JHWCR JOURNAL OF HEALTH, WELLNESS AND COMMUNITY RESEARCH 3007-0570

DOI: https://doi.org/10.61919/bja2pg30

Original Article

Artificial Intelligence for Early Prediction of Poor Compliance in Clear Aligner Therapy

Neha Saeed[1], Muhammad Zeshan Shabbir[2], Wasim Akram[3], Ahmad Waqar[4], Anila Zanib[5], Sara Shahid Abbas[6]

[1] MS (Food and Nutrition), Department of Nutrition and Health Promotion, University of Home Economics, Lahore, Pakistan
[2] Assistant Professor, Department of Computer Science, Government Graduate College of Science, Wahdat Road, Lahore, Pakistan
[3] Perfect Solution Company, Al Khobar, Saudi Arabia
[4] Student, Beaconhouse Potohar Campus, Islamabad, Pakistan
[5] MSPT, Consultant Physiotherapist, Al-Hadi Clinic, Bahawalpur, Pakistan
[6] Dental Surgeon, Easy Life Health Clinic, Lahore, Pakistan

***Corresponding author: Neha Saeed, nehajunaid1122@gmail.com**

"Cite this Article" Received: 05 January 2026; **Accepted**: 05 March 2026; **Published**: 30 March 2026

Author Contributions: Concept: NS, MZS; Design: NS, MZS, SSA; Data Collection: NS, WA, AW, SSA; Analysis: MZS, WA; Drafting: NS, MZS, WA, AW, SSA. **Ethical Approval**: Dental Department, Hameed Latif Hospital, Lahore, Pakistan. **Informed Consent**: Written informed consent was obtained from all participants; **Conflict of Interest**: The authors declare no conflict of interest. **Funding**: No external funding; **Data Availability**: Available from the corresponding author on reasonable request; **Acknowledgments**: NA

ABSTRACT

Background: Successful clear aligner therapy depends on sustained patient adherence, but inadequate wear is often recognised only after treatment progression has been affected. Objective: To develop and internally validate models for predicting poor compliance during weeks 5–12 using baseline characteristics and behavioural data collected during the first four weeks of treatment. Methods: This prospective observational study included 250 adults receiving clear aligner therapy in dental clinics in Lahore, Pakistan. Demographic and clinical characteristics, daily wear time, digital check-ins, aligner-change behaviour, pain scores, and appointment attendance were recorded. Poor compliance was defined as wear below 20 hours per day on at least 30% of monitored days during weeks 5–12. Logistic regression, random forest, and gradient boosting were evaluated using repeated stratified 10-fold cross-validation. Results: Poor compliance occurred in 72 participants (28.8%). Compared with compliant participants, those with poor compliance had 3.20 fewer hours of early daily wear, more missed digital check-ins, higher odds of delayed aligner changes (OR: 7.50; 95% CI: 4.04–13.92), and higher odds of missed or rescheduled appointments (OR: 3.81; 95% CI: 1.99–7.30). Gradient boosting produced an AUC of 0.90, sensitivity of 83%, specificity of 84%, accuracy of 84%, and Brier score of 0.11. Conclusion: Early behavioural data showed promising internally validated performance for identifying subsequent poor compliance. External validation and comparison with simpler prediction approaches are required before clinical implementation. Keywords: Artificial intelligence; Clear aligners; Clinical prediction model; Gradient boosting; Machine learning; Orthodontic adherence; Wear-time monitoring.

INTRODUCTION

Clear aligner therapy has become an established orthodontic option for selected patients seeking a removable and less conspicuous alternative to fixed appliances. Contemporary aligner systems use digitally planned sequential appliances to produce controlled tooth movement, with clinical effectiveness depending on appropriate case selection, accurate treatment planning, aligner fit, biological response, and sustained patient adherence (1–3). Unlike fixed appliances, clear aligners can be removed by patients; consequently, insufficient daily wear may interrupt the planned sequence of tooth movement and compromise treatment progression.

Patients are generally advised to wear clear aligners for approximately 20–22 hours per day. However, actual wear behaviour varies considerably among individuals and may be affected by discomfort, motivation, lifestyle, treatment expectations, previous orthodontic experience, and engagement with the

treating clinician (4–6). Patients may also delay scheduled aligner changes, miss monitoring submissions, lose or damage aligners, or fail to attend planned appointments. These behaviours may contribute to inadequate aligner seating, loss of tracking, prolonged treatment, repeated digital scans, and the need for additional refinement aligners (2,3).

Reliable assessment of aligner adherence remains challenging. Self-reported wear time and patient diaries are readily available but are vulnerable to recall error and socially desirable reporting. Microelectronic wear-time sensors provide more objective information and have demonstrated substantial differences between prescribed and actual wear duration among patients using removable orthodontic appliances (5,6). Mobile applications and remote-monitoring platforms may provide additional behavioural indicators, including digital check-ins, scan submission, reminder responses, aligner-change timing, fit assessments, and communication with the clinic (7,8). Electronic reminders and feedback have also been associated with improved adherence, suggesting that timely identification of emerging behavioural problems may enable corrective intervention before clinically evident treatment failure develops (9).

