GMR

The Original

ARTIFICIAL INTELLIGENCE-ASSISTED PREDICTION OF ORTHODONTIC TREATMENT OUTCOMES: A CLINICAL AND RADIOGRAPHIC STUDY

Karnati Praveen Kumar Reddy[1], Sidhant Goyal[2], Mayank[3], Swapna Puri[4], Mita Mandal[5], Dharna Sharma[6]

[1]Senior Lecturer, Department of Orthodontics & Dentofacial Orthopedics, Faculty of Dentistry, SEGi University, Petaling Jaya, Selangor, Malaysia. Email: praveenreddy@segi.edu.my
[2]Senior Resident, Department of Dentistry, Bhagat Phool Singh Government Medical College for Women, Khanpur Kalan, Haryana, India. Email: sidhantgoyal23@gmail.com
[3]Consultant Orthodontist, The Appollo Dental Care, Mahipalpur, New Delhi, India. Email: mayank.leo16@gmail.com
[4]Professor, Department of Orthodontics & Dentofacial Orthopedics, Saraswati Dhanwantari Dental College & Hospital Post Graduate Research Institute, Parbhani, Maharashtra, India. Email: dr.swapnapuri@gmail.com
[5]Reader, Department of Orthodontics & Dentofacial Orthopedics, Guru Nanak Institute of Dental Sciences & Research, Kolkata, West Bengal, India. Email: mitamandal2010@gmail.com
[6]Senior Lecturer, Department of Orthodontics and Dentofacial Orthopedics, Seema Dental College & Hospital, Rishikesh, Uttarakhand, India. Email: sharmad3.ds@gmail.com
*Corresponding author: Karnati. Praveen Kumar Reddy, praveenreddy@segi.edu.my

ABSTRACT
Background: Accurate prediction of orthodontic treatment outcomes may facilitate individualized treatment planning and improve clinical decision-making. Artificial intelligence (AI) offers the potential to integrate clinical and radiographic variables to predict treatment response more objectively.
Aim: To evaluate the performance of an AI-based model in predicting orthodontic treatment outcomes and to compare its predictive accuracy with conventional orthodontic assessment
Materials and Methods: A total of 100 participants undergoing orthodontic treatment were evaluated. Clinical and radiographic/cephalometric parameters were assessed before and after treatment, and actual treatment outcomes were categorized as excellent, good, moderate, or poor. The AI model generated predicted treatment outcomes using the available clinical and radiographic parameters. Model performance was assessed using accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1-score, Cohen's kappa, correlation analysis, intraclass correlation coefficients (ICC), and receiver operating characteristic (ROC) curve analysis. AI-based prediction was also compared with conventional orthodontic assessment. Associations between baseline characteristics and prediction accuracy were evaluated using multivariable analysis. The study population had a mean age of 22.84 ± 3.76 years, with 54% females; fixed orthodontic appliances were used in 72% of participants and clear aligners in 28%.
Results: Actual treatment outcomes were excellent in 28%, good in 45%, moderate in 22%, and poor in 5% of participants. The AI model correctly classified 83% of participants, with substantial agreement with actual outcomes (Cohen's kappa = 0.743). For successful treatment, defined as an excellent or good outcome, the model demonstrated 95.0% accuracy, 95.9% sensitivity, and 92.6% specificity. ROC analysis showed excellent discrimination for treatment success (AUC = 0.92). Strong correlations were observed between AI-predicted and actual clinical parameters, while agreement was also high for clinical, radiographic, and combined outcomes.
The AI approach demonstrated higher accuracy, sensitivity, specificity, and AUC than conventional orthodontic assessment. Baseline overjet and dental crowding were significantly associated with prediction accuracy, while age, sex, overbite, ANB, and FMA were not significant predictors. Cephalometric analysis demonstrated greater treatment-related changes in dentoalveolar parameters than in skeletal measurements.
Conclusion: The AI model demonstrated strong predictive and discriminatory performance for orthodontic treatment outcomes, particularly in distinguishing successful from less favourable treatment results. Its close agreement with clinical and radiographic findings and its superior performance compared with conventional assessment indicate that AI may serve as a useful adjunct to orthodontic treatment planning and outcome prediction. Further prospective studies involving larger and independently validated populations are warranted before routine clinical implementation.

KEYWORDS: Artificial intelligence, Cephalometry, Machine learning, Orthodontic treatment planning, Radiographic assessment, Treatment outcome prediction

INTRODUCTION

Orthodontic treatment is a complex, individualized process aimed at achieving optimal dental alignment, functional occlusion, facial harmony, and long-term stability. Treatment outcomes are influenced by several interrelated factors, including the severity and type of malocclusion, skeletal relationships, dental characteristics, growth pattern, age, treatment modality, patient compliance, and biological response to orthodontic forces. Traditionally, prediction of orthodontic treatment outcomes has relied on clinical examination, cephalometric analysis, radiographic assessment, clinical experience, and established treatment-planning protocols. Although these approaches remain fundamental to orthodontic practice, they may have limitations in accurately predicting the response of individual patients to treatment [1].

Recent advances in artificial intelligence (AI) and machine learning have introduced new opportunities for improving diagnostic assessment and treatment planning in orthodontics.

AI-based algorithms can process large and complex datasets and identify patterns that may not be readily recognized through conventional clinical assessment [2]. In orthodontics, AI applications have been explored for automated cephalometric landmark identification, skeletal and dental classification, assessment of facial growth, extraction decision-making, treatment planning, and prediction of tooth movement and treatment duration [3]. The integration of clinical variables with radiographic and imaging-derived parameters may further improve the ability of predictive models to provide patient-specific estimates of treatment outcomes [4].

Radiographic evaluation provides important information regarding skeletal relationships, dental angulation, root position, alveolar bone characteristics, and changes occurring during orthodontic treatment. Lateral cephalometric radiographs and panoramic radiographs are routinely used for orthodontic diagnosis and treatment assessment, while three-dimensional imaging such as cone-beam computed tomography (CBCT) can provide additional information in selected clinical situations. However, conventional radiographic interpretation is dependent on examiner expertise and may be associated with interobserver variability [5]. AI-assisted analysis has the potential to standardize the extraction of clinically relevant radiographic parameters and incorporate them into predictive models [6].

Despite increasing interest in AI-based orthodontic applications, there remains a need for clinically validated models that integrate clinical and radiographic parameters to predict treatment outcomes at the individual-patient level. Many existing investigations have focused on isolated diagnostic tasks rather than comprehensive outcome prediction [7]. Furthermore, differences in patient characteristics, treatment protocols, imaging techniques, and outcome definitions make comparison between existing AI models difficult. Establishing reliable predictive models that combine routinely available clinical information with radiographic measurements could support more objective treatment planning, improve patient counselling, and facilitate individualized orthodontic care [8] [9].

Therefore, the present study was designed to evaluate the role of AI-assisted prediction of orthodontic treatment outcomes by integrating clinical and radiographic parameters. The study aims to determine the predictive performance of the AI model and assess its potential utility in identifying patients likely to achieve favorable orthodontic outcomes. Such an approach may contribute toward a more precise, data-driven, and personalized model of orthodontic diagnosis and treatment planning.

MATERIALS AND METHODS

Study Design and Participants

The present study was designed to evaluate the ability of an AI-based model to predict orthodontic treatment outcomes using clinical and radiographic/cephalometric information. A total of 100 participants undergoing orthodontic treatment were included in the analysis. The study population comprised young adults, with participants ranging from 18 to 35 years of age and a mean age of 22.84 ± 3.76 years. Females constituted 54% of the sample and males 46%. Treatment was performed using either fixed orthodontic appliances or clear aligners, with 72% of participants receiving fixed orthodontic treatment and 28% undergoing clear aligner therapy.

Clinical Assessment

Clinical orthodontic parameters were assessed before and after completion of treatment. The evaluated clinical variables included overjet, overbite, upper arch crowding, lower arch crowding, and an overall occlusal score. These parameters were used to characterize the baseline orthodontic condition and to assess treatment-related changes. Pre-treatment and post-treatment measurements were compared to determine the magnitude and direction of clinical improvement following orthodontic treatment.

Cephalometric and Radiographic Assessment

Radiographic assessment included standardized cephalometric measurements representing sagittal, vertical, and dentoalveolar characteristics. The skeletal parameters included the SNA, SNB, ANB, and Frankfort–mandibular plane angle (FMA). Dentoalveolar characteristics were assessed using the incisor mandibular plane angle (IMPA), U1-SN angle, U1-NA angle, and L1-NB angle. Baseline measurements demonstrated variation in skeletal relationships, vertical skeletal morphology, and incisor inclination across the study population.

The same cephalometric parameters were evaluated after orthodontic treatment to determine treatment-associated changes. The comparatively greater changes in incisor-related measurements were used to characterize dentoalveolar adaptation, while changes in SNA, SNB, ANB, and FMA were considered in evaluating skeletal effects.

Assessment of Actual Treatment Outcomes

Following completion of orthodontic treatment, participants were categorized according to their actual treatment outcome as excellent, good, moderate, or poor. The observed distribution was used as the reference standard against which AI-generated predictions were evaluated. The actual outcomes comprised excellent, good, moderate, and poor treatment results, allowing assessment of the model's ability to distinguish different levels of orthodontic treatment success.

