

Prognosis Prediction for Class III Malocclusion Treatment by Feature Wrapping Method

Bo-Mi Kim^a; Bo-Yeong Kang^b; Hong-Gee Kim^c; Seung-Hak Baek^d

ABSTRACT

Objective: To use the feature wrapping (FW) method to identify which cephalometric markers show the highest classification accuracy in prognosis prediction for Class III malocclusion and to compare the prediction accuracy between the FW method and conventional statistical methods such as discriminant analysis (DA).

Materials and Methods: The sample set consisted of 38 patients (15 boys and 23 girls, mean age 8.53 ± 1.36 years) who were diagnosed with Class III malocclusion and received both first-phase (orthopedic) and second-phase (fixed orthodontic) treatments. Lateral cephalograms were taken before (T0) and after first-phase treatment (T1) and after second-phase treatment and retention (T2). Based on the measurements taken at the T2 stage, the patients were allocated into good ($n = 20$) or poor ($n = 18$) prognosis groups. Forty-six cephalometric variables on T0 lateral cephalograms were analyzed by the FW method to identify key determinants for discriminating between the two groups. Sequential forward search (SFS) algorithm and support vector machine (SVM) were used in conjunction with the FW method to improve classification accuracy. To compare the prediction accuracy of the FW method with conventional statistical methods, DA was performed for the same data set.

Results: AB to mandibular plane angle ($^{\circ}$) and A to N-perpendicular (mm) were selected as the most accurate cephalometric predictors by both the FW and DA methods. However, classification accuracy was higher with the FW method (97.2%) compared with DA (92.1%), because the FW method with SFS and SVM has a more precise classification algorithm.

Conclusions: The FW method, which uses a learning algorithm, might be an effective alternative to DA for prognosis prediction. (*Angle Orthod.* 2009;79:683–691.)

KEY WORDS: Class III malocclusion; Prognosis prediction; Feature wrapping method; Sequential forward search algorithm; Support vector machine

INTRODUCTION

Skeletal Class III malocclusion occurs because of undergrowth of the maxilla, overgrowth of the mandibular

bone, or both.¹ Furthermore, if patients also have an anterior crossbite during the growth period, skeletal discrepancies can be worsened.² Although it is difficult to completely change individual growth patterns,³ first-phase (orthopedic) and/or second-phase (fixed orthodontic) treatments can provide a better treatment outcome in cases in which there is a favorable growth potential (Figure 1A). In contrast, if patients have an unfavorable growth potential, the problem cannot be corrected with first- and/or second-phase treatments (Figure 1B). Therefore, predicting the prognosis for Class III malocclusion at the diagnostic stage is important for choosing an effective treatment plan.

Numerous studies have been performed to identify better methods of prognosis prediction for Class III malocclusion.^{4–18} Although discriminant analysis (DA) and the logistic regression methods have been used to investigate cephalometric predictors, it is difficult to find such predictors with consensus and high accuracy. In addition, obtaining a sufficient volume of patient data and long-term follow-up results remains a chal-

^a Graduate MS student, Department of Orthodontics, School of Dentistry, Dental Research Institute, Seoul National University, Seoul, South Korea.

^b Assistant Research Professor, Department of Dental Management and Informatics, Dental Research Institute, Seoul National University, Seoul, South Korea.

^c Associate Professor, Department of Biomedical Knowledge Engineering, School of Dentistry, Seoul National University, Seoul, South Korea.

^d Associate Professor, Department of Orthodontics, School of Dentistry, Seoul National University, Seoul, South Korea.

Corresponding author: Dr Seung-Hak Baek, Department of Orthodontics, School of Dentistry, Dental Research Institute, Seoul National University, Yeonkum-dong #28, Jongro-ku, Seoul, South Korea 110-768 (e-mail: drwhite@unitel.co.kr)

Accepted: August 2008. Submitted: July 2008.

© 2009 by The EH Angle Education and Research Foundation, Inc.

