New approach for the diagnosis of extractions with neural network machine learning

Article *in* American journal of orthodontics and dentofacial orthopedics: official publication of the American Association of Orthodontists, its constituent societies, and the American Board of Orthodontics · January 2016

DOI: 10.1016/j.ajodo.2015.07.030

AJO-DO

New approach for the diagnosis of extractions with neural network machine learning

Seok-Ki Jung[a] and Tae-Woo Kim[b]
Ansan and Seoul, Korea

Introduction: The decision to extract teeth for orthodontic treatment is important and difficult because it tends to be based on the practitioner's experiences. The purposes of this study were to construct an artificial intelligence expert system for the diagnosis of extractions using neural network machine learning and to evaluate the performance of this model. **Methods:** The subjects included 156 patients. Input data consisted of 12 cephalometric variables and an additional 6 indexes. Output data consisted of 3 bits to divide the extraction patterns. Four neural network machine learning models for the diagnosis of extractions were constructed using a back-propagation algorithm and were evaluated. **Results:** The success rates of the models were 93% for the diagnosis of extraction vs nonextraction and 84% for the detailed diagnosis of the extraction patterns. **Conclusions:** This study suggests that artificial intelligence expert systems with neural network machine learning could be useful in orthodontics. Improved performance was achieved by components such as proper selection of the input data, appropriate organization of the modeling, and preferable generalization. (Am J Orthod Dentofacial Orthop 2016;149:127-33)

The most important part of orthodontic treatment is to determine the treatment plan.[1] An important part of treatment planning is the decision about extractions and the teeth to be extracted, because extractions are irreversible. Therefore, a prudent decision about extractions is required. A wrong decision could result in many problems during the orthodontic treatment. Undesirable results could be obtained, or in the worst-case scenario, the treatment might not be finished. Problems could include failure of anchorage control, abnormal inclination of the anterior teeth, unfavorable profile, improper occlusion, inadequate overjet and overbite, and difficulties in the closure of extraction spaces. Generally, most orthodontists make a decision with data from the clinical evaluations, photographs, dental models, and radiographs based on their experience and knowledge. Since there is no formula for the treatment plan, the decision depends on the practitioner's heuristics in many cases.[2] This often causes intraclinician and interclinician variability in the treatment planning process.[3] In addition, different records used for the diagnosis can cause differences in the treatment plan.[4-6] Moreover, differences in treatment planning can occur between experienced and less-experienced practitioners.[7] In particular, differences in extraction decisions could be critical. While allowing inexperienced practitioners to learn from the decisions of experienced practitioners would be helpful, decisions cannot be standardized with these combinations of measurements. Thus, another approach is needed.

Recently, there have been many studies about artificial intelligence and bioinformatics.[8-10] One approach is machine learning using a neural network system.[11,12] This emulates human learning in a situation that cannot be formulized or standardized. The human neural system consists of neurons that are linked at the synapse to send information. By repeated learning, each synapse linkage can be reinforced or weakened. In machine learning with the neural network, neurons link the input to the output, and each neuron is linked at the synapse. In each synapse, information of the input neurons is collected by a weighting technique. Weighted values are adjusted through iterative learning (Fig 1). Excessive iterative learning can elevate goodness-of-fit of the training set. However, errors of the test set can also be increased; this is called

[a]Clinical instructor, Department of Orthodontics, Korea University Ansan Hospital, Ansan, Korea.
[b]Professor, Department of Orthodontics, School of Dentistry, Dental Research Institute, Seoul National University, Seoul, Korea.
All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest, and none were reported.
Address correspondence to: Tae-Woo Kim, Department of Orthodontics, School of Dentistry, Dental Research Institute, Seoul National University, 101 Daehakro, Jongro-Gu, Seoul 110-749, Republic of Korea; e-mail, taewoo@snu.ac.kr.
Submitted, March 2015; revised and accepted, July 2015.
0889-5406/$36.00

http://dx.doi.org/10.1016/j.ajodo.2015.07.030

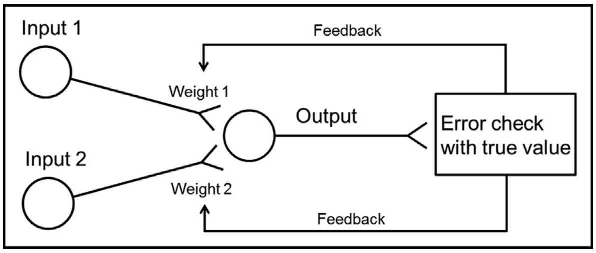

Fig 1. Schematic diagram of the neural network machine learning and weighting adjustment.

