

Predicting cessation of orthodontic treatments using a classification-based approach

Abstract

In recent years, dental care has received increasing attention from people across the globe. With growing living conditions, people are more aware of preventable conditions that might be avoided. Malocclusion is one among the most studied problems in orthodontics. The statistical predictive model building plays a vital role in dentistry particularly, for clinical decision making. Developing a model for predicting the factors affecting for discontinuation of treatment is a vital step in assessing the therapeutic effect of treatment, resource management and cost reduction in the healthcare industry. Logistic regression and Probit regression models are considered as a successful widely used approach to analyze a classification problem with factor predictor variables. In this study, Naïve Bayes classifier and random forest classification models are introduced to predict discontinuation of orthodontic treatments of dental patients. Based on this study the duration of active treatment was the most significant factor affecting the discontinuation of the treatment. When comparing the four approaches, random forest classifier showed the highest accuracy and specificity, while Naïve Bayes model indicated the highest sensitivity on the prediction of discontinuation of the treatment. Besides, the classification-based approach with modern predictive algorithms shows a robust result for orthodontic data.

Keywords: Dental malocclusion, classification, logistic Regression, probit Models, naïve bayes, random forests

Volume 9 Issue 2 - 2020

R.A.I.H. Dharmasena,¹ Lakshika S. Nawarathna,² V.S.N.Vithanaarachchi³

¹Department of Statistics, University of Manitoba Winnipeg, Canada

²Department of Statistics and Computer Science, University of Peradeniya, Sri Lanka

³Division of Orthodontics, Faculty of Dental Sciences, University of Peradeniya, Sri Lanka

Correspondence: Lakshika S. Nawarathna, Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka, Email lakshikas@pdn.ac.lk

Received: February 10, 2020 | **Published:** April 30, 2020

Introduction

Malocclusion of the teeth is a misalignment condition where teeth deviate from ideal occlusion that can cause serious aesthetic issues and oral health complications. The teeth will not be able to perform important functions when they are misaligned. Malocclusions are mainly resulted due to environmental and genetic factors. It can be inherited in nature which means, it can be passed down from one generation to the next. But this can cause some oral habits too.¹ Specially thumb or finger sucking, pacifier use for a longer period and mouth breathing are most common oral habits that can cause malocclusion. Sports injuries, automobile and fall accidents can also lead to this.²

Malocclusion is neither a sickness nor a life-threatening condition and usually is not serious enough to require treatment. But there has been a considerable demand for orthodontic care.^{3,4} It is usually diagnosed through routine dental examination. In a child's life, the period of eruption in permanent teeth must be considered critical.^{5,6} Depending on the classification of malocclusion, the symptoms of the disorder may be subtle or severe. Moreover, the treatment of malocclusion places a considerable burden on health care resources nationally and globally, significantly when treatments are funded by public means.⁷ Malocclusions are one amongst the most studied problems in orthodontics, using completely different classifications in several populations, usually to find out about its prevalence, causes and establishing treatment procedures.⁸ The selection from potential alternatives treatments should ideally be based on well-known effective treatments, rather than be dependent on visible clinical impression.

Depending on the type of malocclusion, the orthodontists recommend various treatments. These can include applying braces, wires or plates to correct the position of the teeth, enhance the jaw growth with functional orthopaedic devices and to stabilize the

jawbone with surgical procedures. To evaluate the effectiveness of the treatment, it is necessary to use both valid and reliable measures of results.⁹ The treatment of this condition in children and adults usually ends up in correcting the misalignment and early treatment is cost effective and it reduces the duration of the treatment.¹⁰

Statistical methodologies and applications play a major role in dentistry and dental research mainly in evidence-based dentistry. Clinical trials, designing experiments on treatments end up with data which are needed to be analyzed properly to get the most use out of it. Statistic based approach is the most reliable and widely used method to interpret information gained by clinical data.¹¹ Statistical predictive model building is a common application of statistics to dentistry mainly for clinical decision making.¹² Logistic regression and Probit models are some of the most widely used predictive models in bioinformatics for decision making.^{13,14} With the advancement of development of computational power within the last decades, evolutionary search algorithms and unsupervised learning algorithms emerged as important heuristic optimization techniques for decision making.¹⁵ These studies have vital importance when addressing the therapeutic goals in the completion of orthodontic treatment. On recent studies, application of unsupervised methods like Naïve Bayes models and random forest models in bioinformatics are not rare.

