Development of Risk Prediction Models for Severe Periodontitis in a Thai Population: Statistical and Machine-Learning Approaches

Htun Teza, Anuchate Pattanateepapon, Attawood Lertpimonchai, Prin Vathesatogkit, Gareth J. McKay, John Attia, Ammarin Thakkinstian

Submitted to: JMIR Formative Research
on: April 20, 2023

Table of Contents

Development of Risk Prediction Models for Severe Periodontitis in a Thai Population: Statistical and Machine-Learning Approaches

Htun Teza[1] MSci; Anuchate Pattanateepapon[1] DEng; Attawood Lertpimonchai[2] DDS, PhD; Prin Vathesatogkit[3] MD; Gareth J. McKay[4] PhD; John Attia[5] MD, PhD; Ammarin Thakkinstian[1] PhD

[1]Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine Ramathibodi Hospital, Mahidol University Bangkok TH
[2]Department of Periodontology, Faculty of Dentistry, Chulalongkorn University Bangkok TH
[3]Department of Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University Bangkok TH
[4]Centre for Public Health, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast Belfast GB
[5]School of Medicine and Public Health, and Hunter Medical Research Institute, University of Newcastle NSW AU

Corresponding Author:
Anuchate Pattanateepapon DEng
Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine Ramathibodi Hospital, Mahidol University
270 RAMA VI Road, Rachathevi,
Bangkok
TH

Abstract

Background: Chronic periodontitis affects 26% in Thai adults and 11.2% globally, resulting in loss of tooth and quality of life. Although periodontal probing is the gold standard, it is time and resource-consuming, so a screening model to identify chronic periodontitis can aid in reducing workload for dentists. While cross-sectional logistic regression is common to apply, optimal performance depends on feature selection and engineering. Machine learning recently has been applied due to their complex yet powerful performances.

Objective: We aim to compare the performance of screening models developed on statistical and machine-learning principles.

Methods: This study used data from the prospective Electricity Generating Authority of Thailand (EGAT) cohort. Dental examinations were performed for 2008 and 2013 surveys. The outcome of interest was periodontitis diagnosed by the Centre for Disease Control – American Academy of Periodontology defined criteria. Risk prediction models were developed using mixed-effects logistic regression (MELR), recurrent neural networks (RNN), mixed-effects support vector machine (ME-SVM), and mixed-effects decision tree (ME-DT).

Results: Of 2,086 subjects, data were split into 80% and 20% for development and validation sets, respectively. As a result, 1,759 and 327 subjects were used for development and validation, with a prevalence of periodontitis of 34.4% and 34.1%, respectively. MELR (AUC and F-Score of 98.0% and 86.9%) performed better than RNN, ME-SVM and ME-DT (74.7% and 57.3%; 76.1% and 56.4%; 69.5% and 50.9% respectively) in identifying severe periodontitis.

Conclusions: MELR model might be potentially applied as a screening model to evaluate the need for further periodontal evaluation. However, the model requires further external validation.

(JMIR Preprints 20/04/2023:48351)
DOI: https://doi.org/10.2196/preprints.48351

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
Please make my preprint PDF available to anyone at any time (recommended).
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
✓ **No, I do not wish to publish my submitted manuscript as a preprint.**
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Htun Teza, MSc (ORCID: 0000-0002-1076-9513)
Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
Email: htun.tez@mahidol.edu

Anuchate Pattanateepapon, DEng (ORCID: 0000-0003-1246-9482)
Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
Email: anuchate.gab@mahidol.ac.th

Attawood Lertpimonchai, DDS PhD (ORCID: 0000-0003-2501-1534)
Department of Periodontology, Faculty of Dentistry, Chulalongkorn University, Bangkok, Thailand
Email: attawood.l@chula.ac.th

Prin Vathesatogkit, MD (ORCID: 0000-0003-0433-2072)
Department of Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
Email: prin.vat@mahidol.ac.th

Gareth J. McKay, PhD (ORCID: 0000-0001-8197-6280)
Centre for Public Health, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast, UK
Email: g.j.mckay@qub.ac.uk

John Attia, MD PhD (ORCID: 0000-0001-9800-1308)
School of Medicine and Public Health, and Hunter Medical Research Institute, University of Newcastle, New Lambton, NSW, Australia
Email: john.attia@newcastle.edu.au

Ammarin Thakkinstian, PhD (ORCID: 0000-0001-9991-386X)
Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
Email: ammarin.tha@mahidol.ac.th

Corresponding author:
Anuchate Pattanateepapon, DEng (ORCID: 0000-0003-1246-9482)
Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
Email: anuchate.gab@mahidol.ac.th