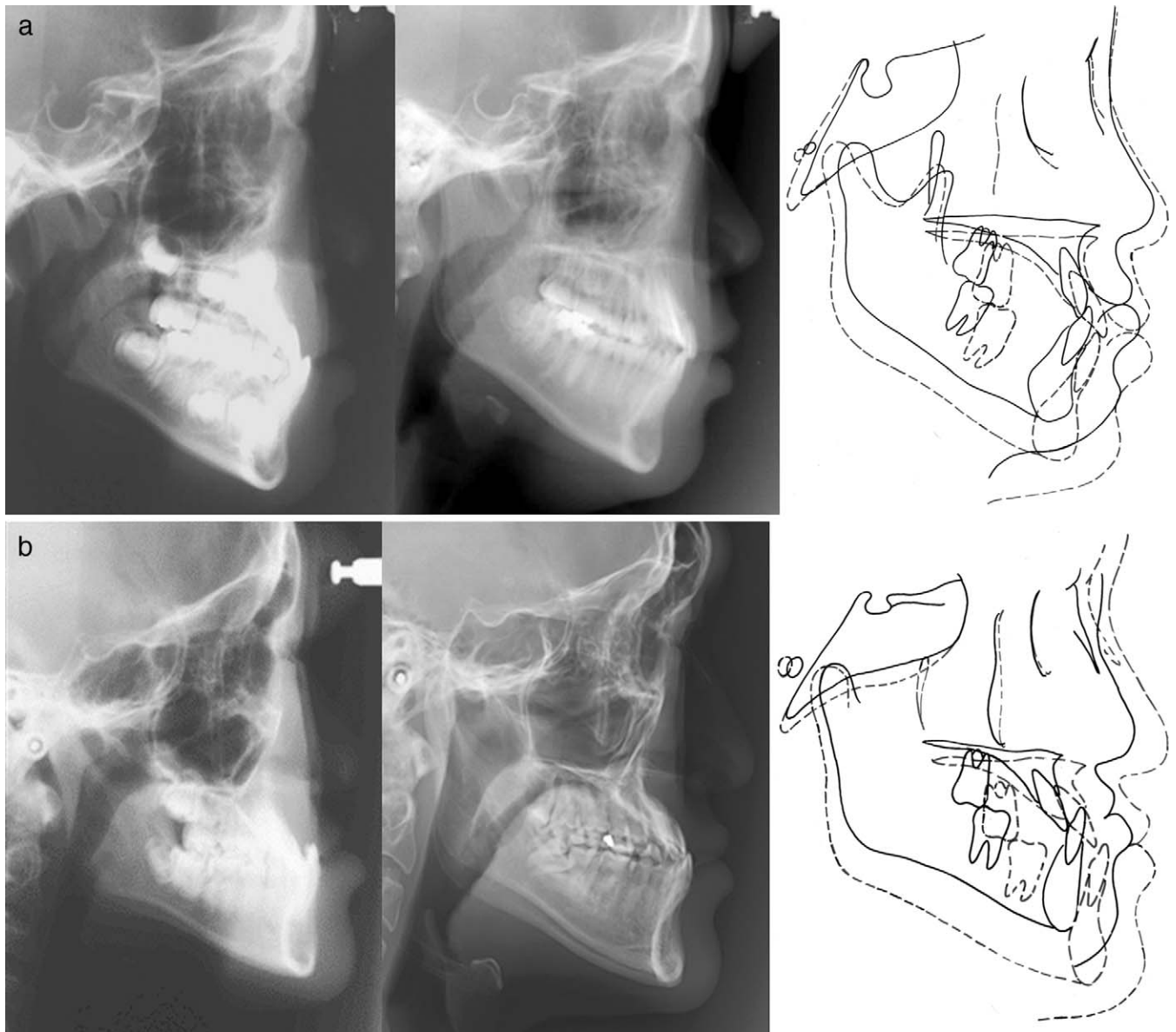


Figure 1. Lateral cephalograms before the first-phase (orthopedic) treatment (solid line) and after the second-phase (fixed orthodontic) treatment and retention (dotted line). (A) Good prognosis group. (B) Poor prognosis group.

lenge. Thus, a new methodology that can improve prediction accuracy based on a relatively small amount of patient data is needed.

The feature wrapping (FW) method employs a learning algorithm that can evaluate every set of features generated from original features in subjects and select the subset of features that show the highest classifi-

cation accuracy. FW works by identifying a small subset of necessary and sufficient features that can serve as input for the underlying predictor method.¹⁹ In the present study, the sequential forward search (SFS) algorithm and support vector machine (SVM) were used to improve classification accuracy in conjunction with the FW method.²⁰ Indeed, SVM is widely used in a

Table 1. Demographic Data for the Two Groups

	Good Prognosis Group (n = 20; 7 Males and 13 Females)			Poor Prognosis Group (n = 18; 8 Males and 10 Females)			Significance
	T0	T1	T2	T0	T1	T2	
Mean age, y	8.18 ± 1.38	11.30 ± 2.15	17.35 ± 1.76	8.71 ± 1.32	12.21 ± 1.76	17.86 ± 1.58	NS

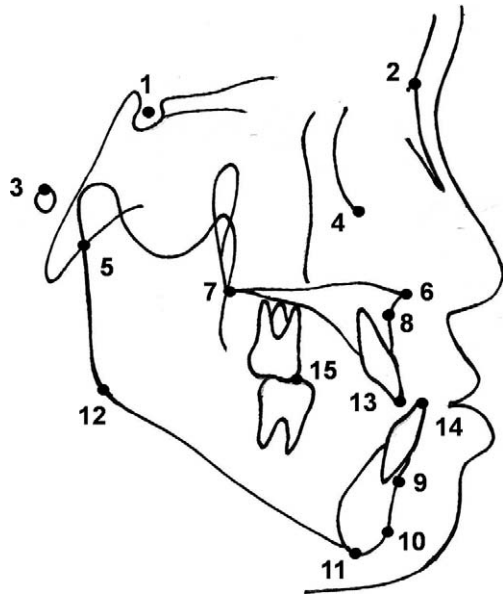


Figure 2. Cephalometric landmarks used in this study: 1, sella; 2, nasion; 3, porion; 4, orbitale; 5, articulare; 6, anterior nasal spine; 7, posterior nasal spine; 8, point A; 9, point B; 10, pogonion; 11, menton; 12, gonion; 13, incisal tip of upper central incisor; 14, incisal tip of lower central incisor; 15, point between the tips of the mesiobuccal cusps of all fully erupted upper and lower permanent first molars or deciduous second molars in primary dentition.

variety of fields in the medical sciences. For example, Bullinger et al²¹ successfully used SVM to discriminate between healthy control subjects and patients suffering from breast cancer by extracting nucleosides in urine samples. Likewise, Kawai et al²² used SVM to predict the pleiotropic effects of drugs, while Judson et al²³ used SVM to classify chemical toxicities.

