

Artificial Neural Network Modeling for Deciding if Extractions Are Necessary Prior to Orthodontic Treatment

Xiaoqiu Xie^a; Lin Wang^b; Aming Wang^c

ABSTRACT

Objective: To construct a decision-making expert system (ES) for the orthodontic treatment of patients between 11 and 15 years old to determine whether extraction is needed by using artificial neural networks (ANN). Specifically, we will uncover the factors that affect this decision-making process.

Methods: A total of 200 subjects were chosen; among them, 120 were accepted for extraction treatments, and 80 were chosen for nonextraction treatments. For each case, 23 indices were selected. A 23-13-1 Back Propagation (BP) ANN model was constructed, and the data for 180 patients were aggregated to constitute the training set. Data for the other 20 patients were used as the testing set.

Results: When data from the 180 patients that had been trained were tested, the result was 100%, as expected. The untrained data from 20 patients in the testing set were 80% correct (ie, 16 cases were forecasted successfully). In the meantime, the relative contributions of the 23 input indices to the final output index (extraction/nonextraction) were calculated. "Anterior teeth uncovered by incompetent lips" and "IMPA (L1-MP)" were the two indices that gave the biggest contributions sequentially; the index of FMA (FH-MP) gave the smallest contribution.

Conclusions: (1) The constructed artificial neural network in this study was effective, with 80% accuracy, in determining whether extraction or nonextraction treatment was best for malocclusion patients between 11 and 15 years old; (2) when the clinician is predicting whether an orthodontic treatment requires extraction, the indices "anterior teeth uncovered by incompetent lips" and "IMPA (L1-MP)" should be taken into consideration first. (*Angle Orthod.* 2010;80:262–266.)

KEY WORDS: Artificial neural network; Malocclusion; Extraction

INTRODUCTION

Morbidity associated with malocclusion has increased recently. Orthodontic treatments for malocclusion can be classified as extraction treatments and nonextraction treatments. The decision to extract is based primarily on

such factors as cast measurements, cephalometry, and growth. Because such consideration is dependent on only a few factors, it usually cannot provide general guidance to the practitioner; rather the decision to extract requires a multiple-factor analysis, which often includes the clinical experiences of the orthodontist.

Currently, many multiple-factor analysis methods are available for use. Among these, the most frequently used is the statistical process known as fuzzy grouping analysis. Fuzzy grouping analysis regroups multiple factors based on their closeness in affecting the extraction decision. Classification by this algorithm is applicable to many patients. The aim of this study was to construct a decision-making expert system (ES) for orthodontic treatment by using a new approach (ie, the artificial neural networks [ANN]). With the use of such a system, a suggested decision can be made as to whether orthodontic treatment in patients ranging from 11 to 15 years old requires an extraction.

The ANN is a computational or mathematical model based on the biological signal processing of the

^a Associate Chief Physician, Department of Orthodontics, Stomatology Hospital Affiliated with Medical School, Nanjing University, Nanjing, PR China; Graduate student, Department of Orthodontics, School of Stomatology, Nanjing Medical University, Nanjing, PR China.

^b Professor and Dean, Department of Orthodontics, School of Stomatology, Nanjing Medical University, Nanjing, PR China.

^c Professor and Department Chair, Department of Mathematics and Physics, Xuzhou Medical College, Xuzhou, PR China.

Corresponding author: Dr Lin Wang, Professor and Dean, Department of Orthodontics, School of Stomatology, Nanjing Medical University, 136 Han Zhong Road, Nanjing 210029, PR China
(e-mail: lw603@njmu.edu.cn)

Accepted: July 2009. Submitted: November 2008.

© 2010 by The EH Angle Education and Research Foundation, Inc.

cerebrum. The human cerebrum comprises neurons and the interconnections between them. Construction of the ANN model is based on analysis and learning of the structure, mechanism, and function of biological neural networks. An understanding of biological neural networks allows for the construction of an ANN that can help model complex relationships or establish patterns within a group of datum points. When intuition and concrete thinking information are processed by ANN models, much better results can be obtained than with the use of traditional processing modalities. The ANN can process nonlinear relationships and can exhibit learning ability. ANN provides researchers with advantages such as large-scale parallelism, distribution representation of knowledge, robustness, and self-organization, all of which offer a new approach to complex problems.

