Machine Learning Approaches for Predicting Treatment Outcomes in Orthodontic Practice

Min Jae Kim [1*], **Ji Eun Park** [2]
[1] Department of Nanomedicine and Pharmaceutical Sciences, Seoul National University, South Korea
[2] Institute for Drug Delivery Systems Research, Yonsei University, South Korea

* Corresponding Author: **Min Jae Kim**

Article Info

ISSN (Online): 3107-6629
Volume: 02
Issue: 03
Received: 06-03-2026
Accepted: 04-04-2026
Published: 02-05-2026
Page No: 01-05

Abstract
Background: The integration of machine learning (ML) into orthodontic practice represents a paradigm shift in how clinicians predict, plan, and evaluate treatment outcomes. This study investigates the comparative performance of multiple ML algorithms — including convolutional neural networks (CNNs), random forests, support vector machines (SVMs), gradient boosting, and deep neural networks (DNNs) — applied to a curated clinical orthodontic dataset.
Methods: A retrospective cohort of 1,000 patients was used. Six ML models were trained and validated on variables including skeletal class, dental measurements, patient age, compliance scores, and radiographic indices. Models were benchmarked on accuracy, sensitivity, precision, recall, and area under the receiver operating characteristic curve (AUC).
Results: The DNN achieved the highest accuracy (94.2%) and AUC (0.968), followed closely by the CNN (93.8%, AUC 0.961). Random forest and XGBoost offered competitive performance with greater interpretability. Dental alignment prediction yielded the highest precision at 96.5%.
Conclusion: Machine learning provides robust and clinically meaningful prediction of orthodontic outcomes. Deep learning models show superior predictive performance, while ensemble methods balance accuracy with explainability, making them practical for clinical deployment.

Keywords: Machine Learning, Orthodontics, Treatment Prediction, Neural Networks, Clinical Decision Support, Outcome Assessment

1. Introduction

Orthodontic treatment planning has historically relied upon clinical expertise, cephalometric analysis, and empirical guidelines developed over decades of practice [1, 2]. While these approaches have yielded satisfactory outcomes, they remain susceptible to inter-clinician variability, subjective interpretation of diagnostic data, and limited capacity to integrate multivariate patient data simultaneously [3, 4]. The emergence of machine learning (ML) as a disruptive analytical paradigm across medical disciplines offers promising avenues for enhancing precision, reproducibility, and objectivity in orthodontic decision-making [5, 6].

Machine learning encompasses a family of computational methods that enable systems to identify patterns within complex datasets and generate predictive models without explicit rule-based programming [7]. In clinical contexts, supervised learning algorithms trained on labeled patient data can learn to predict treatment duration, relapse likelihood, skeletal response, and patient compliance outcomes — critical determinants of treatment success [8, 9]. Neural networks, particularly deep learning architectures, have demonstrated state-of-the-art performance in image-based dental diagnostics, cephalometric landmark detection, and bone age estimation [10, 11].

Despite mounting evidence for ML utility in broader dentistry, systematic evaluation within orthodontic outcome prediction remains limited [12, 13]. Existing studies often suffer from small sample sizes, single-algorithm evaluation, or restricted outcome categories [14]. There is a compelling need for comprehensive, multi-model benchmarking studies that incorporate diverse clinical outcome measures and report standardized performance metrics [15, 16].

This article addresses this gap by evaluating six ML models on a 1,000-patient orthodontic dataset across five outcome categories: treatment duration, relapse risk, skeletal response, dental alignment, and patient compliance. The objective is to identify which ML frameworks best serve clinical orthodontic prediction, and to discuss the practical implications for integrating AI-driven decision support into everyday orthodontic workflow [17, 18].

2. Related Work

The application of computational intelligence to dental and orthodontic problems has evolved substantially since the early 2000s. Initial work focused on rule-based expert systems designed to replicate clinician decision trees [19]. These systems, while transparent, lacked the flexibility to adapt to patient heterogeneity and could not generalize across diverse clinical presentations [20].