The purposes of this study were, therefore, to identify which cephalometric markers show the highest

classification accuracy in prognosis prediction for Class III malocclusion by the FW method and to compare the prediction accuracies between the FW method and conventional statistical methods such as DA.

MATERIALS AND METHODS

The sample set consisted of 38 patients (15 boys and 23 girls, mean age 8.53 ± 1.36 years) who were diagnosed with Class III malocclusion and received both first-phase (orthopedic) and second-phase (fixed orthodontic) treatment at the Department of Orthodontics, Seoul National University Dental Hospital (Seoul, Korea). The first-phase treatment included orthopedics such as a chin cup or face mask with rapid palatal expansion according to the skeletal pattern. A chin cup or face mask was used for 12 to 14 hours per day with a force of 300 to 500 g per side. After orthopedic treatment, all subjects were treated with fixed appliances and preadjusted brackets. Fixed lingual retainers for the upper and lower anterior teeth, as well as a removable retainer for upper dentition, were used for retention. Demographic data are described in Table 1.

Inclusion criteria for this study was as follows: (1) existence of an anterior crossbite at the initial state that was corrected by first-phase treatment, (2) use of fixed appliance therapy for second-phase treatment, (3) follow-up performed until little craniofacial growth remained, and (4) lack of congenital deformities such as a cleft lip and palate.

Subjects were allocated into two groups according to final occlusal status. The good prognosis group (group 1) consisted of subjects who maintained favorable occlusal status with a normal overbite (>1.4 mm) and overjet (>2 mm). On the contrary, subjects who experienced relapse of the anterior crossbite (overjet

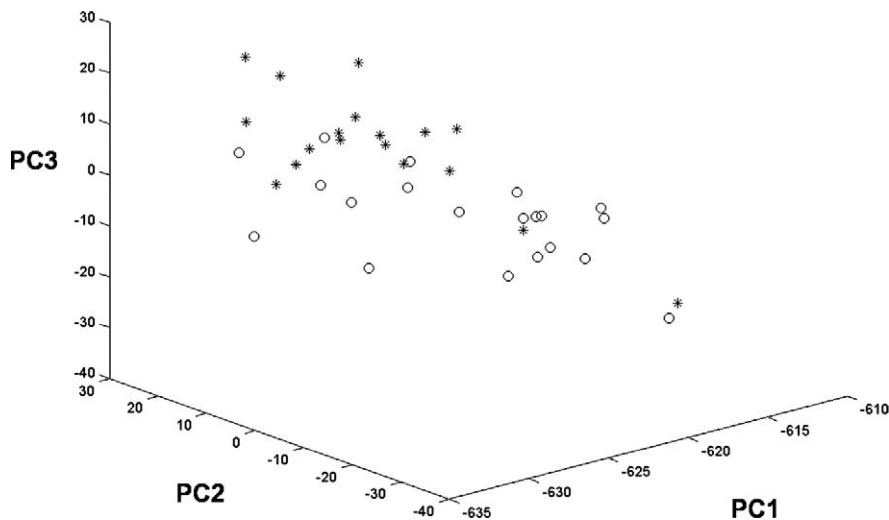


Figure 3. Three-dimensional principal component analysis plot of the data from the 38 patients. Circles indicate good prognosis; stars, poor prognosis.

Table 2. Definition of the Variables

Variable	Definition
Anteroposterior relationship	
SNA (°)	Angle between the anterior cranial base (SN) and NA line
SNB (°)	Angle between the anterior cranial base (SN) and NB line
ANB (°)	Angle between NA and NB lines
A-B plane angle (°)	Angle between the AB and facial planes (N-Pog)
AB to mandibular plane (°)	Angle between the AB plane and mandibular plane (Go-Me)
AB to occlusal plane (°)	Angle between the AB and occlusal planes
Facial plane angle (°)	Angle between the FH plane and facial plane (N-Pog)
A-N perp (mm)	Perpendicular distance from A to the N perp line to FH plane
Pog-N perp (mm)	Perpendicular distance from B to the N perp line to FH plane
APDI (°)	Sum of the facial plane to FH plane angle, AB to facial plane angle, and palatal plane (ANS-PNS) to FH plane angle
Ant. cranial base (mm)	Length of the anterior cranial base (SN)
Pog to NB (mm)	Perpendicular distance from Pog to the N-B line
Ar-Pog (mm)	Length from Ar to Pog
Body length (mm)	Mandibular body length (Go-Me)
Wits appraisal (mm)	Perpendicular distance from AO (perpendicular point from point A to the occlusal plane) to BO (perpendicular point from point B to the occlusal plane)
Vertical relationship	
Saddle angle (°)	Angle constructed by the SN plane and the S-Ar line
Articular angle (°)	Angle constructed by the S-Ar and Ar-Go lines
Gonial angle (°)	Angle constructed by the Gn-Go and Go-Ar lines
Upper gonial angle (°)	Angle constructed by the N-Go and Go-Ar lines
Lower gonial angle (°)	Angle constructed by the N-Go and Go-Gn lines
Bjork sum (°)	Sum of the saddle, articular, and gonial angles
Facial axis (°)	Angle between the Pt-Gn and cranial axes (Ba-N)
Y-axis to SN (°)	Angle between the SN plane and Y-axis (S-Gn)
Palatal plane angle (°)	Angle between the FH plane and palatal plane (ANS-PNS)
SN-GoGn (°)	Angle between the SN plane and Go-Gn line
FMA (°)	Angle between the FH and mandibular planes
Occlusal plane to SN (°)	Angle between the SN and occlusal planes
Occlusal plane to mandibular plane angle (°)	Angle between the occlusal and mandibular planes
ODI (°)	Sum of the AB to mandibular plane angle and palatal plane (ANS-PNS) to FH plane angle