Attawood Lertpimonchai, DDS PhD (ORCID: 0000-0003-2501-1534)
Department of Periodontology, Faculty of Dentistry, Chulalongkorn University, Bangkok, Thailand
Email: attawood.l@chula.ac.th

Author Contributions:

Htun Teza - Conceptualised the study design, Data curations, Data analysis, Data interpretations, Drafted the manuscript
Anuchate Pattanateepapon - Conceptualised the study design, Data curations, Data analysis, Data interpretations, Review & Editing
Attawood Lertpimonchai - Conceptualised the study design, Data curations, Data analysis, Data interpretations, Review & Editing
Prin Vathesatogkit – Review & Editing
Gareth J. McKay – Review & Editing
John Attia – Review & Editing
Ammarin Thakkinstian - Conceptualised the study design, Data curations, Data analysis, Data interpretations, Review & Editing

Development of Risk Prediction Models for Severe Periodontitis in a Thai Population: Statistical and Machine-Learning Approaches

Abstract

Background: Severe periodontitis affects 26% in Thai adults and 11.2% globally, characterised by loss of alveolar bone height. Full-mouth examination by periodontal probing is the gold standard for diagnosis, but it is time and resource intensive. Therefore, a screening model to identify those at high risk of severe periodontitis would offer a more targeted approach and aid in reducing workload for Dentists. While statistical modelling by a logistic regression is commonly applied, optimal performance depends on feature selections and engineering. Machine learning recently has been recently gaining favour given their potential discriminatory power, capable in dealing with multi-way interactions without requirements of linear assumptions.

Objective: We aim to compare the performance of screening models developed on statistical and machine-learning approaches for risk prediction of severe periodontitis.

Methods: This study used data from the prospective Electricity Generating Authority of Thailand (EGAT) cohort. Dental examinations were performed for 2008 and 2013 surveys. Oral examinations (i.e., number of teeth and oral hygiene index/plaque score), periodontal pocket depth (PPD) and gingival recession were performed by well-trained Dentists. The outcome of interest was severe periodontitis diagnosed by the Centre for Disease Control – American Academy of Periodontology, which defined as two or more interproximal sites with a clinical attachment level > 6 mm (on different teeth) and 1 or more interproximal sites with PPD > 5 mm. Risk prediction models were developed using a mixed-effects logistic regression (MELR), recurrent neural networks (RNN), mixed-effects support vector machine (ME-SVM), and mixed-effects decision tree (ME-DT). A total of 21 features were considered as predictive features including 6 demographic characteristics, 5 underlying, 2 risk behaviours, 2 oral features, and 6 laboratorial features.

Results: Of 2,086 subjects, data were split into 80% (n=1759) and 20% (n=327) for development and validation sets, with a corresponding prevalence of periodontitis of 34.4% and 34.1%. The final MELR model contained six features (i.e., sex, education, smoking, diabetes mellitus, number of teeth, and plaque score) with the AUC (95% CI) and positive likelihood ratios [LR+ (95%CI)] of 0.983 (0.977 – 0.989) and 11.9 (8.8 – 16.3) respectively. Machine learnings yielded much lower performance than the MELR model, with these corresponding values of 0.712 (0.669 – 0.754) and 2.1 (1.8 – 2.6) for RNN model, 0.698 (0.681 – 0.734) and 2.1 (1.7 – 2.6) for the ME-SVM model, 0.662 (0.621 – 0.702) and 2.4 (1.9 -3.0) for ME-DT.

Conclusions: MELR model might be useful relative to machine learnings in a large-scale screening to identify those at high risk of severe periodontitis for further periodontal evaluation. However, an external validation using data from other centres is required to evaluate the generalizability of the MELR model.

Keywords: Periodontitis; prediction; machine learning; repeated measures; panel data

Introduction

Periodontitis, one of the most common oral diseases, is the major cause of tooth loss in adult life [1] with a prevalence of 11.2% globally and 15.0-20.0% in Asia [2]. It is a complex inflammatory disease affecting supportive structures around the tooth, resulting in loosening and eventual loss [3]. This leads to decreased dental occlusion, digestive ability, and quality of life. In addition to oral manifestations, it is also associated with other inflammatory or systemic diseases [4], including atherosclerotic vascular disease [5], diabetes mellitus, chronic kidney disease [6], chronic obstructive pulmonary disease, rheumatoid arthritis, Alzheimer's disease, and erectile dysfunction [7].

Severe periodontitis is characterized by loss of alveolar bone height, which is asymptomatic until the tooth becomes mobile. Radiographs are usually used as the standard tool for diagnosis along with a full mouth examination by dentists, both of which are time and resource-intensive, especially in public health sectors constrained by the large number of subjects that require examination. The impact on resource allocation can in part, be addressed through the use of a screening tool such as risk prediction models to identify those at high risk of severe periodontitis.

