SCREENING MODELS FOR PERIODONTITIS

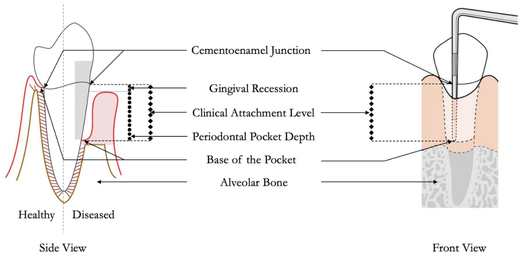
Htun Teza [a], Anuchate Pattanateepapon [a], Attawood Lertpimonchai [b], Ammarin Thakkinstian [a]

a. Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand.
b. Department of Periodontology, Faculty of Dentistry, Chulalongkorn University, Bangkok, Thailand.

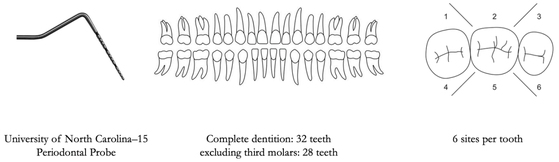
Background and Rationale

Periodontitis is one of the most common oral disease and leading cause of tooth loss in adults. 743 million people affected worldwide with the prevalence of 11.2% globally.

It is characterized by resorption of the alveolar bone, and Centers for Disease Control and Prevention/American Academy of Periodontology (CDC/AAP) working group considers subject with ≥2 interproximal sites with Clinical attachment level (CAL) ≥6 mm in different teeth and 1 site with Periodontal pocket depth (PD) ≥5 mm as severe periodontitis.

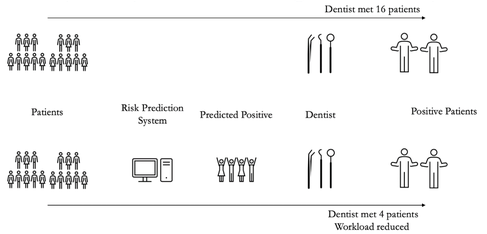


Without radiographs; increase in clinical attachment level is examined. The distance is manually measured using periodontal probes at six sites per tooth and every present teeth excluding third molars: 28 teeth; 168 sites per person.



University of North Carolina–15 Periodontal Probe | Complete dentition: 32 teeth excluding third molars: 28 teeth | 6 sites per tooth

It is a time consuming process, requiring trained personnel to examine. Risk prediction system can help as screening models; especially in cases with high number of patients.



Objectives

To construct and compare the performance of risk prediction models developed by traditional statistical and machine-learning approaches for severe periodontitis.

Materials and Methods

This study used data from the prospective Electricity Generating Authority of Thailand cohort study collected at 2008 and 2013. The outcome of interest was severe periodontitis defined by CDC-AAP aforementioned.

Multivariate imputation with chain equations were applied for imputation. Since repeatedly measured data was applied, multilevel predictive mean matching method was used with miceadds-3.13-12 R library. Mixed-effects logistic regression (MELR), recurrent neural networks (RNN), and a mixed-effects support vector machine (ME-SVM) were performed on imputed data using STATA 16.0, Python 3.8.2, and R 4.02.

21 features including demographic characteristics, underlying disease, risk, oral features, and laboratory features were considered as predictors.

2,271 subjects were examined in 2008, and 2,086 subjects were followed up 5 years later. Data was split into 80% (1,759) for development and 20% (327) for validation sets with prevalence of 34.6% and 34.4%, respectively.

Model Development

For the MELR, feature selection was performed based on forward and backward selection. Interactions between significant features were considered but none was significant. All 21 features were considered in ML models.

For RNN and ME-SVM, hyperparameter optimization was done using a random-search followed by grid-search. This process can be subject to random noise in train data, so a validation dataset was used to assess and if overfitting was observed, the hyperparameters were readjusted

Model Development (contd.)

The RNN model was developed using Keras-2.4.3 and TensorFlow-2.3.1. The final model had 4 hidden layers, 62, 72, 72, 62 simple RNN nodes, Tanh activation function, feed-forward order, dropout of 0.2 between layers, sigmoid output node, binary cross entropy as loss function, accuracy as a monitor metric, learning rate of 0.01, batch size of 64 for mini-batch optimization in 1000 epochs were used with early stopping due to time and resource constraints.

The ME-SVM included support vector regression developed within the e1071-1.7.4 R library framework for fixed-effects, and a linear mixed model developed with lme4-1.1.26 for random-effects. Support vector regression here applies nu-regression, nu value of 0.5, cost value of, and radial basis kernel function with gamma value of 0.3.

