castro AB, Danneels M, Schelfhout A, Teughels W, Willems G. Risk factors for gingival recessions after orthodontic treatment: a systematic review. *Eur J Orthod.* 2023;45(5):528-544. [CrossRef]
3. Morris JW, Campbell PM, Tadlock LP, Boley J, Buschang PH. Prevalence of gingival recession after orthodontic tooth movements. *Am J Orthod Dentofacial Orthop.* 2017;151(5):851-859. [CrossRef]
4. Gebistorf M, Mijuskovic M, Pandis N, Fudalej PS, Katsaros C. Gingival recession in orthodontic patients 10 to 15 years posttreatment: a retrospective cohort study. *Am J Orthod Dentofacial Orthop.* 2018;153(5):645-655. [CrossRef]
5. Gül İ, Çolak R, Cicek O. Evaluation of the effect of periodontal health and orthodontic treatment on gingival recession: a cross-sectional study. *BMC Oral Health.* 2025;25(1):1069. [CrossRef]
6. Pernet F, Vento C, Pandis N, Kiliaridis S. Long-term evaluation of lower incisors gingival recessions after orthodontic treatment. *Eur J Orthod.* 2019;41(6):559-564. [CrossRef]
7. Tepedino M, Franchi L, Fabbro O, Chimenti C. Post-orthodontic lower incisor inclination and gingival recession-a systematic review. *Prog Orthod.* 2018;19(1):17. [CrossRef]
8. Matsumoto K, Sherrill-Mix S, Boucher N, Tanna N. A cone-beam computed tomographic evaluation of alveolar bone dimensional changes and the periodontal limits of mandibular incisor advancement in skeletal class II patients. *Angle Orthod.* 2020;90(3):330-338. [CrossRef]
9. Kalina E, Zadurska M, Górski B. Postorthodontic lower incisor and canine inclination and labial gingival recession in adult patients : a prospective study. *J Orofac Orthop.* 2021;82(4):246-256. [CrossRef]
10. Bin Bahar BSK, Alkhalidy SR, Kaklamanos EG, Athanasiou AE. Do orthodontic patients develop more gingival recession in anterior teeth compared to untreated individuals? A systematic review of controlled studies. *Int Orthod.* 2020;18(1):1-9. [CrossRef]
11. Lee JB, Baek SJ, Kim M, Pang EK. Correlation analysis of gingival recession after orthodontic treatment in the anterior region: an evaluation of soft and hard tissues. *J Periodontal Implant Sci.* 2020;50(3):146-158. [CrossRef]
12. Kloukos D, Koukos G, Pandis N, Doulis I, Stavropoulos A, Katsaros C. Effect of orthodontic treatment with fixed appliances on the

development of gingival recession. A prospective controlled study. *Eur J Orthod*. 2025;47(3):cjaf022. [CrossRef]

13. Amid R, Kadkhodazadeh M, Moscowchi A, et al. Effect of gingival biotype on orthodontic treatment-inducedperiodontal complications: a systematic review. *J Adv Periodontol Implant Dent*. 2020;12(1):3-10. [CrossRef]
14. Küçükoğlu Çolak M, Çolak R, Cicek O. Fractal dimension analysis of mandibular trabecular architecture in gingival recession during orthodontic retention: a cross-sectional study. *Diagnostics*. 2025;15(8):1013. [CrossRef]
15. Çiçek O, Yılmaz H, Melayim H. Evaluation of changes in alveolar bone remodeling after orthodontic treatment using fractal analysis: a retrospective study. *Int Dent Res*. 2024;14(1):19-26. [CrossRef]
16. Cohen J. Statistical power analysis for the behavioral sciences. *(2nd ed.). Routledge; New York. 1988*. [CrossRef]
17. Steiner CC. Cephalometrics for you and me. *American Journal of Orthodontics*. 1953;39(10):729-755. [CrossRef]
18. Casko JS, Vaden JL, Kokich VG, et al. Objective grading system for dental casts and panoramic radiographs. American Board of Orthodontics. *Am J Orthod Dentofacial Orthop*. 1998;114(5):589-599. [CrossRef]
19. Alkan Ö, Kaya Y, Tunca M, Keskin S. Changes in the gingival thickness and keratinized gingival width of maxillary and mandibular anterior teeth after orthodontic treatment. *Angle Orthod*. 2021;91(4):459-467. [CrossRef]
20. Allais D, Melsen B. Does labial movement of lower incisors influence the level of the gingival margin? A case-control study of adult orthodontic patients. *Eur J Orthod*. 2003;25(4):343-352. [CrossRef]
21. Frost NA, Mealey BL, Jones AA, Huynh-Ba G. Periodontal biotype: gingival thickness as it relates to probe visibility and buccal plate thickness. *J Periodontol*. 2015;86(10):1141-1149. [CrossRef]

DOI: 10.4274/TurkJOrthod.2026.2025.103

TURKISH JOURNAL OF ORTHODONTICS

Original Article

Bibliometric Mapping of Digitalization in Orthodontics: Evaluating Scientific Contributions, Collaboration Networks, and Future Directions

Damla Dalman Şimşek

Uşak University Faculty of Dentistry, Department of Orthodontics, Uşak, Türkiye

Cite this article as: Dalman Şimşek D. Bibliometric mapping of digitalization in orthodontics: evaluating scientific contributions, collaboration networks, and future directions. *Turk J Orthod.* 2026; 39(3): 169-179

Main Points

- Digital technologies such as cone-beam computed tomography, intraoral scanners, computer-aided design and computer-aided manufacturing, and three-dimensional printing have significantly improved diagnostic accuracy and treatment planning in orthodontics.
- Clear aligners, virtual surgical planning, and artificial intelligence -based tools are increasingly shaping daily orthodontic practice.
- The coronavirus disease 2019 pandemic accelerated the adoption of digital workflows, thus enhancing patient monitoring, remote care, and treatment efficiency.
- Despite clinical advantages, high costs, technical requirements, and limited accessibility remain barriers to the widespread use of digital orthodontics.
- Understanding global research trends in digital orthodontics can help clinicians to adapt evidence-based innovations into practice.

ABSTRACT

Objective: To map scholarly trends in digital orthodontics by quantifying publication growth, identifying influential contributors and collaboration networks, and delineating emerging research themes.

Methods: A bibliometric analysis was conducted using the records published between January 2000 and May 2024 from the Web of Science Core Collection; VOSviewer was used to analyze co-authorship networks, keyword co-occurrence, institutional productivity, and citation patterns.

Results: A total of 1,567 publications were analyzed. Research output markedly increased after 2020 and peaked in 2023. The United States, Italy, and China were the leading contributing countries. Key institutions included Peking University, Radboud University, and the University of São Paulo. The most frequently cited authors were Xu, Kuijpers-Jagtman, and Lo Giudice. Major thematic clusters were centered on three-dimensional printing, artificial intelligence, and digital modeling.

Conclusion: In recent years, digital orthodontics research has considerably accelerated, thus expanding international collaboration and rapid technological innovation. A limitation of this study is that coverage may differ from that of multi-database approaches. In addition, future bibliometric assessments should be complemented by qualitative appraisals and clinical outcome- oriented evaluations to strengthen translational relevance.