Identification of severe periodontitis risk factors has been achieved largely through cross-sectional investigations that have evaluated demographic features, risk behaviours, and oral characteristics [8-13]. Inclusion of both demographic and oral features as predictors has been reported to out-perform models composed of either feature alone [9]. Furthermore, addition of saliva biomarkers to established risk factors further improved performance[11], but including such parameters necessitates the requirement for oral examination, which is contradictory to the purpose of a screening tool (for reducing time and resources). However, the majority of studies have used cross-sectional data, which fails to capture the complex relationship between the features and outcomes in contrast to longitudinal investigations that reflect both inter-individual and intra-individual dynamics [14].

Machine learning approaches for disease risk prediction have been proposed, which might be better in dealing with multi-dimensional interactions, collinearity between features, and non-linear relationships better than the traditional statistical models [15-17]. Several machine learning algorithms such as support vector machines, decision trees and artificial neural networks have also been reported to improve diagnosis of periodontal disease [18, 19]. The performance of artificial neural networks were also considered [20] with including probing pocket depth (PPD) as one of their predictors. However, PPD is required a comprehensive periodontal examination, which is time and resource intensively consume through the oral examination process. Applying machine learning screening models without PPD may help reduce the number of patients requiring dental examination and associated resource commitments. As such, the aim of this study was to use longitudinal data to compare the performance of statistical and machine learning approaches for periodontitis risk prediction.

Material and Methods

Setting and Study Population

This study used data from the prospective Electricity Generating Authority of Thailand (EGAT) cohort study [21]. Dental examinations were performed in 2008 and 2013 surveys.

Patient Eligibility

All participants were included if they had received periodontal examinations in both surveys (2008 and 2013) regardless having periodontitis at baseline survey. Some participants were excluded if they did not receive periodontal examinations due to (1) a refusal to participate, (2) systemic conditions that required antibiotic prophylaxis before dental examination such as congenital heart disease or valvular heart disease, a previous history of bacterial endocarditis or rheumatic fever, total joint replacement, and end-stage renal disease, and/or (3) being fully edentulous.

Clinical Features

Demographic and laboratory characteristics

General demographic data (i.e., age, gender, educational level, income, and marital status), behavioural data (i.e., smoking status, alcohol consumption, and exercise/physical activity), family history of illness, and underlying diseases (e.g., diabetes mellitus, hypertension, chronic kidney disease, etc.) were collected by self-administered questionnaires at both time points. Physical examination (i.e., blood pressure, heart rate, blood glucose level, weight, height, and waist & hip circumference) was also performed at the survey site. Laboratory tests included fasting state glucose, lipid profile, renal and liver function, and a complete blood count.

Oral features

Oral examinations included the number of teeth and oral hygiene index (plaque score) [22] which were carried out by the Department of Periodontology, Faculty of Dentistry, Chulalongkorn University. Periodontal pocket depth (PPD) and gingival recession (RE) were measured at six sites (i.e., buccal/labial, lingual/palatal, mesio-buccal, mesio-lingual, disto-buccal, and disto-lingual sites) on all fully erupted teeth, except third molars and retained roots. Centre for Disease Control – American Academy of Periodontology (CDC-AAP) criteria were used to classify severe periodontitis. PPD was defined as the distance from the coronal point of the gingival margin to the tip of a periodontal probe, and the RE as the distance to the cementoenamel junction, with the clinical attachment level (CAL) calculated by subtracting the RE from the PPD.

Outcome

The primary outcome of interest was severe periodontitis as defined by the CDC-AAP guidelines [23] at two or more interproximal sites with a CAL $\geq$ 6 mm (on different teeth) and 1 or more interproximal sites with PPD $\geq$ 5 mm.

Model Development

Among the included subjects, missing data rate was relatively low, which ranged from 0.03% to 9.3% (Supplementary Table 1). Multiple imputation with chain equations (MICE) were applied to impute missing data assuming data were missing at random, see more detail in Supplementary document. Given that repeatedly measured data was applied, multilevel predictive mean matching method was applied for all continuous features using miceadds-3.13-12 R library. Features used to impute the missing data are presented in Supplementary Table 2. A total 5 imputations for MICE were constructed with 8 iterations each for estimation, see model convergence in Supplementary Figure 1. Distributions of features in complete-case and imputed data are almost the same (Supplementary Table 3 and Figure 2).

A total of 21 features including demographic characteristics (age, gender, education level, income, body mass index, waist to hip ratio), underlying disease (diabetes mellitus, hypertension,

dyslipidaemia, chronic kidney disease), risk behaviours (smoking status, alcohol drinking habits), oral features (number of teeth, plaque score), and laboratory features (lymphocytes, uric acid, triglyceride, cholesterol, high density lipoprotein, low density lipoprotein and lipid lowering drug), were considered as predictors. Of them, nine features were included as categorical data with the rest as continuous data.