Decision threshold was set at 0.35 to reflect observed prevalence. Model performance was evaluated by comparisons of estimates for sensitivity, specificity, accuracy, positive likelihood ratio, area under receiver operating curves (ROC), and F-score. In addition, discrimination and calibration performances were also assessed using area under ROC and Brier scores.

Table 1. Estimation of regression coefficients and odds ratio of the features in MELR

Features	β (SE)	P value	Odds ratios (95% CI)
Gender			
Male	0.97 (0.23)	< 0.001	2.63 (1.68 to 3.81)
Female	ref		ref
Education level			
High school graduate or lower	2.03 (0.39)	< 0.001	7.59 (3.56 to 13.79)
Vocational school graduate	1.37 (0.36)	< 0.001	3.92 (1.94 to 6.78)
Bachelor's degree graduate	0.30 (0.34)	0.393	1.35 (0.66 to 2.61)
Above Bachelor's degree	ref		ref
Smoking			
Current smoker	1.65 (0.25)	< 0.001	5.38 (3.64 to 9.06)
Ex-smoker	0.73 (0.21)	0.001	2.09 (1.33 to 2.82)
Non-smoker	ref		ref
Diabetes Mellitus	0.51 (0.23)	0.024	1.66 (1.11 to 2.48)
Number of present/remaining teeth	-0.06 (0.02)	< 0.001	0.94 (0.91 to 0.96)
Plaque Score	0.03 (0.004)	< 0.001	1.026 (1.019 to 1.033)

Abbreviations: CI: Confidence Interval; SE: Standard Error; ref: Reference covariate group; β: Regression Coefficient.

Results

The mean age was 54.4 years (SD=5.1); 70.6% were males, 46.1% were educated to a bachelor's degree level or higher, and 69.9% earned >50,000 THB (1,500 USD) per month. 50.1% consumed alcohol and 17.1% were current smokers. The prevalence of diabetes, hypertension and dyslipidaemia were 13.3%, 46.4% and 75.4% respectively.

For MELR model, ROC (95% CI) for development data were 0.980 (0.977 – 0.984) and for the validation set 0.983 (0.977 – 0.989). RNN model yielded an 0.747 (0.727 – 0.766) and 0.712 (0.669 – 0.754) respectively. The ROCs for the ME-SVM model were 0.761 (0.754 – 0.766) and 0.698 (0.681 – 0.734).

Table 2. Model Performances

Models	Mixed Effects Logistic Regression		Recurrent Neural Networks		Mixed Effects Support Vector Machine	
Data	Development	Validation	Development	Validation	Development	Validation
Metrics						
%Sensitivity	91.2 (89.4 – 92.8)	89.4 (85.0 – 92.8)	61.6 (58.3 – 64.9)	54.9 (48.0 – 61.7)	52.8 (49.5 – 56.0)	46.1 (39.1 – 53.2)
%Specificity	90.3 (88.9 – 91.6)	92.5 (89.9 – 94.7)	72.9 (70.9 – 75.0)	74.4 (70.1 – 78.3)	82.7 (80.9 – 84.4)	78.2 (74.2 – 81.8)
%Accuracy	90.6 (89.5 – 91.6)	91.4 (89.2 – 93.3)	69.3 (67.6 – 71.1)	68.2 (64.5 – 71.7)	72.7 (71.0 – 74.4)	68.6 (65.0 – 72.1)
ROC	0.980 (0.977 – 0.984)	0.983 (0.977 – 0.989)	0.747 (0.727 – 0.766)	0.712 (0.669 - 0.754)	0.761 (0.754-0.766)	0.698 (0.681-0.734)
F score	0.869	0.878	0.573	0.543	0.564	0.467
Brier score	0.061	0.058	0.181	0.187	0.198	0.200
LR*	9.4 (8.2 – 10.8)	11.9 (8.8 – 16.3)	2.3 (2.1 – 2.5)	2.15 (1.8 – 2.6)	3.1 (2.7 – 3.4)	2.1 (1.7 – 2.6)

Abbreviations: LR*: Positive Likelihood Ratio.

Conclusion

MELR approach performed excellently, and to our knowledge represents one of the best screening models for severe periodontitis. Machine learning approaches failed to demonstrate similar performance despite their ability to estimate non-linear relationships.

These models can be applied in health information systems to monitor oral health. With further independent model validation, such a tool could be evaluated in a primary care setting to assist dental professionals in the screening of severe periodontitis to improve and direct resource allocation to where it is needed most.

This work is part of the manuscript currently under review at Journal of Clinical Periodontology.