Original Article

Orthodontic treatment outcome predictive performance differences between artificial intelligence and conventional methods

Sung Joo Cho^a; Jun-Ho Moon^b; Dong-Yub Ko^c; Ju-Myung Lee^a; Ji-Ae Park^d; Richard E. Donatelli^e; Shin-Jae Lee^f

ABSTRACT

Objectives: To evaluate an artificial intelligence (AI) model in predicting soft tissue and alveolar bone changes following orthodontic treatment and compare the predictive performance of the AI model with conventional prediction models.

Materials and Methods: A total of 1774 lateral cephalograms of 887 adult patients who had undergone orthodontic treatment were collected. Patients who had orthognathic surgery were excluded. On each cephalogram, 78 landmarks were detected using PIPNet-based AI. Prediction models consisted of 132 predictor variables and 88 outcome variables. Predictor variables were demographics (age, sex), clinical (treatment time, premolar extraction), and Cartesian coordinates of the 64 anatomic landmarks. Outcome variables were Cartesian coordinates of the 22 soft tissue and 22 hard tissue landmarks after orthodontic treatment. The AI prediction model was based on the TabNet deep neural network. Two conventional statistical methods, multivariate multiple linear regression (MMLR) and partial least squares regression (PLSR), were each implemented for comparison. Prediction accuracy among the methods was compared.

Results: Overall, MMLR demonstrated the most accurate results, while AI was least accurate. AI showed superior predictions in only 5 of the 44 anatomic landmarks, all of which were soft tissue landmarks inferior to menton to the terminal point of the neck.

Conclusions: When predicting changes following orthodontic treatment, AI was not as effective as conventional statistical methods. However, AI had an outstanding advantage in predicting soft tissue landmarks with substantial variability. Overall, results may indicate the need for a hybrid prediction model that combines conventional and AI methods. (*Angle Orthod*. 2024;94:557–565.)

KEY WORDS: Artificial intelligence; Multivariate multiple linear regression; Partial least squares; Orthodontic treatment; Profile change; Machine learning

^a Graduate Student (PhD), Department of Orthodontics, Seoul National University, Seoul, Korea.

^b Private Practice, Cheonan, Korea.

^c Research Scientist, AI Research Center, DDH Inc, Seoul, Korea.

^d Clinical Lecturer, Department of Orthodontics, Seoul National University Dental Hospital, Seoul, Korea.

^e Private Practice, West Palm Beach, Florida, USA.

^f Professor, Department of Orthodontics and Dental Research Institute, Seoul National University School of Dentistry, Seoul, Korea.

Corresponding author: Dr Shin-Jae Lee, Professor, Department of Orthodontics and Dental Research Institute, Seoul National University School of Dentistry, Jongro-Gu, Seoul 03080, Korea

(e-mail: nonext.shinjae@gmail.com)

Accepted: March 2024. Submitted: November 2023.

Published Online: May 10, 2024

 $\ensuremath{\textcircled{\sc 0}}$ 2024 by The EH Angle Education and Research Foundation, Inc.

INTRODUCTION

With the advent of high-speed computer technology, the use of artificial intelligence (AI) in research has become popular. With a wealth of new AI literature emerging, orthodontics is no exception to the rapid influx of AI. Recently, several studies have used AI to predict facial soft tissue changes following orthodontic treatment. However, these AI studies seem to simply repeat analyses using AI, which have already been analyzed using conventional statistical methods (Table 1).^{1–3} Since Al requires significant computing resources even for simple tasks, it must demonstrate superior effectiveness over traditional methods to justify its use. If an AI model does not perform better than conventional methods, it is impractical and unnecessarily costly. To determine its practicality, it might be essential to compare the accuracy of AI predictions with traditional methods.

