Title page

Manuscript ID：KJO-21-255

Title: Prediction of patient experience of Invisalign treatment using artificial neural networks

Lin Xu[1], Li Mei[2], Ruiqi Lu[3], Yuan Li[1], Hanshi Li[1], Yu Li[1*]

[1]State Key Laboratory of Oral Diseases, National Clinical Research Center for Oral Diseases, Department of Orthodontics, West China Hospital of Stomatology, Sichuan University, Chengdu, China.

[2]Discipline of Orthodontics, Department of Oral Sciences, Sir John Walsh Research Institute, Faculty of Dentistry, University of Otago, Dunedin, New Zealand.

[3]Department of Electronic Engineering, Tsinghua University, Beijing, China.

Correspondence:

Yu Li, Department of Orthodontics, State Key Laboratory of Oral Diseases, National Clinical Research Center for Oral Diseases, West China Hospital of Stomatology, Sichuan University, 14#, 3rd Section, South Renmin Road, Chengdu 610041, China. E-mail: yuli@scu.edu.cn. Phone number: 028-85503645.

Funding

This work was supported by National Natural Science Foundation of China (NSFC) [grant number 31971247].

Data availability

The repository of the source code underlying this article is available at https://github.com/lurqTHU/clinic.

Prediction of patient experience of Invisalign treatment using artificial neural networks

ABSTRACT

Objectives: Poor experience of Invisalign treatment will affect patient compliance, and thus the treatment outcome. Knowing potential discomfort level for a specific case in advance can help orthodontists better prepare for the patient to overcome the tough stage. The aim of this study was to construct artificial neural networks (ANNs) for predicting patient experience in the early stage of Invisalign treatment.

Methods: A total of 196 patients were included in the study. Data collection included the questionnaires of pain, anxiety, and quality of life (QoL). Four-layer fully connected multilayer perceptions (MLP) with three back propagation were constructed for predicting patient experience of the treatment. Input data consisted of 17 clinical features. The partial derivatives method was used to calculate relative contributions of each input in ANNs.

Results: The success rates of prediction were 87.7% for pain, 93.4% for anxiety, and 92.4% for QoL. The ANNs of predicting pain, anxiety, and QoL yielded the area under the curve of 0.963, 0.992, and 0.982, respectively. The number of teeth with lingual attachments was found to be the most important factor affecting the outcome of negative experience, followed by the number of lingual buttons, and upper incisors with attachments.

Conclusions: The preliminary study has constructed ANNs which show good accuracy in predicting patient experience (*i.e.* pain, anxiety and QoL) of Invisalign treatment, bearing a potential for clinical use with further enlarged learning in future.

KEY WORDS: Computer algorithm, Pain, Compliance, Aligners.

INTRODUCTION

Artificial intelligence (AI) has been developing rapidly and making remarkable achievements in various domains of medicine and dentistry.[1] Compared with the traditional logistic regression model, artificial neural network (ANN) with multilayer perceptions (MLP) has shown an advantage in modelling complicated nonlinear relationships, with higher sensitivity, specificity, and accuracy of medical diagnostic.[2] In a study determining the cervical vertebrae stages in orthodontics, ANN demonstrated the most stable algorithm with a high accuracy value compared with k-nearest neighbors (k-NN), Naive Bayes (NB), decision tree (Tree), support vector machine (SVM), and random forest (RF).[3] Because of its capabilities of modelling nonlinear relationships in a high-dimensional data set, ANN has provided new approaches for orthodontists in automated cephalometric analysis and cone beam computed tomography image segmentation, accurate diagnosis, and treatment planning.[4]

Invisalign, one of the fastest developing orthodontic appliances in dentistry, translates orthodontic treatment plan into a series of clear aligners to align teeth.[5] Although Invisalign is more esthetic and comfortable compared with the traditional fixed appliances,[6, 7] in our clinical practice, some patients still complain about a varying degree of discomfort and anxiety.[8, 9] Both traditional appliances and Invisalign have shown, to some extent, oral dysfunction, mucosal irritation, difficulty in chewing, and swollen throat or tongue.[10, 11] This could reduce the wear time of aligners and compliance, thus influencing the treatment outcome;[12] even a very small number of patients give up treatment because of terrible experience.[13, 14] Therefore, the attention to mental status should be considered in the treatment plan for the best possible patient-centered care.

Theoretically speaking, the complexity of appliances may directly affect patients' comfort level. However, the impact and relationship between different designs of aligners are unclear. And clinical evidence in predicting patient experience using ANN is still lacking.

Therefore, an AI system was construct for patient comfort prediction, which can be applied in software later, helps orthodontists to predict the likely comfort level of designed aligners. If a significant discomfort is detected, some modifications and health education in advance can be considered to improve patient's comfort level and compliance. The aim of this study was to construct ANNs to predict patient experience (*i.e.* pain, anxiety and QoL) of

Invisalign treatment based on different designs of Invisalign treatment. The findings of the study can help clinicians detect individuals who might be at risk of poor patient experience with a reduced compliance during treatment.