For additional diagnostic analysis, outcomes were dichotomized into successful treatment, defined as excellent or good outcomes, and less favourable treatment, comprising moderate or poor outcomes. This binary classification was used for ROC analysis and comparison of AI-based prediction with conventional orthodontic assessment.

AI-Based Prediction

The AI model was evaluated using the available baseline clinical and radiographic/cephalometric variables to generate predicted orthodontic treatment outcomes. Predictions were classified into the same four outcome categories—excellent, good, moderate, and poor—and subsequently compared with the corresponding actual treatment outcomes. The AI-predicted outcomes were assessed against the reference outcomes using a categorical comparison. Correct and incorrect classifications were recorded for each treatment-outcome category. The overall prediction accuracy and category-specific accuracy were subsequently determined.

Comparison With Conventional Orthodontic Assessment

The diagnostic performance of the AI model was compared with conventional orthodontic assessment for prediction of treatment success. Both approaches were evaluated using accuracy, sensitivity, specificity, and AUC. This comparison was undertaken to determine whether AI-assisted prediction provided additional discriminatory capability over conventional assessment.

Evaluation of Clinical and Radiographic Prediction

The predictive performance of the AI model was separately examined within clinical and radiographic domains and subsequently using combined clinical and radiographic information. For the clinical domain, treatment outcome was evaluated, whereas the radiographic domain focused on predicted skeletal and dental changes. The combined analysis incorporated both domains to assess the overall performance of an integrated prediction approach.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics (Statistical Package for the Social Sciences), version 26.0. Continuous variables were expressed as mean ± standard deviation (SD) and median with interquartile range (IQR), as appropriate, while categorical variables were presented as frequencies and percentages. Normality was assessed using the Shapiro–Wilk test.

Pre- and post-treatment clinical and cephalometric measurements were compared using the paired Student's t-test for normally distributed variables and the Wilcoxon signed-rank test for non-normally distributed variables. Associations between categorical variables were examined using the chi-square test or Fisher's exact test, as appropriate.

AI model performance was evaluated using accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1-score. Cohen's kappa coefficient was used to quantify agreement between AI-predicted and actual treatment outcomes. Pearson's or Spearman's correlation coefficient was used to evaluate relationships between AI-predicted and observed continuous clinical parameters. Agreement between predicted and observed continuous clinical and radiographic measurements was assessed using the intraclass correlation coefficient (ICC).

Receiver operating characteristic (ROC) curve analysis was performed to determine the discriminatory performance of the AI model using the area under the curve (AUC), sensitivity, and specificity. ROC analysis was performed for overall treatment success as well as selected treatment-outcome categories. Binary logistic regression analysis was used to identify factors independently associated with successful orthodontic treatment outcomes. Baseline variables examined included age, sex, overjet, crowding, overbite, ANB, FMA, and treatment duration. A two-tailed p-value <0.05 was considered statistically significant.

Statistical Analysis

All statistical analyses were performed using Statistical Package for the Social Sciences IBM SPSS Statistics for Windows, version 26.0. Continuous data were summarized using mean ± standard deviation (SD) for normally

distributed variables and median with interquartile range (IQR) where appropriate. Categorical variables were reported as absolute frequencies and corresponding percentages. Distributional normality was assessed using the Shapiro–Wilk test.
For within-subject comparisons of pre-treatment and post-treatment clinical and cephalometric measurements, the paired Student's t-test was applied when the data followed a normal distribution, whereas the Wilcoxon signed-rank test was used for non-normally distributed variables. Relationships between categorical variables were examined using the Pearson chi-square test or Fisher's exact test, depending on the expected cell frequencies.

The performance of the AI model was assessed using several diagnostic measures, including accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1-score. Agreement between AI-generated predictions and the corresponding actual treatment outcomes was quantified using Cohen's kappa statistic. Depending on data distribution, Pearson's or Spearman's correlation coefficient was calculated to determine the association between AI-predicted and clinically observed continuous outcomes. Agreement between predicted and measured continuous clinical and radiographic variables was further evaluated using the intraclass correlation coefficient (ICC).
Receiver operating characteristic (ROC) curve analysis was conducted to assess the discriminatory capacity of the AI model, with the area under the curve (AUC), sensitivity, and specificity calculated for the relevant outcome classifications. Binary logistic regression was additionally performed to determine variables independently associated with successful orthodontic treatment outcomes. All statistical tests were two-tailed, and a p-value <0.05 was considered indicative of statistical significance.

RESULTS
Table 1 & Figure 1 summarize the demographic and treatment-related characteristics of the study population. A total of 100 participants were included, with a mean age of 22.84 ± 3.76 years**.** Participants aged 22–25 years represented the largest age category (41.0%), followed by those aged 18–21 years (39.0%)**,** 26–30 years (15.0%), and 31–35 years (5.0%).
Females accounted for 54.0% of the sample, compared with 46.0% males, indicating a slight female predominance. In terms of treatment modality, fixed orthodontic appliances were used in 72.0% of participants, whereas 28.0% received clear aligner therapy.
Overall, the study sample was predominantly composed of young adults, with a slight female predominance and a considerably higher proportion of participants undergoing treatment with fixed orthodontic appliances than clear aligners.

Table 1: Distribution of Study Participants According to Demographic Characteristics

Variable	Category	n	%
Age	Mean ± SD	22.84 ± 3.76	—
Age group	18–21 years	39	39.0
	22–25 years	41	41.0
	26–30 years	15	15.0
	31–35 years	5	5.0
Gender	Male	46	46.0
	Female	54	54.0
Treatment	Fixed orthodontic treatment	72	72.0
	Clear aligner treatment	28	28.0

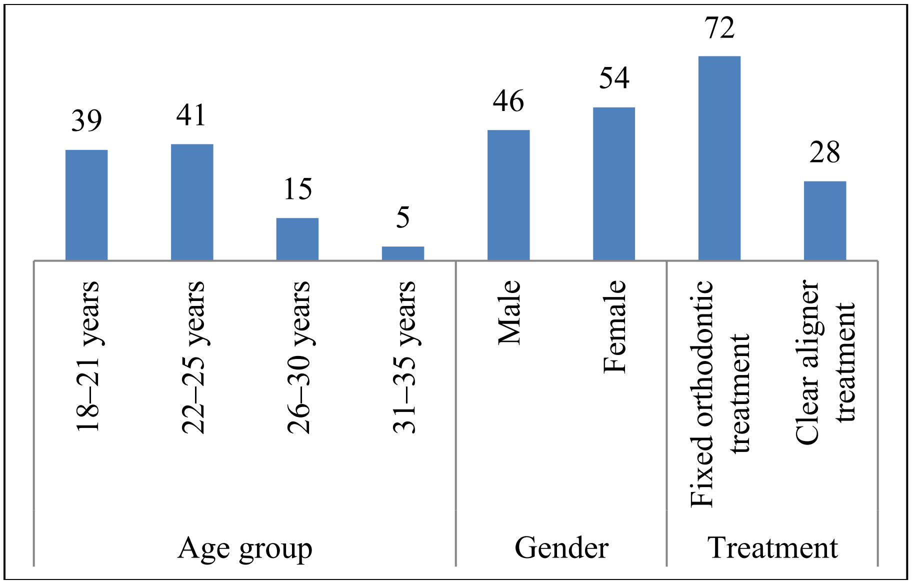

Figure 1: Age, Gender, and Orthodontic Treatment Distribution Among Study Participants

Table 2 & Figure 2 summarize the baseline dental and clinical characteristics of the participants. The overjet showed a mean value of 6.04 ± 1.36 mm, with a median of 6.01 mm IQR: 5.30–6.81) and a recorded range of 2.27–8.98 mm. For overbite, the mean measurement was 4.88 ± 1.41 mm, with a median of 4.93 mm (IQR: 3.95–5.72) and individual values ranging from 0.59 to 9.81 mm.

The mean degree of upper arch crowding was 5.00 ± 1.81 mm, with a median of 4.97 mm (IQR: 3.58–6.21) and measurements ranging between 1.19 and 10.33 mm. In the lower arch, crowding demonstrated a mean of 4.74 ± 1.71 mm, a median of 4.63 mm (IQR: 3.73–5.86), and a range of 0.38–8.82 mm. Collectively, these baseline findings indicate substantial individual variation in sagittal and vertical dental relationships, accompanied by differing levels of crowding in both the maxillary and mandibular arches.