To the best of our knowledge, there are no explicit guidelines for sample size estimation for machine learning models, but previous studies have recommended this should be based on disease prevalence estimates [24]. According to the 8th National Oral Health Survey of Thailand (2017) [25], the adult prevalence of severe periodontitis in the Thai population was 26%. A total of 296 participants would therefore be required assuming a type one error rate of 5% and confidence interval of 95%, with 77 having severe periodontitis. Our data included 2,086 participants that underwent periodontal examination, with 721 characterised with severe symptoms, providing sufficient power.

Four models were considered: a mixed-effects logistic regression (MELR), recurrent neural networks (RNN), a mixed-effects support vector machine (ME-SVM), and a mixed-effects decision tree (ME-DT). For mixed-effects approaches, a random intercept was fitted considering effect of subjects as random. The framework for model development is shown in Supplementary Figure 3. For the MELR [26], feature selection was performed based on suggestion by Hosmer-Lemeshow [27] as following steps: First, univariate analysis of MELR was performed and indicated that 15 out of 21 features (i.e., age, gender, education level, income, waist to hip ratio, diabetes mellitus, hypertension, smoking status, alcohol drinking habits, number of teeth, plaque score, lymphocytes, uric acid, triglyceride, and high density lipoprotein) had p-values ≤0.10 (Supplementary Table 4), these were then considered simultaneously in a multivariate MELR. Second, a stepwise with forward selection was applied by including each of 15 features which was the most significant into the MELR model one by one, only significant features were finally kept in the final model. Third, six non-significant features (i.e., BMI, dyslipidaemia, CKD, cholesterol, LDL and lipid lowering drug) in the univariate analysis were also re-considered to add in the final multivariate MELR model that contained only significant features; but none of them was significant, thus they were omitted. Fourth, an interaction between significant features (e.g., smoking and gender, smoking and plaque score, plaque score and diabetes) was considered but none was significant. Finally, odds ratios (OR) and 95% confidence intervals (CI) of all significant features were estimated based on the final model.

A total of 21 features were considered in ML models. For RNN [28], ME-SVM, and ME-DT [29], hyperparameter optimization was done using a random-search of the hyperparameter sets followed by grid-search procedures. This process can be subject to unfocused random noise in data development and a failure to generalize. As such, a validation dataset was used to assess model performance and the hyperparameters were readjusted if overfitting was present. The RNN model was developed using Keras-2.4.3 and TensorFlow-2.3.1. The final model specifications for RNN were four hidden layers with 62, 72, 72, 62 simple RNN nodes with a Tanh activation function in feed-forward order with a dropout of 0.2 allocated between hidden and output layers. The output layer had one sigmoid node for binary classification. Binary cross entropy represented a loss function, with accuracy as a monitor metric. A learning rate of 0.01 and a batch size of 64 were applied for mini-batch optimization. A total of 10,000 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 [30], with a nu value of 0.5, cost value (C) of 0.1 as the penalty parameter for misclassifications, and radial basis kernel function with a

gamma value of 0.3. Similarly, ME-DT used the rpart-4.1.16 library framework with a maximum tree depth of 18 and a minimum number of subjects for splitting of 20. Hyperparameter tuning was performed by leave-one-out (K-1) cross-validation.

Probability of being periodontitis was estimated by each model. Patients were classified as positive if the estimated probability was $\geq$ 0.35 as for the prevalence of periodontitis of our data. A contingency 2x2 table was then constructed comparing positive and negative classifications with actual periodontitis. Model performance of each model was further evaluated by estimating sensitivity, specificity, accuracy, positive likelihood ratio (LR^+), and F-score. In addition, discrimination and calibration performances were also assessed using area under receiver operating curves (AUC) and Brier scores. Values range from 0 to 1 for both with a higher score being preferable for AUC in contrast to a lower score for Brier.

All analyses were performed based on imputed data using STATA version 16.0 for MELR, Python version 3.8.2 for RNN, and R version 4.02 for ME-SVM and ME-DT.

Ethical Considerations

The study was approved by the Human Research Ethics Committee, Faculty of Medicine Ramathibodi Hospital, Mahidol University (COA.MURA2020/1560). For the prospective EGAT cohort, all participations were voluntary and they gave written informed consents, including permission for secondary analyses of collected data for necessary further studies. Identifications and personal information were encrypted and kept in databases where only principal investigators could access.