Table 2. Continued

Variable	Definition
Go-Ar (mm)	Ramus height
Posterior facial height (mm)	Length from S to Go
Anterior facial height (mm)	Length from N to Gn
Lower anterior facial height (mm)	Length from ANS to Me
Proportion	
Facial height ratio	Posterior facial height/anterior facial height
Lower facial height ratio	ANS-Me/Nasion-Me
Dental relationship	
U1 to SN (°)	Angle between the maxillary incisor axis line and the S-N plane
U1 to FH (°)	Angle between the maxillary incisor axis line and the FH plane
L1 to A-Pog (mm)	Perpendicular distance from the mandibular incisor tip to the A-Pog line
L1 to occlusal plane (°)	Angle between the mandibular incisor axis line and the occlusal plane
FMIA (°)	Angle between the FH plane and the mandibular incisor axis line
IMPA (°)	Angle between the mandibular incisor axis line and the mandibular plane
Interincisal angle (°)	Angle between the maxillary incisor axis line and mandibular incisor axis line
OB (mm)	Overbite
OJ (mm)	Overjet
Soft tissue	
Upper lip E-plane (mm)	Perpendicular distance from the most anterior point of upper lip to the E-line (nose tip-chin tip)
Lower lip E-plane (mm)	Perpendicular distance from the most anterior point of lower lip to the E-line (nose tip-chin tip)

<0 mm) were classified into the poor prognosis group (group 2).

Lateral cephalograms (magnification factor = 10%) were taken before (T0) and after first-phase treatment (T1) and after second-phase treatment and retention (T2). The mean treatment time between T0 and T1 was 3.19 ± 1.78 years and between T1 and T2 was 5.97 ± 1.59 years. The mean age between the two groups during the three stages did not differ significantly. Fifteen landmarks and 46 skeletal and dental

Table 3. Classification Accuracy for Given Patient Data^a

System Extracted	Answer	
	Yes	No
Yes	True positive (TP)	False positive (FP)
No	False negative (FN)	True negative (TN)

^a Accuracy = (TP + TN)/(TP + FP + TN + FN). (Equation 1)

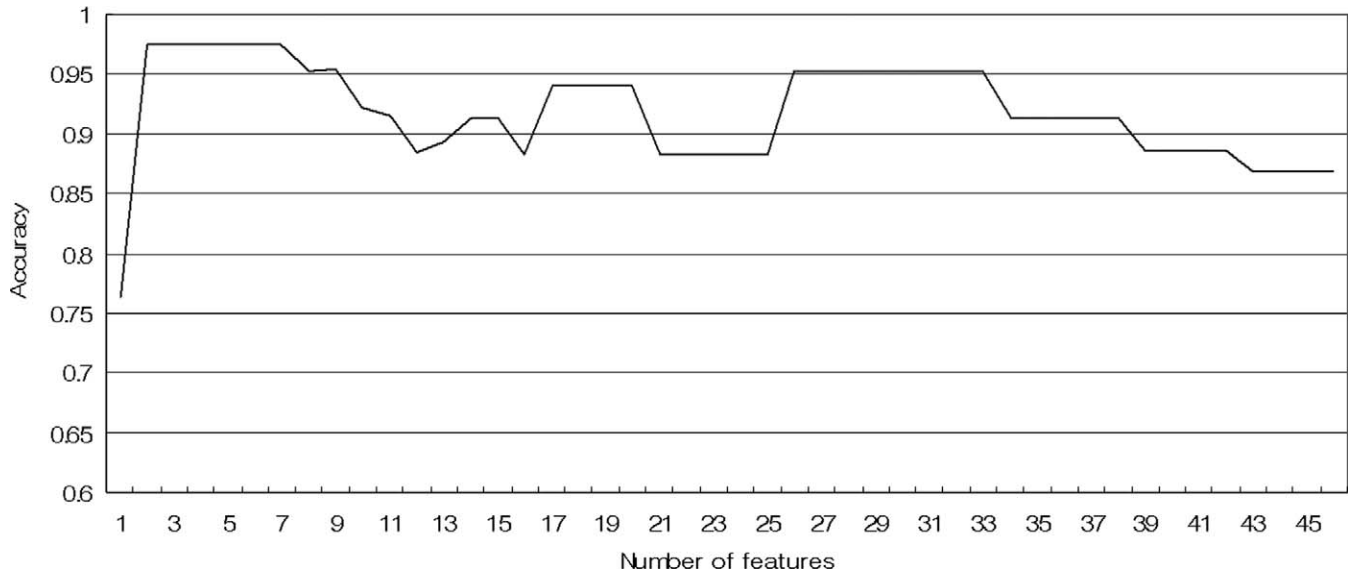


Figure 4. Classification accuracy for the 38 patients that was achieved with feature wrapping (SVM and SFS). SVM indicates support vector machine; SFS, sequential forward search algorithm.