Keywords: Digital orthodontics, bibliometrics, collaboration networks, artificial intelligence

Corresponding author: Damla Dalman Şimşek, **e-mail:** damla.dalman@usak.edu.tr **ORCID:** orcid.org/0000-0002-7585-5746

Received: September 03, 2025 **Accepted:** June 23, 2026 **Publication Date:** September 25, 2026

INTRODUCTION

In recent years, the integration of digital technologies has significantly transformed healthcare, and conventional orthodontic methods are being increasingly replaced by digital applications to enhance diagnostic precision, treatment predictability, and personalization. Two-dimensional imaging techniques, such as panoramic and cephalometric radiographs, have long been used in orthodontics. However, due to magnification, distortion, and superimposition, these methods are limited in accurately assessing anatomical structures.[1,2] Three-dimensional (3D) imaging modalities, particularly computed tomography (CT) and cone-beam CT (CBCT), have addressed some of these limitations by enabling detailed visualization and reconstruction of craniofacial structures.[3] In particular, CBCT has become increasingly popular due to its low radiation dose and high-resolution imaging capabilities. At present, CBCT plays an important diagnostic tool in selected complex cases, including localizing impacted teeth, assessing root resorption, evaluating temporomandibular joint disorders, and planning orthognathic surgery.[4]

Intraoral scanners have replaced conventional impression techniques, offering a more comfortable and efficient digital modeling process.[5] By producing high-resolution digital models, intraoral scanners enable a wide range of analyses, including arch width and length measurements, Bolton analysis, tooth segmentation, and occlusion evaluation.[6,7] Virtual treatment planning software facilitates digital manipulation of teeth to their ideal positions, supporting indirect bonding, custom arch wires, and aligner therapy.[8] Similarly, virtual surgical planning in orthognathic surgery enables osteotomy simulation, evaluation of surgical strategies, prediction of postoperative facial profiles, and fabrication of surgical splints and fixation plates.[9,10] Computer-aided design and computer-aided manufacturing (CAD/CAM) systems and 3D printing technologies have enabled high-precision customized fabrication of orthodontic appliances.[11] These technologies support the production of various devices such as retainers, functional appliances, expansion devices, indirect bonding trays, splints, labial and lingual bracket systems, and clear aligners.[6,12] While clear aligner therapy relies on digitally planned sequential corrections, lingual orthodontics benefits from robotically bent arch wires and customized brackets designed using CAD/CAM technology.

More recently, artificial intelligence (AI) offers the potential to enhance speed, accuracy, and individualization in clinical processes such as diagnosis, treatment planning, and patient monitoring. AI is particularly valuable in orthognathic surgical decision-making.[13] At present, AI is being explored for decision-making in areas such as camouflage versus surgical treatment, extraction planning, identification of cephalometric landmarks, and skeletal maturation assessment based on cervical vertebrae.[14,15]

Despite the rapid expansion of digital technologies in orthodontics, a comprehensive understanding of how this field has evolved at the scientific level remains limited. Bibliometric analysis is a quantitative method used to evaluate relationships and impact among published works within a specific research domain.[16,17] Unlike conventional literature reviews, bibliometric methods allow for the rapid and systematic identification of prominent research directions and knowledge structures, guiding future research efforts.[18] Accordingly, this study aimed to map the evolution of digitalization in orthodontic research by analyzing publication trends and citation patterns; identifying the contributions and collaboration networks of authors, institutions, and countries; determining influential journals and highly cited publications; and characterizing established and emerging research themes through keyword and reference analyses.

METHODS

Data Collection

A comprehensive literature search on the use of digital technologies in orthodontics was conducted using the Web of Science (WoS) core database. An advanced search was performed using the query TS=[(digital and orthodontics) or (digital* and orthodont*)] without any field restrictions. The publication period for the articles was set to the period from January 1, 2000, to May 31, 2024. A preliminary screening indicated that the number of eligible publications prior to 2000 was negligible; therefore, to ensure a consistent and meaningful time window for trend evaluation, the analysis was restricted to the records published from 2000 onwards. Only original research articles and review articles were included. No language restrictions were applied since the search was restricted to WoS, the coverage may differ from that of multi-database approaches; therefore, publications indexed exclusively in Scopus, PubMed/MEDLINE, and regional databases were not captured.

This study was a bibliometric analysis based exclusively on publicly available, previously published bibliographic data retrieved from the Web of Science Core Collection. No human participants, patient records, or animal subjects were involved, and no personal or identifiable data were collected. Therefore, ethics committee approval and informed consent were not required.

Data Transfer and Processing

The retrieved records were exported from WoSCore Collection (WoSCC) as plain-text files compatible with VOSviewer (Centre for Science and Technology Studies, Leiden University, The Netherlands; version 1.6.20). The dataset included metadata such as authorship, affiliations, abstracts, citations, and keywords. Before network analysis, the data were reviewed and standardized to reduce inconsistencies in terminology.

Bibliometric and Network Analysis

VOSviewer was used to generate co-authorship maps (authors, countries, institutions), citation networks (journals, references), and keyword co-occurrence clusters. All metrics were calculated using full counting methods. Total link strength was interpreted as a measure of influence within the network. To improve map interpretability, the following threshold criteria were applied: for collaboration maps, only authors, institutions, and countries with ≥5 publications were included; for keyword co-occurrence analysis, only author-provided keywords with a minimum occurrence of ≥10 were included. Keyword mapping was based on author-provided keywords, and thesaurus-based harmonization was applied to merge synonyms and spelling variants (e.g., "AI" and "artificial intelligence").

Statistical Analysis

The analyses were primarily descriptive and bibliometric; therefore, no inferential statistical tests were performed. However, relative comparisons (e.g., growth rates, citation trends) were discussed to contextualize patterns across regions and years.

RESULTS

General Information

As a result of the search query, a total of 1567 articles were identified. After using VOSviewer software to apply the specified filters to the articles, the analyses were conducted. Of the identified studies, 1464 (94.6%) were original research articles and 103 (6.6%) were review articles. The highest number of publications was observed in 2023, with a peak of 218 (Table 1). Most publications (78.4%) were categorized within the WoS research category Dentistry, Oral Surgery and Medicine (Table 2).

Countries and Institutions

The United States (USA) was the most productive country, with 282 publications, followed by Italy with 201, and China with 150 (Table 3). According to countries' total link strength scores, the USA ranked first. In the country-based citation analysis, the highest number of citations was observed in the USA, Italy, the Netherlands, Germany, and the United Kingdom (Table 3). With regard to institutional distribution, the top five institutions in terms of publication volume were Peking University, University of São Paulo, University of Ferrara, University of Michigan, and Radboud University. The top five institutions with the highest number of citations were Radboud University, University of Alberta, University of Michigan, University of Groningen, and Peking University (Table 4).

Authors

With respect to the number of published articles, the leading authors were Tianmin Xu, Anne Marie Kuijpers-Jagtman, Antonino Lo Giudice, Guilherme Janson, and Vincenzo Ronsivalle (Table 5). Based on the author citation analysis, the most cited author was Carlos Flores-Mir, with 627 citations. This

Table 1. Annual distribution of publications

Years	Number of articles
2000	5
2001	4
2002	7
2003	7
2004	6
2005	19
2006	13
2007	25
2008	38
2009	30
2010	37
2011	45
2012	46
2013	37
2014	41
2015	56
2016	53
2017	67
2018	86
2019	91
2020	138
2021	180
2022	212
2023	218
2024	106

Table 2. Distribution of publications by web of science research category

	Journal type	Records	% of 1567
1	Dentistry oral surgery medicine	1228	78.4
2	Medicine general internal	99	6.3
3	Material science multidisciplinary	54	3.5
4	Physics applied	51	3.3
5	Chemistry multidisciplinary	47	3

Table 3. Country-wise publication output, citation counts, and total link strength

	Country	Article	% of 1567 article	Citation counts	Total link strength
1	USA	282	18	6539	1409
2	Italy	201	12.8	2972	716
3	China	150	9.6	1402	412
4	Brazil	123	7.8	1465	567
5	Germany	121	7.7	2011	679
6	India	103	6.6	458	319
7	Türkiye	97	6.2	1028	381
8	England	74	4.7	1667	562
9	South Korea	72	4.6	1611	587
10	Sweden	58	3.7	990	358

Table 4. Institution-wise publication output, citation counts, and total link strength