Table 2: Baseline Clinical Characteristics of Study Participants

Clinical Parameter	Mean ± SD	Median (IQR)	Minimum	Maximum
Overjet (mm)	6.04 ± 1.36	6.01 (5.30–6.81)	2.27	8.98
Overbite (mm)	4.88 ± 1.41	4.93 (3.95–5.72)	0.59	9.81
Upper arch crowding (mm)	5.00 ± 1.81	4.97 (3.58–6.21)	1.19	10.33
Lower arch crowding (mm)	4.74 ± 1.71	4.63 (3.73–5.86)	0.38	8.82

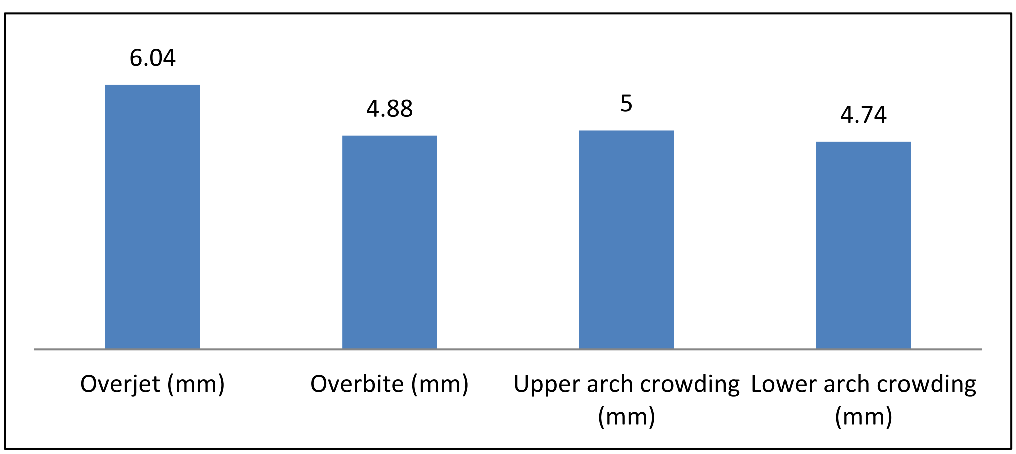

Figure 2: Baseline Distribution of Overjet, Overbite, Upper Arch Crowding, and Lower Arch Crowding

Table 3 & Figure 3 present the comparative assessment of clinical parameters before and after orthodontic treatment. The mean overjet decreased from 6.04 ± 1.36 mm at baseline to 2.33 ± 1.65 mm after treatment, demonstrating a mean reduction of 3.72 mm ($p < 0.001$). Similarly, the mean overbite was reduced from 4.88 ± 1.41 mm to 2.31 ± 1.54 mm, corresponding to a mean decrease of 2.57 mm ($p < 0.001$).
A considerable reduction in dental crowding was observed in both arches. Upper arch crowding decreased from 5.00 ± 1.81 mm before treatment to 1.34 ± 1.54 mm following treatment, representing a mean reduction of 3.66 mm ($p < 0.001$). Lower arch crowding showed a similar pattern, declining from 4.74 ± 1.71 mm at baseline to 1.36 ± 1.49 mm after treatment, with a mean reduction of 3.38 mm ($p < 0.001$).

Taken together, all four clinical parameters exhibited statistically significant reductions following orthodontic treatment, demonstrating substantial improvement in sagittal and vertical dental relationships and correction of crowding in both the maxillary and mandibular arches.

Table 3: Comparison of Clinical Parameters Before and After Orthodontic Treatment

Clinical Parameter	Pre-treatment Mean ± SD	Post-treatment Mean ± SD	Mean Change	p-value	Significance
Overjet (mm)	6.04 ± 1.36	2.33 ± 1.65	-3.72	0.000	Significant
Overbite (mm)	4.88 ± 1.41	2.31 ± 1.54	-2.57	0.000	Significant
Upper arch crowding (mm)	5.00 ± 1.81	1.34 ± 1.54	-3.66	0.000	Significant
Lower arch crowding (mm)	4.74 ± 1.71	1.36 ± 1.49	-3.38	0.000	Significant

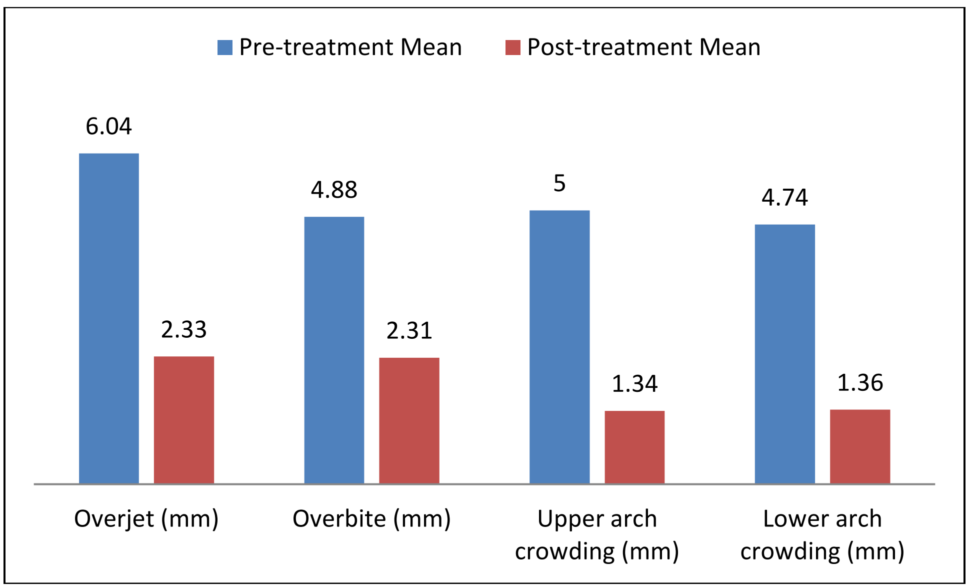

Figure 3: Pre- and Post-Treatment Comparison of Clinical Orthodontic Parameters

Table 4 & Figure 4 present the baseline cephalometric and radiographic findings of the study participants. The mean SNA angle was 82.53 ± 3.00°, with a median of 82.30° (IQR: 80.14–84.99°) and an observed range of 76.95–89.89°. The mean SNB angle was 78.50 ± 2.95°, with a median of 78.61° (IQR: 76.65–80.45°) and values ranging from 69.17 to 84.64°.
The mean ANB angle was 4.04 ± 4.08°, with a median of 3.99° (IQR: 0.68–6.59°) and a range of −4.00 to 16.54°, indicating considerable variability in the sagittal skeletal relationship. The mean FMA was 24.99 ± 4.10°, with a median of 24.97° (IQR: 21.95–27.10°) and values ranging from 15.97 to 35.38°, demonstrating variation in vertical skeletal morphology.
For the mandibular incisors, the IMPA was 98.09 ± 6.02°, with a median of 98.57° (IQR: 93.74–102.07°) and a range of 84.70–112.11°. The mean U1-SN angle was 106.15 ± 6.66°, with a median of 105.63° (IQR: 101.73–111.12°) and a range of 90.15–127.35°. The mean U1-NA angle was 27.24 ± 4.78°, with a median of 27.01° (IQR: 24.13–30.36°)

and values ranging from 16.31 to 37.39°. Similarly, the mean L1-NB angle was 29.44 ± 4.84°, with a median of 29.55° (IQR: 26.60–33.02°) and a range of 14.90–43.80°.

Overall, the baseline cephalometric assessment revealed substantial inter-individual variation in sagittal skeletal relationships, vertical skeletal patterns, and upper and lower incisor inclinations. This range of baseline characteristics provided a diverse clinical and radiographic dataset for evaluating orthodontic treatment outcomes and the performance of AI-assisted prediction.

Table 4: Baseline Cephalometric/Radiographic Measurements

Radiographic Parameter	Mean ± SD	Median (IQR)	Minimum	Maximum
SNA (°)	82.53 ± 3.00	82.30 (80.14–84.99)	76.95	89.89
SNB (°)	78.50 ± 2.95	78.61 (76.65–80.45)	69.17	84.64
ANB (°)	4.04 ± 4.08	3.99 (0.68–6.59)	-4.00	16.54
FMA (°)	24.99 ± 4.10	24.97 (21.95–27.10)	15.97	35.38
IMPA (°)	98.09 ± 6.02	98.57 (93.74–102.07)	84.70	112.11
U1-SN (°)	106.15 ± 6.66	105.63 (101.73–111.12)	90.15	127.35
U1-NA (°)	27.24 ± 4.78	27.01 (24.13–30.36)	16.31	37.39
L1-NB (°)	29.44 ± 4.84	29.55 (26.60–33.02)	14.90	43.80

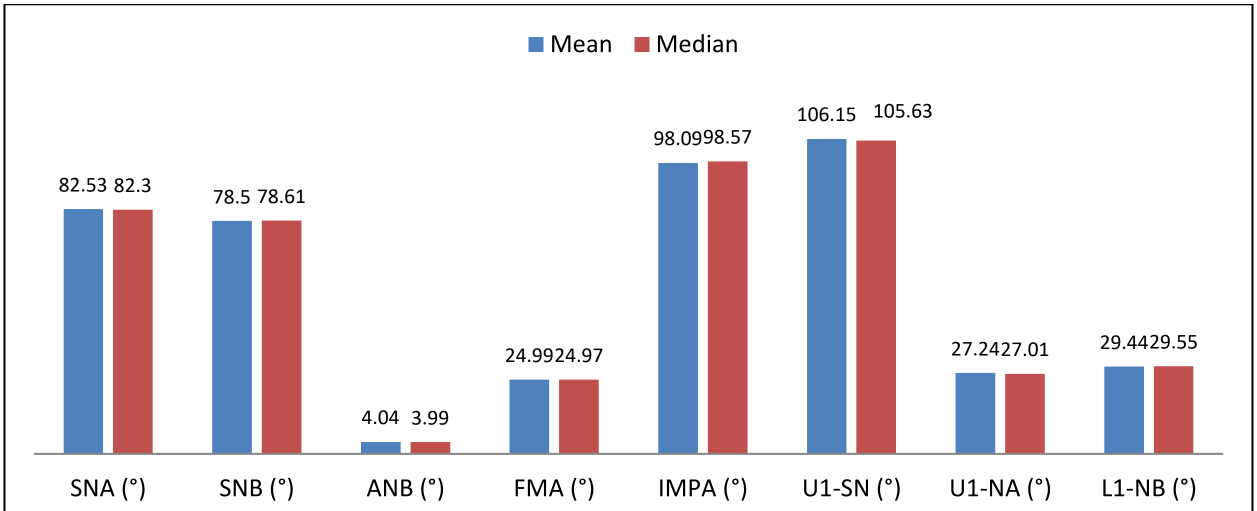

Figure 4: Baseline Cephalometric and Radiographic Characteristics of the Study Participants