variables were used in this study (Figure 2; Table 2). Method errors were calculated by Dahlberg’s formula,²⁴ $ME = \sqrt{\sum d^2/2n}$, where $\sum d^2$ is the sum of the squared differences between the two mean values, and n is the number of double measurements. The method errors for linear and angular measurement were not statistically significant and did not exceed 0.6 mm and 0.8°, respectively, for any variables.

For prognosis prediction, the FW method with the SFS algorithm²⁵ and SVM^{22,26,27} and principal component analysis (PCA)²⁸ were used.

PCA was used to examine the characteristics of the patients with good and poor prognosis. To simplify a data set, PCA transforms features in a multivariate data set into salient features that are not correlated with each other. Therefore, features representing patient samples can be reduced to a smaller number of features that are referred to as the principal components (PC). The largest variance for the data set is set as the first axis (the first PC) in the coordinate system. Likewise, the second greatest variance is set as the

second axis (the second PC), and so on. Therefore, PCA has the unique capacity of being able to perform an optimal linear transformation that maintains the subspace with the largest variance.

The SFS algorithm for feature selection works as follows²⁵: first, it chooses the single best feature that has the highest classification accuracy. Next, it forms all possible two-dimensional feature vectors that contain the best feature from the first step and chooses the best two-variable feature set that has the highest classification accuracy. This process then continues until a prespecified criterion is met, such as the dimension of the feature vector. The selected cephalometric markers were used to build the optimal classi-

Table 4. Cephalometric Markers and Prediction Accuracy of SVM for Classifying the 38 Patient Samples^a

SFS Step	Features	Accuracy
1	ABMP	0.7632
2	A-N Perp	0.9737
3	ODI	0.9737
4	LGA	0.9737
5	L1APg	0.9737
6	LL	0.9737
7	OPMP	0.9737

^a SVM indicates support vector machine; SFS, sequential forward search.

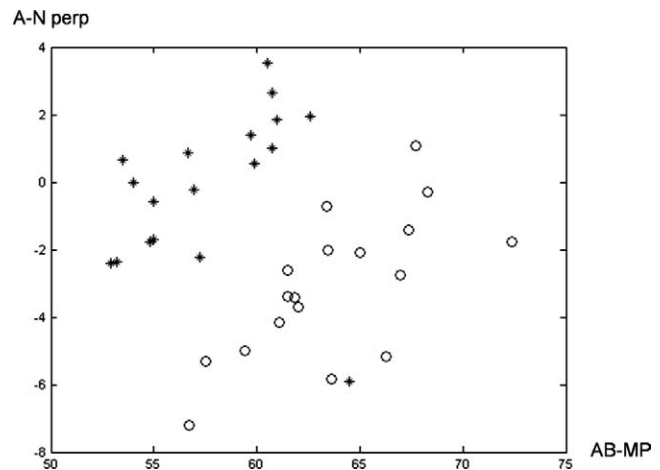


Figure 5. Two-dimensional plot of the 38 patient samples with two selected features, AB to mandibular plane angle (AB-MP, °) and A-N perp (mm). Circle indicates good prognosis; star, poor prognosis; x, values of AB-MP; y, values of A-N perp.

Table 5. Classification Results of the Stepwise Discriminant Analysis^a

Actual Group	Number of Cases	Predicted Membership		
		Group 1	Group 2	Sum
1	20	18 90.9%	2 10.0%	20 100%
2	18	1 5.6%	17 94.4%	18 100%

^a Percentage of original grouped cases correctly classified: 92.1%.

fier for discriminating between the patient samples. After completion of the feature selection step, the selected key markers were applied to the prediction module to generate the optimal classifiers that were the best predictors of prognosis.

The SVM method was used as a classifying and learning algorithm for the development of an FW method to identify cephalometric variables and prediction modules (Appendix). The classification accuracy of the model was estimated as the ratio of the number of correctly classified samples in all of the generated test samples from the leave-one-out cross-validation (LOOCV) procedure, divided by the total number of patient samples (Table 3; equation 1). In LOOCV, one test sample is extracted from a total of n samples. This test sample is then used for computing the classification accuracy of the remaining n – 1 training samples, and this process is repeated n times. By viewing the input data as two sets of vectors in an n-dimensional space, an SVM is able to construct a separating hyperplane in that space, one that maximizes the margin between the two data sets. To calculate the margin, two parallel hyperplanes, one on each side of the separating hyperplane, are pushed up against the two

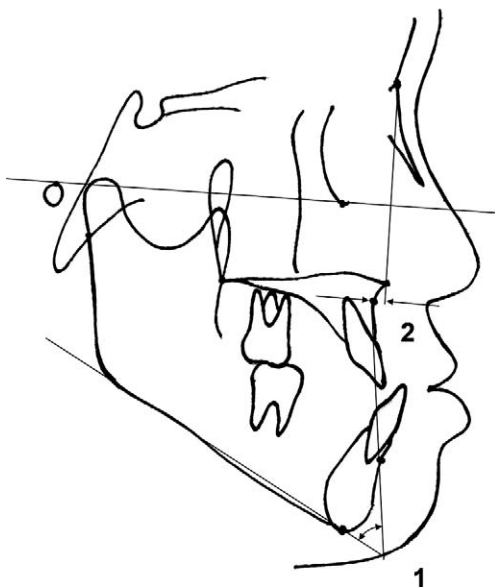


Figure 6. The most significant variables. 1, AB-MP (°); 2, A-N Perp (mm).