MATERIALS AND METHODS

Subject selection

The study was approved by the Ethics Committee of West China Hospital of Stomatology, Sichuan University (WCHSIRB-D-2019-073). Written informed consent was obtained from each patient. The study was designed as a prospective cohort study. A total of 196 patients wearing Invisalign clear aligners (Align Technology, USA) were recruited at the Department of Orthodontics, West China Hospital of Stomatology, Sichuan University, Chengdu, China, between 2018 and 2021. The sample size was decided based on practical grounds (existing study cohort), and also referred to similar studies concerning artificial intelligence systems.[15-17]

Inclusion criteria were: (1) Patients between 18-50 years old; (2) Going to wearing Invisalign clear aligners; (3) No history of major dentoalveolar diseases. Exclusion criteria were: (1) Dental and oral diseases, *e.g.* caries, periodontal diseases, and temporomandibular joint disorders; (2) Severe systemic diseases; (3) Psychological and mental disorders; (4) Taking medications that treat or cause pain and mental diseases. One patient was included only once. Tooth extraction surgeries and the placement of TADs would interfere with, even cover up the aligner designs on the patients' self-report discomfort. So those surgeries should be done at least one week before or after the questionnaire investigation in order to avoid potential interferences.

All patients wore clear aligners following the same wearing protocol (22 hours/day for 10 days).

Data collection

Patients were asked to fill in the questionnaires daily for eight days, briefly included before and the first seven days after wearing the first set of aligners. Questionnaires included the Visual Analogue Scale (VAS) of pain, Self-Rating Anxiety Scale (SAS), and Oral Health Impact Profile-14 (OHIP-14) based on the literature.[18-20] VAS: The degree of pain was assessed by marking the pain level on a 10-mm straight line, ranging from 0 mm (no pain) to 10 mm (worst pain). SAS: The level of anxiety was assessed with 20 questions. Each question was graded as "occasionally", "sometimes", "often" and "always". OHIP-14: The QoL was assessed with 14 items, with 5 options “never”, “hardly”, “occasionally”, “fairly often” and “very often”. These questionnaires were validated with good reliability in the literature.

Patients’ clinical records were obtained from the hospital databases. The animation scheme and treatment design of the Invisalign treatment (with the patient’s personal information hidden) was collected from ClinCheck system (Align Technology, California, USA). A total of 17 clinical features were collected from medical records and ClinCheck (Table 1). Given the complexity of clinical practice, the features were not grouped together based on any experiences, such as grouping the elastics and precision cut together, and were maintained as original as possible.

Dataset pre-processing

All elements of input of clinical features were normalized in the range of 0-1. The normalization of input is shown in Table 1. For integral and continuous feature, maximum minimum normalization (MMN), a linear and hyperparameter-free method, was utilized for normalization. Non-linear method, such as dividing subgroups (analysis of data grouping) and non-linear curve, would introduce more hyperparameters, and increase the complexity of ANN model and risk of over-fitting. For binary feature, the raw value was already normalized and therefore kept unchanged.

The label of patient experience was 0 or 1. After collecting questionnaire scores of patient experience, the difference of pain, anxiety, and QoL level were calculated between the highest and lowest scores to address inter-patient subjectivity. Higher difference indicated more negative patient experience of Invisalign treatment. Then the differences were all binarized using thresholds (3.0 for pain, 6.5 for anxiety, and 7.0 for quality of life). These thresholds were determined as the averages of difference of pain, anxiety, and QoL level in training and validation set. Patients with differences higher than the threshold were positive samples (label = 1), and lower were negative (label = 0). Table 2 shows the numbers of patients with positive and negative labels distributed in training, validation and test set. Because the numbers of positive and negative samples differed for different prediction targets of pain, anxiety, and QoL, samples with different labels were randomly partitioned among datasets. The positive and negative samples were of similar proportion in the training set, validation set, and test set in each ANN. The binarized values were the final prediction targets.

Artificial neural networks (ANNs) construction

Figure 1 demonstrates the analysis process of ANNs. The above 17 clinical features (Table 1) for each patient were collected as input. Three ANNs were constructed to predict whether negative experience occurred or not in patients undergoing the first set of aligners. ANNs were four-layer fully connected multilayer perceptions (MLP), with 17 input nodes, two hidden layers with nine hidden nodes per layer, and one output node. The relu function was chosen as the activation function for non-linearity after each hidden layer, and its definition is $\mathrm{relu}(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$, where x is the value calculated by linear operations before the activation function.[21]