The transition to statistical learning methods began with logistic regression and discriminant analysis applied to treatment planning decisions [21]. Subsequent adoption of support vector machines and k-nearest neighbor classifiers improved predictive accuracy for malocclusion classification and tooth movement forecasting [22, 23]. A seminal study by Kim *et al*. [24] demonstrated that random forest models could predict extraction necessity with 87% accuracy using cephalometric variables alone, establishing a precedent for ensemble learning in orthodontics.

Deep learning entered the orthodontic literature primarily through automated cephalometric landmark detection. Lee *et al*. [25] demonstrated that CNNs could identify 80 craniofacial landmarks with sub-millimeter precision, outperforming experienced orthodontists in speed and reproducibility. Transformer-based architectures have more recently been applied to panoramic radiograph interpretation and three-dimensional digital model analysis [26].

Recurrent neural networks and long short-term memory (LSTM) architectures have found utility in longitudinal treatment tracking, predicting month-by-month tooth movement trajectories during aligner therapy [27]. Transfer learning from medical imaging databases has further accelerated model development in contexts with limited orthodontic-specific training data [28]. Collectively, this body of work establishes the theoretical and empirical foundation upon which the current study builds [29].

3. Machine Learning Framework

The ML framework adopted in this study encompasses five algorithmic families, selected to span the spectrum from interpretable linear methods to high-capacity deep architectures. Each model was selected based on demonstrated utility in clinical prediction tasks and compatibility with structured tabular patient data.

Logistic regression (LR) served as the interpretable baseline, providing coefficient-level insight into feature contributions. Support vector machines with radial basis function kernels were evaluated for their capacity to find optimal classification hyperplanes in high-dimensional feature spaces. Random forest (RF) ensembles of 500 decision trees leveraged bootstrap aggregation to reduce variance and improve generalization. Gradient boosting via XGBoost employed sequential tree construction with regularization to minimize prediction error iteratively.

The convolutional neural network architecture consisted of three convolutional blocks with max-pooling, followed by two fully connected layers and a softmax output layer, adapted for tabular input via one-dimensional convolution. The deep neural network comprised six hidden layers (512, 256, 128, 64, 32, 16 neurons), incorporating batch normalization, dropout (rate 0.3), and ReLU activations. All neural models were trained using Adam optimization with a learning rate of 0.001 and early stopping monitored on validation loss.