-			
Research Group	Sample Sizes	No. Predicted Landmarks	Prediction Methods
Present study, 2024	887	22 soft tissue and 22 skeletal landmarks	Multivariate multiple linear regression, par- tial least squares regression, and TabNet deep neural network
Park et al., ¹ 2022	312	8 soft tissue landmarks	Conditional generative adversarial network
Tanikawa and Yamashiro, ² 2021	65	18 soft tissues landmarks	Geometric morphometric method and deep neural networks
Park et al., ³ 2021	284	14 soft tissue and 15 skeletal landmarks	U-Net-based convolutional neural networks

Table 1. Summary of Orthodontic Prediction Research Using Artificial Intelligence

To develop a clinically applicable method for predicting smooth soft tissue curves, it is necessary to analyze multiple predictor and outcome (response) variables of the soft tissue landmarks.^{4,5} Multivariate multiple linear regression (MMLR), which produces the ordinary least squares estimator, is one of the conventional methods available to do so.⁶ However, MMLR has limitations when there are numerous variables and those variables are significantly correlated. Partial least squares regression (PLSR) applies dimensional reduction latent modeling and has preferably been used to provide more accurate prediction results after combined surgical-orthodontic treatment^{5,7–10} or in predicting facial growth changes.¹¹

In developing AI models, deep-learning algorithms based on convolutional neural network (CNN) architecture have been the most popular. One of the latest CNNs, TabNet deep neural network (DNN), has been used to develop an individualized facial growth prediction model¹² and to predict soft tissue changes following orthognathic surgery.¹³ TabNet DNN can model complex nonlinear relationships, incorporating multiple predictors and outcome variables.¹⁴ Previously, AI showed effectiveness when automatically identifying cephalometric landmarks and subsequent analyses¹⁵⁻²⁰ and for predicting facial growth in growing children.¹² Contrary to the initial assumption that AI could provide a universal solution to diverse challenges, AI has not always been effective. For example, when predicting soft tissue changes following orthognathic surgery, the Al prediction was not as effective as the conventional PLSR method, particularly when predicting areas with small surgical changes.¹³

The purpose of this study was to develop and evaluate an AI model for predicting changes following orthodontic treatment. The specific aim was to compare the predictive performance of the AI prediction model with conventional prediction methods, MMLR and PLSR.

MATERIALS AND METHODS

Subjects

The institutional review board for the protection of human subjects of the Seoul National University School of Dentistry approved the research protocol (S-D20200036).

The subjects were 887 patients (604 females and 283 males; mean age = 24.2 ± 8.5 years) who had undergone orthodontic treatment at the Department of Orthodontics, Seoul National University Dental Hospital, Seoul, Korea, from January 2013 to December 2022. The inclusion criteria were (1) females aged > 15 years, males > 17 years, to exclude subjects during major growth spurts, and (2) treated with comprehensive orthodontic treatment using fixed appliances. The exclusion criteria were (1) a history of orthognathic surgery and (2) the presence of craniofacial syndromes.

Predictors and Outcome Variables

For all patients, lateral cephalograms taken before (T1) and after (T2) orthodontic treatment were collected.



Scattor plots

Real case comparison

Figure 1. The experimental design.



Figure 2. The reference planes and 78 cephalometric landmarks used in this study: (A) pretreatment image, skeletal landmarks in capital letters; (B) posttreatment image, soft tissue landmarks in lowercase letters.

On 1774 images from 887 patients, 46 skeletal and 32 soft tissue landmarks were identified using automated landmark detection software (Ceppro, DDH Inc, Seoul, Korea) based on the PIPNet algorithm by Jin et al. (2021).²¹

The prediction models comprised 132 predictors and 88 outcome variables (Figure 1). The predictor variables were demographics (age, sex), clinical (treatment time, premolar extraction), and Cartesian coordinates of the 64 anatomic landmarks.

Landmarks were chosen to reflect orthodontic treatment changes of the incisors (16 variables), the molars (24 variables), the soft tissue from subnasale to the terminal point of the neck (44 variables), the alveolar bone (12 variables), and the hard tissue of the mandible (32 variables).

The outcome variables included Cartesian coordinates of the 22 soft tissue landmarks from subnasale to the terminal point of the neck, 6 alveolar bone landmarks, and 16 skeletal landmarks on the mandible that could undergo changes according to orthodontic tooth movement (Figure 2; Table 2).

AI Prediction Model

The AI prediction model applied in the present study was based on the TabNet DNN algorithm by Arik and

Pfister (2021).¹⁴ The training and testing were performed using Python (Python Software Foundation, Wilmington, Del) on a desktop computer run on Ubuntu 22.04 LTS of Linux distribution.