The positive probability was obtained by applying a non-linear sigmoid function to the value of output node. Its definition is:

$$\text{sigmoid}(x) = \frac{1}{1+\exp(-x)} \tag{1}$$

where x is the result of the output node.[22] The other two ANNs for anxiety and QoL prediction shared the same model structure but were trained separately and therefore the values of parameters were different. During the training stage, random dropout with the probability of 0.5 was adopted in the hidden layer, which randomly set the activation values of certain number of hidden nodes to be 0 to increase the training stability. The Binary Cross-Entropy loss was used to calculate the difference between the ground truth and the predicted result. Its definition is:

$$\text{BCE}(x) = -[(1-y)\log(1-x) + y\log(x)] \tag{2}$$

where y is the ground truth label and x is the predicted result of the ANN. The algorithm of Back Propagation was used to update the parameters of the neural network based on equation (2). The learning rate was set as 0.1 according to the recent literature. Adaptive moment (Adam) estimation optimizer was adopted for updating the parameters of ANNs.[23]

Training and evaluation of ANNs

The dataset of 196 patients was split into training, validation and test set with a ratio of 3:1:1.[24, 25] Although the back propagation method could be used to train the parameters of

ANNs, there were still some untrainable hyper-parameters, *e.g.* the learning rate, the total number of training steps and the number of nodes in the hidden layer. To determine these hyper-parameters, four-fold cross validation method was used in the training and validation set. [17, 26] For each fold of validation, the samples within training and validation set were firstly combined together and then randomly partitioned into two parts with ratio of 3:1. The larger part was used to optimize the parameters of ANNs, and the smaller part was used to monitor the training process and check if over-fitting happened. Normally the loss on the smaller part will firstly decrease for some training steps and then start to increase at a point, producing a minimum. Such procedure was repeated four times for each set of hyper-parameters, and the average loss was calculated. The set of hyper-parameters with the smallest loss were chosen. After all hyper-parameters were determined through validation, the training and validation set were combined together again to train the final model. The test set was held-out and not available during all the processes stated above and was only used to evaluate the success rate of the final model.

The label of each sample was either 0 or 1, but the predicted probability of ANN using equation (1) ranged from 0 to 1. Therefore, a threshold was needed to determine if a sample suffered from bad experiences. We defined a determination of suffering pain, anxiety, and decreased QoL for each patient as the predicted probability being higher than the corresponding optimum diagnostic cutoff value derived from the Receiver Operating Characteristic (ROC) curves.[27] Using the cutoff value, the predicted probability of each sample could be binarized into 0 and 1. If the prediction of one sample equals its label, then it's counted as a success sample, otherwise a failure sample. Success rate of prediction, sensitivity, specificity and the area under the ROC curve (AUC) were used to evaluate ANNs performance. As training and validation sets were all used during training the model, the success rate on them was obviously high. However, the test set was held-out during the training process in order to simulated the real situation where the model predicts experiences for new patients. Therefore, if the success rate was also high on test set, then we can assume with certainty that the model could perform well under real scenes.

Analysis of contributions of input features

The partial derivatives method, which is widely applied in giving the contribution profile of input factors, was used to calculate the relative contributions of inputs and rank them in order.[28] Contribution of each input to each ANN was calculated, indicating the influence of each feature on the output of pain, anxiety, and QoL. Total contribution of three ANNs for each input was also calculated, and higher values of total contribution denoted higher influence of each input factor on the output of overall negative experience with equal weights.

RESULTS

Accuracy of ANN prediction of patient experience

The learning curves of the three ANNs during cross validation are shown in Figure 2a-c. The training and validation loss were demonstrated, where the training loss was being optimized during training and the validation loss was used to determine when the learning was sufficient and whether to stop the training process. The training loss decreased slowly with fluctuation. The validation loss decreased relative fast at the early stage and quickly saturated. After training for certain epochs, the validation loss stopped decreasing and started to increase. It meant that though the model behaved better on the training set, the accuracy on the validation set did not improve. It was a sign of over-fitting and the training procedure was therefore stopped at the lowest point of the validation loss curve. It could also be seen that the validation loss was consistently higher than the training loss for the prediction of anxiety and QoL, because the model was optimized only on the training set. According to the learning curves, in the experiment, the training for pain, anxiety and QoL were stopped at 25, 24 and 22 epochs, respectively.

Receiver Operating Characteristic (ROC) and area under the curve (AUC) is an effective and comprehensive measure for assessing the inherent validity of a diagnostic test and the overall performance of the ROC curve. The AUC, sensitivity and specificity of predicting pain, anxiety, and QoL were shown in Figure 2d-f and Table 3. It demonstrated a satisfactory performance of the three ANNs in distinguishing whether patients undergoing discomfort.

The success rates of ANNs were calculated according to ROC curves. The overall success rate of ANN for pain prediction was 87.7%, and success rate of the training set, validation set, and test set are 87.9% (95% CI 83.6%-90.5%), 86.1% (95% CI 83.3%-91.7%), and 88.6% (95% CI 84.1%-93.2%), respectively. The total success rate of anxiety prediction was 93.4%, and success rate of the training set, validation set, and test set are 94.6% (95% CI 87.5%-97.3%), 94.9% (95% CI 89.7%-97.4%), and 88.9% (95% CI 80.0%-91.1%), respectively. The overall success rate of ANN for predicting QoL was 92.4%, and success rate of the training set, validation set, and test set are 91.9% (95% CI 83.7%-96.2%), 94.3% (95% CI 80.0%-97.1%), and 92.1% (95% CI 81.6%-97.4%), respectively. The accuracy on test set was consistent with those on the training and validation sets, indicating negligible over-fitting. Notably, the test set was not available during the learning process until the final evaluation of success rate. This demonstrated the constructed ANNs could prospectively predict the discomfort level for new patients with clinical features of treatment plan.