	Institution	Article	% of 1567 article	Citation counts	Total link strength
1	Radboud University	31	2	1360	238
2	University of Michigan	34	2.2	862	212
3	Peking University	74	4.7	426	153
4	Groningen University	12	0.8	679	138
5	São Paulo University	46	2.9	295	138
6	University of Alberta	21	1.3	894	137
7	University of Bern	18	1.1	235	123
8	Aarhus University	13	0.8	202	115
9	Yonsei University	15	0.9	352	104
10	Case Western Reserve University	17	1.1	375	102

Table 5. Author-wise publication output

	Author	Article
1	Xu, Tianmin	21
2	Kuijpers-Jagtman, Anne Marie	20
3	Lo Giudice, Antonino	19
4	Janson, Guilherme	19
5	Ronsivalle, Vincenzo	16
6	Garib, Daniela	16
7	Cevidanes, Lucia Soares	15
8	Cossa, Paola	15
9	Leonardi, Rosalia	14
10	Cattaneo, Paolo	14

Table 6. Author-wise citation counts

	Author	Citation counts
1	Flores-Mir, Carlos	627
2	Ren, Yijin	606
3	Major, Paul W	540
4	Heo, Giseon	458
5	Kuijpers-Jagtman, Anne Marie	422
6	Fields, Henry W	401
7	Hatcher, David	353
8	Huang, John C	353
9	Maki, Koutaro	353
10	Miller, Arthur J	349

author was followed by Yijin Ren, Paul W. Major, Giseon Heo, and Anne Marie Kuijpers-Jagtman (Table 6). In terms of total link strength, Luca Lombardo, Antonino Lo Giudice, and Vincenzo Ronsivalle had the highest values in the author collaboration network (Supplementary Table 1).

Journals

The top five journals where the identified articles were most frequently published included the American Journal of Orthodontics and Dentofacial Orthopedics (AJODO) (320), European Journal of Orthodontics (147), Orthodontics and Craniofacial Research (73), Korean Journal of Orthodontics (57), and Journal of Orthodontics (51). According to the journal citation analysis, the three journals with the highest number of citations were AJODO, European Journal of Orthodontics, and The Angle Orthodontist. In terms of total link strength scores, the top five journals were AJODO, European Journal of Orthodontics, Orthodontics and Craniofacial Research, The Angle Orthodontist, and Korean Journal of Orthodontics (Supplementary Table 2).

Keywords

Following the keywords "orthodontics" and "digital orthodontics," the most frequently used keywords were digital models, digital dentistry, 3D printing, accuracy, and clear aligners. In terms of total link strength scores, the top five keywords were 3D printing, digital dentistry, digital models, accuracy, and AI (Supplementary Table 3).

Reference Impact

The most cited references included Fleming et al.,[19] Stevens et al.,[20] and Santoro et al.[21] These studies focused on the comparative accuracy of digital vs. conventional measurements (Supplementary Tables 4 and 5). More recent studies are increasingly cited, suggesting that digital orthodontics is a field driven by innovation and evolving clinical paradigms.

DISCUSSION

The thematic structure revealed by the bibliometric mapping has direct implications for everyday orthodontic practice. The prominence of clusters related to intraoral scanning and digital models aligns with the ongoing clinical transition from conventional impressions to digital data acquisition, facilitating more efficient storage and retrieval of patient records and potentially improving interdisciplinary communication and case transfer within digital workflows.[22] In their review on the future of digital orthodontics, Tarraf and Ali[4] emphasized the transformative impact of the widespread adoption of 3D imaging and intraoral scanners on clinical practice. In parallel, the integration of technologies such as CBCT, intraoral scanners, and CAD/CAM systems into routine care has likely contributed to the growth of scientific output in these domains.[23] The prominence of 3D printing and additive manufacturing represents another major component of this digital transformation. These technologies have enabled the fabrication of patient-specific orthodontic appliances and auxiliaries-such as aligner-related workflows, retainers, indirect bonding trays, surgical splints, and customized adjunctive devices-with potential gains in turnaround time, standardization, and reproducibility.[24] Likewise, the rapid increase in AI-focused themes indicates the growing clinical interest in automated image interpretation and clinical decision-support applications (e.g., cephalometric landmark identification, decision-making regarding extractions in orthodontic treatment planning, and growth monitoring based on cervical vertebrae).[15,25,26] Collectively, these developments may enhance consistency across diagnostic and planning stages and reduce operator-dependent variability. Overall, the observed trends suggest that "digitalization" is evolving beyond standalone tools towards integrated, end-to-end digital workflows encompassing diagnosis, treatment planning, appliance fabrication, and follow-up monitoring.

The acceleration in publication output observed after 2020 may reflect not only technological maturation but also the pandemic-driven need for digital data sharing, remote communication, and care strategies that minimize physical contact. During the period analyzed,the observed increase in attention to tele-orthodontics and remote patient monitoring suggests that digital workflows have expanded beyond diagnosis and planning to also include follow-up stages, such as supporting control appointments and enabling remote treatment monitoring. Beyond this acute phase, hybrid care models, in which in-office procedures are complemented by digitally supported monitoring and follow-up processes, are likely to be adopted on a more permanent basis. While routinization of digital records and standardized data formats may facilitate remote case assessment, interdisciplinary consultation, and patient communication,[27,28] emerging AI-based analytics could support triage and prioritization in high-volume clinical settings.[29] However, the long-term sustainability of this transformation will depend on clinician training, interoperability across software ecosystems, data-quality standards, and evidence regarding clinical effectiveness and cost-effectiveness. Accordingly, future research should not only map publication dynamics but also evaluate the real-world impact of digital workflows on clinical outcomes, efficiency, and patient-reported experiences in orthodontic care.

Countries and Institutions

In terms of publication output, the USA, Italy, China, Brazil, and Germany were the most productive countries. In co-authorship analysis, total link strength reflects the intensity of international research collaborations (Table 3). The co-authorship network map (Figure 1A) highlights the central role of the USA. The results suggest that the USA maintains extensive collaborations with multiple countries, supported by the country's high output and strong link strength. Furthermore, Italy, Brazil, Switzerland, and Germany follow the USA in collaborative output. Although China has substantially increased its research output in certain fields, it remains behind the Euro-American axis in citation impact.[30] Despite a high number of publications, China's comparatively low link strength suggests a more limited role in international co-authorship networks. According to the results of citation analysis, the most frequently cited countries were the United States (6539 citations), Italy (2972 citations), and Germany (2011 citations) (Table 3). International collaborations are considered an effective means of enhancing the quality of scientific research and the resulting scholarly output.[31] Moreover, internationally co-authored publications generally achieve a greater impact than those arising solely from national collaborations.[32,33] The overlay visualization of country citations (Figure 1B) shows a steady increase in U.S. citations since 2018. Following the COVID-19 pandemic, the widespread use of online communication tools appears to have facilitated cross-continental collaboration. As shown in Figure 1B, the constraining effect of geographic distance on collaboration has diminished since 2020, enabling broader international research networks. China, Italy, Sweden, and Germany have shown marked increases in citations during the post-pandemic period.