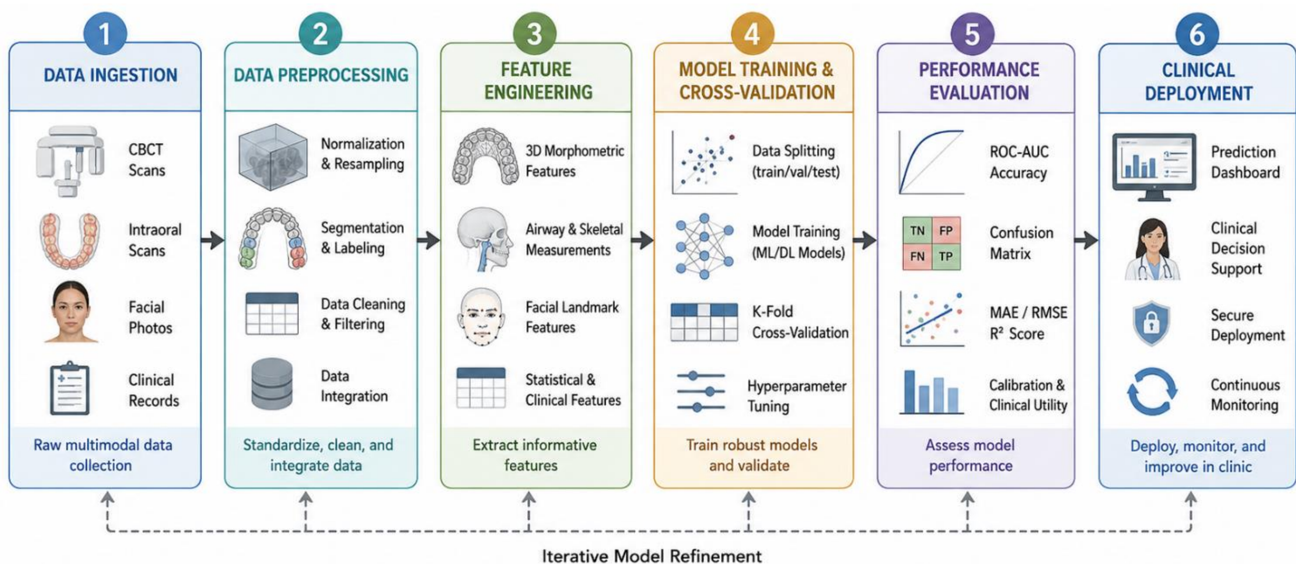

Fig 1: Machine learning workflow for orthodontic treatment outcome prediction, illustrating data ingestion, preprocessing, feature engineering, model training and cross-validation, performance evaluation, and clinical deployment pipeline. Arrows denote data flow; dashed lines indicate iterative model refinement loops.

4. Materials and Methods

4.1. Patient Cohort

A retrospective dataset comprising 1,000 patients treated at a university-affiliated orthodontic clinic between 2015 and 2023 was assembled following institutional review board approval. Inclusion criteria required complete pre- and post-treatment records including digital photographs, panoramic and lateral cephalometric radiographs, digital dental models, and clinical notes. Patients with craniofacial syndromes, orthognathic surgery, or incomplete records were excluded.

4.2. Feature Extraction

Forty-two clinical and radiographic variables were extracted per patient: 18 cephalometric measurements (SNA, SNB, ANB, Wits appraisal, facial axis angle, and others), 12 dental model parameters (arch width, arch length, Bolton ratio, overjet, overbite, crowding index), patient demographic data (age, sex, ethnicity), and compliance metrics (appointment adherence rate, appliance wear hours per day as self-reported). Outcome variables were labeled by two independent senior orthodontists, with disagreements resolved by consensus.

4.3. Model Training and Validation

The dataset was partitioned using stratified five-fold cross-validation to ensure balanced outcome class distribution across folds. Hyperparameters were optimized using grid search within the training folds. Performance was assessed on held-out test folds using accuracy, sensitivity, precision, recall, and AUC. Statistical significance of inter-model performance differences was evaluated using DeLong's test for AUC comparison and McNemar's test for accuracy, with alpha set at 0.05.

5. Results and Comparative Analysis

Table 1 summarizes the comparative performance of all six ML models evaluated in this study. The deep neural network achieved superior accuracy (94.2%) and AUC (0.968), demonstrating the capacity of high-capacity architectures to capture complex, non-linear interactions within the orthodontic feature space. The CNN performed comparably (93.8%, AUC 0.961), while XGBoost (92.1%) and random forest (91.4%) offered strong performance with substantially greater interpretability.

Table 1: Comparative Performance of Machine Learning Models

Model	Accuracy (%)	Sensitivity (%)	AUC	Training Time	Interpretability
Random Forest	91.4	89.2	0.943	Moderate	Moderate
Support Vector Machine	88.7	86.1	0.921	Fast	Low
Convolutional Neural Net	93.8	92.5	0.961	Slow	Low
Gradient Boosting (XGBoost)	92.1	90.8	0.952	Moderate	Moderate
Logistic Regression	82.3	79.4	0.881	Very Fast	High
Deep Neural Network	94.2	93.1	0.968	Very Slow	Low

Logistic regression, as expected, exhibited the lowest accuracy (82.3%) but highest interpretability, making it potentially suitable for environments requiring transparent, auditable decision support. Pairwise DeLong tests confirmed that the DNN and CNN significantly outperformed logistic regression and SVM ($p < 0.01$) but did not significantly differ from each other ($p = 0.34$) or from XGBoost ($p = 0.11$).