The objective of this study is to predict the continuation or discontinuation of orthodontic treatment for dental malocclusion by identifying the factors affecting the decision of discontinuing the treatment. Moreover, we identify the most suitable predictive model to address this scenario using several different learning algorithms by comparing the accuracies of classical approaches with the Naïve Bayes models and random forest models.

This article is organized as follows. In Section 2, we discuss the statistical theory behind the two data mining algorithms and the conventional models used in this research together with the model

reduction techniques under the materials section. Next, in Section 3 we illustrate the methodology by analyzing the clinical records obtained from the Division of Orthodontics, University Dental Hospital, Peradeniya, Sri Lanka. To sum up in Section 4, the article is concluded with a discussion. The statistical software R and Waikato Environment for Knowledge Analysis (Weka) were used for all the statistical computations in this article.

Materials and Methods

To build a predictive model for discontinuation of orthodontic

treatments of dental patients, a clinical dataset was used from the clinical records obtained from Division of Orthodontics, University Dental Hospital, Peradeniya, Sri Lanka. This dataset consisted of 310 records of clinical treatments for dental malocclusion. The variable discontinuation of orthodontic treatment was considered as the dependent variable. People treated more than 5 years were diagnosed as those who continue the treatment. Further, 12.903% of patients were diagnosed as to continue the treatment while 87.096% patients were diagnosed to discontinue the orthodontic treatments. There were no missing data in the dataset and all the variables were recorded to a common Likert scale as illustrated in Table 1.

Table 1 Likert Scale recoding of variables used in the analysis

Variable	Likert Scale					
	1	2	3	4	5	6
Discontinuation of treatment (Y)	Discontinue	Continue				
Age (X ₁)	1 – 10	11 – 20	21 – 30	31 – 40	41 – 50	
Gender (X ₂)	Male	Female				
Type of malocclusion (X ₃)	Class I	Class II Division I	Class II Division 2	Class III		
Severity of malocclusion (X ₄)	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	
Treatment indicated (X ₅)	Non-extraction	Extraction deciduous tooth	Extraction permanent tooth			
Simple removable appliance (X ₆)	No	Yes				
Fixed appliance (X ₇)	No	Single arch	Both arches			
Growth modification appliance (X ₈)	No	Twin block	Head gear	Face mask	Other	
Cost of treatment in LKR (X ₉)	No	200-400	400-1000	1100-3500	3600-7500	Above 7500
Stage of treatment at cessation (X ₁₀)	Record taking	Treatment planning	Appliance fitting	Review visits	End of active treatment	Retention phase
Duration of active treatment (X ₁₁)	< 6 months	6 – 12 months	1 – 2 years	2 – 5 years	> 5 years	

Actual clinical data were used to build up several predictive models using different learning algorithms namely Naïve Bayes, Random Forest, Logistic Regression and Probit model and the accuracy and reliability of each model were compared.

Prediction model

In this study, two data mining algorithms, Naïve Bayes and Random Forest were introduced beside the most generally used statistical methods Logistic regression and Probit model¹⁶ to develop the prediction models for predicting cessation of orthodontic treatments.

Naïve Bayes classifier: Naïve Bayesian is a specialized form of the Bayesian network which is a simple probabilistic classifier based on Bayesian theory. All Naïve Bayes classifiers assumed that the predictive variables are conditionally independent given the class and no hidden or latent attributes influence the prediction method.¹⁷

Let $X = (x_1, \dots, x_n)$ where $n = (1, 2, \dots, 11)$; a vector representing 11 features (independent variables) which assigns to $p(C_k | x_1, \dots, x_n)$ instance probabilities for each

$$p(C_k | x_1, \dots, x_n) = \frac{p(C_k)p(X|C_k)}{p(X)} \tag{1}$$

of k possible outcomes ('0' or '1') or classes C_k . Using Bayes' Theorem, now the joint probability model can be defined as,

$$p(C_k, x_1, \dots, x_n) = p(x_1, \dots, x_n, C_k) = p(x_1 | x_2, \dots, x_n, C_k) p(x_2 | x_3, \dots, x_n, C_k) \dots p(x_{n-1} | x_n, C_k) p(x_n | C_k) p(C_k) \tag{2}$$

With respect to the independence assumption of the Naïve Bayes classifier, the conditional distribution over the class variable C is,

$$p(C_k | x_1, \dots, x_n) = \frac{1}{Z} p(C_k) \prod_{i=1}^n p(x_i | C_k) \tag{3}$$

where the evidence $Z = p(X)$ is a scaling factor dependent only on x_1, \dots, x_n . Therefore, the Naïve Bayes classifier is the function that assigns a class label $\hat{y} = C_k$ as follows.