ANN models are beneficial when one is researching medical problems because of their ability to process complicated problems of uncertainty, nonconfiguration, nonlinearity, and multiple-factor interactions. As a result, the application of ANN shows great potential as a support system and management system in medical decision making. ANN also allows for the utilization of multiple factors to solve problems¹⁻⁴ such as medical prognostication, classification, pattern recognition, and image processing.

ANN shows promise in the field of orthodontics. We can see this in the human craniofacial growth study of Lux et al⁵ by ANN, which concluded that ANN were advantageous in classifying and analyzing previously unknown child cases with respect to growth patterns in children. For this reason, in our study, ANN was chosen to investigate whether extraction is an appropriate treatment for malocclusion. ANN can help researchers to extract meaningful information from complicated primary data sets of subjects, and to extract new medical information. With the help of ANN, predictions of diagnosis, treatment, and outcome for a single patient can be made directly. Researchers do not need to search for rules of the data before ANN to make the individualized decision.⁶⁻¹¹

SUBJECTS AND METHODS

Subject Selection

We conducted a detailed investigation of the population that visited our department during the years from 1999 to 2005. It was found that most malocclusion patients with permanent dentition accepted for fixed appliance orthodontic treatments were between 11 and 15 years of age. Thus, patients in this age group were selected for this study. The numbers of extraction and nonextraction patients were calculated, and their ratio was approximately 6:4. For this study,

we chose 200 patients between the ages of 11 and 15 years. Among those chosen, 120 were selected for extraction treatment and 80 were selected for non-extraction treatment, matching the 6:4 ratio. This project, in which patient records were used, was started after institutional approval was received.

Index Determination

Through literature review and clinical case study regarding the frequently used varieties of indices, 25 indices were selected for screening of subjects. Two of these were nonquantification indices, which included the situation of heredity and protruded anterior teeth uncovered by incompetent lips. Among the quantifiable indices, 5 were derived from cast measurement, 13 from hard tissue cephalometrics, and 5 from soft tissue cephalometrics. Cast measurement indices included crowding in the upper and lower dental arches, overbite, overjet, and the space needed to correct the Spee's curve. Hard tissue cephalometric indices consisted of ANB, Wits (mm), FMA (FH-MP), FMIA (L1-FH), IMPA (L1-MP), L1-NB, L1-NB (mm), U1-SN, U1-NA, U1-NA (mm), and U1-L1, respectively; soft tissue cephalometric indices included NLA (Cm-Sn-UL), Ns-Sn-Pos, UL-Eplane (mm), LL-Eplane (mm), and Z angle, respectively.

Setup of the Subject Database

Of total data sets from 200 patients, 180 were aggregated to constitute the training set for the artificial neural network. The remaining 20 were used as the testing set.

The input data were preprocessed first. Nonquantification indices were converted into figures between 0 and 1 using the encoding method; quantification indices were encoded and normalized. Both of the encoding methods employed nonlinear transformations. Because of ANB=SNA-SNB in the data preprocessing method, SNA and SNB were omitted for training samples. As a result, all of the 23 indices were quantified. After the data were preprocessed, all input indices were valued at between 0 and 1. Because the output index was extraction or nonextraction, quantification was processed as 0.99 for "yes" and 0.01 for "no." Based on the data processing method described above, a case database was constructed for predicting whether or not orthodontic treatment for malocclusion patients between 11 and 15 years old required extraction.

Construction of the ANN Model

The ANN model was constructed to predict whether malocclusion patients between 11 and 15 years old required orthodontic extraction treatment. The con-

structured ANN model had 23 neurons in the input layer and 1 neuron in the output layer; this corresponded to the use of extraction or nonextraction treatments. The model was implemented using the FORTRAN programming language, which is based on the principle of artificial neural networks.

This Back Propagation (BP) ANN employs the error backward propagation learning algorithm.¹² The basic principle of the BP algorithm is the propagation of errors from the output layer backward to the input layer by each layer that “shares” the error with neurons of each layer. Thus the reference errors of each layer of neuron are obtained for use in adjusting the corresponding connection weights, to make the error function diminish as far as possible. To enhance the performance of BP networks, a suitable learning parameter η and momentum parameter ε should be chosen properly.