In routine practice, inadequate adherence is often recognised only after a patient misses an appointment, reports delayed aligner changes, or presents with poor aligner fit or loss of tracking. At that stage, several weeks of treatment may already have been affected. Information available during the initial treatment period may therefore provide an opportunity to estimate the probability of subsequent poor adherence. Baseline demographic and clinical characteristics may contribute to risk assessment, but early treatment behaviours—including objectively recorded wear time, missed digital check-ins, delayed aligner changes, appointment attendance, and discomfort—may provide more direct and clinically actionable information.

Clinical prediction models combine multiple patient characteristics to estimate the probability of a defined future outcome. Logistic regression provides a transparent statistical approach, whereas machine-learning methods such as random forest and gradient boosting can model nonlinear relationships and complex interactions among predictors. Machine learning has increasingly been investigated in orthodontic diagnosis, treatment planning, image analysis, remote monitoring, and prediction of clear-aligner outcomes (10–12). Nevertheless, high apparent discrimination does not necessarily indicate clinical usefulness. Prediction models require clearly separated predictor and outcome periods, reproducible outcome definitions, appropriate handling of missing data, assessment of both discrimination and calibration, protection against overfitting and information leakage, and validation in populations distinct from those used for model development (13–15).

Evidence concerning prediction of clear-aligner adherence in Pakistani clinical settings remains limited. Local treatment behaviour may be influenced by education, occupational commitments, travel distance, financial constraints, family responsibilities, smartphone access, digital literacy, and expectations regarding orthodontic treatment. Previous studies from Pakistan have demonstrated that psychosocial expectations and understanding of orthodontic care vary across patient populations, although these investigations did not specifically develop models for predicting clear-aligner adherence (16,17). A locally developed model using prospectively collected clinical and behavioural data may therefore provide preliminary evidence concerning the feasibility of early risk stratification. The present study aimed to develop and internally validate models for predicting poor compliance during weeks 5–12 of clear aligner therapy using demographic and clinical characteristics recorded at baseline and behavioural information collected during the first four weeks of treatment. Logistic regression, random forest, and gradient-boosting approaches were compared. It was hypothesised that early behavioural indicators, particularly objectively recorded wear time and digital engagement, would provide greater predictive information than demographic characteristics alone.

MATERIALS AND METHODS

A prospective observational clinical prediction-model study was conducted in the orthodontic departments of selected public and private dental clinics in Lahore, Pakistan. Participants were observed during the first 12 weeks of clear aligner therapy without alteration of the treatment prescribed by their orthodontists. Baseline characteristics and information collected during weeks 1–4 constituted the predictor data, whereas poor compliance during weeks 5–12 was the prespecified outcome. This temporal separation was used to ensure that information occurring during the outcome period was not included among the model predictors. The study was planned with reference to the TRIPOD+AI recommendations for prediction models developed using regression or machine-learning methods, and potential sources of bias were considered according to the principal domains of PROBAST (14,15).

Patients beginning clear aligner treatment during the recruitment period were screened consecutively. Men and women aged 18 years or older were eligible when they had permanent dentition, owned or had regular access to a smartphone, were able to understand the treatment and digital-monitoring instructions, and agreed to follow the prescribed clear-aligner and monitoring procedures. Patients were excluded when they had craniofacial anomalies, severe periodontal disease, active temporomandibular joint disorders, medical conditions likely to affect orthodontic tooth movement, previous clear-aligner treatment, a requirement for orthognathic surgery, or an inability to attend the planned follow-up appointments. Eligible patients received information about the study, participation was voluntary, and written informed consent was obtained before data collection. Refusal to participate did not affect routine orthodontic care.

A non-probability consecutive sampling technique was used, and every eligible patient initiating treatment during the recruitment period was invited to participate. The recruitment target was 275 patients, allowing for an anticipated attrition of approximately 10%. Of the 275 patients assessed, 12 did not meet the eligibility criteria, eight declined participation, and five did not complete the baseline assessment. The final analytical sample comprised 250 participants. Because 72 participants experienced the outcome, the complexity of the developed models was restricted in relation to the available number of poor-compliance events to reduce the risk of overfitting.

Baseline information was collected using a structured data-collection form. Demographic variables included age, sex, educational level, occupation, marital status, travelling distance to the clinic, and approximate monthly household income. Previous orthodontic treatment, smoking status, familiarity with smartphone applications, and the principal reason for choosing clear aligners were also recorded. A trained orthodontist conducted the baseline clinical examination. Recorded clinical characteristics included malocclusion type, Angle classification, severity of crowding, overjet, overbite, extraction status, number of attachments, use of elastics, total number of planned aligners, anticipated treatment duration, and treatment complexity as determined from the initial clinical assessment and digital treatment plan.