Results

A total of 2,271 subjects were initially included in the cohort, but 2,086 subjects were followed up 5 years later in 2008. The key characteristics of the 2,086 participants between those with and without periodontitis are reported in Table 1. The mean age was 54.4 years (SD=5.0) with the youngest being 43.7 years and the oldest being 70.3; 71% of the subjects were males, 47.2% were educated to a bachelor's degree level or higher, and the majority (70.8%) earned > 50,000 baht (approximately 1,500 USD) per month. Approximately 67.8% consumed alcohol and 16.7% were current smokers. The prevalence of diabetes, hypertension and dyslipidaemia were 12.9%, 44.8% and 71.5% respectively. Data from both surveys were split into 80% for development and 20% for validation sets [31] at subject-levels to prevent data leakage, i.e., subjects were present in either the development or the validation set only. As a result, 1,759 and 327 subjects were used for development and validation, with a prevalence of periodontitis of 34.4% and 34.1%, respectively (supplementary Table 6).

Table 1: Demographics and clinical characteristics for subjects with observed severe periodontitis and those without

	Total number of subjects/observations (n=2,086, N=3,883)	Severe periodontitis (n=721, N=1,333)	Non-severe periodontitis (n=1,365, N=2,550)
Characteristics			
Age, mean (SD), years	54.4 (5.0)	55.0 (5.1)	54.0 (5.0)
Gender, n (%)			
Male	1,482 (71.0)	591 (82.0)	891 (65.3)
Female	604 (29.0)	130 (18.0)	474 (34.7)
Education level, N (%)			
High school graduate or lower	767 (19.8)	406 (30.4)	361 (14.2)
Vocational school graduate	1,282 (33.0)	522 (39.2)	760 (29.8)
Bachelor's degree graduate	1,519 (39.1)	341 (25.6)	1,178 (46.2)
Above Bachelor's degree	315 (8.1)	64 (4.8)	251 (9.8)
Income, N(%), Baht pm			
Less than 20,000	306 (7.9)	149 (11.2)	157 (6.2)
Between 20,000 and 49,999	828 (21.3)	365 (27.4)	463 (18.1)
More than 50,000	2,749 (70.8)	819 (61.4)	1,930 (75.7)
Body mass index, mean (SD)	24.9 (3.7)	24.9 (3.7)	24.8 (3.7)
Waist to hip ratio, mean (SD)	0.9 (0.1)	0.9 (0.1)	0.9 (0.1)
Diabetes mellitus, N (%)	499 (12.9)	234 (17.5)	265 (10.4)
Hypertension, N (%)	1,741 (44.8)	676 (50.7)	1,065 (41.8)
Dyslipidaemia, N (%)	2,775 (71.5)	954 (71.6)	1,821 (71.4)
Chronic kidney disease, N (%)	293 (7.5)	105 (7.9)	188 (7.4)
Smoking, N (%)			
Non-smoker	2,092 (53.9)	498 (37.3)	1,594 (62.5)
Ex-smoker	1,143 (29.4)	446 (33.5)	697 (27.3)
Current smoker	648 (16.7)	389 (29.2)	259 (10.2)
Alcohol, N (%)			
Non-consumer	1,249 (32.2)	309 (23.2)	940 (36.8)

Occasional consumer	695 (17.9)	242 (18.1)	453 (17.8)
Frequent consumer	1,939 (49.9)	782 (58.7)	1,157 (45.4)
Number of present/remaining teeth, mean (SD)	23.4 (4.9)	22.1 (5.4)	24.1 (4.5)
Plaque score, mean (SD), percentage	70.9 (21.5)	78.5 (18.8)	66.9 (21.7)
Lymphocytes, mean (SD), mm3	2,156.3 (623.0)	2,224.5 (636.9)	2,120.6 (612.7)
Uric acid, mean (SD)	5.9 (1.4)	6.1 (1.4)	5.8 (1.5)
Triglyceride, mean (SD), mg/dL	147.6 (96.2)	159.6 (109.5)	141.4 (87.8)
Cholesterol, mean (SD), mg/dL	225.1 (43.3)	223.2 (44.4)	226.0 (42.6)
High density lipoprotein, mean (SD) , mg/dL	54.1 (14.3)	51.5 (13.8)	55.4 (14.5)
Low density lipoprotein, mean (SD) , mg/dL	147.7 (39.4)	145.7 (40.1)	148.7 (39.0)
Taking lipid lowering medications, N (%)	960 (24.7)	328 (24.6)	632 (24.8)

Abbreviations: mg/dL: Milligrams per Decilitre; mm3: Per Cubic Millimetre; n: number of subjects; N: number of observations; pm: Per Month; SD: Standard Deviation.