To develop an optimal AI prediction model, various AI training circumstances (also called hyperparameters) were tested, and the optimal conditions were selected by comparing prediction errors of numerous combinations of training hyperparameters. Regarding the early stopping number of training epochs, 50, 100, 1000, and 10,000 were tested. Subsequently, the AI model trained through 10,000 epochs was selected as the optimal AI model (Figure 3A).

The oversampling method based on the synthetic minority oversample technique (SMOTE)²² was implemented. SMOTE values of 0.05, 0.1, 0.2, and 0.3 were tested. However, the results from varying values of SMOTE did not show significant differences (Figure 3B).

Two Conventional Statistical Prediction Models

MMLR is based on the ordinary least squares estimator. The stepwise variable selection method was used in constructing the MMLR prediction model.

PLSR combines the benefits of principal component analysis and MMLR through dimensional reduction

	Mea	Mean \pm Standard Deviation			Superior Models ^b		
Landmarks	MMLR	PLSR	AI	MMLR	PLSR	AI	<i>P</i> Value ^c
Soft tissue							
Subnasale	0.79 ± 0.50	0.81 ± 0.50	2.16 ± 1.28				<.0001
Point A	0.79 ± 0.49	0.78 ± 0.48	1.98 ± 1.17				<.0001
Superior labial sulcus	1.01 ± 0.59	1.05 ± 0.60	1.80 ± 1.09				<.0001
Labrale superius	1.18 ± 0.71	1.20 ± 0.70	1.95 ± 1.16				<.0001
Upper lip	1.27 ± 0.78	1.28 ± 0.74	1.95 ± 1.13				<.0001
Upper lip adjunct contour	1.34 ± 0.81	1.32 ± 0.77	1.84 ± 1.09				<.0001
Stomion superius	1.20 ± 0.69	1.20 ± 0.67	1.61 ± 0.96				<.0001
Stomion inferius	1.56 ± 1.01	1.51 ± 1.01	1.75 ± 1.06				<.0001
Lower lip adjunct contour	1.71 ± 1.05	1.66 ± 1.05	1.84 ± 1.09				.0600
Lower lip	1.92 ± 1.15	1.90 ± 1.14	2.01 ± 1.15				1.0000
Labrale inferius	1.97 ± 1.23	1.96 ± 1.25	2.10 ± 1.24				1.0000
Inferior labial sulcus	1.70 ± 1.08	1.68 ± 1.07	1.78 ± 1.09				1.0000
Point B	1.86 ± 1.31	1.93 ± 1.28	2.13 ± 1.32				<.0001
Protuberance menti	$\textbf{2.10} \pm \textbf{1.45}$	2.09 ± 1.43	2.29 ± 1.48				.1400
Pogonion	2.49 ± 1.69	2.47 ± 1.64	2.79 ± 1.76				<.0001
Gnathion	2.01 ± 1.27	2.09 ± 1.30	2.38 ± 1.44	•			<.0001
Menton	1.55 ± 1.03	1.74 ± 1.07	1.99 ± 1.18	•			<.0001
Menton adjunct contour	3.26 ± 2.08	3.18 ± 1.96	1.94 ± 1.37				<.0001
Cervical point	5.57 ± 3.36	5.29 ± 3.19	2.98 ± 2.15				<.0001
Anterocervical contour	4.96 ± 3.02	4.69 ± 2.87	2.56 ± 1.91				<.0001
Posterocervical contour	5.93 ± 3.55	5.63 ± 3.38	3.09 ± 2.29				<.0001
Terminal point	7.52 ± 4.81	7.34 ± 4.58	5.53 ± 3.47				<.0001
Hard tissue							
Point A	0.48 ± 0.31	0.72 ± 0.43	1.84 ± 1.04	•			<.0001
Point A contour	0.23 ± 0.13	0.49 ± 0.26	1.69 ± 1.00	•			<.0001
Supradentale	0.40 ± 0.23	0.61 ± 0.35	1.68 ± 1.00	•			<.0001
Infradentale	0.28 ± 0.16	0.54 ± 0.31	1.50 ± 0.81				<.0001
Point B contour	0.18 ± 0.11	0.66 ± 0.39	1.47 ± 0.82				<.0001
Point B	0.74 ± 0.46	1.08 ± 0.62	1.65 ± 0.93	•			<.0001
Protuberance menti	0.89 ± 0.58	1.12 ± 0.70	1.69 ± 0.97				<.0001
Pogonion	1.09 ± 0.82	1.31 ± 0.88	2.04 ± 1.25	•			<.0001
Gnathion	1.12 ± 0.86	1.27 ± 0.91	1.98 ± 1.22	•			<.0001
Menton	1.15 ± 0.89	1.39 ± 0.96	2.04 ± 1.24	•			<.0001
Gonion, constructed	1.48 ± 1.11	1.45 ± 1.13	2.16 ± 1.40	•	-		<.0001
Mn. Body contour 1	1.17 ± 0.76	1.31 ± 0.80	1.83 ± 1.11	•			<.0001
Mn. Body contour 2	1.14 ± 0.72	1.28 ± 0.77	1.83 ± 1.05	•			<.0001
Mn. Body contour 3	1.25 ± 0.86	1.37 ± 0.87	1.88 ± 1.15	•			<.0001
Gonion, anatomic	1.52 ± 1.14	1.48 ± 1.14	2.17 ± 1.43	•	•		<.0001
Gonion contour 1	1.46 ± 1.10	1.43 ± 1.11	2.06 ± 1.35	•	•		<.0001
Gonion contour 2	1.53 ± 1.14	1.50 ± 1.13	2.13 ± 1.43				<.0001
Articulare	0.84 ± 0.52	0.83 ± 0.49	2.61 ± 1.44				<.0001
	0.86 ± 0.53	0.84 ± 0.50	2.04 ± 1.15				< .0001
Hamus contour 2	1.18 ± 0.75	1.18 ± 0.75	1.80 ± 1.11				< .0001
Condylion Romuo tin	0.79 ± 0.45	0.79 ± 0.44	2.40 ± 1.33				< .0001
namus tip	1.09 ± 0.07	1.03 ± 0.05	2.20 ± 1.20	•	•		<.0001