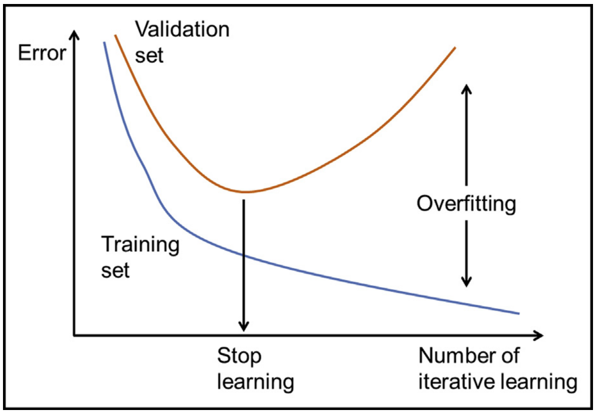

Fig 2. Learning curve of the training and validation sets.

overfitting. To avoid this, a validation set is introduced to stop learning and to make a generalized model (Fig 2). The generalized decision-making model can be formed through these procedures.

The aim of this study was to make an artificial intelligence decision-making model for the diagnosis of extractions using neural network machine learning. In addition, we wanted to evaluate the validity and accuracy of this model.

MATERIAL AND METHODS

The subjects consisted of 156 patients who had visited Seoul National University Dental Hospital in Seoul, South Korea, for an orthodontic consultation. Exclusion criteria were persons with unerupted permanent teeth or missing teeth (except for third molars), malformed teeth, previous orthodontic treatment history, maxillofacial deformities, and orthognathic surgery. Inclusion criteria were persons included in 5 treatment plan groups: nonextraction, maxillary and mandibular first premolar extractions (Ext_type_44-44), maxillary and mandibular second premolar extractions (Ext_type_55-55), maxillary first premolar and mandibular second premolar extractions (Ext_type_44-55), and maxillary first premolar extractions only (Ext_type_44-00) (Table I). For all subjects, the treatment plans were determined by 1 orthodontic specialist (T-W.K.), who had more than 10 years of experience.

Lateral cephalograms were collected as orthodontic records for all subjects. All tracings were made by 1 investigator (S-K.J.) and repeated twice at intervals of 2 weeks to analyze measurement errors. The reference points were digitized with the V-ceph program (version 5.3; Osstem, Seoul, Korea). Twenty-six landmarks and 12 measurements were chosen (Fig 3).

From this sample, 96 persons were assigned to the learning set, and 60 persons were assigned to the test set (Table I). The test set was used only for evaluation of the models. Two-thirds of the learning set was assigned to the training set and one-third of the learning set was assigned to the validation set. To find the optimal model, sliding window validation was performed. This is the validation technique to choose a validation set through the window moving sideways from the serial data.[13] To prevent overfitting, iterative learning was stopped at the minimum error point of the validation set. Next, through evaluation of the test set, the adequacy and accuracy were evaluated, and the best-fit model was chosen.

Table I. The subjects' sex, age, and other characteristics

Variable	*n*	*Mean age (y)*	*SD*
Sex			
Female	94	25	7
Male	62	23	6
Type of extractions			
No extractions	62		
Ext_type_44-44	20		
Ext_type_44-55	36		
Ext_type_55-55	25		
Ext_type_44-00	13		
Type of learning			
Learning set	96		
Test set	60		
Total	156		

A 2-layer neural network including 1 hidden layer was selected for the machine learning. There were 4 hidden nodes in the hidden layer. Hidden nodes play the role of interneurons in the artificial neural network system, and learning is performed through their weighted-values adjustment. Twelve measurements were selected for the input data: ANB angle, overjet, Björk sum, overbite, maxillary central incisor to SN angle, maxillary central incisor to occlusal plane angle,