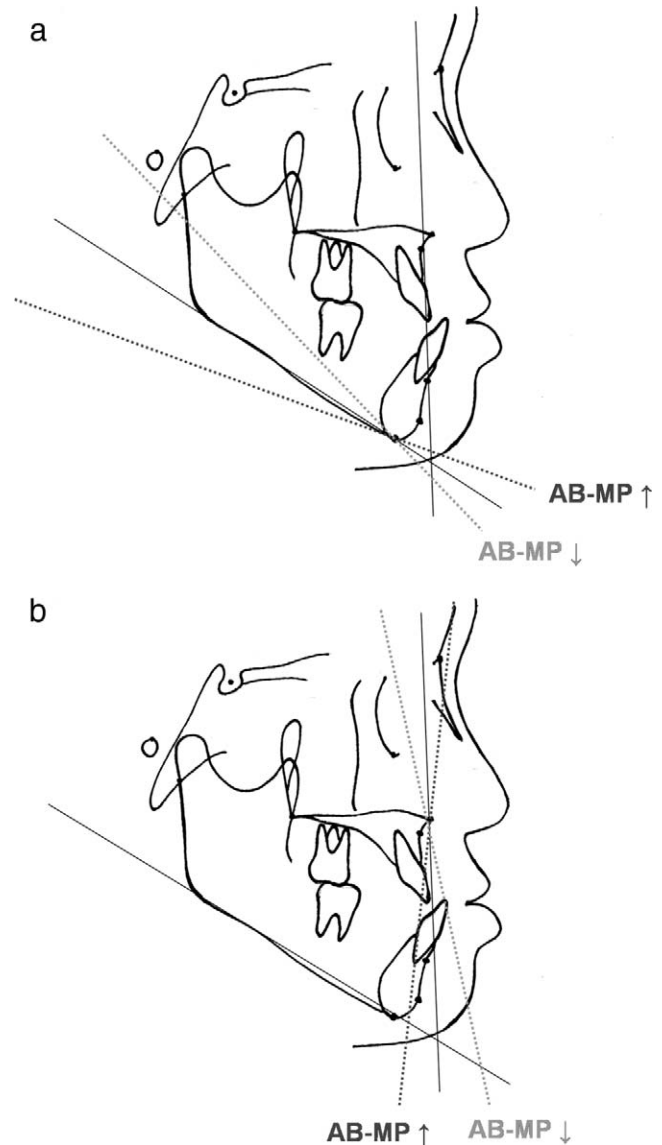


Figure 7. Relationship between AB-MP and skeletal pattern. If the mandibular plane is steep (A) or point B is anteriorly positioned (B), the patient has a low AB-MP value.

data sets. Intuitively, good separation is achieved by the hyperplane that has the largest distance to the neighboring data points of both classes. A tutorial on SVMs has been produced by Burges,²⁶ and a comparison of SVM to other classifiers has been made by Meyer et al.²⁷

RESULTS

Principal Component Analysis

By applying PCA to the feature matrix of patient samples, the features were projected onto a three-dimensional coordinate system composed of PCs 1, 2, and 3. Figure 3 illustrates the distribution of patient samples on this coordinate system. From this plot, we

expected to develop a classifier that could distinguish between good and poor prognosis samples. These two classes can be discriminated more effectively by the FW method.

FW Method

The SVM and SFS algorithms combine all possible two-dimensional feature sets into the best single feature and continue to search for an optimal subset that has the highest classification accuracy. Regarding prediction accuracy for classifying the 38 patient samples, the best accuracy that was achieved with a single feature (AB-MP) was 76.32%, while prediction analysis with two features (AB-MP and A-N perp) produced an accuracy of 97.37% compared with the LOOCV set (Figure 4; Table 4). Therefore, the highest classification accuracy was obtained using only two features, AB-MP and A-N perp.

The results of two-dimensional plotting of AB-MP and A-N perp features (Figure 5) indicated that most patients with good or poor prognosis for Class III malocclusion treatment could be clearly separated based on the values of these features. This finding implies that high classification accuracy was achieved, although a small number of features were used for the FW approach.

To compare our method with a conventional statistical method, stepwise DA was performed for the same data (Table 5). Although the same two variables, AB-MP and A-N perp, were selected from stepwise selection, the classification accuracy was 92.1%, which was lower than for the FW method (97.2%).