$$\hat{y} = \underset{k \in \{1, \dots, k\}}{\operatorname{argmax}} p(C_k) \prod_{i=1}^n p(x_i | C_k) \tag{4}$$

Random Forest classifier: Unlike single classification trees, random forest creates many classification trees which classify a new object from an input vector by inserting to all trees and select the trees which classify the best out of the trees in the forest.¹⁸ Random forest classifier does not overfit although the number of trees is increased, and it creates the model fast with large databases without changing or deleting variables. When dealing with random forest classifiers, it is not needed to cross-validate data or uses a separate test to get an unbiased estimate of the prediction since test set errors are calculated internally on the run of the random forest classifier on a dataset.¹⁹ For the dataset, a random forest with a maximum of 2000 trees were created and measured the classification accuracy.

Logistic Regression models: Logistic regression is used especially in the case that the model contains a binary categorical dependent variable, that is the output can take only two values, '0' or '1'. Here, the dependent variable of the predictive model is a disconnection of the treatment (Y) which has only two outcomes namely 'Yes' or 'No' and is categorical which enable the opportunity to employ a logistic regression model.²⁰ The general logistic function $\sigma(t)$ where $t = \beta_0 + \sum_{i=1}^{11} \beta_i x_i$ can be defined as,

$$\sigma(t) = \frac{1}{1 + e^{-t}} \tag{5}$$

Then the proposed logistic regression model is defined as,

$$y = \operatorname{logit}(p) = \log \frac{p}{1-p} = \beta_0 + \sum_{i=1}^{11} \beta_i x_i \tag{6}$$

where p is the probability of the dependent variable equaling a "success" and β_1, \dots, β_n be the regression coefficients.

Probit models: As in logistic regression models, Probit models are also used when the dependent variable is dichotomous. It employs a Probit link function which mostly estimated using the maximum likelihood procedure. Assuming the dependent variable (Y) is binary, with a vector of X variables which influences Y . Then the model takes the form,

$$p(Y = 1 | X) = \phi(X^T \beta) \tag{7}$$

where ϕ is the Cumulative Distribution Function(CDF) of the standard normal distribution. The parameters β 's are typically estimated by maximum likelihood estimator.¹³ The proposed Probit model is as follows.

$$p(Y = 1 | X) = \phi\left(\beta_0 + \sum_{i=1}^{11} \beta_i x_i\right)$$

Model Reduction: To get the optimum model for the logistic regression model, model reduction using backward elimination and bidirectional elimination were used to fit the model. Elimination was done based on the Akaike Information Criterion(AIC) and Bayesian Information Criterion(BIC) values which are estimators of the relative quality of statistical models. Models end up with minimum relative AIC and BIC values are considered to be the best model in model reductions. Additionally, adjusted R-squared (R^2) values were also obtained to compare the model performance in reduced models.²¹

Estimation for model performance

10-Fold Cross Validation method: k-Fold Cross Validation is a model validation technique which partition the dataset into k equal partitions and keep one partition for testing and use the rest of ($k-1$) partitions to train the model as the testing set. This is done k times (number of folds) and the average of the estimates are taken as the final estimation.²² In this study, 10-fold cross-validation (i.e., $k = 10$) was used to validate logistic regression, Probit models and Naïve Bayes classifier.

Confusion matrix: Confusion matrix or error matrix is often used in statistical modelling to evaluate and visualize the model performance. As shown in Table 2, it is a two by two matrix which can be used to obtain sensitivity, specificity and accuracy of predictive model classifications.¹⁶

Table 2 Confusion matrix to validate the results

Discontinuation of orthodontic treatment (Y)		Actual class	
		Discontinue ('No')	Discontinue ('Yes')
Predicted class	Discontinue ('No')	True Negatives (TN)	False Positive (FP)
	Discontinue ('Yes')	False Negatives (FN)	True Positive (TP)

True Positive (TP) is the number of dental patients who were predicted to continue the treatment and they do continue the dental treatments. True Negative (TN) means the number of dental patients who were predicted to discontinue the treatments are actually has been discontinued the treatment. False Positive (FP) can be taken as the numbers of dental patients predicted to continue the treatment are needed to discontinue the treatment. False Negative (FN) gives the number of dental patients predicted to discontinue the treatment actually need to continue the treatment. The sensitivity, specificity and accuracy were calculated by TP, TN, FP and FN values by the confusion matrix. Sensitivity or Recall is the probability that the model can correctly predict the discontinuation of a dental patient.