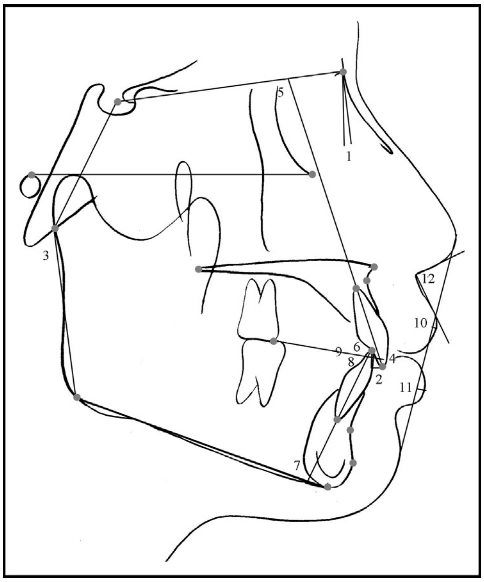

Fig 3. Linear and angular measurements used in this study: *1*, ANB angle; *2*, overjet; *3*, Björk sum; *4*, overbite; *5*, maxillary central incisor to SN angle; *6*, maxillary central incisor to occlusal plane angle; *7*, IMPA; *8*, mandibular central incisor to occlusal plane angle; *9*, interincisal angle; *10*, upper lip to E-line; *11*, lower lip to E-line; and *12*, nasolabial angle.

IMPA, mandibular central incisor to occlusal plane angle, interincisal angle, upper lip to E-line, lower lip to E-line, and nasolabial angle. These had clinical relevance such as anteroposterior relationships, vertical relationships, tooth inclinations, and soft tissue characteristics. In addition, 6 indexes—maxillary arch length discrepancy index, mandibular arch length discrepancy index, molar key index, large overjet index, protrusion index, and chief complaint index for protrusion—were included in the input data (Table II). The input data consisted of 18 elements in this manner. Maximum-minimum normalization was chosen for normalization of the input data in the range of 0 to 1. The learning rate was 0.9, and the sigmoid function was chosen as the activation function. The language R program (http://www.r-project.org/) was used for coding to construct machine learning models.[14] A back-propagation algorithm was applied to adjust the weighted values.

The output data were composed of 3 bits. Dx_ext was the index about whether extractions were needed. The value of 0 meant nonextraction, and 1 meant extractions. Dx_diff was the index about whether differential extractions between the maxillary and mandibular arches were needed. The value of 0 meant identical extractions such as Ext_type_44-44 and Ext_type_55-55. The value of 1 meant differential extractions such as Ext_type_44-55 and Ext_type_44-00. Dx_more was the index about whether more retraction was needed. The value of 0 meant mild-to-moderate retraction, such as in Ext_type_55-55 and Ext_type_44-55. The value of 1 meant moderate-to-severe retraction, such as in Ext_type_44-44 and Ext_type_44-00 (Table III).

Table II. Descriptions for the 6 additional indexes

Index	*Weighting*	*Criterion (mm)*
Arch length discrepancy		
Spacing	0	ALD > 0
Normal	0.25	$-1 <$ ALD ≤ 0
Mild crowding	0.5	$-3 <$ ALD ≤ -1
Moderate crowding	0.75	$-5 <$ ALD ≤ -3
Severe crowding	1	ALD ≤ -5
Molar key		
Class III key	0	
Super Class I key	0.25	
Class I key	0.5	
End-on key	0.75	
Class II key	1	
Large overjet		
Not severe	0	Overjet ≤ 5
Severe	1	Overjet > 5
Protrusion		
Concave profile	0	
Normal profile	0.25	
Mild protrusion	0.5	
Moderate protrusion	0.75	
Severe protrusion	1	
Chief complaint for protrusion		
No protrusion in chief complaint	0	
Protrusion in chief complaint	1	

ALD, Arch length discrepancy.