All participants received routine verbal and written instructions regarding aligner insertion, removal, cleaning, storage, and scheduled replacement. They were advised to wear the aligners for approximately 20–22 hours per day and to remove them principally for eating, drinking where appropriate, and oral hygiene. The research protocol did not introduce reminders or behavioural interventions beyond those routinely used by the participating clinics.

Daily wear time was obtained from the electronic wear-time record available through the aligner-monitoring system or from the clinic-approved mobile-monitoring application. The available digital records were used to determine mean daily wear time, missed digital check-ins, delayed scan submissions, aligner-change timing, and responses to routine reminders. Appointment information was extracted from clinic records and included attendance, cancellation, rescheduling, late arrival, and

failure to attend. Predictor variables derived from digital and appointment records were restricted to information recorded during the first four weeks of treatment.

Participants completed a brief weekly assessment concerning pain, discomfort, speech difficulty, eating problems, satisfaction, motivation, and self-reported wear time. During routine follow-up, treating orthodontists examined aligner seating and tooth tracking using a standard clinical checklist intended to promote consistency across participating clinics. Information from these assessments was recorded as part of the prospective study database.

The primary outcome was poor compliance during weeks 5–12 of treatment. A participant was classified as having poor compliance when objectively recorded wear time was below 20 hours per day on at least 30% of monitored days during the outcome period. Wear-time observations recorded during weeks 1–4 were treated only as predictors and were not used to determine the outcome. Delayed aligner changes and appointment behaviour during weeks 1–4 were similarly considered predictor variables. Outcome classification was therefore based on subsequent wear-time behaviour during weeks 5–12, maintaining temporal separation between the predictors and predicted outcome. This operational definition reflected the clinical requirement for sustained daily use of removable orthodontic appliances and previous evidence demonstrating heterogeneous adherence to prescribed wear schedules (4–6,9).

Data were entered into a password-protected database using coded participant identifiers. Names and contact details were stored separately from the analytical dataset. Data-entry accuracy was assessed by comparing a randomly selected 10% of database records with the corresponding source forms. Records were also reviewed for implausible values, internal inconsistencies, and duplicate entries before analysis.

Continuous variables were summarised using mean and standard deviation when their distributions were approximately symmetrical and median and interquartile range when distributions were skewed. Categorical variables were reported as frequencies and percentages. Baseline and early behavioural characteristics were described for the complete sample and according to subsequent compliance status. Group comparisons were used for descriptive interpretation rather than automatic predictor selection. Candidate predictors were identified before model development on the basis of clinical relevance, temporal availability, and evidence from previous clear-aligner and adherence research.

Three prediction approaches were evaluated. Logistic regression was used as the reference model because of its interpretability, while random forest and gradient boosting were evaluated as machine-learning alternatives capable of identifying nonlinear and higher-order relationships. Predictor preprocessing, including treatment of missing predictor values, was conducted within the model-training process so that information from validation observations did not contribute to model fitting. Outcome values were not imputed. Class weighting was applied during model training to account for the lower frequency of poor-compliance events relative to compliant observations.

Internal validation was performed using repeated stratified 10-fold cross-validation. Participants were divided into 10 folds while preserving the relative frequency of poor compliance, and each fold was used in turn for validation while the remaining folds were used for model training. All predictor preprocessing and model fitting were repeated within the corresponding training data to reduce information leakage. Model performance was assessed using out-of-fold predictions.

Discrimination was evaluated using the area under the receiver operating characteristic curve. At the prespecified classification threshold used for model comparison, sensitivity, specificity, accuracy, positive predictive value, and negative predictive value were calculated. Overall probabilistic accuracy was assessed using the Brier score. Calibration was examined by comparing predicted and observed probabilities and by estimating the calibration slope. The comparative interpretation of the models considered discrimination, calibration, classification performance, model complexity, and clinical interpretability rather than relying exclusively on the area under the curve.

RESULTS

A total of 275 patients were assessed for eligibility. Twelve did not meet the eligibility criteria, eight declined participation, and five did not complete the baseline assessment. The final analysis therefore included 250 participants. The mean age was 26.4 ± 5.7 years, with an age range of 18–43 years. Of the participants, 158 (63.2%) were female and 92 (36.8%) were male. Poor compliance during weeks 5–12 was identified in 72 participants, corresponding to an event rate of 28.8%, while 178 (71.2%) remained compliant with the prescribed wear schedule.

Table 1. Demographic and Baseline Clinical Characteristics According to Subsequent Compliance Status

Variable	Total (n = 250)	Compliant (n = 178)	Poor Compliance (n = 72)	Effect Estimate (95% CI)	p-Value
Age, years	26.4 ± 5.7	27.0 ± 5.8	24.8 ± 5.3	MD −2.20 (−3.70 to −0.70)	0.006
Female sex	158 (63.2)	115 (64.6)	43 (59.7)	OR 0.81 (0.46–1.43)	0.469
Secondary education or below	82 (32.8)	50 (28.1)	32 (44.4)	OR 2.05 (1.16–3.62)	0.013
Moderate or severe treatment complexity	116 (46.4)	71 (39.9)	45 (62.5)	OR 2.51 (1.43–4.41)	0.001
Extraction treatment	54 (21.6)	36 (20.2)	18 (25.0)	OR 1.31 (0.69–2.51)	0.405

Values are presented as mean ± standard deviation or n (%). CI: confidence interval; MD: mean difference, calculated as poor-compliance group minus compliant group; OR: odds ratio for poor compliance.