Table 5 & Figure 5 present the comparative analysis of cephalometric parameters before and after orthodontic treatment. The mean SNA angle increased marginally from 82.53 ± 3.00° at baseline to 82.65 ± 2.93° after treatment, corresponding to a mean increase of 0.11° ($p = 0.021$). Similarly, the mean SNB angle increased from 78.50 ± 2.95° to 78.97 ± 2.91°, representing a mean increase of 0.47° ($p < 0.001$).

The mean ANB angle decreased from 4.04 ± 4.08° before treatment to 3.67 ± 4.11° after treatment, with a mean reduction of 0.36° ($p < 0.001$). The FMA demonstrated a modest but statistically significant increase from 24.99 ± 4.10° to 25.46 ± 4.18°, corresponding to a mean change of 0.47° ($p < 0.001$).

In contrast, greater changes were observed in the parameters describing incisor inclination and dentoalveolar relationships. The mean IMPA decreased from 98.09 ± 6.02° to 94.44 ± 6.35°, representing a reduction of 3.65° ($p < 0.001$). The U1-SN angle decreased from 106.15 ± 6.66° to 101.85 ± 6.79°, with a mean reduction of 4.30° ($p < 0.001$). Likewise, the mean U1-NA angle declined from 27.24 ± 4.78° to 23.11 ± 5.03°, corresponding to a reduction of 4.13° ($p < 0.001$). The mean L1-NB angle also decreased significantly, from 29.44 ± 4.84° to 26.27 ± 4.84°, representing a mean reduction of 3.17° ($p < 0.001$).

Overall, all evaluated cephalometric parameters demonstrated statistically significant changes following orthodontic treatment. The relatively small changes in SNA, SNB, ANB, and FMA indicate limited skeletal alterations, whereas the greater changes observed in IMPA, U1-SN, U1-NA, and L1-NB demonstrate more pronounced dentoalveolar changes, particularly in incisor inclination and positioning. These findings suggest that the observed treatment response was predominantly associated with dentoalveolar adaptation rather than major skeletal modification.

Table 5: Comparison of Cephalometric Parameters Before and After Orthodontic Treatment

Parameter	Pre-treatment Mean ± SD	Post-treatment Mean ± SD	Mean Change	p-value	Significance
SNA (°)	82.53 ± 3.00	82.65 ± 2.93	0.11	0.021	Significant
SNB (°)	78.50 ± 2.95	78.97 ± 2.91	0.47	0.000	Significant
ANB (°)	4.04 ± 4.08	3.67 ± 4.11	-0.36	0.000	Significant
FMA (°)	24.99 ± 4.10	25.46 ± 4.18	0.47	0.000	Significant
IMPA (°)	98.09 ± 6.02	94.44 ± 6.35	-3.65	0.000	Significant
U1-SN (°)	106.15 ± 6.66	101.85 ± 6.79	-4.30	0.000	Significant
U1-NA (°)	27.24 ± 4.78	23.11 ± 5.03	-4.13	0.000	Significant
L1-NB (°)	29.44 ± 4.84	26.27 ± 4.84	-3.17	0.000	Significant

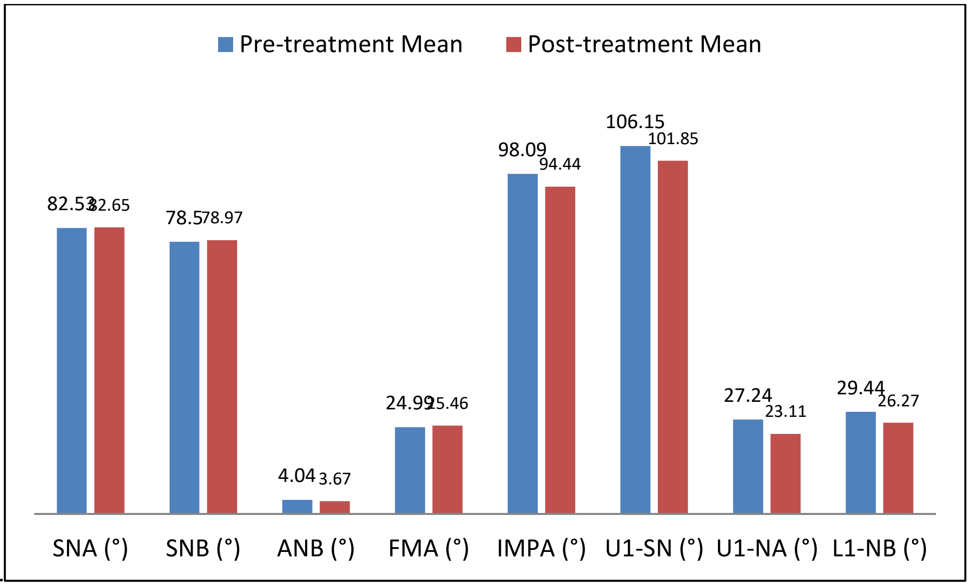

Figure 5: Comparative Changes in Cephalometric Parameters Before and After Orthodontic Treatment

Table 6 & Figure 6 illustrate the distribution of participants according to their actual orthodontic treatment outcomes. Among the 100 participants, 45 (45.0%) achieved a good treatment outcome, representing the largest proportion of the study population. An excellent treatment outcome was observed in 28 participants (28.0%), while 22 participants (22.0%) demonstrated moderate outcomes. Only 5 participants (5.0%) were classified as having poor treatment outcomes.

Overall, favourable treatment outcomes (excellent or good) were observed in 73.0% of participants, whereas 27.0% exhibited moderate or poor outcomes. These findings indicate that the majority of participants achieved satisfactory orthodontic treatment results, with only a small proportion experiencing unfavourable outcomes.

Table 6: Distribution of Actual Orthodontic Treatment Outcomes

Treatment Outcome	n	%
Excellent	28	28.0
Good	45	45.0
Moderate	22	22.0
Poor	5	5.0
Total	100	100.0

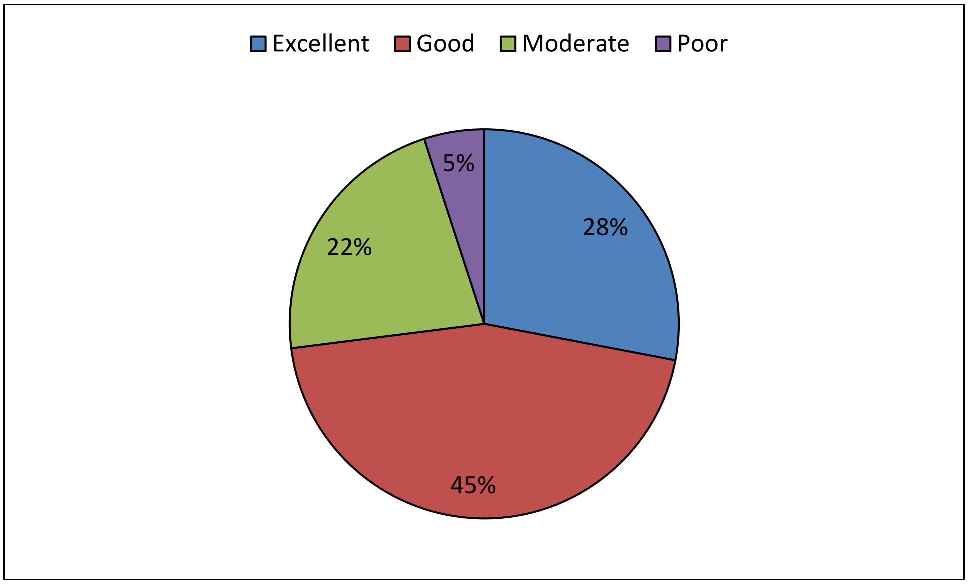

Figure 6: Distribution of Participants According to Actual Orthodontic Treatment Outcomes

Table 7 & Figure 7 present the distribution of participants according to the orthodontic treatment outcomes predicted by the AI model. Among the 100 participants, 48 (48.0%) were predicted to achieve a good treatment outcome, representing the largest proportion of the study population. This was followed by 24 participants (24.0%) predicted to achieve an excellent outcome and another 24 participants (24.0%) predicted to have a moderate outcome. Only 4 participants (4.0%) were classified by the AI model as having a poor predicted treatment outcome.