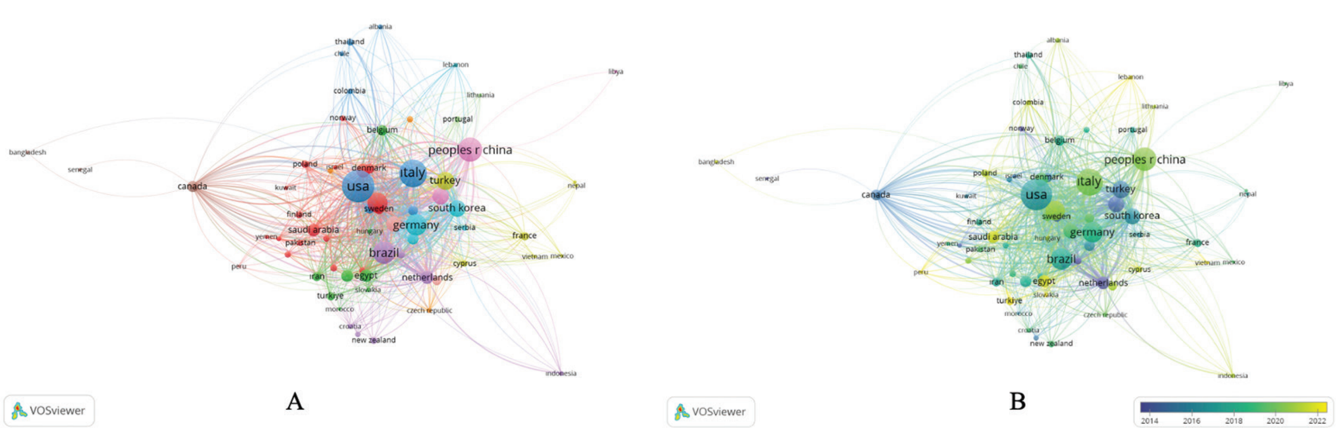


Figure 1. A) Network visualization of co-authorship among countries **B)** overlay visualization of country citations.

Since institutional prestige is closely associated with the quantity and quality of scientific publications, scholars at such institutions tend to produce more publications and receive higher citation counts.[34] At the institutional level, Peking University had the highest publication volume, indicating that a significant portion of China's contributions to digital orthodontics originates from this institution. Furthermore, the University of São Paulo and the University of Michigan stand out as key global research hubs. In particular, the University of São Paulo emerged as a major contributor in Latin America, reflecting Brazil's scientific productivity. Despite a lower publication count, Radboud University, with 1360 citations and a total link strength of 238, ranked highest in citation impact and link strength, indicating its influential role in the field (Table 4). This conclusion is supported by the institutional co-authorship network (Figure 2). The University of Michigan (862 citations, 212 link strength) and Peking University (426 citations, 153 link strength) also demonstrated considerable academic influence. Groningen (679 citations; link strength 138) and São Paulo (295 citations; link strength 138) demonstrated strength in both citations and collaborative networks. Groningen's citation performance further underscores the Netherlands' contribution to this domain. Taken together, these findings suggest that institutional performance should be interpreted using multiple indicators (output, citation impact, and collaboration), rather than publication volume alone.

Authors

Authorship analysis revealed that the most prolific contributors in this field were Xu Tianmin (21 publications), Kuijpers-Jagtman (20) publications, and Lo Giudice (19) publications (Table 5). While high productivity indicates sustained scholarly engagement, publication counts alone do not fully capture academic influence. An author's influence in the literature depends not only on the quantity of their output, but also on the number of citations their work receives and the extent of their scholarly collaborations. In this context, authors with high publication volume should also be evaluated based on their citation frequency and collaborative networks.

When examining citation counts, Flores-Mir (627), Ren Yijin (606), and Major (540) emerged as the most highly cited authors (Table 6). A comparison of publication volumes and citation counts revealed that some authors, despite having relatively few publications, achieved substantial academic influence. For instance, Kuijpers-Jagtman ranked highly by publication count, and they were also among the most frequently cited authors. This highlights that not only the quantity but also the quality of scientific work are critical factors in shaping academic success.

The total link strength values presented in Supplementary Table 1 are important indicators of the intensity of academic collaborations. Authors such as Luca Lombardo (20.0), Antonino Lo Giudice (17.0), and Vincenzo Ronsivalle (16.0) had high link strength scores, reflecting their active engagement in extensive academic collaboration networks. A high link

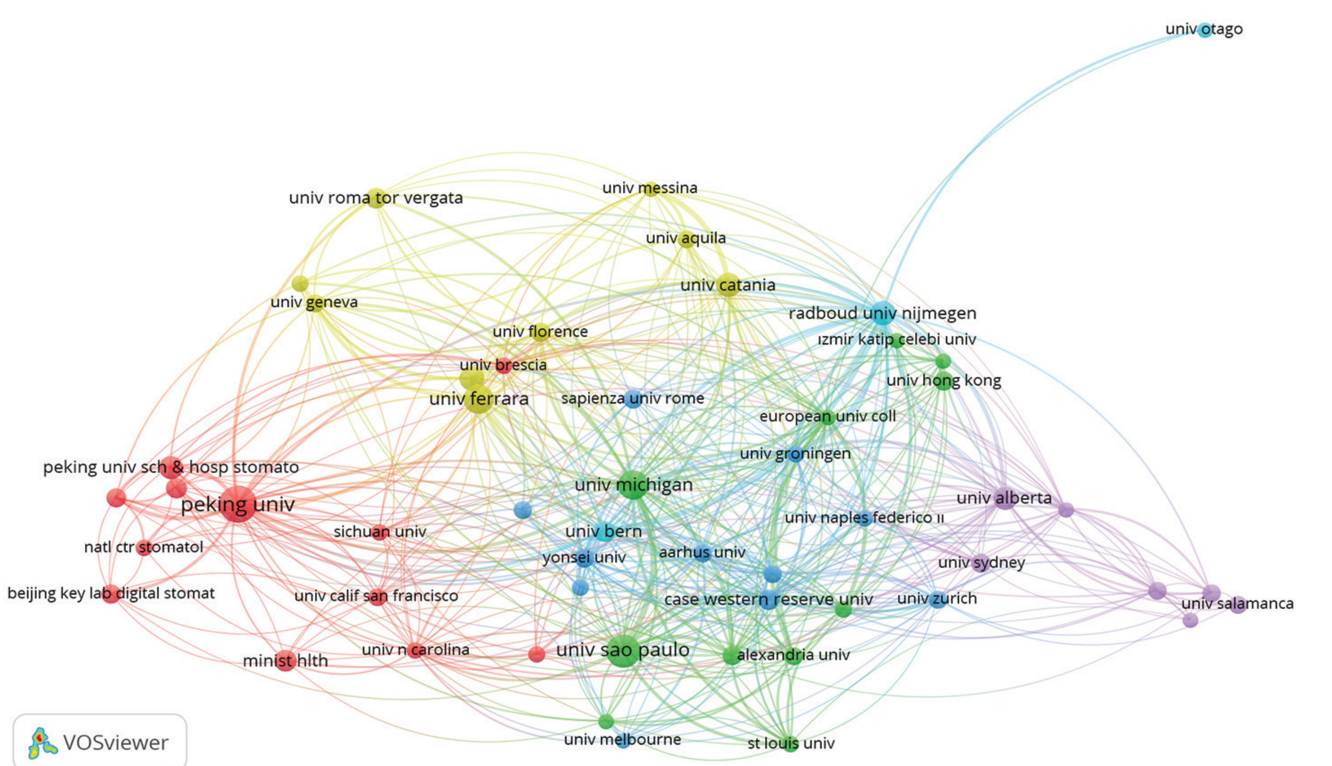


Figure 2. Network visualization of co-authorship among institutions.