Participants who subsequently demonstrated poor compliance were, on average, 2.20 years younger than compliant participants (95% CI: −3.70 to −0.70; p = 0.006). Secondary education or below was associated with approximately twice the unadjusted odds of poor compliance (OR: 2.05; 95% CI: 1.16–3.62), while moderate or severe treatment complexity was associated with 2.51-fold higher odds (95% CI: 1.43–4.41). Sex and extraction status were not clearly associated with the outcome because their confidence intervals included the null value.

Table 2. Early Behavioural and Treatment Characteristics According to Subsequent Compliance Status

Variable	Total (n = 250)	Compliant (n = 178)	Poor Compliance (n = 72)	Effect Estimate (95% CI)	p-Value
Mean early wear time, hours/day	20.2 ± 1.8	21.1 ± 0.8	17.9 ± 1.4	MD −3.20 (−3.55 to −2.85)	<0.001
Missed digital check-ins, median (IQR)	1 (0–3)	0 (0–1)	4 (2–6)	—	<0.001
Delayed aligner changes	70 (28.0)	28 (15.7)	42 (58.3)	OR 7.50 (4.04–13.92)	<0.001
Missed or rescheduled appointment	49 (19.6)	23 (12.9)	26 (36.1)	OR 3.81 (1.99–7.30)	<0.001
Early pain score	4.3 ± 1.9	3.9 ± 1.8	5.2 ± 1.8	MD 1.30 (0.80–1.80)	<0.001

Values are presented as mean ± standard deviation, median (interquartile range), or n (%). CI: confidence interval; IQR: interquartile range; MD: mean difference, calculated as poor-compliance group minus compliant group; OR: odds ratio for poor compliance. Effect estimates are unadjusted and were derived from the reported aggregate data. An effect estimate for missed digital check-ins could not be calculated from the reported medians and interquartile ranges. Early behavioural differences were more pronounced than baseline demographic differences. Participants who subsequently developed poor compliance wore their aligners for an average of 3.20 fewer hours per day during the first four weeks than compliant participants (95% CI: −3.55 to −2.85; p < 0.001). The poorly compliant group also recorded a median of four missed digital check-ins compared with zero in the compliant group.

Delayed aligner changes were reported in 42 of 72 participants with poor compliance and 28 of 178 compliant participants. This corresponded to an unadjusted odds ratio of 7.50 (95% CI: 4.04–13.92). Similarly, participants who missed or rescheduled an early appointment had 3.81-fold higher unadjusted odds of subsequent poor compliance (95% CI: 1.99–7.30). Mean early pain scores were 1.30 points higher in the poor-compliance group than in the compliant group (95% CI: 0.80–1.80; p < 0.001).

Table 3. Internally Validated Performance of Models for Predicting Poor Compliance

Model	AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)	Brier Score
Logistic regression	0.74	69	72	71	49	85	0.18

Model	AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)	Brier Score
Random forest	0.85	78	80	79	61	90	0.14
Gradient boosting	0.90	83	84	84	68	92	0.11

AUC: area under the receiver operating characteristic curve; NPV: negative predictive value; PPV: positive predictive value. A lower Brier score indicates better probabilistic accuracy. Performance estimates were obtained through repeated stratified 10-fold cross-validation. Confidence intervals and cross-validation variability were not available in the reported analytical output. Logistic regression demonstrated an AUC of 0.74, sensitivity of 69%, specificity of 72%, and overall accuracy of 71%. The random-forest model showed higher discrimination, with an AUC of 0.85, and achieved a sensitivity of 78%, specificity of 80%, and accuracy of 79%.

Gradient boosting produced the highest reported discrimination, with an AUC of 0.90. At the classification threshold used in the analysis, the model achieved a sensitivity of 83%, specificity of 84%, and accuracy of 84%. Its positive predictive value was 68%, whereas its negative predictive value was 92%. The Brier score was 0.11, compared with 0.14 for random forest and 0.18 for logistic regression. These point estimates indicated better overall predictive performance for gradient boosting within the internal cross-validation procedure, although the statistical uncertainty surrounding differences between the models could not be evaluated from the available output.

Table 4. Comparative Interpretation of Model Performance

Performance Domain	Logistic Regression	Random Forest	Gradient Boosting
Discrimination, AUC	0.74	0.85	0.90
Poor-compliance cases identified, sensitivity (%)	69	78	83
Compliant cases correctly classified, specificity (%)	72	80	84
Overall correct classification, accuracy (%)	71	79	84
Low-risk classification reliability, NPV (%)	85	90	92
Probabilistic error, Brier score	0.18	0.14	0.11

AUC: area under the receiver operating characteristic curve; NPV: negative predictive value. Values reproduce the internally validated performance estimates reported in the study.