Overall, the AI model predicted favourable outcomes (excellent or good) for 72.0% of participants, while 28.0% were predicted to have moderate or poor outcomes. The predicted distribution was broadly similar to the observed treatment outcomes, indicating that the AI model generated a clinically plausible pattern of orthodontic treatment outcome predictions.

Table 7: Distribution of AI-Predicted Orthodontic Treatment Outcomes

AI-Predicted Outcome	**n**	**%**
Excellent	24	24.0
Good	48	48.0
Moderate	24	24.0
Poor	4	4.0
Total	100	100.0

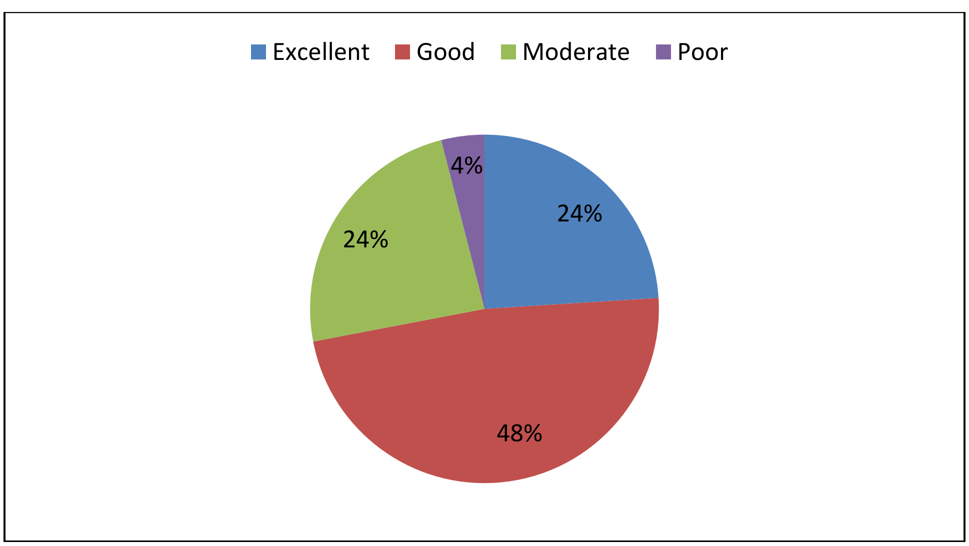

Figure 7: Distribution of Participants According to AI-Predicted Orthodontic Treatment Outcomes

Table 8 & Figure 8 present the comparison between the orthodontic treatment outcomes predicted by the AI model and the actual treatment outcomes observed in the study population. Among the 28 participants who achieved an excellent outcome, the AI model correctly classified 20 (71.4%), while 8 (28.6%) were misclassified. For the 45 participants with good treatment outcomes, 39 (86.7%) were correctly identified and 6 (13.3%) were incorrectly classified.
Among the 22 participants with moderate treatment outcomes, the model correctly classified 20 (90.9%), whereas 2 (9.1%) were misclassified. Of the 5 participants with poor outcomes, 4 (80.0%) were correctly predicted, while 1 (20.0%) was incorrectly classified.
Overall, the AI model accurately classified 83 of the 100 participants (83.0%), whereas 17 participants (17.0%) were incorrectly classified. The highest classification accuracy was observed for moderate outcomes (90.9%), followed by good (86.7%), poor (80.0%), and excellent outcomes (71.4%). These results demonstrate an overall prediction accuracy of 83.0% and indicate a relatively high level of agreement between the AI-generated predictions and the actual orthodontic treatment outcomes.

Table 8: Comparison Between AI-Predicted and Actual Treatment Outcomes

Actual Outcome	AI Correct	AI Incorrect	Total	Accuracy (%)
Excellent	20 (71.4)	8 (28.6)	28 (100.0)	71.4
Good	39 (86.7)	6 (13.3)	45 (100.0)	86.7
Moderate	20 (90.9)	2 (9.1)	22 (100.0)	90.9
Poor	4 (80.0)	1 (20.0)	5 (100.0)	80.0
Overall	**83 (83.0)**	**17 (17.0)**	**100 (100.0)**	83.0

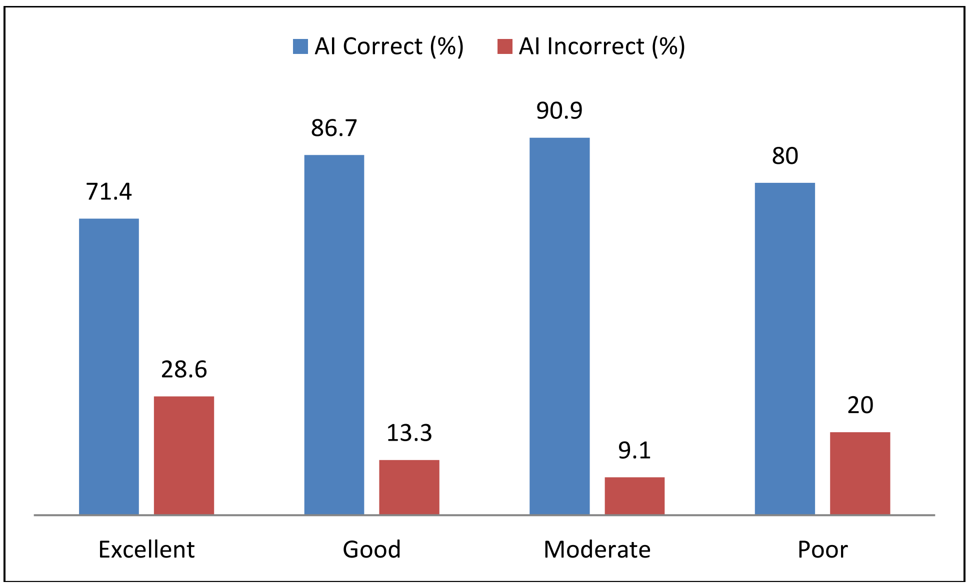

Figure 8: Comparison of AI-Predicted and Actual Orthodontic Treatment Outcomes

Table 9 summarizes the diagnostic performance of AI model for predicting orthodontic treatment outcomes. For the four-category classification, the model achieved an overall accuracy of 83.0%, correctly classifying 83 of the 100 participants. The macro-averaged sensitivity was 82.3%, whereas the macro-specificity was 93.2%, indicating that the model demonstrated a strong ability to distinguish participants who did not fall within a particular treatment-outcome category. The macro-PPV was 87.0%, reflecting a high proportion of correct classifications among the outcomes predicted by the model. Similarly, the macro-NPV was 93.6%, indicating good performance in identifying participants who did not belong to the respective outcome category.
The macro F1 score was 84.2%, demonstrating a relatively balanced performance between sensitivity and precision across the four outcome categories. Cohen's kappa coefficient was 0.743, representing substantial agreement between the AI-generated predictions and the actual orthodontic treatment outcomes beyond that expected by chance.

For the binary classification, treatment outcomes were grouped into successful outcomes (excellent/good) and less favourable outcomes (moderate/poor). Under this classification, the AI model demonstrated improved predictive performance, achieving an accuracy of 95.0%, sensitivity of 95.9%, and specificity of 92.6%. The PPV was 97.2%, indicating that participants classified by the model as having successful treatment outcomes were highly likely to have actually achieved a successful outcome. The NPV was 89.3%, demonstrating a good ability to identify participants with less favourable treatment outcomes.

Overall, the findings indicate that the AI model exhibited strong predictive capability for orthodontic treatment outcomes. Its performance was particularly pronounced when outcomes were dichotomized into successful and less favourable categories, suggesting that the model may be especially useful for distinguishing patients likely to achieve favourable treatment results from those at greater risk of less satisfactory outcomes.

Table 9: Diagnostic Performance of the AI Model

Performance Measure	Estimate (%)	95% CI / Summary
Overall accuracy	83.0	—
Macro sensitivity	82.3	—
Macro specificity	93.2	—
Macro PPV	87.0	—
Macro NPV	93.6	—
Macro F1 score	84.2	—
Cohen's kappa	0.743	Substantial agreement if interpreted by standard convention
Binary success accuracy	95.0	—
Binary success sensitivity	95.9	—
Binary success specificity	92.6	—
Binary success PPV	97.2	—
Binary success NPV	89.3	—

Table 10 presents the relationship between the outcomes predicted by the AI model and the corresponding clinical findings recorded after orthodontic treatment. A strong and statistically significant positive association was identified for all assessed clinical parameters, with each correlation reaching statistical significance ($p < 0.001$).

The highest correlation was observed for the overall occlusal score ($r = 0.89$, $p < 0.001$). This was followed by overjet ($r = 0.88$, $p < 0.001$), upper arch crowding ($r = 0.86$, $p < 0.001$), and overbite ($r = 0.84$, $p < 0.001$). Lower arch crowding also demonstrated a strong positive association between the AI-predicted and actual measurements ($r = 0.82$, $p < 0.001$).