$$Sensitivity = \frac{TP}{TP + FN} \tag{9}$$

Specificity can be taken as the probability that the model can correctly predict discontinuation of orthodontic treatment.

$$Specificity = \frac{TN}{FP + TN} \tag{10}$$

Accuracy is the probability that the model can correctly predict continuation of discontinuation of orthodontic treatments of a dental patient.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

Receiver Operating Characteristic (ROC) Curve: The Receiver Operating Characteristic (ROC) curve is a popular method to evaluate model performance. It is based on sensitivity and specificity where the x-axis is 1-specificity (False Positive Rate) and the y-axis is sensitivity (True Positive Rate) of a given model. Area Under Curve(AUC) is a measure that can be obtained to interpret the ROC plots easily. It is the area under ROC curve of a model. The value of AUC ranges from 0-1 where 1 is the perfect fit or the perfect classifier. This method is convenient to compare the performance of multiple models.²³

Results and discussion

We present details on descriptive analysis, model fitting, model validation, and a detailed discussion on the factors affecting discontinuation of the orthodontic treatments. Three hundred and ten patient’s records were analyzed, and their age range was 7-30 years.

Table 3 shows the estimated partial regression coefficients corresponding to each explanatory variable mentioned in the Table 1, the standard errors, the z-statistics and the p-values used for testing the significance of each coefficient. Moreover, the summary of coefficients of reduced logistic model which is obtained by backward elimination and bidirectional elimination is shown in Table 4.

Table 3 Model coefficients of logistic regression full model

Variable	Estimate	Std. Error	z value	p-value
(Intercept)	40.40	8520	0.01	0.996
Female	0.83	0.65	1.27	0.205
Age (11 – 20)	0	0.79	0	0.997
Age (21 – 30)	-3.30	1.68	-1.96	0.050
Age (31 – 40)	-0.38	17900	0	1
Age (41 – 50)	19.20	17700	0	0.999
Type of malocclusion (Class I I-Division I)	1.17	0.83	1.41	0.159
Type of malocclusion (Class I I-Division2)	0.84	1.10	0.77	0.443
Type of malocclusion (Class I I I)	-0.13	0.80	-0.17	0.868
Severity of malocclusion (Grade 2)	-16.20	8100	0	0.998
Severity of malocclusion (Grade 3)	-20.50	8100	0	0.998
Severity of malocclusion (Grade 4)	-20.60	8100	0	0.998
Severity of malocclusion (Grade 5)	-19.00	8100	0	0.998

Table Continued

Variable	Estimate	Std. Error	z value	p-value
Simple removable appliance (Yes)	2.62	1.90	1.38	0.167
Fixed appliance (Single arch)	12.90	7460	0	0.999
Fixed appliance (Both arches)	-0.29	1.87	-0.15	0.878
Growth modification appliance (Twin block)	1.00	1.88	0.53	0.597
Growth modification appliance (Head gear)	-4.45	6.77	-0.66	0.511
Growth modification appliance (Face mask)	1.53	2.05	0.75	0.454
Growth modification appliance (Other)	-4.50	3.52	-1.28	0.201
Stage of treatment at cessation (Treatment planning)	-23.20	2660	-0.01	0.993
Stage of treatment at cessation (Appliance fitting)	-4.63	3990	0	0.999
Stage of treatment at cessation (Review visits)	-23.90	2660	-0.01	0.993
Stage of treatment at cessation (End of active treatment)	-25.40	2660	-0.01	0.992
Stage of treatment at cessation (Retention phase)	-25.90	2660	-0.01	0.992
Treatment indicated (Extraction deciduous tooth)	-2.08	0.92	-2.27	0.023
Treatment indicated (Extraction permanent tooth)	0.21	0.91	0.23	0.819
Cost of treatment in LKR (200-400)	3.72	3.33	1.12	0.264
Cost of treatment in LKR (400-1000)	5.88	3.59	1.64	0.101
Cost of treatment in LKR (1100-3500)	6.95	3.67	1.90	0.058
Cost of treatment in LKR (3600-7500)	4.81	3.72	1.29	0.196
Cost of treatment in LKR (Above 7500)	6.77	3.76	1.80	0.072
Duration of active treatment (6 – 12 months)	2.94	1.54	1.91	0.056
Duration of active treatment (1 – 2 years)	1.44	1.20	1.20	0.232
Duration of active treatment (2 – 5 years)	0.44	1.14	0.39	0.696
Duration of active treatment (> 5 years)	-4.55	1.30	-3.49	0.001