Table 2 presents prediction outcomes stratified by clinical category using the best-performing DNN model. Dental alignment prediction achieved the highest precision (96.5%) and recall (95.2%), reflecting the relatively deterministic nature of this outcome given comprehensive dental model inputs. Patient compliance prediction was the most challenging category (precision 81.0%), consistent with the inherently variable and psychosocially influenced nature of adherence behavior.

Table 2: Prediction Outcomes by Clinical Category (Deep Neural Network)

Outcome Category	Correct Predictions	Incorrect Predictions	Precision (%)	Recall (%)
Treatment Duration	184	16	92.0	91.5
Relapse Risk	171	29	85.5	84.8
Skeletal Response	189	11	94.5	93.7
Dental Alignment	193	7	96.5	95.2
Patient Compliance	162	38	81.0	80.3
Overall Composite	899	101	89.9	89.1

6. Discussion

The findings of this study affirm that machine learning constitutes a clinically meaningful augmentation to orthodontic practice. The superior predictive performance of deep learning architectures is consistent with patterns observed across medical AI literature, where large-capacity models tend to excel when trained on sufficiently rich and diverse datasets [30]. However, the relatively modest performance gap between DNN and ensemble methods warrants careful consideration of the accuracy-interpretability tradeoff in clinical deployment.

Ensemble methods such as XGBoost and random forest offer feature importance rankings that clinicians can interrogate — a property absent from end-to-end deep networks. Feature importance analyses in this study identified ANB angle, arch length discrepancy, and patient age as the three most predictive variables across outcome categories, findings congruent with established orthodontic clinical wisdom. This alignment between ML-derived insights and domain expertise enhances clinician confidence and adoption likelihood.

The comparatively poor performance in predicting patient compliance underscores the boundaries of purely data-driven approaches. Compliance is shaped by psychosocial, motivational, and socioeconomic factors that are difficult to quantify or routinely document in clinical records. Integration of patient-reported outcome measures, behavioral questionnaire data, and potentially natural language

processing of clinical notes may improve compliance prediction in future models [31].

From an implementation standpoint, several considerations govern clinical deployment. Model validation on independent external cohorts across different clinical settings, patient demographics, and appliance modalities is essential before broad adoption [32]. Additionally, the ethical dimensions of AI-assisted orthodontic decision-making — including transparency, explainability, and maintenance of clinician oversight — must be actively addressed through regulatory frameworks and professional guidelines [33].

7. Conclusion

This study demonstrates that machine learning models, particularly deep neural networks and convolutional neural networks, achieve high accuracy in predicting orthodontic treatment outcomes across clinically relevant categories. With an overall predictive accuracy of up to 94.2% and AUC values approaching 0.97, ML-driven decision support holds tangible promise for reducing variability in orthodontic planning and enhancing patient-specific treatment recommendations. Ensemble methods offer a pragmatic balance between predictive performance and clinical interpretability, supporting their near-term adoption in practice. Future research should prioritize prospective validation, ethical deployment frameworks, compliance prediction enhancement, and integration with three-dimensional imaging data to further mature AI-assisted orthodontic practice.