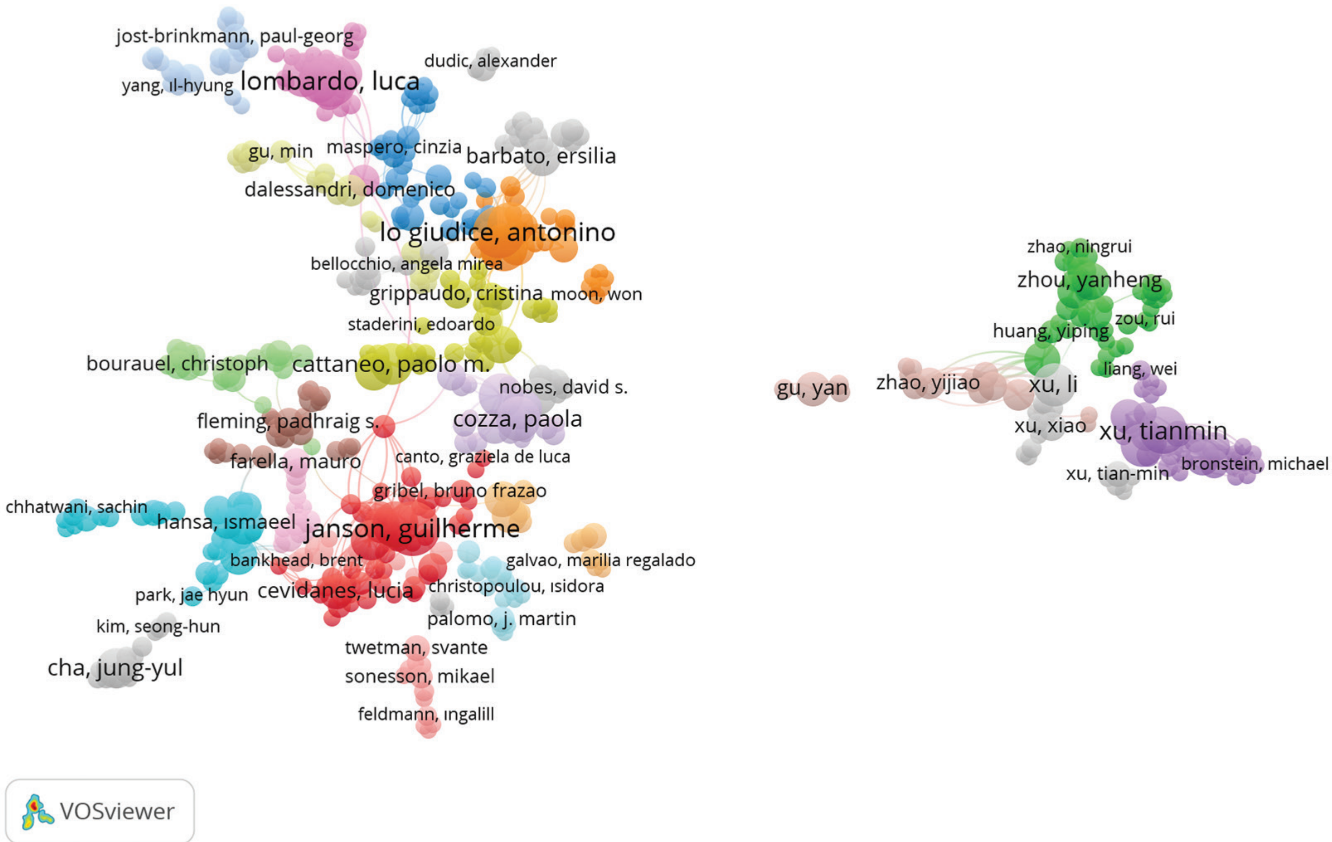


Figure 3. Network visualization of co-authorship among authors.

strength score suggests that these authors frequently collaborate with other researchers and participate in joint research projects. For instance, Lo Giudice appears to play a key role not only by contributing individually to the literature, but also by supporting international academic collaborations, given this author's high scores in both publication volume and link strength.

The co-authorship network map in Figure 3 shows that geographic and institutional proximity plays an important role in collaboration preferences among authors. Academic communities based predominantly in Europe and North America tended to establish broader collaborations and foster global interactions. Although Chinese researchers form a distinct cluster, their connections with Western academic communities remain limited. Authors appearing as isolated nodes may represent smaller teams or comparatively independent lines of work. As shown in Figure 4, future growth in international collaboration may increase interactions among these major clusters.

Journals

AJODO emerged as the most productive and widely preferred journal. While the European Journal of Orthodontics predominantly features Europe-based research, Orthodontics & Craniofacial Research focuses on more specialized and multidisciplinary studies in orthodontics. Although The Angle Orthodontist and Orthodontics & Craniofacial Research publish fewer articles, their high citation rates suggest a selective publication policy and a substantial scientific impact. AJODO leads the field not only in volume and number of citations, but also in collaborative strength. Since 2014, as shown in Figure 5, AJODO, European Journal of Orthodontics, Journal of Orthodontics, and The Angle Orthodontist have been prominent in citation trends. Since 2020, Orthodontics & Craniofacial Research and Progress in Orthodontics have experienced increased visibility. The recent rise in open-access journals, such as Progress in Orthodontics and Journal of Digital Orthodontics, has likely facilitated broader dissemination of research in digital orthodontics.

Keywords

The results of the keyword analysis highlight the prominent research foci in this field. After the terms "orthodontics" and "digital orthodontics," the most frequently used keywords related to digital orthodontics and associated technologies were "digital models" (52 occurrences), "digital dentistry" (45), and "3D printing" (44) (Supplementary Table 3). The frequent appearance of these three keywords suggests that research in the field of digital orthodontics has increasingly focused on digital modeling, digital dentistry applications, and 3D printing techniques. Terms such as "accuracy" (35) and "clear aligners" (31) emphasize the relationship between digital approaches and the precision and fit achieved in treatment processes. In

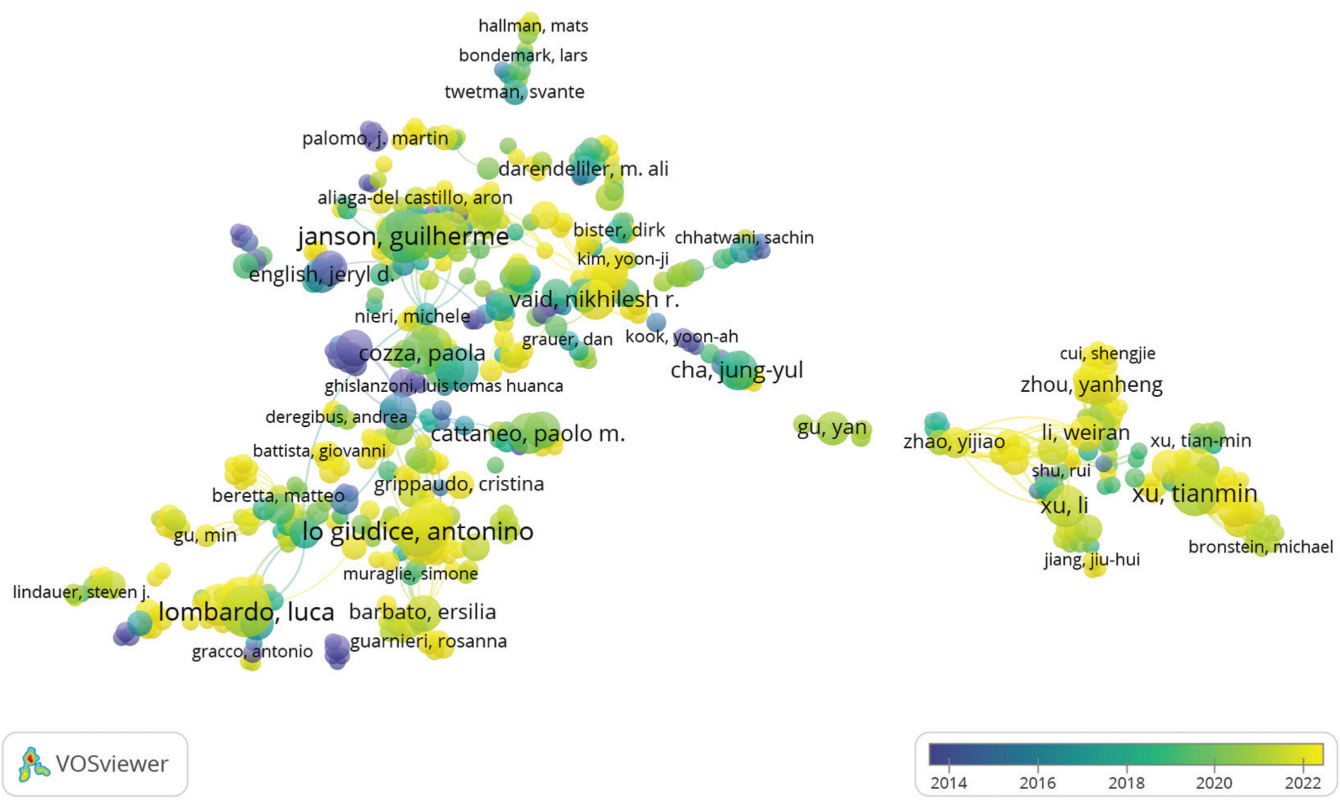


Figure 4. Overlay visualization of co-authorship among authors.