The gradient-boosting model demonstrated the most favourable point estimates across all reported performance measures. Compared with logistic regression, gradient boosting had an absolute AUC difference of 0.16, an absolute sensitivity difference of 14 percentage points, an absolute specificity difference of 12 percentage points, and an absolute accuracy difference of 13 percentage points. Its Brier score was 0.07 lower than that of logistic regression, indicating lower average prediction error within the internal validation procedure.

Compared with random forest, gradient boosting had an absolute AUC difference of 0.05 and improved sensitivity, specificity, and accuracy by five, four, and five percentage points, respectively. The Brier score was 0.03 lower. These differences should be interpreted as descriptive because confidence intervals and paired comparisons between the cross-validated model predictions were not available.

Mean aligner wear time during the first four weeks was reported as the most influential predictor in the gradient-boosting model. Missed digital check-ins ranked second, followed by delayed aligner changes and appointment behaviour. Baseline demographic variables contributed less predictive information than early treatment behaviour. The reported calibration slope for the gradient-boosting model was 0.96, indicating close agreement between predicted and observed risks across much of the probability range. A small degree of risk overestimation was observed in the highest predicted-risk group.

Overall, 72 of 250 participants developed poor compliance during weeks 5–12. Reduced early wear time represented the largest observed behavioural difference between the outcome groups, while delayed aligner changes and missed or rescheduled appointments showed the strongest unadjusted categorical associations. Among the three evaluated approaches, gradient boosting produced the highest internally validated discrimination and the lowest reported Brier score. These results support the predictive

relevance of early behavioural data but do not, in the absence of external validation and uncertainty estimates, establish superiority or clinical transportability of the gradient-boosting model.

Figure 1 Early predictors and internally validated model performance for poor compliance during clear aligner therapy. Panel A presents the compliance outcome during weeks 5–12, with poor compliance observed in 72 of 250 participants (28.8%). Panel B displays unadjusted odds ratios and 95% confidence intervals for categorical predictors. Delayed aligner changes demonstrated the strongest association with subsequent poor compliance (OR: 7.50; 95% CI: 4.04–13.92), followed by missed or rescheduled appointments (OR: 3.81; 95% CI: 1.99–7.30), moderate or severe treatment complexity (OR: 2.51; 95% CI: 1.43–4.41), and secondary education or below (OR: 2.05; 95% CI: 1.16–3.62). Panel C presents mean differences between participants with poor compliance and compliant participants. The poor-compliance group was younger by 2.20 years, wore aligners for 3.20 fewer hours per day during the first four weeks, and had pain scores 1.30 points higher. Panel D compares internally validated model performance. Gradient boosting produced the highest AUC (0.90), sensitivity (83%), specificity (84%), accuracy (84%), and negative predictive value (92%), together with the lowest Brier score (0.11), compared with random forest and logistic regression.

DISCUSSION

The present study evaluated whether information available during the first four weeks of clear aligner therapy could predict poor compliance during weeks 5–12. The principal finding was that early treatment behaviour provided substantially greater predictive information than most baseline demographic characteristics. Participants who subsequently demonstrated poor compliance had lower mean daily aligner wear time, more missed digital check-ins, more frequent delays in aligner changes, more missed or rescheduled appointments, and higher early pain scores. Among the evaluated approaches, gradient boosting produced the most favourable internally validated point estimates, with an area under the receiver operating characteristic curve of 0.90, sensitivity of 83%, specificity of 84%, accuracy of 84%, and a Brier score of 0.11. These findings indicate that patterns associated with later poor compliance were already detectable during the initial treatment period; however, the results represent internal model performance and do not establish transportability to other orthodontic populations.

Poor compliance was identified in 72 of the 250 participants, corresponding to an event rate of 28.8%. Previous research has similarly demonstrated substantial variation in adherence to clear aligner therapy and other removable orthodontic appliances (4–6). Direct comparison of event rates across studies remains difficult because prescribed wear duration, monitoring technology, follow-up period, and operational definitions of adherence differ. Nevertheless, the frequency observed in the present study

reinforces the clinical importance of monitoring actual treatment behaviour rather than relying solely on retrospective patient reports.

Mean aligner wear time during the first four weeks represented the largest continuous difference between the outcome groups. Participants who subsequently developed poor compliance wore their aligners for an average of 3.20 fewer hours per day than compliant participants. Electronic monitoring studies have demonstrated that actual wear duration may differ substantially from the duration reported by patients or prescribed by clinicians (5,6). The current findings extend this evidence by suggesting that early wear behaviour may also predict the persistence of inadequate wear during subsequent treatment. This relationship should be interpreted carefully because early and later wear time represent temporally separated measurements of the same underlying behaviour. The model may therefore identify behavioural persistence rather than an independent biological or clinical risk mechanism.