DISCUSSION

Several studies on prognosis prediction of treatment for Class III malocclusion have identified various predictors selected from statistical methods.^{4,14-16,18} Among these predictive variables, Yang and Kim⁴ presented the Björk sum, gonial angle, and occlusal plane to AB plane angle; Ko et al¹⁴ presented the lower incisor to occlusal plane angle and AB to mandibular plane angle, among others; Baccetti et al¹⁵ presented the mandibular ramus, cranial base angle, and mandibular plane angle; Moon et al¹⁶ presented the AB to mandibular plane angle and A-N perp; and Ghiz et al¹⁸ presented the gonial angle and mandibular length. However, despite these studies, a consensus predictor for treatment of Class III malocclusion has been elusive because of the difficulty in the collection of long-term follow-up samples, different grouping criteria, relatively low classification accuracies, and diversified variables.

In this study, cephalometric variables were investigated to build an optimal classifier for discriminating

between patient samples. AB-MP and A-N perp were selected as prognosis predictors, which is in accordance with the study by Moon et al.¹⁶ AB-MP is a variable used to describe the relationship between the anterior border of the maxillary and mandibular alveolar bone and the mandibular plane (Figure 6). A low AB-MP value indicates that the skeletal pattern is hyperdivergent and that the degree of mandibular prognathism is severe (Figure 7). Therefore, treatment results with a low AB-MP can be predictive of poor prognosis. It also implies that although face mask or chin cup therapy can correct an anterior crossbite, poor prognosis can still be determined according to continuous mandibular growth.

The second predictor, A-N perp, describes the anteroposterior position of the maxilla (Figure 6) and plays a decisive role in prognosis prediction because the anteroposterior position of the maxilla relative to the mandible is important. The percentage of Class III malocclusions due to retrusive point A in Koreans (18%) is lower than in whites (57%).^{29,30} Therefore, A-N perp was selected by a second SFS step following AB-MP (Table 4); however, the importance of A-N perp in samples of white patients will need to be investigated to verify this result.

Conventional statistical methods such as DA can also be used to analyze the data in this study; however, incorporation of new data into DA (Table 5) showed lower accuracy than the SVM method (Table 4). The SVM and SFS algorithms are a set of related and supervised learning methods used for classification and regression. Because they can minimize empirical classification errors and maximize the geometric margin, they are also known as maximum margin classifiers. Therefore, unlike conventional methods, they can produce the same or better classification accuracy when new data are inputted, thereby creating a valuable diagnostic program for prognosis prediction.

The FW method, as described here, may be an effective tool for prognosis prediction of treatment for Class III malocclusion. However, because of several limitations of this study, such as insufficient accumulation of long-term cases and lack of a multicenter analysis for evaluating ethnic differences, further studies are needed before the results of this study can be used in a clinical setting.

CONCLUSIONS

- AB-MP and A-N perp were selected as the most accurate cephalometric predictors for prognosis prediction of Class III malocclusion by both FW and DA methods.
- The classification accuracy was higher with the FW method (97.2%) than with DA (92.1%) because of

the FW method's sophisticated classification and learning algorithm.

- FW might be an effective alternative to DA for prognosis prediction.