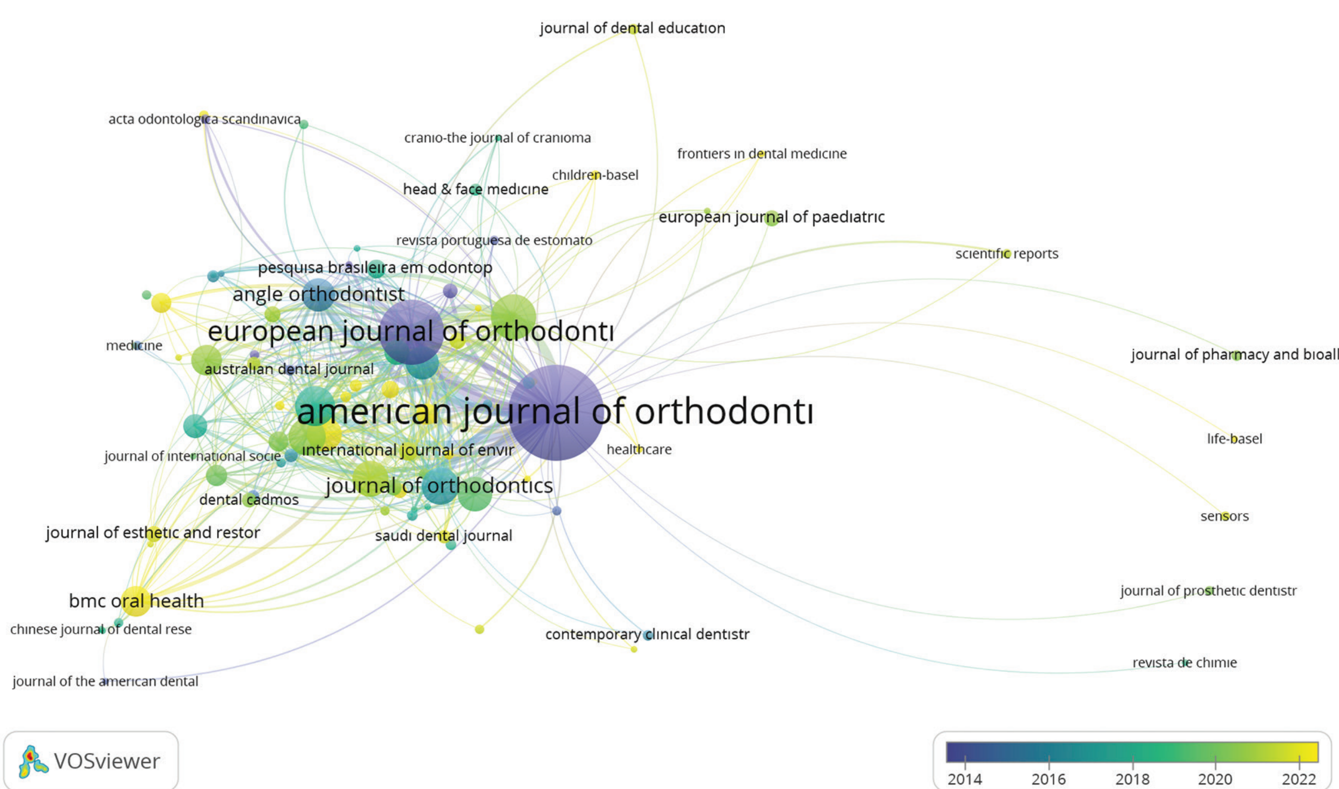


Figure 5. Overlay visualization of journal citations.

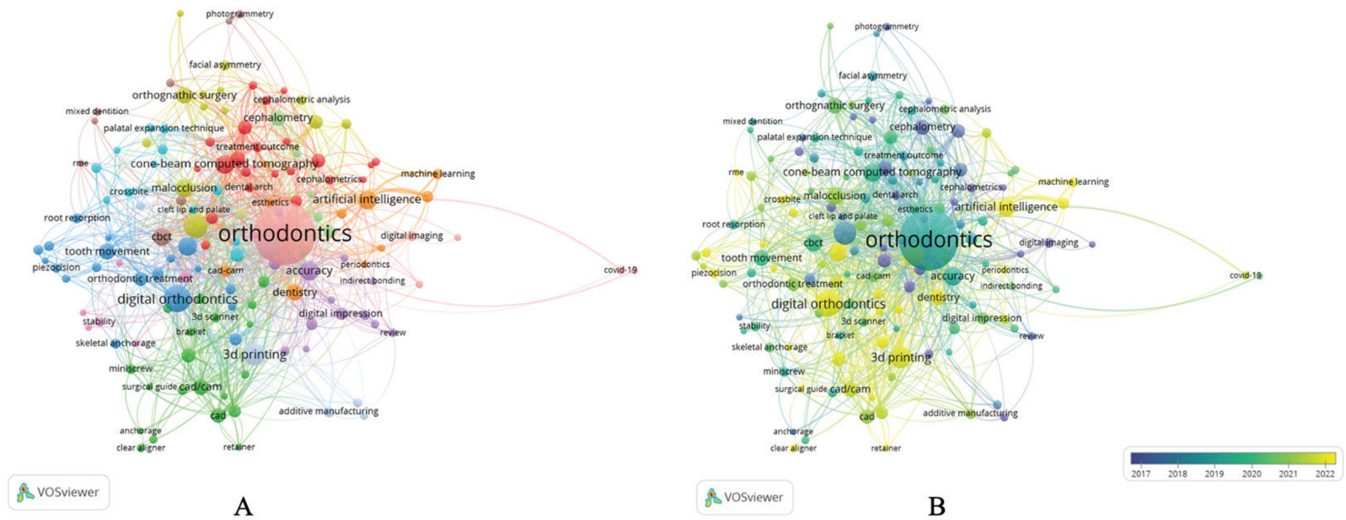


Figure 6. A) Keyword co-occurrence analysis **B)** overlay visualization of keyword co-occurrence.

addition, "AI" (30) and "CBCT" (29) are additional significant terms that reflect the expanding applications of digital technologies in dentistry and orthodontics.

As shown in the keyword co-occurrence map in Figure 6A, the 144 nodes correspond to 144 unique keywords used across 1567 articles. A total of 995 links were identified among these nodes, indicating that pairs of keywords co-occurred in articles 995 times. The central placement of the terms "3D printing," "digital dentistry," and "AI" on the map, combined with these keywords' high total link strength scores, confirms that the corresponding concepts possess the strongest connections within the network (Supplementary Table 3). This finding demonstrates that the digitalization of orthodontic applications is no longer confined to imaging techniques but also extends to production processes and decision-support systems. Indeed, Choi et al.[14] emphasized the significant contribution of AI-based models in orthognathic surgery decision making.