Data from the 180 patients-in-training set were used to train the ANN model described above. Data from 20 patients were used to test the accuracy of the ANN model. When η was chosen as 0.9 and ε as 0.7, and the number of neurons in the hidden layer was 13, the model had a nice learning effect. The 20 test samples proved successful in evaluating factors that affect the decision-making process.

RESULTS

Correlation Factors of Extraction or Nonextraction Orthodontic Treatment

The constructed 23-13-1 BP ANN was used to train the data from 180 subjects. After 61,198 instances of training iterations, the error was smaller than 0.005, which met the requirement. When the number of iterations was set at 200,000 for training, the error was even smaller than 0.001944.

Contributions of the 23 input layer indices to the output layer index were analyzed through the method of neural network data processing. The connection strengths of each neuron in the input layer with each neuron in the hidden layer were used to represent the values of contribution from every input index. The values of a new index $F(i)$ were calculated respectively to represent the contributions. These new indices were ordered by their magnitude, with the largest on top. The new index described the contributions from every input index to the result, as is shown in Table 1.

According to the new index $F(i)$, input indices “anterior teeth uncovered by incompetent lips” and “IMPA (L1-MP)” were the two that were associated with the biggest contributions sequentially; the index FMA (FH-MP) gave the smallest contribution. The contributions of other indices differed a little, and the differences did not show strict quantitative relations.

Table 1. Analysis of Contributions of Every Input Index

Input Index	$F(i)$	Order
Anterior teeth uncovered by incompetent lips	14042.44	1
IMPA (L1-MP)	11833.00	2
Overjet	6693.52	3
Crowding in the upper dental arch, mm	6135.12	4
Space for correcting Spee's curve, mm	5948.78	5
ANB	5891.38	6
Overbite	5689.83	7
LI-NB	5640.86	8
LL-E plane, mm	4697.43	9
Soft tissue convexity (Ns-Sn-Pos)	4033.89	10
Interincisal angle (UI-LI)	4000.84	11
UI-NA, mm	3967.02	12
Z angle	2872.43	13
Wits, mm	2664.37	14
NLA (Cm-Sn-UL)	2620.67	15
UI-SN	2447.43	16
UL- E plane, mm	2250.07	17
UI-NA	2199.48	18
Heredity	2190.95	19
Crowding in the lower dental arch, mm	2173.83	20
FMIA (L1-FH)	1966.10	21
L1-NB, mm	1520.10	22
FMA (FH-MP)	410.39	23

Use of the Decision-Making BP ANN in Determining Whether Orthodontic Treatment of Malocclusion Patients 11 to 15 Years Old Required an Extraction

To check the accuracy of this model, the data from 180 samples that had been trained were tested first. The rate of accuracy was 100%, which demonstrated that the constructed ANN could make correct decisions regarding the data of the trained 180 samples. Then, the data of 20 testing set samples that had not been trained were tested. Table 2 shows the results. When predefined network error was set at ≤ 0.3 (generally, this value ought to be at least ≤ 0.5), 16 samples were predicted successfully. This demonstrated that the rate of accuracy was 80%. As for the marginal cases, a different ANN model (eg, wavelet ANN model) could be employed, so that nonlinear mapping, classification, and identification performance could be enhanced.

DISCUSSION

Correlation Factors of Extraction or Nonextraction Orthodontic Treatment

This study was successful in determining the serial order of correlation factors in the determination of extraction or nonextraction orthodontic treatment through the method of ANN (Table 1). Results implied that in judging whether an orthodontic treatment needs extraction or not, the two indices “anterior teeth uncovered by incompetent lips” and “IMPA (L1-MP)” should be considered first.