The final multivariate MELR model included six features: sex, education, smoking, diabetes mellitus, number of teeth, and plaque score. The regression coefficients and ORs for each feature are reported in Figure 1 and Supplementary Table 5. The odds of men having severe periodontitis were 2.63 times higher relative to women. Education level was significantly associated with severe periodontitis with lower levels of education, resulting those educated to vocational level, and high school graduates associated with 3.92 and 7.59 times greater likelihood of severe periodontitis compared to those educated above than the bachelor's degree. Current and ex-smokers had 5.38 and 2.09 times higher odds of severe periodontitis than non-smokers. Participants with diabetes had a 66% greater risk of severe periodontitis compared to no diabetes. The risk of periodontitis increased by 3% per unit increase in plaque score, in contrast to a 6% reduction in risk, for every remaining tooth.

Model performance was evaluated with both development and validation data (Table 2). For the development dataset, AUC (95% CI), F1 and Brier scores for the MELR model were 0.980 (0.977 – 0.984), 0.869, and 0.061, respectively. The corresponding values for the validation set were 0.983 (0.977 – 0.989), 0.878, and 0.058 indicating the model performed well in both datasets; the corresponding positive likelihood ratios (at the threshold of 0.35) were 9.4 (8.2 – 10.8) and 11.9 (8.8 – 16.3) respectively. This could be interpreted that patients were about 9-folds more likely to have periodontitis than non-periodontitis given the model classified them as positive (i.e., estimated probability ≥0.35).

The RNN model yielded an AUC value of 0.747 (0.727 – 0.766) and 0.712 (0.669 – 0.754) for the development and validation datasets, respectively. The corresponding LR^+s were 2.3 (2.1 - 2.5) and 2.1 (1.8 – 2.6), which were much lower compared to the MELR. The AUCs for the ME-SVM model were 0.761 (0.754 – 0.766) and 0.698 (0.681 – 0.734) for both the development and validation datasets with corresponding LR+s of 3.1 (2.7 – 3.4) and 2.1 (1.7 – 2.6). For ME-DT, AUC and LR+ values were 0.695 (0.677 - 0.714) and 2.4 (2.1 – 2.6) for development dataset, while the corresponding values for validation dataset were 0.662 (0.621 – 0.702) and 2.4 (1.9 -3.0). The receiver operating characteristics curves for all models are shown in Supplementary Figure 4.

Table 2: Performances of the predictive models in development and validation cohorts

Models	Mixed Effects Logistic Regression		Recurrent Neural Networks		Mixed Effects Support Vector Machine		Mixed Effects Decision Tree	
Data	Development	Validation	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)	47.0 (44.0 – 50.1)	44.5 (38.4 – 50.7)
%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)	80.2 (78.4 – 81.9)	81.3 (77.6 – 84.6)
%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)	68.8 (67.1 – 70.4)	68.7 (65.3 – 72.0)
AUC	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)	0.695 (0.677 – 0.714)	0.662 (0.621 – 0.702)
F score	0.869	0.878	0.573	0.543	0.564	0.467	0.509	0.493
Brier score	0.061	0.058	0.181	0.187	0.198	0.200	0.236	0.240
LR^+	9.4 (8.2 – 10.8)	11.9 (8.8 – 16.3)	2.3 (2.1 – 2.5)	2.1 (1.8 – 2.6)	3.1 (2.7 – 3.4)	2.1 (1.7 – 2.6)	2.4 (2.1 – 2.6)	2.4 (1.9 – 3.0)

Abbreviations: AUC: Area Under Receiver Operating Characteristic Curve; LR^+: Positive Likelihood Ratio.

Discussion

Principal Results

Our study developed risk prediction models for periodontitis using traditional statistical and machine learning approaches. The MELR model performed best with an AUC value of 0.983, in comparison to both machine learning approaches, RNN and ME-SVM, which had fair performances with AUC values of 0.712 and 0.698, respectively. In addition, the Brier scores for the RNN and ME-SVM were similarly high at 0.187 to 0.200, in contrast to a value of 0.058 for the MELR which reflects an overfitting for both machine learning models compared to the MELR. Furthermore, a LR+ value as low as 2-3 for the machine learning approaches contrasted to the high value of 11.9 for the MELR.

Comparison with Prior Work

Our MELR model also performed better than previous predictive models that applied logistic regression (AUC = 0.71) [10], and was superior even to those that included salivary biomarkers such as Chitinase and Protease activity (AUC = 0.91) [18]. Our analyses suggest that the mixed model approach performs better than logistic regression because the former considers latent subject-specific variability and thus better captures information about population-average effects of the risk features than regression approaches that use cross-sectional data.