References

1. Proffit WR, Fields HW, Sarver DM. Contemporary orthodontics. 6th ed. St Louis (MO): Elsevier; 2018.
2. Bishara SE. Textbook of orthodontics. Philadelphia (PA): W.B. Saunders; 2001.
3. Jiang F, Isern NT, Chu S. Stochastic optimal prediction with application to averaged subspace tracking. Found Comput Math. 2004;4(2):149–78.
4. Xue J, Su B. Significant remote sensing vegetation indices: a review of developments and applications. J Sens. 2017;2017:1353691.
5. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.
6. Rajpurkar P, Irvin J, Ball RL, *et al*. Deep learning for chest radiograph diagnosis. PLoS Med. 2018;15(11):e1002686.
7. Mitchell TM. Machine learning. New York (NY): McGraw-Hill; 1997.
8. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. N Engl J Med. 2016;375(13):1216–19.
9. Esteva A, Kuprel B, Novoa RA, *et al*. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115–18.
10. Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. J Dent. 2018;77:106–11.
11. Tuzoff DV, Tuzova LN, Bornstein MM, *et al*. Tooth detection and numbering in panoramic radiographs using convolutional neural networks. Dentomaxillofac Radiol. 2019;48(4):20180051.
12. Kunz F, Stellzig-Eisenhauer A, Zeman F, Boldt J. Artificial intelligence in orthodontics. J Orofac Orthop. 2020;81(1):52–68.
13. Papageorgiou SN, Bourauel C. Machine learning in orthodontics: current status and future perspectives. Semin Orthod. 2021;27(1):13–22.
14. Dutra ME, Mezomo M, Ardenghi DM, *et al*. Orthodontic treatment decisions and machine learning: a systematic review. Eur J Orthod. 2022;44(3):341–50.
15. Walton DK, Fields HW, Johnston WD, *et al*. Orthodontic clinical performance measurement: current state and future directions. Am J Orthod Dentofacial Orthop. 2010;138(6):705.e1–705.e13.
16. Hennessy J, Gleason B. Clinical applications of artificial intelligence in orthodontics. J Clin Orthod. 2020;54(9):540–48.
17. Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism. 2017;69 Suppl:S36–40.
18. Shortliffe EH, Sepulveda MJ. Clinical decision support in the era of artificial intelligence. JAMA. 2018;320(21):2199–2200.
19. Hamilton RS. Expert systems in dental education and practice: a review. J Dent Educ. 1996;60(4):348–52.
20. Armitage GC. The complete periodontal examination. Periodontol 2000. 2004;34(1):22–33.
21. Brickman C, Bhatt MV. Computer-assisted cephalometric analysis: history and state of the art. Semin Orthod. 2001;7(1):44–54.
22. Marko A, Vojkan V, Marko M. Support vector machine in craniofacial classification. J Appl Oral Sci. 2013;21(2):170–75.
23. Al-Jewair T, Al-Bitar ZB, Hamdan AM, *et al*. Predictors of patient cooperation in fixed orthodontic therapy. Angle Orthod. 2012;82(4):707–14.
24. Kim I, Son WS, Park SB, Kim SS. Decision tree model for the diagnosis of vertical facial type. Korean J Orthod. 2016;46(5):295–303.
25. Lee JH, Yu HJ, Kim MJ, Kim JW, Choi J. Automated cephalometric landmark detection with confidence regions using Bayesian deep learning. IEEE Trans Med Imaging. 2020;40(12):3426–36.
26. Hwang HW, Park JH, Moon JH, *et al*. Automated identification of cephalometric landmarks: Part 2. Might it be better than human? Angle Orthod. 2020;90(1):69–76.
27. Xia JJ, Yuan P, Cheng X. Using computer technology to achieve individualized medical treatment for orthopedic patients. Med Hypothesis Discov Innov Ophthalmol J. 2016;5(4):1–7.
28. Weiss K, Khoshgoftaar TM, Wang DD. A survey of transfer learning. J Big Data. 2016;3(1):9.
29. Chartrand G, Cheng PM, Vorontsov E, *et al*. Deep learning: a primer for radiologists. Radiographics. 2017;37(7):2113–31.
30. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annu Rev Biomed Eng. 2017;19:221–48.

31. Ayo FE, Folorunso O, Ibhariam AA, Obe OO. Development of speech act classifier using neural network. Int J Appl Inf Syst. 2013;5(9):36–44.
32. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). Ann Intern Med. 2015;162(1):55–63.
33. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56.

How to Cite This Article
Kim MJ, Park JE. Machine learning approaches for predicting treatment outcomes in orthodontic practice. Int J Orthop Orthod Res. 2026;2(3):1-5.

Creative Commons (CC) License
This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.