Missed digital check-ins were also more frequent among participants who subsequently demonstrated poor compliance. Failure to complete digital monitoring may reflect reduced treatment engagement, forgetfulness, competing responsibilities, difficulties using the monitoring platform, unreliable internet access, disabled notifications, or limited digital literacy. Remote-monitoring systems can improve communication and provide timely information regarding aligner fit and treatment progression, but their effectiveness depends on sustained patient interaction with the technology (7,8). Electronic reminders and feedback have previously been associated with improved clear-aligner adherence, indicating that digital engagement may represent both a predictor and a potentially modifiable treatment behaviour (9). Clinicians should therefore explore the reason for missed submissions rather than interpreting them automatically as unwillingness to cooperate.

Delayed aligner changes demonstrated the strongest unadjusted categorical association with subsequent poor compliance. Participants with delayed changes had 7.50-fold higher unadjusted odds of poor compliance than those without delayed changes. This finding is clinically plausible because delayed replacement may occur when the preceding aligner has not been worn for the prescribed duration, has been lost, does not fit adequately, or has caused discomfort. Inadequate tracking and discrepancies between planned and achieved tooth movement may increase the need for additional refinement aligners (18). Systematic reviews have also shown that the predictability of individual tooth movements varies across clear-aligner protocols, even under appropriate treatment conditions (1–3,19,20). Irregular wear or delayed progression through the aligner sequence may further increase the difference between planned and achieved movement.

Missed or rescheduled appointments were associated with 3.81-fold higher unadjusted odds of poor compliance. Attendance and aligner wear may both reflect the participant's ability to maintain treatment routines, although appointment disruption should not be interpreted exclusively as poor motivation. Travel distance, employment obligations, treatment costs, family responsibilities, traffic conditions, and clinic scheduling may interfere with attendance. Psychosocial expectations and experiences of orthodontic treatment have been shown to vary among Pakistani patients (16,17). Appointment behaviour should therefore be interpreted in its social and logistical context, particularly when prediction models are used to guide patient communication.

Participants with poor compliance were younger, more frequently had secondary education or below, and more often had moderate or severe treatment complexity. However, the observed differences in early behaviour were larger and more clinically direct than the demographic differences. Previous studies have reported associations between adherence and patient characteristics, treatment demands, motivation, and personality traits (4,21). Demographic variables may help describe groups at increased risk but should not be used in isolation to label individual patients. Behaviour measured during treatment may provide a more immediate basis for supportive intervention, although the potential for digital exclusion and unequal access to monitoring technology must be considered.

Early pain scores were 1.30 points higher among participants who subsequently demonstrated poor compliance. Discomfort may reduce wear duration, delay reinsertion after meals, and discourage progression to the next aligner. However, pain is subjective and may be influenced by treatment complexity, individual pain sensitivity, anxiety, aligner fit, and communication with the clinician. The association should therefore be interpreted as a predictive pattern rather than evidence that pain independently caused poor compliance. Early identification and management of discomfort may nevertheless be clinically appropriate, particularly when reduced wear time and delayed aligner changes occur concurrently.

Gradient boosting produced better reported point estimates than logistic regression and random forest. Machine-learning methods can capture nonlinear relationships and interactions without requiring all functional forms to be prespecified, which may be advantageous when several behavioural variables jointly contribute to risk (10–12). However, the present results do not demonstrate that nonlinear relationships were responsible for the apparent performance difference because interaction analyses and model-explanation outputs were not reported. In addition, selecting the best-performing model from several evaluated algorithms may produce optimism unless model tuning and performance estimation are completely separated within the resampling procedure.

The incremental clinical value of the gradient-boosting model also requires careful consideration. Early mean wear time was the dominant predictor, and a simple threshold based on wear duration or a parsimonious regression model may provide performance close to that of a more complex algorithm. A complex model should be preferred only when it provides reproducible improvement in discrimination, calibration, or clinical utility over simpler approaches. Future analyses should therefore compare the full gradient-boosting model with early wear time alone, a small clinically selected logistic-regression model, and routine clinical assessment.

The gradient-boosting model had a reported calibration slope of 0.96 and a Brier score of 0.11, suggesting close agreement between predicted and observed risks within the internally validated dataset. Calibration is essential because a model may distinguish higher-risk from lower-risk participants while systematically overestimating or underestimating absolute risk (13–15). The small degree of risk overestimation reported in the highest-risk group may be clinically relevant if the model is used to intensify monitoring or counselling. Calibration intercepts, confidence intervals, calibration plots, and assessment in external populations are required before the model can be used to estimate individual risk in routine practice.