Table 2. Comparison of Orthodontic Treatment Prediction Errors (mm) from Multivariate Multiple Linear Regression (MMLR), Partial Least Squares Regression (PLSR) Method, and the TabNet Artificial Intelligence (AI) Algorithm^a

^a Values are the Euclidean distance between prediction results and actual changes in mm units.

^b For a given landmark, the model that showed more accurate prediction results is indicated by the symbol **■**.

^c Results from analysis of variance with Bonferroni correction.

latent modeling.²³ The PLSR prediction model with 40 latent variables was selected.

Validation and Evaluation of Predictive Performance

To validate the prediction models and to avoid overfitting, the leave-one-out cross-validation, which is known for its superiority compared with other test/validation methods, was used.²⁴ During the validation process, 2661 prediction models were built, with each algorithm (AI, MMLR, and PLSR) having 887 models that excluded one subject during model building.

To compare predictive performance, analysis of variance was conducted. Scatterplots with 95%



Figure 3. Searching for optimal artificial intelligence (AI) model training conditions by comparing 95% confidence ellipses of the AI prediction errors at the upper lip and lower lip: (A) according to the number of training epochs; (B) according to the amount of oversampling.

confidence ellipses were used to evaluate the prediction errors in two dimensions.²⁵ Changes in the 22 skeletal and 22 soft tissue landmarks after orthodontic treatment were connected using spline curves overlaid on real patient photos and cephalometric images (Figure 1).

RESULTS

Among 887 patients, 31% had undergone premolar extraction treatment; 53.6%, 34.8%, and 11.6% had Class I, II, and III malocclusions, respectively. The mean treatment duration was 32 months.

The pooled average prediction errors of the 44 anatomical landmarks were 1.69 mm, 1.74 mm, and 2.12 mm from the MMLR, PLSR, and AI prediction methods, respectively. Overall, MMLR demonstrated the most accurate results in all of the alveolar bone and skeletal landmarks. However, AI demonstrated superiority over MMLR and PLSR in predicting 5 among 22 soft tissue landmarks, all of which were landmarks on the face below menton to the terminal point of the neck that had been poorly predicted by MMLR and PLSR (Table 2).