Predictors and their contributions to patient experience

Contributions of the 17 inputs to the output target were analyzed through partial derivatives method. The results of contribution of inputs in each ANN are illustrated in Table 4. The top three most important inputs were listed below for different prediction categories of patient experience.

The total contributions for 17 input features were ranked in order, which is shown in Figure 3. The number of teeth with lingual attachments was found to be the most important factor affecting outcome of negative experience, followed by the number of lingual buttons, and the number of upper incisors with attachments. Whether bonding attachments instantly had minimal impact to the overall patient experience. The treatment stage was a negligible feature to predict pain, had mild impact on anxiety, while had moderate impact on QoL. This means the treatment stage more or less impact the patient experience.

DISCUSSION

Patient experience of treatment is clinically important. Generally speaking, orthodontists might mainly consider outcome of treatment, but are not particularly clear about the impact of appliance designs on patients' comfort level. Patient experience and response to different designs of orthodontic appliance has been found to be complicated, and is not accurate to judge by experience. The relationship between different designs of aligners is nonlinear (e.g. extraction always requires elastics and precision cuts), which is hard to measure with traditional linear models with several parameters, such as correlation analysis and logistic regression. By use of AI, we can investigate the nonlinear relationships in a high-dimensional data set, and find the potential role of each feature on comfort level.

The AI system constructed for patient comfort prediction bears the potential to be applied in the clinic and enhance patients' compliance. If a high risk of discomfort is detected, orthodontists can avoid or delay some use of uncomfortable accessories as long as they do not affect treatment outcome. More importantly, comprehensive explanation of possible discomfort and timely follow-up are recommended, with medical advice such as replacing aligners less frequently (2/3-week per pair), less wearing time (12-hour per day), and arranging a little more appointments or telephone calls in the early stage of treatment. [29, 30] Let the patients have better preparation in advance, thus ensuring patient-specific orthodontic treatment.

To the best of our knowledge, the present study is the first constructed ANNs to predict patient experience based on the Invisalign treatment designs, and the prediction accuracies are comparable to those presented in the literature. For example, a convolutional neural network was incorporated into a 1-step, end-to-end diagnostic system to diagnose skeletal classification with lateral cephalograms, with prediction accuracy between 89% and 96%.[31] A 23-13-1 Back Propagation ANNs was constructed for deciding whether dental extractions are necessary prior to orthodontic treatment with a success rate at 80%, and identified two contributing indices that should be taken into consideration first.[15] A neural network machine learning was utilized for the diagnosis of extractions patterns with an accuracy of 84%.[17] In the present study, clinical performance of ANNs for pain, anxiety and QoL prediction was 87.7%, 93.4%, and 92.4%. This demonstrated a satisfactory performance of the ANNs used in the study. To further improve the accuracy, we could enlarge the sample size for training and evaluation, collect more diagnostic features which may be related to patient experience. And when the sample size is larger, more sophisticated model structure, e.g. more layers and nodes, can be constructed with better performance.

This study investigated the possible predictors of patient experience of Invisalign treatment, and found features with different contributions presented in Table 4. The number of teeth with lingual attachments and buttons were found to be the most important factors affecting outcome of negative experience. One explanation was that these lingual devices could aggravate the irritation of mucosal and tongue during Invisalign treatment. The same is true about the attachments on upper incisors, which would compromise dental esthetics during the treatment.[32] Age and crowding were found to influence the QoL, which were also shown to occur during fixed appliances treatment.[33, 34] Patient experience at different Invisalign treatment stages (i.e. the initial treatment or refinement) is poorly understood. In the present study, treatment stage was found to be a negligible feature to predict pain, and had mild impact on anxiety, while moderate impact on QoL. With the contribution list, orthodontists can pay close attention to patients with high-ranking designs and give more care in advance.

In the present study, we did not choose the gender as an input feature. The majority of included patients were female, but are also indeed in the real clinical situation, and we did not deliberately change the existed gender distribution in the recruitment process. There is no reported difference in QoL between male and females.[33] It is feasible to integrate male and female patients to analyze in many studies concerning expert system.[15, 17, 24]

There are some limitations of the study. Firstly, the dataset was relatively small to perform ANNs analysis, and scores fluctuated up and down around the baseline in a small range, so the training intensity of the model may be not very intense. However, the accuracy on test set was consistent with those on the training and validation sets, indicating negligible over-fitting. Furthermore, substantial efforts were made to enhance the prediction performance. the introduction of random dropout, cross validation, and Adam optimizer reduced over-fitting and increased the learning efficiency of the ANNs, thus compensating for the relatively small sample size in the study.[17] Another limitation is that, the patient self-reported comfort level of wearing clear aligners was relatively subjective and influenced by multiple factors. We defined the difference between the highest and lowest score as output, which reduced the impact of subjective variability. Besides, we will consider involving objective measures to evaluate patient experience, such as physical examination and laboratory-assisted examination. With the continuous accumulation of subsequent clinical data, the ANNs would be further stabilized and sophisticated.

