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.

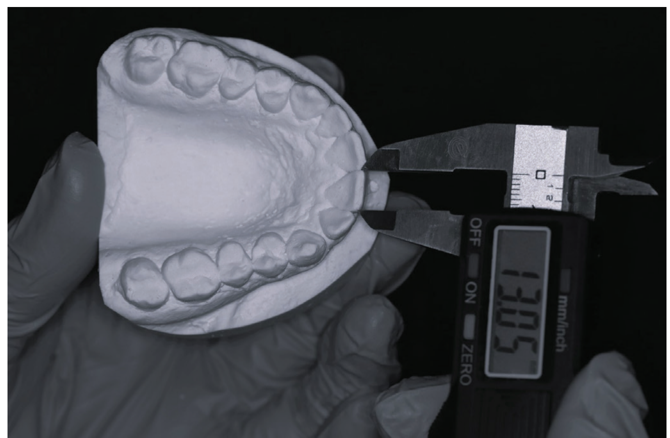

Figure 1. Measurement of mesiodistal tooth dimensions on plaster models using a high-precision digital calliper. The calliper was positioned at the widest mesiodistal diameter of the tooth. Each measurement was performed three times by the same examiner, and the mean value was recorded to ensure intra-examiner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

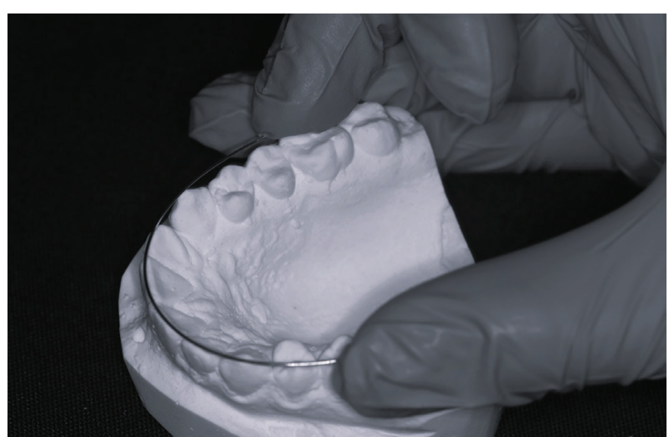

Figure 2. Measurement of arch dimensions on plaster models using digital callipers. Linear measurements, including intercanine width, intermolar width, and arch length, were obtained directly from these conventional models. The caliper beaks were positioned at the predefined anatomical landmarks to measure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

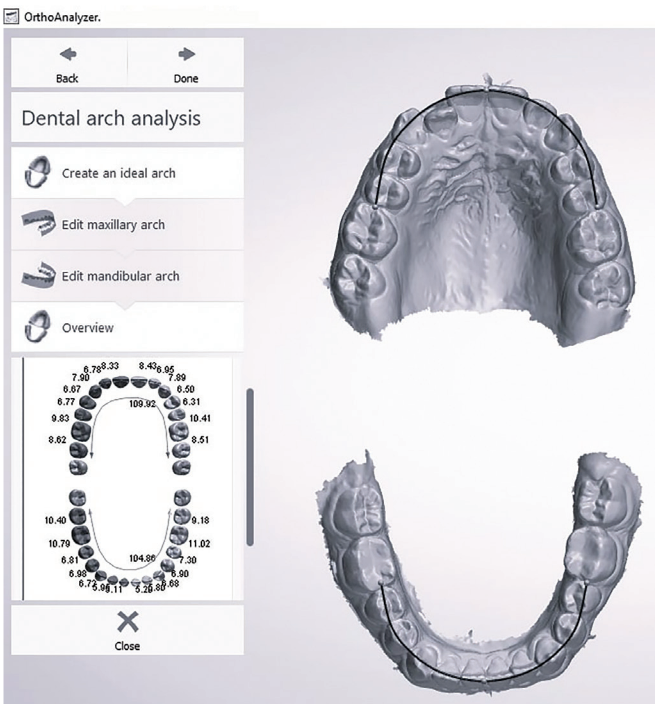

Figure 3. OrthoSystem digital model analysis interface (3Shape, Copenhagen, Denmark) showing maxillary and mandibular arch measurements. The left panel displays the dental arch analysis menu, with options for arch editing and measurements. The schematic diagram in the lower left shows individual tooth 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.

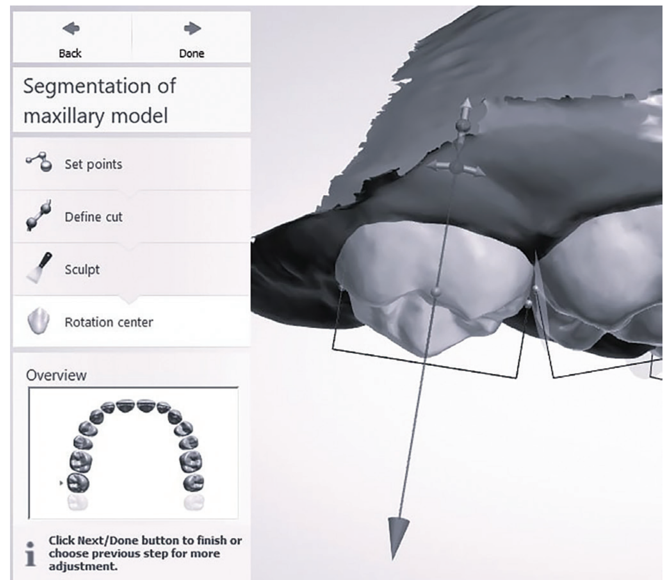

Figure 4. Manual segmentation interface in OrthoSystem software showing the tooth boundary delineation process. The left panel displays the segmentation tools, including set points for landmark placement, define cut for separating individual teeth, sculpt for refining tooth boundaries, and rotation center 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		
Mn	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*
	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*
Mx	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***
	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]