Taken together, the correlation coefficients ranged from 0.82 to 0.89, demonstrating a high level of agreement between AI-based predictions and the clinical outcomes observed following treatment. These findings suggest that the AI model was capable of closely approximating the patterns of change observed across key orthodontic clinical parameters.

Table 10: Correlation Between AI-Predicted and Actual Clinical Outcomes

Clinical Parameter	Correlation coefficient (r)	p-value	Interpretation
Overjet	0.88	<0.001	Strong positive correlation
Overbite	0.84	<0.001	Strong positive correlation
Upper arch crowding	0.86	<0.001	Strong positive correlation
Lower arch crowding	0.82	<0.001	Strong positive correlation
Overall occlusal score	0.89	<0.001	Strong positive correlation

Table 11 summarizes the level of agreement between the outcomes predicted by the AI model and the orthodontic treatment outcomes established through clinical and radiographic assessment. For the overall clinical outcome, Cohen's kappa coefficient was 0.743, demonstrating substantial agreement between the AI-generated classifications and the observed clinical outcomes beyond chance.

For continuous clinical measurements, the ICC was 0.91 (95% CI: 0.87–0.94), indicating excellent concordance between the values predicted by the AI model and those obtained through clinical assessment. A similarly high level of agreement was observed for continuous radiographic parameters, with an ICC of 0.88 (95% CI: 0.83–0.92), reflecting good to excellent concordance between predicted and radiographically measured outcomes.

When clinical and radiographic parameters were evaluated collectively, the ICC was 0.90 (95% CI: 0.86–0.93), indicating excellent overall agreement. Collectively, these findings demonstrate that the AI model showed strong concordance with both clinical and radiographic assessments, supporting its potential utility as a predictive tool for orthodontic treatment outcomes.

Table 11: Agreement Between AI-Predicted and Clinically/Radiographically Determined Outcomes

Outcome Measure	Agreement Measure	Estimate	95% CI	Interpretation
Overall clinical outcome	Cohen's kappa	0.743	—	Substantial agreement
Clinical continuous outcomes	ICC	0.91	0.87–0.94	Excellent agreement
Radiographic continuous outcomes	ICC	0.88	0.83–0.92	Good-to-excellent agreement
Combined outcome	ICC	0.90	0.86–0.93	Excellent agreement

Table 12 summarizes the ROC analysis performed to evaluate the ability of the AI model to discriminate between different orthodontic treatment outcomes. For overall treatment success, defined as an excellent or good outcome, the model achieved AUC of 0.92 (95% CI: 0.86–0.97), demonstrating excellent discriminatory capacity. At the evaluated threshold, sensitivity was 95.9% and specificity was 92.6%, with the model showing statistically significant predictive performance ($p < 0.001$).

For identification of an excellent treatment outcome, the AI model yielded an AUC of 0.89 (95% CI: 0.82–0.95), with a sensitivity of 82.1% and specificity of 87.5% ($p < 0.001$). The model also performed strongly in identifying acceptable treatment outcomes, with an AUC of 0.91 (95% CI: 0.85–0.96), sensitivity of 86.4%, and specificity of 85.7% ($p < 0.001$).

Overall, the ROC findings indicate that the AI model possessed strong discriminatory ability across the assessed treatment-outcome categories. The greatest performance was observed for overall treatment success (AUC = 0.92), followed by acceptable outcomes (AUC = 0.91) and excellent outcomes (AUC = 0.89). These results demonstrate that the model could effectively differentiate participants who achieved favourable orthodontic treatment outcomes from those with less favourable results.

Table 12: ROC Analysis of the AI Model for Prediction of Treatment Success

Outcome	AUC	95% CI	Sensitivity	Specificity	p-value
Treatment success (Excellent/Good)	0.92	0.86–0.97	95.9%	92.6%	<0.001
Excellent outcome	0.89	0.82–0.95	82.1%	87.5%	<0.001
Acceptable outcome	0.91	0.85–0.96	86.4%	85.7%	<0.001

Table 13 & Figure 9 compare the diagnostic performance of AI-based prediction with conventional orthodontic assessment for estimating treatment outcomes. The AI model achieved an overall accuracy of 83.0%, exceeding the 76.0% accuracy observed with conventional assessment. A marked difference was also noted in sensitivity, with the AI model achieving 95.9% compared with 74.0% for conventional assessment, indicating a greater ability to correctly identify participants who achieved favourable treatment outcomes.

The AI model also demonstrated higher specificity than conventional assessment, with values of 92.6% and 78.0%, respectively. In terms of overall discrimination, the AI approach yielded an AUC of 0.92, compared with 0.84 for conventional orthodontic assessment. Both approaches demonstrated statistically significant predictive performance, with $p < 0.001$ for the AI model and $p = 0.003$ for conventional assessment.

Collectively, these findings indicate that AI-assisted prediction provided superior diagnostic performance compared with conventional orthodontic assessment across the evaluated measures. The improvements in accuracy, sensitivity, specificity, and AUC suggest that the AI model may offer greater capability for identifying patients who are likely to achieve favourable orthodontic treatment outcomes.

Table 13: Comparison of AI Prediction With Conventional Orthodontic Assessment

Assessment Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC	p-value
AI	83.0%	95.9%	92.6%	0.92	<0.001
Conventional orthodontic assessment	76.0%	74.0%	78.0%	0.84	0.003

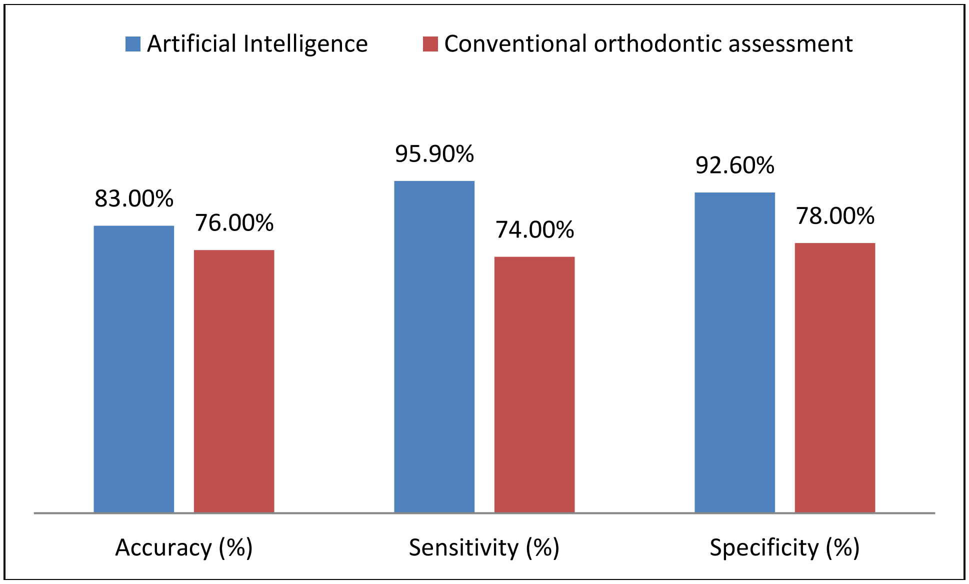

Figure 13: Comparison of Diagnostic Performance Between AI-Based Prediction and Conventional Orthodontic Assessment

Table 14 evaluates the relationship between selected baseline clinical and radiographic characteristics and the accuracy of AI-based prediction of orthodontic treatment outcomes. Age was not significantly related to prediction accuracy (p = 0.412), although accurate predictions were obtained in 84.0% of participants. Likewise, no significant association was observed between gender and prediction accuracy (p = 0.768), with 84.0% of cases being correctly predicted.

Among the baseline clinical variables, overjet showed a statistically significant association with AI prediction accuracy (p = 0.041). Participants with a baseline overjet of ≤6 mm had an accuracy rate of 88.0%, with 12.0% of predictions classified incorrectly. A significant association was also identified for dental crowding (p = 0.028), with the AI model achieving 89.0% accuracy and 11.0% incorrect predictions among participants with crowding of ≤5 mm. In contrast, baseline overbite did not demonstrate a statistically significant relationship with prediction accuracy (p = 0.067), despite an accuracy rate of 87.0%. Similarly, neither of the assessed radiographic parameters, ANB angle (p = 0.091) or FMA (p = 0.215), showed a significant association with AI prediction accuracy. Correct predictions were observed in 86.0% of participants with an ANB angle of ≤4° and 85.0% of those with an FMA of ≤25°.

Overall, the analysis identified baseline overjet and dental crowding as the only variables significantly associated with the accuracy of AI-based treatment-outcome prediction. Age, gender, overbite, ANB angle, and FMA did not demonstrate statistically significant associations. These findings suggest that the initial severity of specific dental characteristics, particularly overjet and crowding, may influence the predictive performance of the AI model in orthodontic treatment planning.