CONCLUSIONS

The three constructed ANNs demonstrated good success rates for predicting pain (87.7%), anxiety (93.4%), and QoL (92.4%) during Invisalign treatment. The number of teeth with lingual attachments was found to be the most important factor affecting the outcome of negative experience, followed by the number of lingual buttons, and the number of upper incisors with attachments.

References

1. Schwendicke F, Samek W, Krois J. Artificial Intelligence in Dentistry: Chances and Challenges. J Dent Res 2020;99:769-74.
2. Shahmoradi L, Safdari R, Mirhosseini MM, Arji G, Jannat B, Abdar M. Predicting risk of acute appendicitis: A comparison of artificial neural network and logistic regression models. Acta Med Iran 2018;56:784-95.
3. Kok H, Acilar AM, Izgi MS. Usage and comparison of artificial intelligence algorithms for determination of growth and development by cervical vertebrae stages in orthodontics. Prog Orthod 2019;20:41.
4. Wang H, Minnema J, Batenburg KJ, Forouzanfar T, Hu FJ, Wu G. Multiclass CBCT Image Segmentation for Orthodontics with Deep Learning. J Dent Res 2021;100:943-9.
5. Djeu G, Shelton C, Maganzini A. Outcome assessment of Invisalign and traditional orthodontic treatment compared with the American Board of Orthodontics objective grading system. Am J Orthod Dentofacial Orthop 2005;128:292–8.
6. Cardoso PC, Espinosa DG, Mecenas P, Flores-Mir C, Normando D. Pain level between clear aligners and fixed appliances: a systematic review. Prog Orthod 2020;21:3.
7. Rossini G, Parrini S, Castroflorio T, Deregibus A, Debernardi CL. Periodontal health during clear aligners treatment: a systematic review. Eur J Orthod 2015;37:539-43.
8. Gao M, Yan X, Zhao R, Shan Y, Chen Y, Jian F, et al. Comparison of pain perception, anxiety, and impacts on oral health-related quality of life between patients receiving clear aligners and fixed appliances during the initial stage of orthodontic treatment. Eur J Orthod 2020.
9. Flores-Mir C, Brandelli J, Pacheco-Pereira C. Patient satisfaction and quality of life status after 2 treatment modalities: Invisalign and conventional fixed appliances. Am J Orthod Dentofacial Orthop 2018;154:639-44.
10. Nedwed V, Miethke RR. Motivation, acceptance and problems of invisalign patients. J Orofac Orthop 2005;66:162-73.
11. Allareddy V, Nalliah R, Lee MK, Rampa S, Allareddy V. Adverse clinical events reported during Invisalign treatment: Analysis of the MAUDE database. Am J Orthod Dentofacial Orthop 2017;152:706-10.
12. Aljudaibi S, Duane B. Non-pharmacological pain relief during orthodontic treatment. Evid Based Dent 2018;19:48-9.
13. Al-Moghrabi D, Salazar F, Pandis N, Fleming P. Compliance with removable orthodontic appliances and adjuncts: A systematic review and meta-analysis. Am J Orthod Dentofacial Orthop 2017;152:17-32.
14. Charavet C, Le Gall M, Albert A, Bruwier A, Leroy S. Patient compliance and orthodontic treatment efficacy of Planas functional appliances with TheraMon microsensors. Angle Orthod 2019;89:117-22.
15. 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.
16. 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.
17. Jung SK, Kim TW. New approach for the diagnosis of extractions with neural network machine learning. Am J Orthod Dentofacial Orthop 2016;149:127-33.
18. He SL, Wang JH. Reliability and validity of a Chinese version of the Oral Health Impact Profile for