From the point of view of statistical significance, MMLR showed more accurate results than PLSR in 14 landmarks. However, when the prediction errors were evaluated in two dimensions, the differences between MMLR and PLSR did not illustrate clinically significant differences (Figure 4). In certain areas, the differences between the AI and conventional methods were noticeable. Figure 4 illustrates several representative scatterplots of prediction errors where AI demonstrated greater prediction errors, such as at the upper



Figure 4. The prediction errors in several soft tissue landmarks obtained from the multivariate multiple linear regression (MMLR, green), partial least squares regression (PLSR, blue), and artificial intelligence (AI, red) prediction methods. In general, AI showed the least accurate prediction results except for the cervical point, where AI showed the smallest ellipses.

lip, lower lip, soft tissue point B, soft tissue pogonion, and soft tissue menton. However, when predicting the cervical point, AI showed fewer prediction errors than the conventional methods (Figure 4).

To provide real case examples, the predicted outcomes were overlaid with the actual changes following orthodontic treatment. Figure 5A displays an orthodontic patient treated with four premolar extractions, which demonstrated that MMLR and PLSR are more accurate than AI when predicting changes in the alveolar process and lip curves. Figure 5B is a treatment case with an open-bite resolved by counterclockwise autorotation of the mandible through the intrusion of the maxillary posterior teeth, which illustrated poor accuracy by the AI model compared with MMLR and PLSR for the prediction of rotational movement of the mandible. Similar variations in the predictive outcomes were observed in all patients, particularly in the lip region and chin tip (Figure 5C,D). In predicting soft tissue curves below soft tissue menton, AI outperformed MMLR and PLSR for all patients (Figure 5).

DISCUSSION

Although recent AI studies have shown promising features of AI in predicting changes following orthodontic treatment,¹⁻³ no studies compared the predictive performance of AI with conventional statistical methods to determine whether AI was superior enough to deserve the spotlight over conventional methods in this area. In addition, previous literature was based on a relatively small sample size with a limited number of outcome variables. So far, the present study appears to be the first to compare the predictive performance of an AI prediction model with conventional prediction methods for orthodontic treatment outcomes, using the largest sample size ever. This study showed that AI was not effective in predicting changes after orthodontic treatment, except for the neck area. So far, it could be conjectured that AI might not be as effective as conventional methods.

Based on literature reviews, AI has not always been effective. According to Hwang et al. (2020),¹⁹ when detecting cephalometric landmarks, AI was poorer than



Figure 5. Comparison between actual changes after orthodontic treatment and prediction results according to multivariate multiple linear regression (MMLR), partial least squares regression (PLSR), and artificial intelligence (AI) prediction methods in patients with (A) Class III anterior crossbite, (B) Class II open-bite, (C) Class I open-bite, and (D) Class II open-bite.

human examiners in accurately identifying the nose tip, the nasal bone tip, incisal edges, and incisal root tips. These landmarks had a relatively clear form and shape that could be visually pinpointed with ease. Recently, according to Moon et al. (2024),¹² when predicting facial growth, AI was the most accurate in 63 out of 78 landmarks (81%). However, when predicting cranial base landmarks, AI showed poorer results than PLSR. Al also did not outperform PLSR in the study by Park et al. (2024)¹³ on predicting changes after orthognathic surgery, AI only provided the most accurate results for 6 out of 32 landmarks (18.8%). In the present study, the percentage of the cases for which AI showed the most accurate results decreased to 5 of 44 landmarks (11.4%), as summarized in Table 3. This suggests that, when changes were limited, variations were minimal, or a clear cause-and-effect relationship existed, conventional statistical prediction methods were more effective than AI. Although predicting changes after orthodontic treatment is complex, orthodontic treatment changes are not as significant as those resulting from natural facial growth over time or orthognathic surgical procedures.