REFERENCES

- Ngan P, Hägg U, Yiu C, Merwin D, Wei SH. Soft tissue and dentoskeletal profile changes associated with maxillary expansion and protraction headgear treatment. *Am J Orthod Dentofacial Orthop.* 1996;109:38–49.
- Westwood PV, McNamara JA Jr, Baccetti T, Franchi L, Sarver DM. Long-term effects of Class III treatment with rapid maxillary expansion and facemask therapy followed by fixed appliances. *Am J Orthod Dentofacial Orthop.* 2003;123:306–320.
- Harris JE, Kowalski CJ, Walker SJ. Dentofacial differences between “normal” sibs of Class II and Class III patients. *Angle Orthod.* 1975;45:103–107.
- Yang WS, Kim BH. The retrospective study on the indication of the chin cap therapy. *Korean J Orthod.* 1995;25:1–12.
- Sung JH, Kwon OW, Kim SD. The retrospective study on the prognosis of Class III malocclusion treatments. *Korean J Orthod.* 1998;28:175–187.
- Schulhof RJ, Nakamura S, Williamson WV. Prediction of abnormal growth in Class III malocclusions. *Am J Orthod.* 1977;71:421–430.
- Battagel JM. Discriminant analysis: a model for the prediction of relapse in Class III children treated orthodontically by a non-extraction technique. *Eur J Orthod.* 1993;15:199–209.
- Battagel JM. Predictors of relapse in orthodontically treated Class III malocclusions. *Eur J Orthod.* 1994;21:1–13.
- Franchi L, Baccetti T, Tollaro I. Predictive variables for the outcome of early functional treatment of class III malocclusion. *Am J Orthod Dentofacial Orthop.* 1997;112:80–86.
- Tahmina K, Tanaka E, Tanne K. Craniofacial morphology in orthodontically treated patients of class III malocclusion with stable and unstable treatment outcomes. *Am J Orthod Dentofacial Orthop.* 2000;117:681–690.
- Zentner A, Doll GM, Peylo SM. Morphological parameter as predictors of successful correction of class III malocclusion. *Eur J Orthod.* 2001;23:383–392.
- Schuster G, Lux CJ, Stellzig-Eisenhauer A. Children with class III malocclusion: development of multivariate statistical models to predict future need for orthognathic surgery. *Angle Orthod.* 2003;93:136–145.
- Alhajia Abu ES, Richardson A. Growth prediction in Class III patients using cluster and discriminant function analysis. *Eur J Orthod.* 2003;25:599–608.
- Ko YI, Baek SH, Yang WS. Determinants of successful chin cup therapy in skeletal class III malocclusion. *Am J Orthod Dentofacial Orthop.* 2004;126:33–41.
- Baccetti T, Franchi L, McNamara JA. Cephalometric variables predicting the long term success or failure of combined rapid maxillary expansion and facial mask therapy. *Am J Orthod Dentofacial Orthop.* 2004;126:16–22.
- Moon YM, Ahn SJ, Chang YI. Cephalometric predictors of long term stability in the early treatment of Class III malocclusion. *Angle Orthod.* 2005;75:747–753.
- Chen F, Terada K, Hanada K. A special method of predicting mandibular growth potential for class III malocclusion. *Angle Orthod.* 2005;75:191–195.
- Ghiz MA, Ngan P, Gunel E. Cephalometric variables to predict future success of early orthopedic Class III treatment. *Am J Orthod Dentofacial Orthop.* 2005;127:301–306.
- Cedeño W, Agrafiotis DK. Using particle swarms for the development of QSAR models based on K-nearest neighbor and kernel regression. *J Comput Aided Mol Des.* 2003;17:255–263.
- Sahiner B, Chan HP, Petrick N, Wagner RF, Hadjiiski L. Feature selection and classifier performance in computer-aided diagnosis: the effect of finite sample size. *Med Phys.* 2000;27:1509–1522.
- Bullinger D, Frohlich H, Klaus F, Neubauer H, Frickschmidt A, Henneges C, et al. Bioinformatical evaluation of modified nucleosides as biomedical markers in diagnosis of breast cancer. *Anal Chim Acta.* 2008;618:29–34.
- Kawai K, Fujishima S, Takahashi Y. Predictive activity profiling of drugs by topological-fragment-spectra-based support vector machines. *J Chem Inf Model.* 2008;48:1152–1160.
- Judson R, Elloumi F, Setzer RW, Li Z, Shah I. A comparison of machine learning algorithms for chemical toxicity classification using a simulated multi-scale data model. *MC Bioinformatics.* 2008;9:241.
- Dahlberg G. *Statistical Methods for Medical and Biological Students.* London, UK: Allen & Unwin; 1940.
- Pudil P, Novovicova J. Novel methods for subset selection with respect to problem knowledge. *IEEE Intelligent Systems.* 1998;10:66–74.
- Burges CJC. A tutorial on support vector machines for pattern recognition: data mining and knowledge. *Discovery.* 1998;2:121–167.
- Meyer D, Leisch F, Hornik K. The support vector machine under test. *Neurocomputing.* 2003;55:169–186.
- Jolliffe IT. *Principal Component Analysis.* New York, NY: Springer-Verlag, 2002;150–166.
- Lee KS. A roentgenocephalometric study on mandibular prognathism. *Korean J Orthod.* 1979;9:85–97.
- Guyer EC, Ellis EE III, McNamara JA Jr, Behrents RG. Components of Class III malocclusion in juveniles and adolescents. *Angle Orthod.* 1986;56:7–30.

APPENDIX

SVM Algorithm Used in This Study: libSVM classification and regression

Classify a binary data with polynomial kernel ($u \cdot v + 1$)³ and C = 10 options:

- s svm_type: set type of SVM (default 0)
0 = C-SVC
1 = nu-SVC
2 = one-class SVM
3 = epsilon-SVR
4 = nu-SVR
- t kernel_type: set type of kernel function (default 2)
0 = linear: $u \cdot v$
1 = polynomial: $(\gamma \cdot u \cdot v + \text{coef0})^{\text{degree}}$
2 = radial basis function: $\exp(-\gamma \cdot |u - v|^2)$
3 = sigmoid: $\tanh(\gamma \cdot u \cdot v + \text{coef0})$
- d degree: set degree in kernel function (default 3)
- g gamma: set gamma in kernel function (default 1/k)

- r coef0: set coef0 in kernel function (default 0)
- c cost: set the parameter C of C-SVC, epsilon-SVR, and nu-SVR (default 1)
- n nu: set the parameter nu of nu-SVC, one-class SVM, and nu-SVR (default 0.5)
- p epsilon: set the epsilon in loss function of epsilon-SVR (default 0.1)
- m cache size: set cache memory size in MB (default 100)
- e epsilon: set tolerance of termination criterion (default 0.001)
- h shrinking: whether to use the shrinking heuristics 0 or 1 (default 1)
- b probability_estimates: whether to train a SVC or

SVR model for the probability estimates 0 or 1 (default 0)

- wi weight: set the parameter C of class i to weight-C, for C-SVC (default 1)

The k in the -g option indicates the number of attributes in the input data. option-v randomly splits the data into n parts and calculates their cross validation accuracy/mean squared error.

The produced toolkit based on the data of 38 patients. Data can be downloaded on the following Web site: <http://bike.snu.ac.kr/?p=42>. There, if the ABMP and A-N perp values are inserted, prognosis prediction results are generated with +1, -1 values (+1 indicates good prognosis; -1, poor prognosis).