edentulous subjects. Qual Life Res 2015;24:1011-6.
19. Zhang Y, Liu R, Li G, Mao S, Yuan Y. The reliability and validity of a Chinese-version Short Health Anxiety Inventory: an investigation of university students. Neuropsychiatr Dis Treat 2015;11:1739-47.
20. Katz J, Melzack R. Measurement of pain. Surg Clin North Am 1999;79:231-52.
21. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. Preprint at https://arxivorg/abs/14091556 2014.
22. Wanto A, Windarto AP, Hartama D, Parlina I. Use of binary sigmoid function and linear identity in artificial neural networks for forecasting population density. IJISTECH 2017;1:43-54.
23. Kingma DP, Ba J. Adam: A method for stochastic optimization. Preprint at https://arxivorg/abs/14126980 2014.
24. Li P, Kong D, Tang T, Su D, Yang P, Wang H, et al. Orthodontic Treatment Planning based on Artificial Neural Networks. Sci Rep 2019;9:2037.
25. Guyon I, editor A Scaling Law for the Validation-Set Training-Set Size Ratio1997.
26. Sug H. The effect of training set size for the performance of neural networks of classification. WSEAS Trans Comput 2010;9:1297-306.
27. Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32-5.
28. Gevrey M, Dimopoulos L, Lek S. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol Model 2003;160:249-64.
29. Keith DJ, Rinchuse DJ, Kennedy M, Zullo T. Effect of text message follow-up on patient's self-reported level of pain and anxiety. Angle Orthod 2013;83:605-10.
30. Bartlett BW, Firestone AR, Vig KW, Beck FM, Marucha PT. The influence of a structured telephone call on orthodontic pain and anxiety. Am J Orthod Dentofacial Orthop 2005;128:435-41.
31. Yu HJ, Cho SR, Kim MJ, Kim WH, Kim JW, Choi J. Automated Skeletal Classification with Lateral Cephalometry Based on Artificial Intelligence. J Dent Res 2020;99:249-56.
32. Thai JK, Araujo E, McCray J, Schneider PP, Kim KB. Esthetic perception of clear aligner therapy attachments using eye-tracking technology. Am J Orthod Dentofacial Orthop 2020;158:400-9.
33. Masood Y, Masood M, Zainul NN, Araby NB, Hussain SF, Newton T. Impact of malocclusion on oral health related quality of life in young people. Health Qual Life Outcomes 2013;11:25.
34. Wang J, Tang X, Shen Y, Shang G, Fang L, Wang R, et al. The correlations between health-related quality of life changes and pain and anxiety in orthodontic patients in the initial stage of treatment. Biomed Res Int 2015;2015:725913.

Figure legends

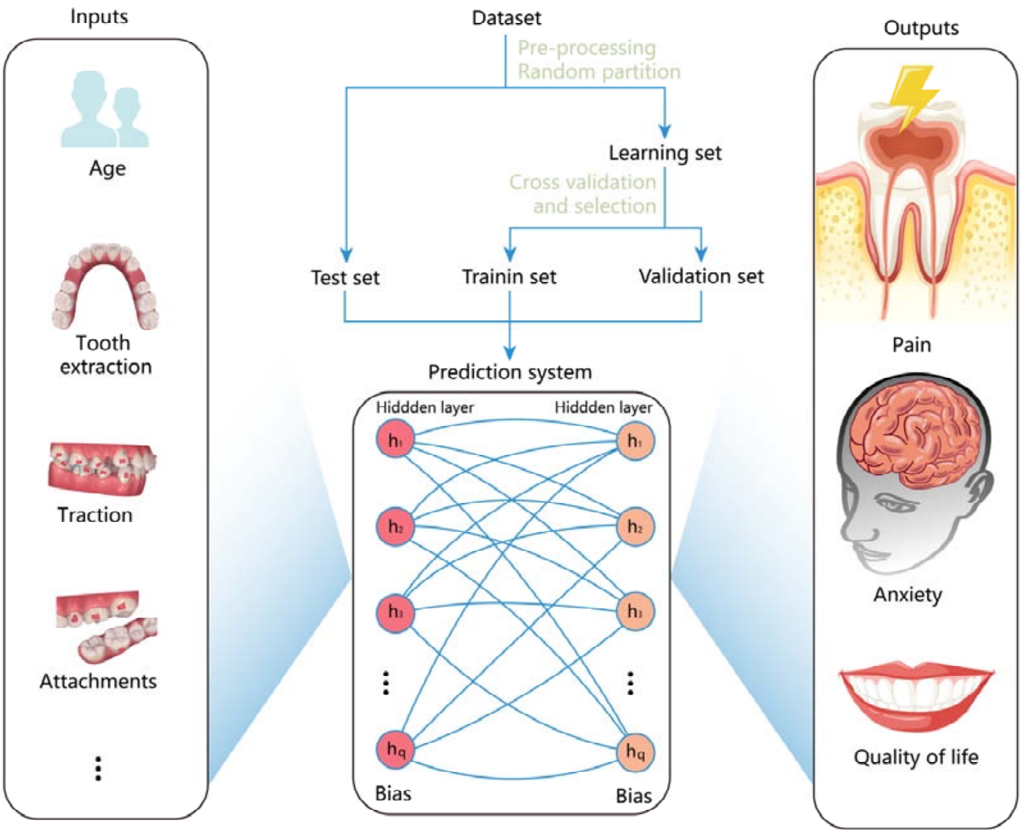

Figure 1. Flow diagram of the analysis process of artificial neural networks. Three artificial neural networks were fully connected, and included two hidden layers with hidden size of nine.