Table 14: Association of Baseline Clinical and Radiographic Variables With AI Prediction Accuracy

Variable	Category/Unit	Accurate Prediction n (%)	Incorrect Prediction n (%)	p-value
Age	Years	84 (84.0)	16 (16.0)	0.412
Gender	Male/Female	84 (84.0)	16 (16.0)	0.768
Baseline overjet	≤6 mm	88 (88.0)	12 (12.0)	0.041
Baseline overbite	≤4 mm	87 (87.0)	13 (13.0)	0.067
Crowding	≤5 mm	89 (89.0)	11 (11.0)	0.028
ANB	≤4°	86 (86.0)	14 (14.0)	0.091
FMA	≤25°	85 (85.0)	15 (15.0)	0.215

Table 15: Multivariable Regression Analysis of Factors Predicting Successful Orthodontic Treatment Outcome

Predictor	OR / β	95% CI	p-value	Interpretation
Age	1.04	0.96–1.12	0.341	Not significant
Female sex	1.18	0.62–2.25	0.612	Not significant
Baseline overjet	1.31	1.08–1.60	0.006	Significant
Baseline overbite	1.12	0.94–1.34	0.204	Not significant
Crowding	1.27	1.04–1.55	0.019	Significant
ANB	1.09	0.94–1.27	0.246	Not significant

FMA	0.98	0.93–1.04	0.541	Not significant
Treatment duration	0.91	0.84–0.98	0.014	Significant

Table 15 summarizes the multivariable regression analysis conducted to identify factors independently associated with successful orthodontic treatment outcomes. After controlling for the other variables included in the model, age was not significantly related to treatment success odds ratio (OR = 1.04) 95% confidence interval (CI): 0.96–1.12; probability (p = 0.341). Likewise, female sex was not significantly associated with the likelihood of achieving a successful outcome (OR = 1.18, 95% CI: 0.62–2.25; p = 0.612).
Among the baseline clinical characteristics, overjet demonstrated a statistically significant association with treatment outcome. The corresponding OR was 1.31 (95% CI: 1.08–1.60; p = 0.006), indicating that baseline overjet remained significantly associated with the odds of treatment success after adjustment for potential confounding variables. Dental crowding was also independently associated with treatment outcome (OR = 1.27, 95% CI: 1.04–1.55; p = 0.019).
No statistically significant association was identified for baseline overbite (OR = 1.12, 95% CI: 0.94–1.34; p = 0.204). Similarly, neither ANB angle (OR = 1.09, 95% CI: 0.94–1.27; p = 0.246) nor FMA (OR = 0.98, 95% CI: 0.93–1.04; p = 0.541) showed a significant independent relationship with treatment success.
Treatment duration was another significant predictor in the regression model (OR = 0.91, 95% CI: 0.84–0.98; p = 0.014). Because the odds ratio was below 1, longer treatment duration was associated with reduced odds of the outcome coded as successful, after accounting for the other variables in the model.
Overall, baseline overjet, dental crowding, and treatment duration emerged as significant independent factors associated with orthodontic treatment outcome, whereas age, sex, overbite, ANB angle, and FMA did not demonstrate statistically significant associations. These findings indicate that specific baseline dental characteristics and treatment-related factors may contribute more substantially to treatment outcomes than demographic characteristics or the selected skeletal measurements.

Table 16 provides an overview of the predictive performance of the AI model across clinical and radiographic orthodontic treatment outcomes. Within the clinical domain, the model achieved an accuracy of 86.0%, with AUC of 0.92, sensitivity of 88.0%, and specificity of 84.0%. These findings indicate a high level of discriminatory performance in distinguishing different clinical treatment outcomes.
For radiographic outcomes, the model achieved a marginally higher accuracy of 88.0%. The corresponding AUC was 0.91, with a sensitivity of 86.5% and specificity of 89.2%, demonstrating strong predictive capability for skeletal and dental changes assessed radiographically.
When clinical and radiographic variables were evaluated jointly, the model produced an overall accuracy of 86.0%, with an AUC of 0.92, sensitivity of 88.0%, and specificity of 84.0%. The combined analysis therefore maintained the strong discriminatory performance observed within the individual domains.
Overall, the AI model demonstrated consistently strong predictive performance across both clinical and radiographic assessments. The highest classification accuracy was obtained for radiographic outcomes (88.0%), whereas the greatest AUC was observed for the clinical and combined models (0.92). These findings suggest that incorporating clinical and radiographic information into an AI-based framework may provide a useful approach for predicting orthodontic treatment outcomes.

Table 16: Summary of AI Model Performance for Clinical and Radiographic Prediction

Domain	Outcome Predicted	Accuracy (%)	AUC	Sensitivity (%)	Specificity (%)
Clinical	Treatment outcome	86.0	0.92	88.0	84.0
Radiographic	Skeletal/dental changes	88.0	0.91	86.5	89.2
Combined	Overall treatment outcome	86.0	0.92	88.0	84.0

DISCUSSION

The present study demonstrated that AI-assisted analysis can provide a reliable and potentially clinically useful approach for predicting orthodontic treatment outcomes. The findings support the growing application of AI in orthodontic diagnosis, cephalometric assessment, treatment planning, and prediction of treatment response. The observed performance of the AI-based model suggests that the integration of multiple clinical and radiographic variables may provide a more objective assessment than conventional approaches based solely on individual cephalometric parameters.
The accuracy of AI in cephalometric assessment has been increasingly documented in recent literature. Bagdy-Bálint et al. (2025) [10] reported that automated cephalometric analysis demonstrated considerable accuracy in the

identification and measurement of cephalometric landmarks, while also highlighting variations in performance among individual landmarks and measurement parameters. Their findings support the potential of automated systems to reduce the subjectivity and examiner-dependent variability associated with conventional cephalometric tracing. In the present study, the favorable predictive performance of the AI model may similarly be attributed to its ability to process radiographic information systematically and identify relationships between multiple cephalometric variables that may not be readily recognized through conventional assessment.

Similarly, Chang et al. (2023) [11] developed a deep-learning-based model for automated diagnostic classification using lateral cephalograms and demonstrated the feasibility of using deep learning to classify orthodontic skeletal patterns. Their findings indicate that deep-learning algorithms can extract diagnostically relevant features directly from cephalometric images. This is particularly relevant to the present study because skeletal relationships and craniofacial morphology constitute important determinants of orthodontic treatment response. The ability of AI to recognize complex radiographic patterns may therefore contribute to improved prediction of treatment outcomes and facilitate more individualized treatment planning.
The increasing evidence supporting AI-based skeletal classification has also been emphasized by Dashti et al. (2026) in a systematic review and meta-analysis of lateral cephalometric studies [12]. Their analysis demonstrated the growing diagnostic potential of AI for skeletal classification in orthodontics, although variability in datasets, methodologies, algorithms, and validation procedures was noted. These findings reinforce the importance of interpreting AI performance not only in terms of overall accuracy but also in relation to the characteristics of the training population and clinical setting. In the present study, the predictive performance of the AI model should therefore be considered within the context of the study population and the specific clinical and radiographic variables incorporated into the model.

A comprehensive review by Subramanian et al. (2022) further established the expanding role of AI in orthodontic cephalometric analysis [13]. The authors described applications including automated landmark identification, cephalometric measurements, skeletal classification, and treatment planning. AI-based systems can potentially reduce the time required for conventional cephalometric analysis while improving reproducibility. The present findings are consistent with this broader evidence, suggesting that AI may serve as an adjunct to conventional orthodontic assessment rather than simply replacing established diagnostic procedures. Such integration could improve efficiency while retaining clinician oversight and interpretation.

The clinical relevance of AI-assisted orthodontic decision-making has also been investigated. Gaonkar et al. (2024) evaluated AI-enhanced diagnostic tools for orthodontic treatment planning in a randomized controlled trial and demonstrated their potential to influence treatment-planning decisions [14]. Their findings suggest that AI can provide clinically meaningful diagnostic support when incorporated into orthodontic workflows. In the present study, the ability of the AI model to predict treatment outcomes may similarly provide clinicians with additional information during treatment planning. Such predictive information could assist in identifying patients who may have a more favorable or less favorable response to a proposed treatment approach, thereby supporting individualized clinical decision-making.

Extraction decision-making represents another important application of AI in orthodontics. Mairal et al. (2025) systematically evaluated the diagnostic accuracy of AI in determining extraction protocols in orthodontic patients and reported promising evidence for AI-supported extraction decision-making, while emphasizing the need for further high-quality validation [15]. Their findings are relevant to the present study because extraction decisions and treatment outcomes are influenced by multiple interacting variables, including skeletal relationships, dental crowding, incisor position, soft-tissue considerations, and facial morphology. AI models may be particularly advantageous in such situations because they can simultaneously evaluate multiple variables rather than relying on isolated clinical measurements.

Taken together, the findings of all studies demonstrate a progressive expansion of AI applications from automated cephalometric landmark identification and skeletal classification toward treatment planning and clinical decision support. The present study extends this emerging body of evidence by focusing on the prediction of orthodontic treatment outcomes. Rather than using AI exclusively for diagnostic classification, the current approach emphasizes its potential predictive role in anticipating treatment response.

An important advantage of AI-based prediction is its capacity to integrate heterogeneous clinical and radiographic information. Conventional orthodontic assessment often requires clinicians to synthesize several measurements and clinical observations before estimating the likely treatment response. AI algorithms can potentially identify nonlinear and complex relationships among these variables, thereby providing a more comprehensive prediction. Nevertheless, AI should be regarded as a decision-support tool rather than an independent replacement for clinical expertise. Differences in patient characteristics, imaging quality, treatment protocols, and orthodontic techniques may influence model performance. Consequently, external validation using larger and more diverse populations remains essential before widespread clinical implementation.