In the longitudinal keyword analysis presented in Figure 6B, an increase in the frequency of use of the terms "3D printing," "clear aligner," "AI," "machine learning," and "digital workflow" has been observed since 2022. This trend can be attributed to the growing adoption of personalized systems produced via digital workflows, such as clear aligners.[35] Similarly, recent clinical developments have become the subject of bibliometric research. The increase in publications on miniscrew-assisted maxillary expansion -following the widespread adoption of miniscrews, the integration of AI into orthodontics, and the growth in literature output accompanying the emergence of clear aligner therapy- has been visualized and quantitatively monitored through bibliometric analyses.[36-38] Collectively, these developments confirm a consistent transition from conventional orthodontic methods toward digital and data-driven approaches.

Reference Analysis

The data presented in Supplementary Table 4[19-21,39,40] and 5[19-21,39,41] indicate that the most frequently cited studies are those by Fleming et al.,[19] Stevens et al.,[20] and Santoro et al.[21] These authors examine the differences between digital and conventional orthodontic measurements, and their studies serve as foundational works in the field. Fleming et al.,[19] emerged as the author with both the highest citation frequency and the greatest total link strength. These findings clearly demonstrate the critical role of Fleming et al.,[19] in shaping and advancing this research area. Stevens et al.,[20] closely follows Fleming et al.,[19] in this field. We also observed that more recent studies tend to receive higher citation counts, indicating that newly developed methods are increasingly preferred over traditional approaches and that there is a growing academic interest in this area.

Over the last decade, intraoral scanning, 3D printing, and related software applications have gained significant interest in the orthodontic literature and among both healthcare providers and patients. Accordingly, it is anticipated that 3D applications will replace two-dimensional methods as the most frequently used measurement techniques.[42] The increasing adoption of 3D technologies in orthodontics has also driven academic research forward. Initially, the primary goal of digital technology in orthodontics was to enhance diagnostic accuracy; however, current advancements have contributed to a more streamlined and efficient workflow. Future technological developments are expected not only to refine treatment outcomes but also to improve patients' access to care and to facilitate better adherence to treatment protocols. In addition, the ease of digital data sharing has strengthened academic collaborations across institutions and countries. The recent rise in open-access publications has further expanded the availability of digital orthodontics research to a broader scientific community, thus accelerating the dissemination of knowledge and contributing to a more dynamic and rapidly evolving body of literature.

While the adoption of digital technology in orthodontic applications has brought significant improvements in diagnostic accuracy and workflow efficiency, several technical and economic challenges remain to be addressed in this transition. First, the acquisition costs and maintenance requirements of digital orthodontic systems (e.g., intraoral scanners, CAD/CAM software, 3D printers) pose certain financial constraints.[43] This may hinder the adoption of digital technology, particularly for institutions and clinics with limited capacity to invest in technological infrastructure. The varying degrees of acceptance of digital technologies among patients demand effective teamwork and prioritization of patient education. Practitioners must also adapt to evolving legal standards and navigate administrative challenges throughout this transition. In addition, ensuring the optimal functioning of digital hardware and software requires regular maintenance and specialized technical support. Despite these challenges, carefully and flexibly implemented digital orthodontic technologies have the potential to significantly enhance patient care and overall efficiency.

Study Limitations

The present study has several limitations. First, the analysis was based exclusively on the WoSCC. Although WoSCC is widely used in bibliometric research, reliance on a single database may have excluded relevant publications indexed only in other databases, such as Scopus, PubMed/MEDLINE, Embase, or regional databases. Consequently, the publication, citation, geographic, and institutional patterns identified in this study should be interpreted within the context of WoSCC coverage.[35]

Second, citation indicators were not adjusted for author- or journal-level self-citations. Highly productive authors, research groups, or journals may accumulate self-citations, potentially affecting citation-based rankings and comparisons.[17] Therefore, citation counts should be interpreted as indicators of citation visibility rather than direct measures of scientific quality or impact.

Third, although no language restrictions were applied, the indexing practices of WoSCC may favor English-language journals and specific publication types, potentially limiting the representation of research from certain countries or regions.

Fourth, bibliometric analysis is inherently quantitative and primarily characterizes publication output, citation patterns, collaboration networks, and thematic structures. It does not evaluate the methodological quality, risk of bias, strength of evidence, or clinical effectiveness of the individual studies included in the analysis. Accordingly, the prominence of a research theme or highly cited publication should not necessarily be interpreted as evidence of greater clinical relevance or scientific quality.

Finally, keyword-based analyses are influenced by variations in terminology because author-provided keywords are not standardized. Conceptually similar terms, such as "clear aligner" and "clear aligners," "CBCT" and "cone-beam computed tomography," or "3D printing" and "additive manufacturing," may therefore fragment related research themes. Although thesaurus-based harmonization was performed to minimize this variability, some differences in terminology and loss of contextual information may have persisted and influenced cluster composition.

Future bibliometric studies could address these limitations by integrating multiple bibliographic databases, performing sensitivity analyses with and without self-citations, and combining bibliometric mapping with qualitative assessments of methodological quality and clinical outcomes. Such approaches may provide a more comprehensive understanding of not only the evolution of digital orthodontic research but also its clinical and translational relevance.

CONCLUSION

The results of this bibliometric study revealed a significant upward trend in literature production regarding the use of digital techniques in orthodontic practices, especially after 2020, with a peak reached in 2023. Furthermore, the results showed that academic collaborations have evolved into a more extensive global network, with a notable rise in internationally co-authored publications in the post-pandemic period. The observation that more recent studies receive higher citation rates suggests that newly developed methods are increasingly adopted over conventional approaches and that scientific interest is increasingly focused on emerging innovations. Despite the potential clinical benefits of digital technologies, the pace and direction of research and publication trends in digital orthodontics can be adversely affected by barriers to widespread implementation, such as costs, infrastructure requirements, training needs, and workflow integration.

Ethics

Ethics Committee Approval: Not applicable. This study is a bibliometric analysis of publicly available published literature and did not involve human participants or animals; therefore, ethics committee approval was not required.

Informed Consent: Not applicable, as no human participants were involved in this study.

Acknowledgments

The author sincerely thanks Prof. Dr. İbrahim Erhan Gelgör for his invaluable guidance and support throughout this study.

Financial Disclosure: The author declared that this study received no financial support.

Supplementary Tables: https://d2v96fxpocvxx.cloudfront.net/cf9d60d6-523c-458a-a2e6-78728d3ffbb0/content-images/8579ddb5-e994-400d-a5ee-1c56177bde9e.pdf

REFERENCES

1. Adams GL, Gansky SA, Miller AJ, Harrell WE Jr, Hatcher DC. Comparison between traditional 2-dimensional cephalometry and a 3-dimensional approach on human dry skulls. *Am J Orthod Dentofacial Orthop.* 2004;126(4):397-409. [CrossRef]