Table 2. Absolute Errors of 20 Test Samples

Sample No.	BP Prediction	Goal Output	Absolute Error
1	0.010	0.010	0.000
2	0.000	0.010	0.010
3	0.000	0.010	0.010
4	0.992	0.010	0.982
5	0.991	0.990	0.001
6	0.992	0.990	0.002
7	0.987	0.990	0.003
8	0.992	0.990	0.002
9	0.992	0.990	0.002
10	0.991	0.990	0.001
11	0.000	0.010	0.010
12	0.000	0.010	0.010
13	0.022	0.010	0.012
14	0.887	0.010	0.877
15	0.012	0.990	0.978
16	0.992	0.990	0.002
17	0.044	0.990	0.946
18	0.991	0.990	0.001
19	0.708	0.990	0.282
20	0.992	0.990	0.002

The main purpose of tooth extraction in orthodontics is to provide relief of crowding. The second important purpose of orthodontic extraction is to diminish the prominence of dental arches and to correct the discrepancy of anteroposterior relationships between arches. Tooth extraction also provides spaces that allow for correcting the discrepancy of vertical dimension. Furthermore, extraction spaces can be utilized to correct width discrepancy between arches (crossbite or scissor bite of posterior teeth), tooth size discrepancy (congenital agomphosis or Bolton index abnormality), and so forth.

As the result of our improved understanding of contemporary orthodontic theory gained by in-depth observation of orthodontic treatments, nonextraction methods tend to be applied more widely. Some researchers like Schwab¹³ object to extraction treatment based on the impact to profile and temporomandibular joint (TMJ) by first premolar extraction. Therefore, nonextraction treatment has been regarded as a viable option in recent years.

“Anterior teeth uncovered by incompetent lips” is a common clinical manifestation that impacts appearance, function, and the potential social psychology of the patient. Bishara et al¹⁴ conducted morphologic comparison research on 91 subjects with Angle Class II division 1 malocclusion who had accepted extraction treatment. They concluded that the degree of protrusion of the anterior lip was paramount in the decision of whether to extract.

IMPA (L1-MP) reflexes the position and inclination to the mandibular plane of the lower incisors. In situations of normal muscle balance, the inclination of the lower

incisors should be within normal limits and should be coordinated with the base bone to ensure the health of the tissues and the stability of the correction. If the inclination of the lower incisors is too big, spaces will be needed to correct it. To gain the spaces, extraction is the favored choice. Therefore, IMPA (L1-MP), which reflexes the inclination of the lower incisors, is another important reference index for deciding whether or not to extract.

FMA (FH-MP) was found to contribute very little in this study, which seemed to be contradictory to the general view that the mandibular plane angle is a significant index for extraction decisions. A possible reason was that the FMA that occurred in most of the subjects in this sample was within the normal range (mean \pm SD). Besides, index IMPA (L1-MP) is actually decided by mandibular morphology and the gradient of the mandibular plane. If the mandibular plane angle is low, the angulation of the lower incisors will be relatively small; the lower incisors are often upright as well. If the mandibular plane angle is high, the angulation of the lower incisors is relatively large; the lower incisors are often labially inclined as well. Therefore, this index, which reflexes labial inclination of the lower incisors, has correlation with the gradient of the mandibular plane. Moreover, it provided the biggest contribution within all quantization indices. In this regard, we also can infer that the mandibular plane angle is a key factor as well when extraction decisions are made. FMA still affected the decision process, but the weight was small. The conclusion from ANN is for reference, and needs a combination with practical orthodontics when applied. Other methods like the pruning algorithm BP neural network also can be applied in the index selection of the ANN model.

Current Application Situation of Expert System in Orthodontics

Expert system (ES) is an important branch of the field of artificial intelligence (AI). AI is a science that strives to make a computer possess or seem to possess intelligence. ES is a computer program system that processes knowledge and information, which is composed primarily of a knowledge base and an inference machine. It simulates the decision-making and working processes of experts and solves actual problems in the field of a single specialty.

Poon et al¹⁵ were the first to use a new approach to knowledge acquisition known as Ripple-Down Rules in Dentistry to develop an ES in clinical orthodontics. This system comprises a knowledge base of 680 rules. Investigators found that such an ES has potential as an interactive advisory tool and is applicable in clinical orthodontic situations.

Hammond et al¹⁶ pointed out in a review that traditional rule-based expert systems had some limitations when applied to orthodontic diagnosis and treatment planning. These limitations may be avoided by using a case-based system, which is a particular type of ES that uses a stored data bank of previously treated cases to provide knowledge for use in solving new treatment problems. Hammond et al¹⁷ also investigated the application of this method in the field of orthodontic diagnosis and treatment planning. A case base of 300 cases was entered into a case-based ES shell. A test set of 30 consecutive cases then was used to test the diagnostic capacity of the system. The computer-generated treatment plan matched the actual treatment plan in 24 of the 30 cases.