Table III. Descriptions of the output data

	Output data		
Group	*Dx_ext*	*Dx_diff*	*Dx_more*
Nonextraction	0		
Ext_type_55-55	1	0	0
Ext_type_44-44	1	0	1
Ext_type_44-55	1	1	0
Ext_type_44-00	1	1	1

Training had 3 stages, and 4 best-fit models were selected through the training. The first classifier (Classifier_1) was the model for determining whether to extract; the output was Dx_ext. The second classifier (Classifier_2) was the model for determining differential extractions; the output was Dx_diff. The third stage was for making the models for determining more retraction; the output was Dx_more. In the third stage, 2 classifiers (Classifier_3 and Classifier_4) regarding identical and

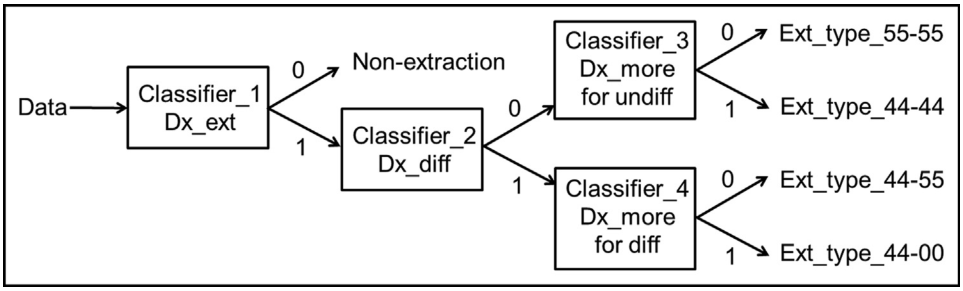

Fig 4. Schematic diagram of the stepwise learning used in this study.

differential extractions were derived (Fig 4). Extraction diagnosis of the total data was performed by the constructed classifiers. In comparison with an actual diagnosis, the decision-making success rates of Dx_ext, Dx_diff, and Dx_more were calculated. Finally, the total success rate of the diagnosis of extractions was calculated.

RESULTS

The results of the decision-making success rates are summarized in Table IV. In addition, each learning and validation curve is shown in Figure 5.

The intraclass correlation coefficient (ICC) was used to evaluate the test-retest reliabilities of the tracings; the values were scored as follows: ICC less than 0.4, poor reliability; ICC between 0.4 and 0.75, moderate reliability; and ICC greater than 0.75, excellent reliability.[15] The ICC values in this study ranged from 0.97 to 0.99, demonstrating excellent reliability.

In the diagnosis of extraction vs nonextraction, the decision-making success rates were 92% in the training set, 94% in the validation set, 93% in the test set, and 93% in total. In the diagnosis of identical vs differential extraction, the success rates were 88% in the training set, 100% in the validation set, 85% in the test set, and 89% in total. In the diagnosis of more retraction in identical extraction, the success rates were 88% in the training set, 75% in the validation set, 85% in the test set, and 84% in total. In the diagnosis of more retraction in differential extraction, the success rates were 95% in the training set, 100% in the validation set, 95% in the test set, and 96% in total. Through the sequential application of decision-making models, the final success rates were 85% in the learning set, 82% in the test set, and 84% in total (Fig 6).

In the analysis of the failed diagnoses, 7 cases were reversed between Ext_type_44-44 and Ext_type_44-55; this was the greatest portion. Next, 6 cases were reversed between Ext_type_55-55 and nonextraction.

Table IV. Decision-making success rates of each classifier (%)

	Learning set			
	Training set	*Validation set*	*Test set*	*Total set*
Classifier_1	92 (59/64)	94 (30/32)	93 (56/60)	93 (145/156)
Classifier_2	88 (35/40)	100 (20/20)	85 (29/34)	89 (84/94)
Classifier_3	88 (21/24)	75 (6/8)	85 (11/13)	84 (38/45)
Classifier_4	95 (20/21)	100 (7/7)	95 (20/21)	96 (47/49)
Total	85 (82/96)		82 (49/60)	84 (131/156)

In the 25 cases of failed diagnosis, unacceptable decisions were found in only 4 cases. The decisions for the other cases were acceptable because they were borderline. When we excluded these cases, the decision-making success rate rose to 97%.