2. Dindaroğlu F, Yetkiner E. Cone beam computed tomography in orthodontics. *Turk J Orthod.* 2016;29(1):16-21. [CrossRef]
3. White SC, Pae EK. Patient image selection criteria for cone beam computed tomography imaging. *Seminars in Orthodontics*. 2009;15(1):19-28. [CrossRef]
4. Tarraf NE, Ali DM. Present and the future of digital orthodontics. *Seminars in Orthodontics*. 2018;24(4):376-385. [CrossRef]
5. Taneva E, Kusnoto B, Evans CA. 3D scanning, imaging, and printing in orthodontics. In: Bourzgui F, ed. *Issues in Contemporary Orthodontics*. InTech; 2015. [CrossRef]
6. Breuning K, Kau C. Digital planning and custom orthodontic treatment. Wiley-Blackwell; 2017. [CrossRef]
7. Kravitz ND, Groth C, Jones PE, Graham JW, Redmond WR. Intraoral digital scanners. *J Clin Orthod*. 2014;48(6):337-347. [CrossRef]
8. Grauer D. Quality in orthodontics: the role of customized appliances. *J Esthet Restor Dent.* 2021;33(1):253-258. [CrossRef]
9. Nadjmi N, Mollemans W, Daelemans A, Van Hemelen G, Schutyser F, Bergé S. Virtual occlusion in planning orthognathic surgical procedures. *Int J Oral Maxillofac Surg.* 2010;39(5):457-462. [CrossRef]
10. Stokbro K, Aagaard E, Torkov P, Bell RB, Thygesen T. Virtual planning in orthognathic surgery. *Int J Oral Maxillofac Surg.* 2014;43(8):957-965. [CrossRef]
11. Christensen LR. Digital workflows in contemporary orthodontics. *APOS*. 2017;7(1):12-18. [CrossRef]
12. Mujagic M, Fauquet C, Galletti C, Palot C, Wiechmann D, Mah J. Digital design and manufacturing of the Lingualcare bracket system. *J Clin Orthod.* 2005;39(6):375-382; quiz 370. [CrossRef]
13. Proffit W, Fields H, Sarver DM. Kombine cerrahi ve ortodontik tedavi. In: *Güncel Ortodonti*. 6th ed. Ankara Nobel Kitabevi; 2013:657-709. [CrossRef]
14. Choi HI, Jung SK, Baek SH, et al. Artificial intelligent model with neural network machine learning for the diagnosis of orthognathic surgery. *J Craniofac Surg.* 2019;30(7):1986-1989. [CrossRef]
15. Kök H, Acilar AM, İzgi MS. Usage and comparison of artificial intelligence algorithms for determination of growth and development by cervical vertebrae stages in orthodontics. *Prog Orthod.* 2019;20(1):41. [CrossRef]
16. Fernandes EC, Nascimento Júnior MB, Paiva Tôrres ACS, Nóbrega FJO, Santos PB. The 100 most-cited articles in orthodontic journals in the last 20 years. *Am J Orthod Dentofacial Orthop.* 2022;161(3):e260-e276. [CrossRef]
17. Tarazona B, Lucas-Dominguez R, Paredes-Gallardo V, Alonso-Arroyo A, Vidal-Infer A. The 100 most-cited articles in orthodontics: a bibliometric study. *Angle Orthod.* 2018;88(6):785-796. [CrossRef]
18. Xu D, Wang YL, Wang KT, et al. A scientometrics analysis and visualization of depressive disorder. *Curr Neuropharmacol.* 2021;19(6):766-786. [CrossRef]
19. Fleming PS, Marinho V, Johal A. Orthodontic measurements on digital study models compared with plaster models: a systematic review. Orthod Craniofac Res. 2011;14(1):1-16. [CrossRef]
20. Stevens DR, Flores-Mir C, Nebbe B, Raboud DW, Heo G, Major PW. Validity, reliability, and reproducibility of plaster vs digital study models: comparison of peer assessment rating and Bolton analysis and their constituent measurements. *Am J Orthod Dentofacial Orthop.* 2006;129(6):794-803. [CrossRef]
21. Santoro M, Galkin S, Teredesai M, Nicolay OF, Cangialosi TJ. Comparison of measurements made on digital and plaster models. *Am J Orthod Dentofacial Orthop.* 2003;124(1):101-105. [CrossRef]
22. Mangano F, Gandolfi A, Luongo G, Logozzo S. Intraoral scanners in dentistry: a review of the current literature. *BMC Oral Health*. 2017;17(1):149. [CrossRef]
23. Kim SH, Kim KB, Choo H. New frontier in advanced dentistry: CBCT, intraoral scanner, sensors, and artificial intelligence in dentistry. *Sensors*. 2022;22(8):2942. [CrossRef]
24. van der Meer WJ, Vissink A, Ren Y. Full 3-dimensional digital workflow for multicomponent dental appliances: a proof of concept. *J Am Dent Assoc.* 2016;147(4):288-291. [CrossRef]
25. Park JH, Hwang HW, Moon JH, et al. Automated identification of cephalometric landmarks: part 1-comparisons between the latest deep-learning methods YOLOV3 and SSD. *Angle Orthod.* 2019;89(6):903-909. [CrossRef]
26. Jung SK, Kim TW. New approach for the diagnosis of extractions with neural network machine learning. *Am J Orthod Dentofacial Orthop.* 2016;149(1):127-133. [CrossRef]
27. Surdu A, Foia CI, Luchian I, et al. Telemedicine and digital tools in dentistry: enhancing diagnosis and remote patient care. *Medicina (Kaunas)*. 2025;61(5):826. [CrossRef]
28. Logan S, Riedy CA, Hargett K, Katebi N. Orthodontists' use of remote monitoring platforms pre-, amid, and post-COVID-19: a survey study. *BMC Oral Health.* 2024;24(1):480. [CrossRef]
29. Guo X, Shao Y. AI-driven dynamic orthodontic treatment management: personalized progress tracking and adjustments-a narrative review. *Front Dent Med.* 2025;6:1612441. [CrossRef]
30. Gomez CJ, Herman AC, Parigi P. Leading countries in global science increasingly receive more citations than other countries doing similar research. *Nat Hum Behav.* 2022;6(7):919-929. [CrossRef]
31. Kohus Z, Demeter M, Szigeti GP, Kun L, Lukács E, Czakó K. The influence of international collaboration on the scientific impact in V4 countries. *Publications*. 2022;10(4):35. [CrossRef]
32. Abramo G, D'Angelo CA, Solazzi M. The relationship between scientists 'research performance and the degree of internationalization of their research. *Scientometrics*. 2011;86(3):629-643. [CrossRef]
33. Lancho-Barrantes BS, Guerrero-Bote VP, De Moya-Anegón F. Citation increments between collaborating countries. *Scientometrics*. 2013;94(3):817-831. [CrossRef]
34. Way SF, Morgan AC, Larremore DB, Clauset A. Productivity, prominence, and the effects of academic environment. *Proc Natl Acad Sci USA*. 2019;116(22):10729-10733. [CrossRef]
35. Küffer M, Drescher D, Becker K. Application of the digital workflow in orofacial orthopedics and orthodontics: printed appliances with skeletal anchorage. *Applied Sciences.* 2022;12(8):3820. [CrossRef]
36. Kiki A, Özdoğan A, Kiki MA, Genç YS. Bibliometric analysis of maxillary expansion publications trends. Turk J Orthod. 2024;37(3):193-200. [CrossRef]
37. Polizzi A, Boato M, Serra S, D'Antò V, Leonardi R. Applications of artificial intelligence in orthodontics: a bibliometric and visual analysis. *Clin Oral Investig.* 2025;29(1):65. [CrossRef]
38. He X, Huang Z, Yang Y, et al. A bibliometric analysis of clear aligner treatment (CAT) from 2003 to 2023. Cureus. 2024;16(6):e63348. [CrossRef]
39. Quimby ML, Vig KW, Rashid RG, Firestone AR. The accuracy and reliability of measurements made on computer-based digital models. *Angle Orthod.* 2004;74(3):298-303. [CrossRef]
40. Houston WJ. The analysis of errors in orthodontic measurements. *Am J Orthod.* 1983;83(5):382-390. [CrossRef]
41. Zilberman O, Huggare JA, Parikakis KA. Evaluation of the validity of tooth size and arch width measurements using conventional and three-dimensional virtual orthodontic models. *Angle Orthod.* 2003;73(3):301-306. [CrossRef]
42. Jedliński M, Mazur M, Grocholewicz K, Janiszewska-Olszowska J. 3D scanners in orthodontics-current knowledge and future perspectives-a systematic review. *Int J Environ Res Public Health.* 2021;18(3):1121. [CrossRef]
43. Pandey R, Kamble R, Kanani H. Revolutionizing smiles: advancing orthodontics through digital innovation. *Cureus.* 2024;16(7):e64086. [CrossRef]