The machine learning models (RNN, ME-SVM and ME-DT) may have performed less well in comparison to the MELR model due to a data imbalance, as one third of our study participants had severe periodontitis. Ling et al. [32] suggests that a classification imbalance may affect model performance if the cost of the two errors (i.e., false positive and false negative in the binary classification) is not the same, or if the class distribution in the validation data is different from that of the development data. The prevalence of severe periodontitis in the development and validation sets was very close, i.e., 34.4% and 34.1% (See Supplementary Table 6), and this was similar to the 8th National Oral Health Survey of Thailand (2017) [25], which reported 26% prevalence in adults and 36% in the elderly. Thus both datasets had similar distributions of severe periodontitis that accurately reflect Thai prevalence.

To simulate the improved performance of the MELR model, a framework to include repeated measures and random effects was applied to the machine learning models, which is a recognized advantage of the ME-ML model, although the model still failed to meet the performance levels of MELR. This may result from differences in optimization and estimation of fixed- and random-effects within these models: the Penalized Quasi-Likelihood (PQL) method was used for MELR [33] and the Expectation-Maximization (EM) was applied in the ME-ML. [29, 34] This framework could be beneficial in estimating non-linear relationships between predictors and outcomes, however further studies are necessary to independently validate this.

Strength and Limitations

A cut off value of 0.35 was selected to reflect the observed prevalence of the condition as the uniform decision threshold and applied to all three screening models to enable cross-model comparisons, but this can be adjusted depending on the objective and outcome. [35] In most

clinical screening or diagnostic tools, it is unlikely that the consequences of a false positive and false negative are similar. By reducing the decision threshold, subjects with a lower probability would be considered positive, increasing the sensitivity (and consequently the number of false positives) but reducing the specificity (number of false negative cases). In mass screening situations, participants identified as positive would be referred for further examination therefore this would lead to increased numbers of subjects requiring comprehensive periodontal probing. However, it would also fulfil the purpose of screening by providing early diagnosis and prompt referral by reducing the number of positive subjects incorrectly identified as negative. Despite the reduced specificity associated with a lower decision threshold, this approach would identify those at greatest risk while reducing the overall workload for examiners and more efficient allocation of resources.

While MELR models can be calculated manually as a linear combination of features, machine learning approaches require to be exported in hierarchical file formats for use in websites or applications which can be developed with a user-friendly interface. Data collection and mining from electronic care records can be combined and reformatted for more complex procedures such as feature engineering and data pre-processing. Risk prediction modelling would therefore be amenable to data updates, facilitating model refinement, with potential portability through web or desktop-based applications that could be provided to healthcare staff.

This study had several limitations. Only internal validation was carried out for model evaluation and external validation using data from other centres or surveys is required. Furthermore, data from other Thai populations, as well as different countries or ethnicities, would help determine the generalizability of the findings. Machine learning approaches in particular would be benefited from further model refinement with larger, better characterized datasets since their performance depends on the data quality of the development sets. Although the clinical features included in the MELR model were relatively easy to examine, the assessment of a plaque score is manually intensive. Although periodontal probing is not required, the examination includes oral rinsing with plaque disclosing solutions and counting of the stained surfaces manually. Self-reporting would not be optimal since improper application and poor assessment would lead to an unreliable scoring and subsequent risk underestimation by the model. The oral features included in developing the models are limited to the factors collected by the EGAT survey. Previous studies have suggested that the inclusion of more relevant oral characteristics such as tooth mobility and gum bleeding would increase the performance even further, although an AUC of 0.98 is considered excellent.

Conclusions

In conclusion, the MELR approach performed excellently, and to our knowledge represents one of the best screening models for severe periodontitis. Machine learning approaches demonstrated fair performance despite their ability to estimate non-linear relationships. Instead of relying on PPD measurements obtained through an extensive periodontal assessment, which can be time-consuming and resource-intensive, the MELR model might be useful in health information systems to monitor oral health, prompting patients to visit a dental professional for comprehensive examination and appropriate treatment. With further independent model external validation, such a tool should 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.

Acknowledgements
This study was part of MSc dissertation for the lead author (Htun Teza) under the Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Thailand. The authors would like to thank Sukanya Siriyotha for her input regarding the data. Htun Teza and Ammarin Thakkinstian received grant from the National Research Council of Thailand (NRCT) N42A640323. The funder was not involved in the design of the study; the collection, analysis, and interpretation of data; writing the report; and did not impose any restrictions regarding the publication of the report.

Data Availability
The data sets generated during analyses for this study are available from the corresponding author on reasonable request. Codes generated for the whole analyses are provided at https://www.rama.mahidol.ac.th/ceb/codes/code_pj1

Conflict of Interest
None declared.

Informed Consent
This study applied secondary data from Electricity Generating Authority of Thailand cohort, which has received informed consent from the participants.