DISCUSSION

For classification problems, machine learning has been used in many studies.[16-21] The decision to extract can be approached as a kind of classification problem. Takada et al[21] reported on a decision-making system for orthodontic treatment planning using a system based on template matching, which means finding a similar case from the established database; this is different from the method used in our study. Similar to this study, Xie et al[22] used artificial neural network modeling for determining extraction or nonextraction. The previous study had determined the necessity for extraction only. However, in this study, we also determined extraction positions. Furthermore, the decision-making success rates were improved. In the previous study, the success rates were 100% in the training set and 80% in the test set. The difference of these success rates could mean overfitting. To minimize overfitting and to verify the fitness of the model, the samples were divided into the learning set and the test set from

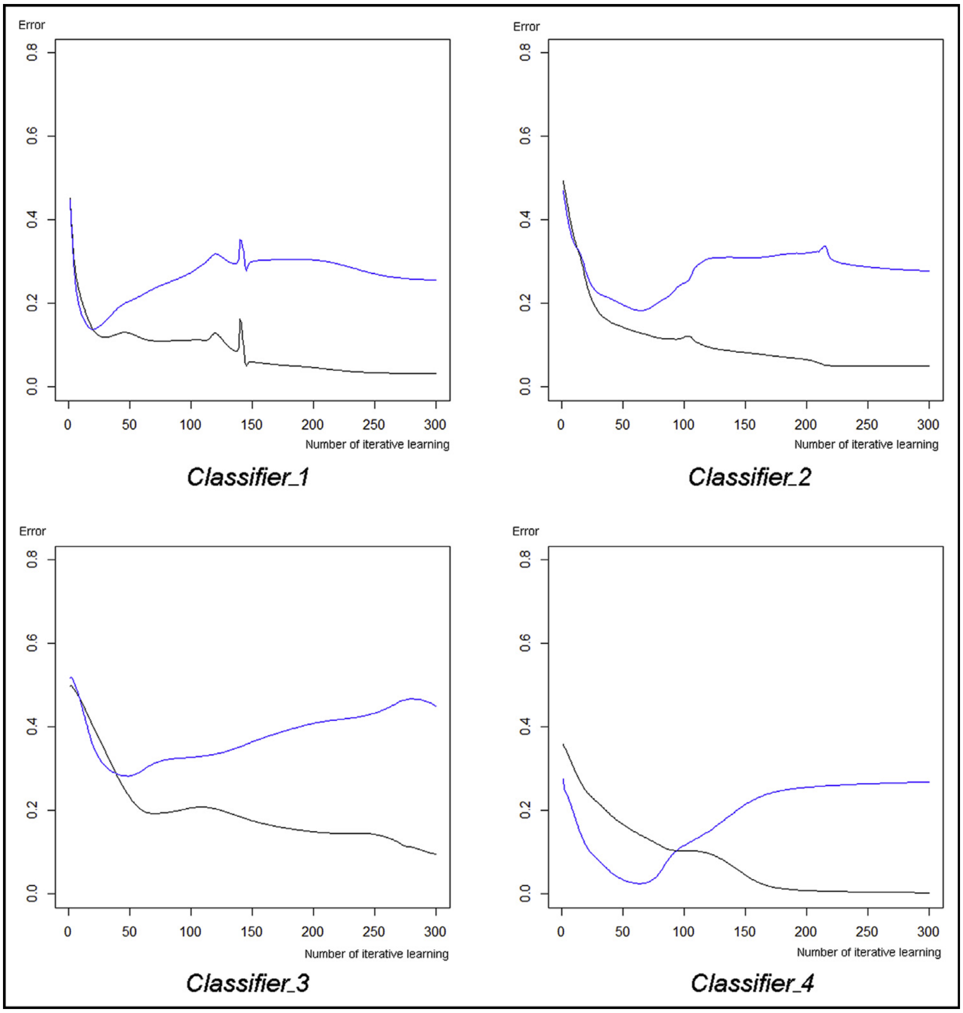

Fig 5. Learning curve (*black*) and validation curve (*blue*) of each classifier.

the beginning in our study. In addition, the learning set was divided into the training set and the validation set to make a generalized model. As a result of this, the success rates of the training set, the validation set, and the test set were similar in this study. It implies that this model was generalized better.

To treat skeletal Class III patients, surgical orthodontic treatments are preferred rather than camouflage treatments for an ideal result. Thus, the diagnosis of extractions was limited to 5 patterns in this study because they included the most nonsurgical orthodontic patients.

The main reasons for extractions are crowding and protrusion.[23] To reflect this, the indexes of arch length discrepancy and protrusive profile were added. In the pilot study, a grouped index showed better performance than a numeric value. The reason might be that the group was more important for the decision of extractions. The chief complaint index for protrusion was added because it could affect the diagnosis in borderline cases. Lastly, the molar key index and the large overjet index were added because they are important components for the diagnosis of differential extractions. For these reasons, the 6 additional indexes were added to the input data.