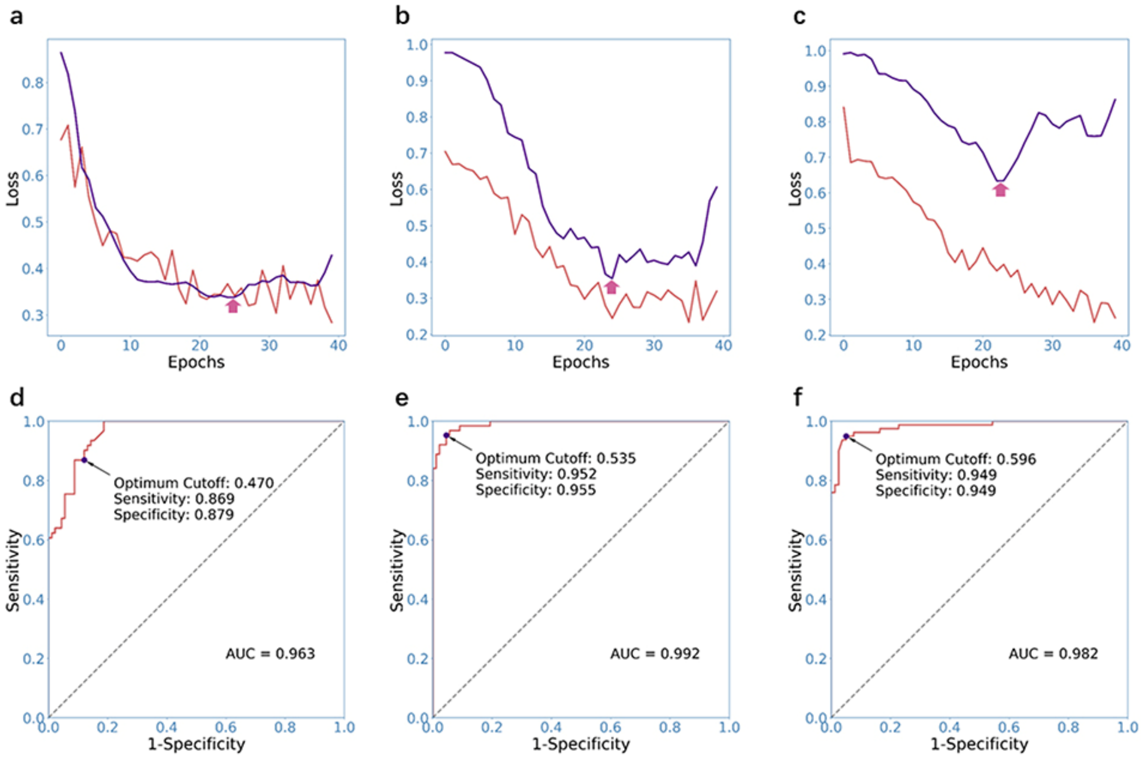

Figure 2. Prediction performance of the artificial neural networks (ANNs). a, b and c indicate the learning curves of ANNs for pain, anxiety, and quality of life, respectively. Red lines represent train loss curve; purple lines represent validation loss curve. Arrows indicate the lowest point of validation loss curve, which means the training procedure for pain, anxiety and quality of life were stopped at 25, 24 and 22 epochs, respectively. **d, e and f** indicate the ROC curves of ANNs for pain, anxiety, and quality of life, respectively. The optimum diagnostic cutoff value is marked as purple points, where the sensitivity and specificity are shown upon the arrows. ANNs, artificial neural networks; ROC, Receiver Operating Characteristic; AUC, area under the curve.

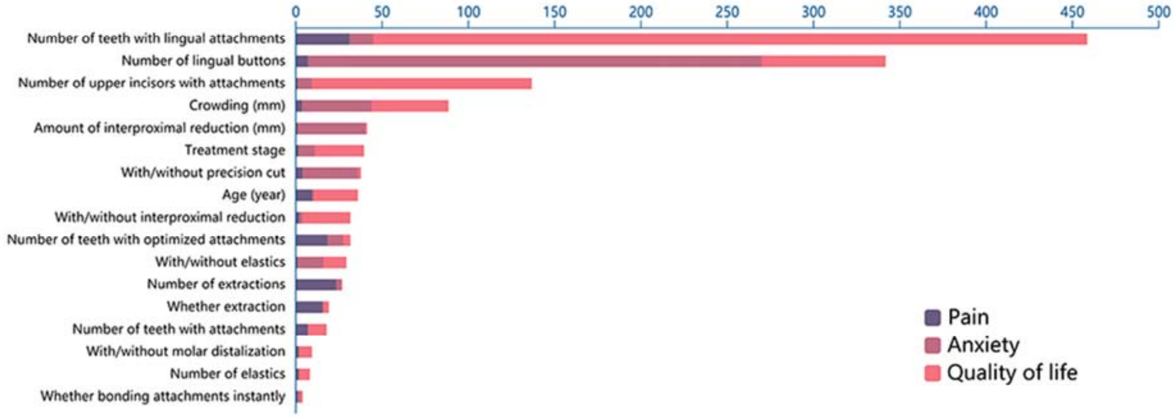

Figure 3. Total contribution of the 17 input features in descending order.