Table 4 Model coefficient information of logistic regression reduced model

Variable	Estimate	Std. Error	z value	p-value
(Intercept)	19.44	1818.23	0.01	0.992
Simple removable appliance (Yes)	1.29	0.61	2.11	0.035
Growth modification appliance (Twin block)	1.99	0.80	2.48	0.013
Growth modification appliance (Head gear)	-1.98	3.16	-0.63	0.531
Growth modification appliance (Face mask)	-0.25	1.46	-0.17	0.863
Growth modification appliance (Other)	-2.74	1.62	-1.70	0.090
Stage of treatment at cessation (Treatment planning)	-17.62	1818.23	-0.01	0.992
Stage of treatment at cessation (Appliance fitting)	-0.12	2688.89	0.00	1.000
Stage of treatment at cessation (Review visits)	-17.61	1818.23	-0.01	0.992
Stage of treatment at cessation (End of active treatment)	-19.30	1818.23	-0.01	0.992

Table Continued

Variable	Estimate	Std. Error	z value	p-value
Stage of treatment at cessation (Retention phase)	-20.18	1818.23	-0.01	0.991
Duration of active treatment (6– 12 months)	2.50	1.38	1.81	0.071
Duration of active treatment(1 –2 years)	1.71	1.21	1.42	0.157
Duration of active treatment (2– 5 years)	0.85	1.05	0.81	0.421
Duration of active treatment (>5 years)	-2.70	1.02	-2.66	0.008

The AIC, BIC and adjusted R-squared values for both full and reduced logistic regression models are shown in Table 5, where the reduced model had the minimum AIC, BIC and adjusted R-squared values than the full model with all predictor variables. Therefore, the reduced model is selected as the better model when compared with the full model.

Table 5 AIC, BIC and R-Squared values for the logistic regression full model and reduced model

Model	AIC	BIC	R-Squared
Full	172.7617	307.2783	0.6682
Reduced	152.6485	208.6971	0.5808

In both full model and reduced models, the variable duration of active treatment more than 5 years was significant at the level of 2-sided alpha 0.05, implying that high rates of discontinuation of treatment with long-term duration of active treatments. Treatment indicated as extraction of deciduous tooth was also significant for the full model while duration of active treatment in between 6 to 12 months was also significant for the reduced model at the confidence level of 95%. Moreover, patients who were treated with simple removable appliances were more prone to discontinue the treatment.

Table 6 shows the 10-Fold cross-validation results of the fitted models. Hence, From the results of model validation, we conclude that the the full model has a high predictive ability compared to the reduced models. Further, Random Forest classifier indicated the highest accuracy and specificity. Moreover, Naïve Bayes classifier had the highest sensitivity even though it indicated the lowest accuracy. Probit model classifier indicated the highest Area Under Curve (AUC) while Naïve Bayes classifier showed the least.

Table 6 Sensitivity, Specificity, Accuracy and AUC value of each classifier

Classifier	Sensitivity	Specificity	Accuracy	AUC
Logistic Regression	70.00%	98.148%	94.52%	95.63%
Reduced Logistic Regression	60.00%	97.407%	92.58%	93.30%
Probit Model	70.00%	97.778%	94.19%	95.70%
Reduced Probit Model	57.50%	97.037%	91.94%	93.94%
Naïve Bayes	93.704%	55.00%	88.70%	83.27%
Random Forest	92.50%	99.63%	98.71%	88.38%

Conclusion

Based on the results of this study it can be concluded that the duration of the treatment for dental malocclusion is the main factor significant when predicting the discontinuation of the orthodontic treatment while the factor treatment indicated had a slight effect on the results. The random forest model showed the highest accuracy and highest specificity while the Naïve Bayes model indicated the highest sensitivity on the prediction of discontinuation of the treatment. Thus, the classification-based approach with modern predictive algorithms shows a robust result for orthodontic data.

Acknowledgment

Authors wish to acknowledge the support from the Faculty of Dental Sciences, University of Peradeniya for data collection and granting us access to utilize data for statistical modelling.

Conflicts of interest

The authors declare that they have no conflict of interest.