The output data were 3 bits, and learning was performed through 4 steps in this study because using an output of 0 or 1 showed better performance in the pilot study. Therefore, 2 bits of output were needed for 4 cases of extraction diagnosis patterns. Another bit was needed to divide extraction and nonextraction.

Although step-by-step learning had the shortcoming of accumulating the errors, the goodness-of-fit was better than 1-step learning because a simpler system might

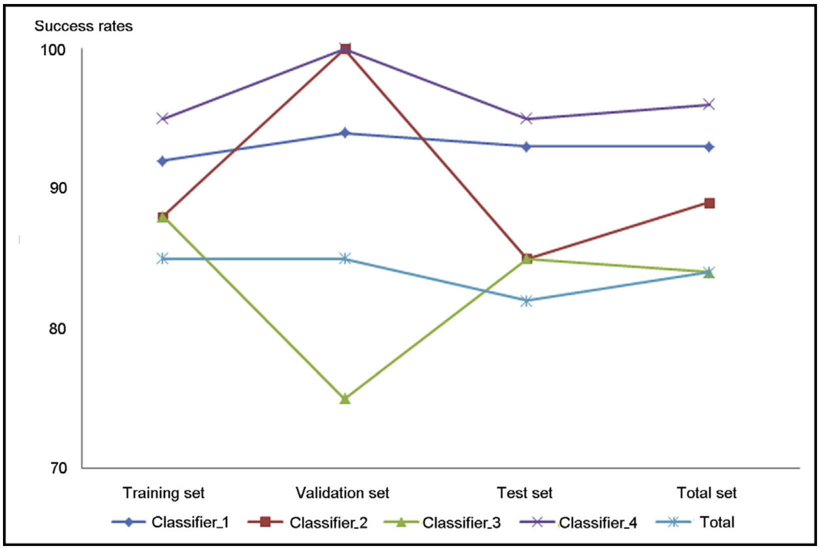

Fig 6. Success rates of each classifier.

have a higher success rate. In addition, a case that failed the previous step tended to fail also with the next step. Thus, accumulation of errors could be minimized.

The limitation of this study was that the diagnosis of extractions was confined to nonsurgical procedures. In addition, the model could not include cases with missing teeth, uncommon extraction patterns, asymmetry, and soft tissue functions. Further study for the diagnosis of surgical procedures and other cases will be planned. Through this, a complete model that covers all cases could be established. Another limitation of this study was the ambiguity of the protrusion index, which could be subjective. It was difficult to express protrusion exactly by the combination of several measurements. However, if the protrusion index is applied consistently, the model could have a reasonable result. If necessary, customized diagnostic learning for each practitioner could be possible to reflect his or her preference.

There is no correct answer for the diagnosis of extractions. The aim of this study was not to find a right answer. By mimicking the decision-making of experienced experts, the artificial intelligence expert system could be a reference for less-experienced practitioners. Clinicians can choose whether to follow that decision. Moreover, it is also possible that expert systems can be made using various philosophies of diagnosis. That is another merit of an artificial intelligence system.

In orthodontics, an expert system can useful.[24,25] The expert system constructed in this study showed high performance. Soon, advanced computer technology could make it possible to automatically measure the data for diagnosis.[26] Thus, an automatic process for treatment planning might be possible.[27,28]

CONCLUSIONS

As a result of making models for the diagnosis of extractions with neural network machine learning, the success rates of the classifiers were 93% for the diagnosis of extraction vs nonextraction and 84% for the detailed diagnosis of the extraction patterns in total. This study suggests that artificially intelligence expert systems with neural network machine learning could be a new approach in orthodontics.