Al does have some disadvantages. First, Al cannot explain how it arrives at solutions, unlike conventional linear regression analysis that can interpret and estimate the relationships between predictor and outcome variables. Since the internal operations of Al cannot be fully understood or interpreted, in this sense, Al could be deemed a black box.¹³ Second, developing an Al model requires significant computation resources and time, ranging from several weeks to months, depending on the sample size and computer specifications.

The large sample size of this study may have influenced the finding that MMLR was more effective than PLSR. Without exception, all of the previous growth prediction studies^{11,12} and orthognathic surgery prediction

Table 3. Comparative Overview of Predictive Performance Results According to the Number of Subjects and Variables Included in the Experimental Design Among the Multivariate Multiple Linear Regression (MMLR), Partial Least Squares Regression (PLSR), and Artificial Intelligence (AI) Prediction Models

AI Study Formulation	No. Subjects (n)	No. Predictors (p)	n/p Ratio	No. Outputs	Most Accurate Method Among AI, PLSR, and MMLR Prediction Models	Ratios of Cases Where AI Showed the Most Accurate Result
Growth prediction ¹²	410	161	2.55	78	AI was the most accurate method	81.1%, 63/78
Orthognathic surgery ¹³	705	254	2.78	32	PLSR, in 16/32, 50.0%	18.8%, 6/32
Orthodontic treatment (present study)	887	132	6.72	44	MMLR, in 34/44, 77.3%	11.4%, 5/44

studies^{5,7-10,13} showed MMLR to have a poorer predictive performance than PLSR. With some caution, it is surmised that this phenomenon might have been related to the ratio of the number of subjects (n) to the number of predictors (p), namely, the n/p ratio. PLSR is advantageous in the case of small n and large p situations,²³ whereas MMLR models would be more robust when the n/p ratio was greater than 5.26 In those studies, the n/p ratios of growth prediction and orthognathic surgery prediction studies were 2.55 and 2.78, respectively. The consequence was that PLSR models were better methods than MMLR models. In contrast, the current study had a larger sample size and fewer variables than previous studies, resulting in an n/p ratio of 6.72. As shown in Table 3, this might have led to the finding that MMLR was the more accurate method. However, further clinical studies or simulations are required to validate this hypothesis.

On a similar note, Al seemed to have shown better predictive performance when the n/p ratio was low, ie, when there were limited numbers of subjects but many predictor variables. This may imply that, in the case of three-dimensional (3D) studies which involve 3 to 4 times more variables than 2D study formulations, Al is likely to play a more meaningful role than conventional statistical methods.^{4,15} Orthodontic treatment prediction results will become more sophisticated in the future as more 3D information is collected.

CONCLUSIONS

- When predicting changes following orthodontic treatment, AI was not as effective as the conventional statistical methods, suggesting that AI might not always be the best option for predicting everything.
- However, this does not necessarily mean that conventional methods should be applied. The strength of the AI prediction method was apparent in predicting the soft tissue changes in the neck, whereas traditional methods had poorly predicted changes in that area.
- Applying multiple methods catered to the anatomic features and variability of response variables may be a viable option to improve predictive performance.

ACKNOWLEDGMENTS

A part of the data presented in the current study was included as a part of a doctoral dissertation (SJC). This study was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (Grant No. HI22C1518), and partly supported by grants (03-2022-0046 and 05-2023-0027) from the SNUDH Research Fund.

DISCLOSURE

The authors report there are no conflicts of interest that should be disclosed.