In the work of Lux et al,⁵ the growth of 43 orthodontically untreated children was analyzed by lateral cephalograms taken at the ages of 7 and 15. For the description of craniofacial skeletal changes, the concept of tensor analysis and related methods were applied. Through the use of an ANN, namely, self-organizing neural maps (SOM), resultant growth data were classified, and relationships of the various growth patterns were monitored. This type of network provided a frame of reference for classifying and analyzing previously unknown cases with respect to their growth pattern.

After studying published work on neural networks, Brickley et al¹⁸ concluded that ANN expert systems may be trained with clinical data only and therefore can be used in cases where “rule-based” decision making is not possible. This is the case in many clinical situations. ANN therefore may become important decision-making tools within dentistry.

CONCLUSIONS

- When the task is to decide whether an orthodontic treatment needs extraction, the indices “anterior teeth uncovered by incompetent lips” and “IMPA (L1-MP)” could be considered first.
- The constructed artificial neural network in this study is able to correctly judge with 80% accuracy whether malocclusion patients between 11 and 15 years of age need extraction.

REFERENCES

1. Fukushima K. Neural network model for selective attention in visual pattern recognition and associative recall. *Applied Optics*. 1987;26:4985–4992.
2. Toulouse G, Dehaene S, Changeux J-P. Spin glass model of learning by selection. *Proc Natl Acad Sci USA*. 1986;83:1695–1698.
3. Yamamoto Y, Nikiforuk PN. A new supervised learning algorithm for multilayered and interconnected neural networks. *IEEE Transactions on Neural Networks*. 2000;11:36–46.
4. Su MC, Chang HT. A new model of self-organizing neural networks and its application in data projection. *IEEE Transactions on Neural Networks*. 2001;12:153–158.
5. Lux CJ, Stellzig A, Volz D, et al. A neural network approach to the analysis and classification of human craniofacial growth. *Growth Dev Aging*. 1998;62:95–106.
6. Dayhoff JE, Dleo JM. Artificial neural networks: opening the black box. *Cancer*. 2001;91(suppl):1615–1635.
7. Bostwick DG, Burke HB. Prediction of individual patient outcome in cancer: comparison of artificial neural networks and Kaplan-Meier methods. *Cancer*. 2001;91(suppl):1643–1646.
8. Montie JE, Wei JT. Artificial neural networks for prostate carcinoma risk assessment. *Cancer*. 2001;91(suppl):1647–1652.
9. Han M, Snow PB. Evaluation of artificial neural networks for the prediction of pathologic stage in prostate carcinoma. *Cancer*. 2001;91(suppl):1661–1666.
10. Snow PB, Kerr DJ. Neural network and regression predictions of 5-year survival after colon carcinoma treatment. *Cancer*. 2001;91(suppl):1673–1678.
11. Gospodarowicz M, Mackillop W. Prognostic factors in clinical decision making: the future. *Cancer*. 2001;91(suppl):1688–1695.
12. Rumelhart DE, McClelland JL. Parallel distributed processing: explorations in the microstructure of cognition. Cambridge: MIT Press; 1986:282–362.
13. Schwab DJ. The borderline patient and tooth removal. *Am J Orthod*. 1971;59(2):126.
14. 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;107:129–135.
15. Poon KC, Freer TJ. EICO-1: an orthodontist-maintained expert system in clinical orthodontics. *Aust Orthod J*. 1999;15:219–228.
16. Hammond RM, Freer TJ. Application of a case-based expert system to orthodontic diagnosis and treatment planning: a review of the literature. *Aust Orthod J*. 1996;14:150–153.
17. Hammond RM, Freer TJ. Application of a case-based expert system to orthodontic diagnosis and treatment planning. *Aust Orthod J*. 1997;14:229–234.
18. Brickley MR, Shepherd JP, Armstrong RA. Neural networks: a new technique for development of decision support systems in dentistry. *J Dent*. 1998;26:305–309.