REFERENCES

1. Proffit WR, Fields HW, Sarver DM. Contemporary orthodontics. 5th ed. St Louis: Mosby; 2013.
2. Ribarevski R, Vig P, Vig KD, Weyant R, O'Brien K. Consistency of orthodontic extraction decisions. Eur J Orthod 1996;18:77-80.
3. Dunbar AC, Bearn D, McIntyre G. The influence of using digital diagnostic information on orthodontic treatment planning—a pilot study. J Healthc Eng 2014;5:411-27.
4. Durao AR, Alqerban A, Ferreira AP, Jacobs R. Influence of lateral cephalometric radiography in orthodontic diagnosis and treatment planning. Angle Orthod 2015;85:206-10.
5. Rischen RJ, Breuning KH, Bronkhorst EM, Kuijpers-Jagtman AM. Records needed for orthodontic diagnosis and treatment planning: a systematic review. PLoS One 2013;8:e74186.
6. Devereux L, Moles D, Cunningham SJ, McKnight M. How important are lateral cephalometric radiographs in orthodontic treatment planning? Am J Orthod Dentofacial Orthop 2011;139:e175-81.
7. Luke LS, Atchison KA, White SC. Consistency of patient classification in orthodontic diagnosis and treatment planning. Angle Orthod 1998;68:513-20.

8. Su MC, Chang HT. A new model of self-organizing neural networks and its application in data projection. IEEE Trans Neural Netw 2001;12:153-8.
9. Yamamoto Y, Nikiforuk PN. A new supervised learning algorithm for multilayered and interconnected neural networks. IEEE Trans Neural Netw 2000;11:36-46.
10. Mackin N, Sims-Williams JH, Stephens CD. Artificial intelligence in the dental surgery: an orthodontic expert system, a dental tool of tomorrow. Dent Update 1991;18:341-3.
11. Brickley MR, Shepherd JP, Armstrong RA. Neural networks: a new technique for development of decision support systems in dentistry. J Dent 1998;26:305-9.
12. Perfetti R, Ricci E. Analog neural network for support vector machine learning. IEEE Trans Neural Netw 2006;17:1085-91.
13. Kennedy RL, Lee Y, van Roy B, Reed CD, Lippman RP. Solving data mining problems through pattern recognition. 1st ed. Upper Saddle River, NJ: Prentice Hall PTR; 1997.
14. R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
15. Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions. Hoboken, NJ: John Wiley & Sons; 2003.
16. Han M, Snow PB, Brandt JM, Partin AW. Evaluation of artificial neural networks for the prediction of pathologic stage in prostate carcinoma. Cancer 2001;91:1661-6.
17. Dayhoff JE, DeLeo JM. Artificial neural networks: opening the black box. Cancer 2001;91:1615-35.
18. Lux CJ, Stellzig A, Volz D, Jager W, Richardson A, Komposch G. A neural network approach to the analysis and classification of human craniofacial growth. Growth Dev Aging 1998;62:95-106.
19. McEvoy FJ, Amigo JM. Using machine learning to classify image features from canine pelvic radiographs: evaluation of partial least squares discriminant analysis and artificial neural network models. Vet Radiol Ultrasound 2013;54:122-6.
20. Yagi M, Ohno H, Takada K. Decision-making system for orthodontic treatment planning based on direct implementation of expertise knowledge. Conf Proc IEEE Eng Med Biol Soc 2010;2010:2894-7.
21. Takada K, Yagi M, Horiguchi E. Computational formulation of orthodontic tooth-extraction decisions. Part I: to extract or not to extract. Angle Orthod 2009;79:885-91.
22. Xie X, Wang L, Wang A. Artificial neural network modeling for deciding if extractions are necessary prior to orthodontic treatment. Angle Orthod 2010;80:262-6.
23. Bishara SE, Cummins DM, Jakobsen JR. The morphologic basis for the extraction decision in Class II, division 1 malocclusions: a comparative study. Am J Orthod Dentofacial Orthop 1995;107:129-35.
24. Hammond RM, Freer TJ. Application of a case-based expert system to orthodontic diagnosis and treatment planning. Aust Orthod J 1997;14:229-34.
25. Stephens C, Mackin N. The validation of an orthodontic expert system rule-base for fixed appliance treatment planning. Eur J Orthod 1998;20:569-78.
26. Forsyth DB, Davis DN. Assessment of an automated cephalometric analysis system. Eur J Orthod 1996;18:471-8.
27. Noroozi H. Introduction of a new orthodontic treatment planning software; a fuzzy logic expert system. Int J Orthod Milwaukee 2006;17:25-9.
28. Noroozi H. Orthodontic treatment planning software. Am J Orthod Dentofacial Orthop 2006;129:834-7.