Funding

There is no funding source.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent

Informed consent was obtained from all individual participants included in the study.

References

- Zhang M, McGrath C, Hagg, U. The impact of malocclusion and its treatment on quality of life: a literature review. *Int J Paediatr Dent.* 2006;16(6):381–387.
- Reid DA, Price AHK. Digital deformities and dental malocclusion due to finger sucking. *British journal of plastic surgery.* 1984;37(4):445–452.
- Jenny J. A social perspective on need and demand for orthodontic treatment. *Int Dent J.* 1975; 25(4):248–256.
- Mohlin B, al-Saadi E, Andrup L, et al. Orthodontics in 12-year old children. Demand, treatment motivating factors and treatment decisions. *Swed Dent J.* 2002;26(2):89–98.
- Vithanaarachchi VS, Nawarathne LS, Wijeyeweera RL. Eruption times and patterns of permanent teeth in Sri Lankan school children in Western province. *Sri Lanka Dent J.* 2018;48(01):25–31.

6. Vithanaarachchi SN, Nawarathna LS. Prevalence of anterior cross bite in preadolescent orthodontic patients attending an orthodontic clinic. *Ceylon Medical Journal*. 2017;62(3):189–192.
7. Petersen PE, Bourgeois D, Ogawa H, et al. The global burden of oral diseases and risks to oral health. *Bull World Health Organ*. 2005;83(9):661–669.
8. Ísper GarbinI AJ, Pereira Perin PC, Saliba Garbin CA, et al. Malocclusion prevalence and comparison between the Angle classification and the Dental Aesthetic Index in scholars in the interior of Sao Paulo state-Brazil. *Dental Press Journal of Orthodontics*. 2010; 15(4):94–102.
9. DeGuzman L, Bahiraei D, Vig KWL, et al. The validation of the Peer Assessment Rating index for malocclusion severity and treatment difficulty. *American Journal of Orthodontics and Dentofacial Orthopedics*. 1995;107(2):172–176.
10. Vithanaarachchi VSN, Nagarathne SPNP, Jayawardena C, et al. Assessment of factors associated with patient's compliance in orthodontic treatment. *Sri Lanka Dental Journal* 2017;47(1):1–12.
11. Shintani A. Primer of statistics in dental research: Part II. *Journal of prosthodontic research* 2014;58(2):85–91.
12. Lin C, Xu L, Chen YX, et al. A statistical model for predicting the retrieval rate of separated instruments and clinical decision-making. *Journal of Dental Sciences*. 2015;10(4):423–430.
13. Bliss CI. The method of probits—a correction. *Sciences*. 1934;79(2053):409–410.
14. Gaddum JH. *Reports on Biological Standards: Methods of Biological Assay Depending on a Quantal Response*. SHM Stationery Office, 1933.
15. Pirooznia M, Yang JY, Yang MQ, et al. A comparative study of different machine learning methods on microarray gene expression data. *BMC Genomics*. 2008;9(1):S13.
16. Endo AS, Takeo and Tanaka H. Comparison of Seven Algorithms to Predict Breast Cancer Survival (i Special Issue; Contribution to 21 Century Intelligent Technologies and Bioinformatics. *International Journal of Biomedical Soft Computing and Human Sciences: the official journal of the Biomedical Fuzzy Systems Association*. 2008;13(2):11–16.
17. John GH, Langley P. Estimating continuous distributions in Bayesian classifiers. *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*. 1995;338–345.
18. Liaw A, Wiener M. Classification and regression by RandomForest. *R news*. 2002;2(3):18–22.
19. Guo L, Ma Y, Cukic B, et al. Robust prediction of fault-proneness by random forests. *Proceedings of the 15th International Symposium on Software Reliability Engineering*. 2004;417–428.
20. Pandit PV, Javali S. Multiple logistic regression model to predict risk factors of oral health diseases. *Romanian Statistical Review*. 2012;5:1–14.
21. Afroughi S, Faghihzadeh S, Khaledi MJ, et al. Dental caries analysis in 3–5 years old children: A spatial modelling. *Arch Oral Biol*. 2010;55(5):374–378.
22. Vanschoren J, Van Rijn JN, Bischl B, et al. OpenML: networked science in machine learning. *ACM SIGKDD Explorations Newsletter*. 2014;15(2):49–60.
23. Beck JR, Shultz EK. The use of relative operating characteristic (ROC) curves in test performance evaluation. *Arch Pathol Lab Med*. 1986;110(1):13–20.