The negative predictive value of 92% suggests that participants classified as low risk were unlikely to demonstrate poor compliance within this study population. This may be useful for prioritising additional support while maintaining routine follow-up for lower-risk patients. However, predictive values depend on outcome prevalence and may change in clinics with different patient characteristics or adherence rates. The classification threshold should also be chosen according to the consequences of false-negative and false-positive classifications. Missing a patient who subsequently becomes poorly compliant may delay intervention, whereas incorrectly classifying a compliant patient as high risk may increase monitoring burden and produce unnecessary concern.

The potential clinical role of the model would be to support, rather than replace, orthodontic assessment. Patients identified as being at increased risk could receive additional counselling, closer monitoring, tailored reminders, earlier assessment of aligner fit, or targeted management of pain and practical barriers. Artificial intelligence in dentistry may improve the organisation and interpretation of complex clinical information, but concerns regarding transparency, generalisability, fairness, privacy, and clinician dependence remain important (22,23). Prediction outputs should therefore be interpreted alongside clinical findings and patient circumstances.

The study had several strengths. Data were collected prospectively, predictors were restricted to information available during the first four weeks, and the outcome was assessed during a subsequent period. The study compared a transparent regression model with two machine-learning approaches and evaluated discrimination, threshold-dependent classification measures, calibration slope, and the Brier score. The inclusion of behavioural, clinical, demographic, digital-monitoring, and appointment variables also reflected the multifactorial nature of treatment adherence.

Several limitations should be considered. Participants were recruited from selected clinics in one city, which limits geographical and institutional generalisability. Wear-time information was obtained from available electronic records or clinic-approved applications, and differences between monitoring systems may have introduced measurement heterogeneity. The number of monitored days and the extent of missing predictor information were not reported in the available results. Internal cross-validation cannot establish performance in a new clinic, and model selection among several algorithms may have produced optimistic estimates. Confidence intervals for performance measures were unavailable, preventing assessment of statistical uncertainty and formal comparison between models. The study did not establish whether the gradient-boosting model added clinically meaningful value beyond early wear time or a parsimonious regression model. Decision-curve analysis was not performed, and the clinical consequences of the selected threshold were not quantified. The follow-up period covered only the early stage of treatment, whereas adherence may change over longer treatment durations. Performance across demographic groups, clinic types, treatment-complexity categories, and monitoring methods was also not evaluated.

The findings support further development of an early adherence-risk model but do not yet justify routine clinical deployment. External validation should be undertaken in independent orthodontic centres and in populations with different demographic, socioeconomic, and treatment characteristics. Future work should report complete model specifications, confidence intervals, calibration intercepts and plots, missing-data patterns, hyperparameter-tuning procedures, and clinically justified risk thresholds. It should also compare complex machine-learning models with simpler alternatives and determine whether risk-guided intervention improves aligner wear, reduces loss of tracking, shortens treatment delays, or decreases the need for refinement.

CONCLUSION

Early treatment behaviour was associated with subsequent poor compliance during clear aligner therapy. Reduced wear time, missed digital check-ins, delayed aligner changes, and missed or rescheduled appointments were more informative than demographic characteristics alone. Gradient boosting produced the most favourable internally validated performance among the evaluated models, but its incremental value over simpler prediction approaches and its performance in independent clinical populations remain uncertain. The model should therefore be regarded as a promising risk-stratification approach requiring transparent reporting, external validation, and assessment of clinical utility before routine implementation.

REFERENCES

1. Robertson L, Kaur H, Fagundes NCF, Romanyk D, Major P, Flores-Mir C. Effectiveness of Clear Aligner Therapy for Orthodontic Treatment: A Systematic Review. Orthod Craniofac Res. 2020;23(2):133-142. doi:10.1111/ocr.12353.

2. Yassir YA, Nabbat SA, McIntyre GT, Bearn DR. Clinical Effectiveness of Clear Aligner Treatment Compared to Fixed Appliance Treatment: An Overview of Systematic Reviews. Clin Oral Investig. 2022;26(4):2353-2370. doi:10.1007/s00784-021-04361-1.

3. Papageorgiou SN, Koletsi D, Iliadi A, Peltomäki T, Eliades T. Treatment Outcome With Orthodontic Aligners and Fixed Appliances: A Systematic Review With Meta-Analyses. Eur J Orthod. 2020;42(3):331-343. doi:10.1093/ejo/cjz094.

4. Timm LH, Farrag G, Baxmann M, Schwendicke F. Factors Influencing Patient Compliance During Clear Aligner Therapy: A Retrospective Cohort Study. J Clin Med. 2021;10(14):3103. doi:10.3390/jcm10143103.

5. Schäfer K, Ludwig B, Meyer-Gutknecht H, Schott TC. Quantifying Patient Adherence During Active Orthodontic Treatment With Removable Appliances Using Microelectronic Wear-Time Documentation. Eur J Orthod. 2015;37(1):73-80. doi:10.1093/ejo/cju012.

6. Schott TC, Ludwig B. Microelectronic Wear-Time Documentation of Removable Orthodontic Devices Detects Heterogeneous Wear Behavior and Individualizes Treatment Planning. Am J Orthod Dentofacial Orthop. 2014;146(2):155-160. doi:10.1016/j.ajodo.2014.04.020.