Table legends

Table 1. Description of input normalization in three artificial neural networks.

Categories	**Data Type**		**Criterion**
Age (year)	Integral		MMN
Treatment stage	Binary	First-time	0
		Refinement	1
Crowding (mm)	Discrete	No	0
		I°	1/3
		II°	2/3
		III°	1
Whether extraction	Binary	Yes	1
		No	0
Number of extractions	Integral		MMN
Whether bonding attachments instantly	Binary	Yes	1
		No	0
With/without molar distalization	Binary	Yes	1
		No	0
With/without elastics	Binary	Yes	1
		No	0
Number of elastics	Integral		MMN
With/without interproximal reduction	Binary	Yes	1
		No	0
Amount of interproximal reduction (mm)	Continuous		MMN
Number of teeth with attachments	Integral		MMN
Number of teeth with optimized attachments	Integral		MMN
Number of teeth with lingual attachments	Integral		MMN
Number of upper incisors with attachments	Integral		MMN
Number of lingual buttons	Integral		MMN
With/without precision cut	Binary	Yes	1
		No	0

MMN, maximum minimum normalization

Table 2. Number of patients with positive and negative tags distributed in training, validation and test set in three artificial neural networks.

Dataset	Pain			Anxiety			Quality of Life		
	Positive	Negative	Total	Positive	Negative	Total	Positive	Negative	Total
Training set	45	71	116	48	64	112	62	61	123
Validation set	16	20	36	16	15	39	17	18	35
Test set	16	28	44	25	20	45	18	28	38
Total set	77	119	196	83	113	196	97	99	196

Changes were calculated as the difference between the highest score and the lowest score. Higher values indicated more negative patient experience of Invisalign treatment. Then they were all binarized using pre-defined thresholds to distinguish positive and negative samples, with 3.0 for pain, 6.5 for anxiety, and 7.0 for quality of life. The binarized values were the final prediction targets.

Table 3. Performance of artificial neural networks (median, 95% CI) for patient experience.

Performance	Pain	Anxiety	Quality of Life
AUC	0.963(0.904, 0.972)	0.992(0.983, 0.995)	0.982(0.950, 0.990)
Sensitivity	0.885 (0.803-0.984)	0.952 (0.921-0.968)	0.937 (0.899-0.975)
Specificity	0.890 (0.813-0.934)	0.955 (0.920-0.977)	0.937 (0.873-0.962)

AUC, area under the curve.

Table 4. Contributions of the 17 inputs for prediction target.

Input Categories	Contribution (median, 95% CI)		
	Pain	Anxiety	Quality of Life
Number of teeth with lingual attachments	30.495 (3.164, 64.621)	13.557 (3.018, 33.124)	414.976 (190.285, 684.003)
Number of lingual buttons	6.139 (1.226, 12.541)	263.655 (145.051, 419.175)	71.548 (28.348, 131.841)
Number of upper incisors with attachments	0.462 (0.068, 4.223)	8.710 (3.520, 15.665)	127.323 (53.239, 223.390)
Crowding (mm)	0.173 (0.004, 1.253)	40.256 (20.196, 66.327)	44.515 (11.818, 84.555)
Amount of interproximal reduction (mm)	0.670 (0.113, 2.692)	39.468 (16.395, 64.707)	0.942 (0.080, 4.211)
Treatment stage	1.451 (0.216, 5.189)	9.740 (5.509, 16.898)	28.250 (13.301, 50.561)
With/without precision cut	3.495 (0.615, 9.066)	32.480 (16.052, 51.128)	1.266 (0.084, 5.304)
Age (year)	9.140 (1.622, 19.141)	0.904 (0.078, 3.351)	25.386 (10.694, 42.722)
With/without interproximal reduction	1.680 (0.300, 7.595)	2.0378 (0.161, 5.573)	27.993 (9.141, 69.337)
Number of teeth with optimized attachments	17.884 (5.501, 42.235)	9.216 (5.308, 17.528)	4.456 (0.872, 12.977)
With/without elastics	0.937 (0.160, 3.142)	14.724 (7.662, 23.509)	13.357 (5.600, 27.314)
Number of extractions	23.077 (7.160, 48.523)	2.976 (0.518, 8.308)	0.827 (0.084, 4.143)
Whether extraction	15.276 (5.932, 38.595)	0.204 (0.020, 1.557)	3.132 (0.441, 8.072)
Number of teeth with attachments	6.380 (1.382, 16.947)	0.544 (0.035, 2.550)	10.554 (3.485, 24.553)
With/without molar distalization	0.962 (0.355, 3.817)	0.925 (0.158, 3.657)	7.119 (1.743, 15.734)
Number of elastics	1.334 (0.198, 5.679)	0.224 (0.039, 1.006)	6.383 (1.064, 20.720)
Whether bonding attachments instantly	0.594 (0.063, 2.863)	2.012 (0.418, 5.371)	0.996 (0.054, 4.459)