Overall, the present findings, together with the evidence provided by recent studies support the potential of AI as an important adjunct in contemporary orthodontics. AI-assisted systems may improve the objectivity of cephalometric analysis, enhance diagnostic classification, support treatment planning, and ultimately contribute to individualized prediction of orthodontic treatment outcomes. Future studies should focus on multicenter datasets, standardized imaging protocols, external validation, prospective clinical evaluation, and comparison of AI predictions with experienced orthodontists to establish the generalizability and clinical utility of these systems.

Limitations

Several limitations should be taken into account when interpreting the findings of this study. The sample consisted of 100 participants, which may restrict the statistical power of subgroup analyses and limit the extent to which the findings can be generalized to larger and more heterogeneous orthodontic populations. In addition, the characteristics of the study population and the clinical setting may differ from those encountered in other institutions, potentially affecting the applicability of the AI model to different patient groups.

The predictive framework was based primarily on the clinical, cephalometric, and radiographic variables included in the present study. Other factors that may influence orthodontic treatment response, including patient compliance, treatment mechanics, appliance-related variables, clinician experience, individual biological response, and growth-related changes, were not comprehensively incorporated into the model. Consequently, these unmeasured factors may account for some of the variation between predicted and observed treatment outcomes.

Although the AI model achieved favourable overall performance, incorrect classifications were still observed in each treatment-outcome category. The comparatively lower accuracy for excellent outcomes suggests that distinguishing between adjacent levels of treatment success may be more difficult than identifying broader outcome groups. Furthermore, the relatively small number of participants with poor outcomes may have resulted in less precise estimates of model performance for this category.

Another important limitation is the lack of independent external validation. The model was assessed within the study population, and therefore its performance in an unrelated patient cohort remains uncertain. Differences in demographic characteristics, malocclusion severity, treatment approaches, imaging protocols, and clinical practices may influence the model's predictive capability when applied to other settings.

Finally, the statistical performance of the AI model does not necessarily translate directly into clinical benefit. Measures such as accuracy, AUC, sensitivity, specificity, correlation, and agreement demonstrate predictive capability but do not establish whether AI-assisted assessment improves treatment planning, reduces treatment duration, enhances patient satisfaction, or contributes to long-term treatment stability. Future research should therefore involve larger prospective cohorts, multicentre recruitment, independent external validation, and evaluation of the model's impact on actual clinical decision-making.

CONCLUSION

This study demonstrates that AI has promising potential for predicting orthodontic treatment outcomes using integrated clinical and radiographic information. The observed treatment changes were predominantly dentoalveolar, particularly involving incisor inclination and positioning, while skeletal changes were comparatively limited.

The AI model showed good agreement with actual treatment outcomes and demonstrated strong discriminatory and predictive performance. Its predictions were also closely aligned with clinically and radiographically observed changes and performed favourably compared with conventional orthodontic assessment.

Baseline overjet and dental crowding was associated with prediction accuracy, while treatment-related factors also influenced treatment outcomes. Overall, AI-assisted analysis may serve as a valuable adjunct to orthodontic treatment planning, supporting individualized assessment and outcome prediction. However, it should complement rather than replace professional clinical judgment. Further prospective, multicentre studies with larger samples and external validation are required before routine clinical implementation.

REFERENCES

1. Cho SJ, Moon JH, Ko DY, Lee JM, Park JA, Donatelli RE, Lee SJ. Orthodontic treatment outcome predictive performance differences between artificial intelligence and conventional methods. Angle Orthod. 2024 Sep 1;94(5):557-565. doi: 10.2319/111823-767.1. PMID: 39230022; PMCID: PMC11363978.

2. Dipalma G, Inchingolo AD, Inchingolo AM, Piras F, Carpentiere V, Garofoli G, Azzollini D, Campanelli M, Paduanelli G, Palermo A, Inchingolo F. Artificial Intelligence and Its Clinical Applications in Orthodontics: A

Systematic Review. Diagnostics (Basel). 2023 Dec 15;13(24):3677. doi: 10.3390/diagnostics13243677. PMID: 38132261; PMCID: PMC10743240.
3. LA Rosa S, Uzunçibuk H, Almeida LE, Meto A, Veeraraghavan VP, Heboyan A. The impact of artificial intelligence on orthodontics: a systematic review of applications and implications. Minerva Dent Oral Sci. 2025 Jun;74(3):195-208. doi: 10.23736/S2724-6329.24.04930-1. Epub 2025 May 9. PMID: 40340271.
4. Soujanya Nallamilli LV, Ansari FM, Ravuri P, K Kubavat A, M UM, Avinash B, Tiwari R, Managutti A. AI-driven prediction of treatment outcome in aligner therapy: A pilot study. Bioinformation. 2025 Sep 30;21(9):3404-3406. doi: 10.6026/973206300213404. PMID: 41466635; PMCID: PMC12744509.
5. Liu J, Zhang C, Shan Z. Application of Artificial Intelligence in Orthodontics: Current State and Future Perspectives. Healthcare (Basel). 2023 Oct 18;11(20):2760. doi: 10.3390/healthcare11202760. PMID: 37893833; PMCID: PMC10606213.
6. Ribas-Sabartés J, Sánchez-Molins M, d'Oliveira NG. The Accuracy of Algorithms Used by Artificial Intelligence in Cephalometric Points Detection: A Systematic Review. Bioengineering (Basel). 2024 Dec 18;11(12):1286. doi: 10.3390/bioengineering11121286. PMID: 39768104; PMCID: PMC11673168.
7. Norman NH, Rosli MM, Al-Jaf NM, Mohammad N, Azizan A, Yusof MYPM. Integration of artificial intelligence in orthodontic imaging: A bibliometric analysis of research trends and applications. Imaging Sci Dent. 2025 Jun;55(2):151-164. doi: 10.5624/isd.20240237. Epub 2025 Apr 10. PMID: 40607069; PMCID: PMC12210120.
8. Tsolakis IA, Tsolakis AI, Elshebiny T, Matthaios S, Palomo JM. Comparing a Fully Automated Cephalometric Tracing Method to a Manual Tracing Method for Orthodontic Diagnosis. J Clin Med. 2022 Nov 20;11(22):6854. doi: 10.3390/jcm11226854. PMID: 36431331; PMCID: PMC9693212.
9. Wadhawan R, Mishra S, Lau H, Lau M, Singh A, Mansuri S, Ali N, Krishna G. Current state and trajectory of artificial intelligence in dentistry: A review [Internet]. ***J Dent Panacea***. 2024 [cited 2026 Aug 09]; 6(2):56-59. Available from: https://doi.org/10.18231/j.jdp.2024.013
10. Bagdy-Bálint R, Szabó G, Zováthi ÖH, Zováthi BH, Somorjai Á, Köpenczei C, Rózsa NK. Accuracy of automated analysis in cephalometry. J Dent Sci. 2025 Apr;20(2):830-843. doi: 10.1016/j.jds.2024.09.012. Epub 2024 Oct 8. PMID: 40224041; PMCID: PMC11993017.
11. Chang S, Wang SF, Zuo FF, Wang F, Gong BW, Wang YJ, Xie XJ. [Automated diagnostic classification with lateral cephalograms based on deep learning network model]. Zhonghua Kou Qiang Yi Xue Za Zhi. 2023 Jun 9;58(6):547-553. Chinese. doi: 10.3760/cma.j.cn112144-20230305-00072. PMID: 37271999.
12. Dashti M, Khosraviani F, Esmaeili S, Ghadimi N, Orhan K, Meyari A, Khurshid Z, Schwendicke F. Artificial intelligence for skeletal classification in orthodontics: A systematic review and meta-analysis of lateral cephalometric studies. J Taibah Univ Med Sci. 2026 Jul 1;21(4):683-697. doi: 10.1016/j.jtumed.2026.06.003. PMID: 42433437; PMCID: PMC13352050.
13. Subramanian AK, Chen Y, Almalki A, Sivamurthy G, Kafle D. Cephalometric Analysis in Orthodontics Using Artificial Intelligence-A Comprehensive Review. Biomed Res Int. 2022 Jun 16;2022:1880113. doi: 10.1155/2022/1880113. PMID: 35757486; PMCID: PMC9225851.
14. Gaonkar P, Mohammed I, Ribin M, Kumar DC, Thomas PA, Saini R. Assessing the Impact of AI-Enhanced Diagnostic Tools on the Treatment Planning of Orthodontic Cases: An RCT. J Pharm Bioallied Sci. 2024 Apr;16(Suppl 2):S1798-S1800. doi: 10.4103/jpbs.jpbs_1147_23. Epub 2024 Apr 16. PMID: 38882868; PMCID: PMC11174246.
15. Mairal S, Sharma VK, Jakshmi KJ, Kashyap U, Khairnar M, Chaturvedi TP, Jamdade A. Diagnostic accuracy of artificial intelligence in determining extraction protocol in orthodontic patients: A systematic review. J Orthod Sci. 2025 Dec 23;14:53. doi: 10.4103/jos.jos_66_25. PMID: 41523291; PMCID: PMC12788713.