7. Hansa I, Katyal V, Ferguson DJ, Vaid N. Outcomes of Clear Aligner Treatment With and Without Dental Monitoring: A Retrospective Cohort Study. Am J Orthod Dentofacial Orthop. 2021;159(4):453-459. doi:10.1016/j.ajodo.2020.02.010.

8. Ferlito T, Hsiou D, Hargett K, Herzog C, Bachour P, Katebi N, et al. Assessment of Artificial Intelligence-Based Remote Monitoring of Clear Aligner Therapy: A Prospective Study. Am J Orthod Dentofacial Orthop. 2023;164(2):194-200. doi:10.1016/j.ajodo.2022.11.020.

9. Timm LH, Farrag G, Wolf D, Baxmann M, Schwendicke F. Effect of Electronic Reminders on Patients' Compliance During Clear Aligner Treatment: An Interrupted Time Series Study. Sci Rep. 2022;12:16652. doi:10.1038/s41598-022-20820-5.

10. Wolf D, Farrag G, Flügge T, Timm LH. Predicting Outcome in Clear Aligner Treatment: A Machine Learning Analysis. J Clin Med. 2024;13(13):3672. doi:10.3390/jcm13133672.

11. Bichu YM, Hansa I, Bichu AY, Premjani P, Flores-Mir C, Vaid NR. Applications of Artificial Intelligence and Machine Learning in Orthodontics: A Scoping Review. Prog Orthod. 2021;22(1):18. doi:10.1186/s40510-021-00361-9.

12. Khanagar SB, Al-Ehaideb A, Vishwanathaiah S, Maganur PC, Patil S, Naik S, et al. Scope and Performance of Artificial Intelligence Technology in Orthodontic Diagnosis, Treatment Planning, and Clinical Decision-Making: A Systematic Review. J Dent Sci. 2021;16(1):482-492. doi:10.1016/j.jds.2020.05.022.

13. Schwendicke F, Samek W, Krois J. Artificial Intelligence in Dentistry: Chances and Challenges. J Dent Res. 2020;99(7):769-774. doi:10.1177/0022034520915714.

14. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods. BMJ. 2024;385:e078378. doi:10.1136/bmj-2023-078378.

15. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019;170(1):51-58. doi:10.7326/M18-1376.

16. Shahzad HB, Iftikhar D, Huda NU, Enver N, Awais F, Hussain S. Psychosocial Impacts of Fixed Orthodontic Treatment in Lahore, Pakistan. Makara J Health Res. 2020;24(3):187-192. doi:10.7454/msk.v24i3.1237.

17. Mahida HK, Memon S, Memon J. Expectations of Prospective Orthodontic Patients Regarding Post-Orthodontic Retention. J Pak Med Assoc. 2024;74(12):2086-2090. doi:10.47391/JPMA.10629.

18. Charalampakis O, Iliadi A, Ueno H, Oliver DR, Kim KB. Accuracy of Clear Aligners: A Retrospective Study of Patients Who Needed Refinement. Am J Orthod Dentofacial Orthop. 2018;154(1):47-54. doi:10.1016/j.ajodo.2017.11.028.

19. Rossini G, Parrini S, Castroflorio T, Deregibus A, Debernardi CL. Efficacy of Clear Aligners in Controlling Orthodontic Tooth Movement: A Systematic Review. Angle Orthod. 2015;85(5):881-889. doi:10.2319/061614-436.1.

20. Papadimitriou A, Mousoulea S, Gkantidis N, Kloukos D. Clinical Effectiveness of Invisalign Orthodontic Treatment: A Systematic Review. Prog Orthod. 2018;19(1):37. doi:10.1186/s40510-018-0235-z.

21. Ghoneim SH, Afif KS. The Effect of Personality Traits on Patient Compliance With Clear Aligners. Cureus. 2024;16(12):e74922. doi:10.7759/cureus.74922.

22. Khanagar SB, Al-Ehaideb A, Maganur PC, Vishwanathaiah S, Patil S, Baeshen HA, et al. Developments, Application, and Performance of Artificial Intelligence in Dentistry: A Systematic Review. J Dent Sci. 2021;16(1):508-522. doi:10.1016/j.jds.2020.06.019.

23. Naureen S. Limitations of Artificial Intelligence in Orthodontics: Literature Review. J Bahria Univ Med Dent Coll. 2025;15(1):53-59. doi:10.51985/JBUMDC2024452.

24. Al-Moghrabi D, Salazar FC, Pandis N, Fleming PS. Compliance With Removable Orthodontic Appliances and Adjuncts: A Systematic Review and Meta-Analysis. Am J Orthod Dentofacial Orthop. 2017;152(1):17-32. doi:10.1016/j.ajodo.2017.03.019.

25. Weir T. Clear Aligners in Orthodontic Treatment. Aust Dent J. 2017;62 Suppl 1:58-62. doi:10.1111/adj.12480.