REFERENCES

- Park YS, Choi JH, Kim Y, et al. Deep learning-based prediction of the 3D postorthodontic facial changes. *J Dent Res.* 2022;101:1372–1379.
- Tanikawa C, Yamashiro T. Development of novel artificial intelligence systems to predict facial morphology after orthognathic surgery and orthodontic treatment in Japanese patients. *Sci Rep.* 2021;11:15853.
- Park JH, Kim Y-J, Kim J, et al. Use of artificial intelligence to predict outcomes of nonextraction treatment of Class II malocclusions. *Semin Orthod*. 2021;27:87–95.
- 4. Kang TJ, Eo SH, Cho H, Donatelli RE, Lee SJ. A sparse principal component analysis of Class III malocclusions. *Angle Orthod*. 2019;89:768–774.
- Suh HY, Lee HJ, Lee YS, Eo SH, Donatelli RE, Lee SJ. Predicting soft tissue changes after orthognathic surgery: the sparse partial least squares method. *Angle Orthod*. 2019; 89:910–916.
- Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York, NY: Springer; 2009.
- Yoon KS, Lee HJ, Lee SJ, Donatelli RE. Testing a better method of predicting postsurgery soft tissue response in Class II patients: a prospective study and validity assessment. *Angle Orthod*. 2015;85:597–603.
- Lee YS, Suh HY, Lee SJ, Donatelli RE. A more accurate soft-tissue prediction model for Class III 2-jaw surgeries. *Am J Orthod Dentofacial Orthop*. 2014;146:724–733.
- 9. Lee HJ, Suh HY, Lee YS, et al. A better statistical method of predicting postsurgery soft tissue response in Class II patients. *Angle Orthod*. 2014;84:322–328.
- Suh HY, Lee SJ, Lee YS, et al. A more accurate method of predicting soft tissue changes after mandibular setback surgery. J Oral Maxillofac Surg. 2012;70:e553–562.
- Moon JH, Kim MG, Hwang HW, Cho SJ, Donatelli RE, Lee SJ. Evaluation of an individualized facial growth prediction model based on the multivariate partial least squares method. *Angle Orthod*. 2022;92:705–713.
- Moon JH, Shin HK, Lee JM, et al. Comparison of individualized facial growth prediction models based on the partial least squares and artificial intelligence. *Angle Orthod*. 2024; 94:207–205.
- Park JA, Moon JH, Lee JM, et al. Does artificial intelligence predict orthognathic surgical outcomes better than conventional linear regression methods? *Angle Orthod*. 2024 in press DOI:10.2319/111423-756.1
- 14. Arik SÖ, Pfister T. TabNet: attentive interpretable tabular learning. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021;35:6679–6687.
- 15. Ghowsi A, Hatcher D, Suh H, et al. Automated landmark identification on cone-beam computed tomography: accuracy and reliability. *Angle Orthod*. 2022;92:642–654.
- Hwang HW, Moon JH, Kim MG, Donatelli RE, Lee SJ. Evaluation of automated cephalometric analysis based on the latest deep learning method. *Angle Orthod*. 2021;91:329–335.
- Moon JH, Hwang HW, Yu Y, Kim MG, Donatelli RE, Lee SJ. How much deep learning is enough for automatic identification to be reliable? *Angle Orthod*. 2020;90:823–830.

- Moon JH, Hwang HW, Lee SJ. Evaluation of an automated superimposition method for computer-aided cephalometrics. *Angle Orthod*. 2020;90:390–396.
- 19. Hwang HW, Park JH, Moon JH, et al. Automated identification of cephalometric landmarks: part 2—might it be better than human? *Angle Orthod*. 2020;90:69–76.
- 20. Park JH, Hwang HW, Moon JH, et al. Automated identification of cephalometric landmarks: part 1—comparisons between the latest deep-learning methods YOLOV3 and SSD. *Angle Orthod*. 2019;89:903–909.
- Jin H, Liao S, Shao L. Pixel-in-pixel net: towards efficient facial landmark detection in the wild. *Int J Comput Vis.* 2021;129:3174–3194.
- 22. Fernández A, Garcia S, Herrera F, Chawla NV. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *J Artif Intell Res.* 2018;61:863–905.
- 23. Kim K, Lee SJ, Eo SH, Cho SJ, Lee JW. Modified partial least squares method implementing mixed-effect model. *Commun Stat Appl Methods*. 2023;30:65–73.
- 24. Donatelli RE, Lee SJ. How to test validity in orthodontic research: a mixed dentition analysis example. *Am J Orthod Dentofacial Orthop.* 2015;147:272–279.
- 25. Moon JH, Lee JM, Park JA, Suh HY, Lee SJ. Reliability statistics every orthodontist should know. *Semin Orthod*. 2024;30:45–49.
- 26. Norman GR, Streiner DL. *Biostatistics: The Bare Essentials*. St. Louis, MO: Mosby Year Book; 1994.