Abbreviations
AUC: Area under Receiver Operating Curves
BMI: Body Mass Index
CAL: Clinical Attachment Level
CDC-AAP: Centre for Disease Control – American Academy of Periodontology
EGAT: Electricity Generating Authority of Thailand
LDL: Low Density Lipoprotein
LR+: Positive Likelihood Ratio
MELR: Mixed-Effects Logistic Regression
ME-DT: Mixed-Effects Decision Tree
ME-SVM: Mixed-Effects Support Vector Machine
mg/dL: Milligrams per Decilitre
mm3: Per Cubic Millimetre
MICE: Multiple Imputation with Chain Equations
n (%): number (percentage) of subjects
N (%): number (percentage) of observations
pm: Per Month
PPD: Periodontal/Probing Pocket Depth
RE: Gingival Recession
RNN: Recurrent Neural Networks
SD: Standard Deviation

References

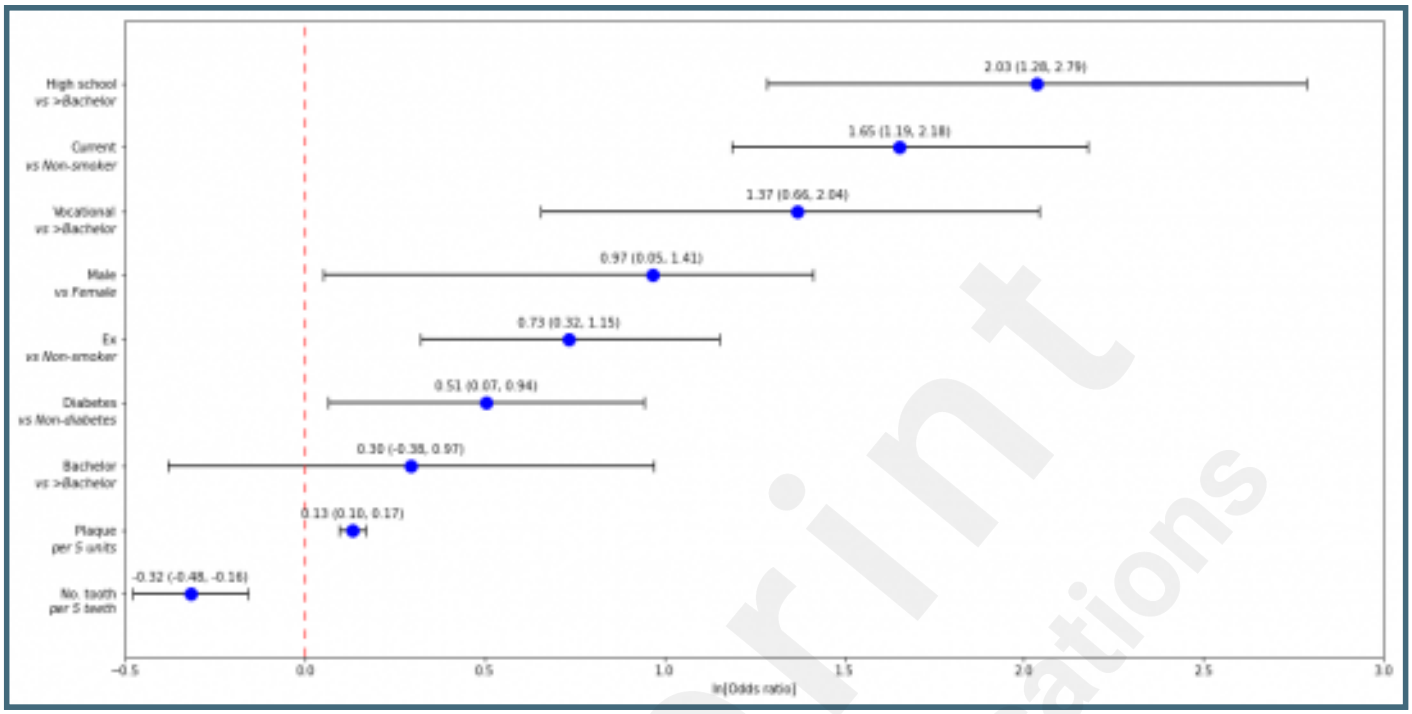
1. Phipps KR, Stevens VJ. Relative contribution of caries and periodontal disease in adult tooth loss for an HMO dental population. J Public Health Dent. 1995 Fall;55(4):250-2. PMID: 8551465. doi: 10.1111/j.1752-7325.1995.tb02377.x.
2. Corbet EF, Leung WK. Epidemiology of periodontitis in the Asia and Oceania regions. Periodontol 2000. 2011 Jun;56(1):25-64. PMID: 21501236. doi: 10.1111/j.1600-0757.2010.00362.x.
3. Tonetti MS, Jepsen S, Jin L, Otomo-Corgel J. Impact of the global burden of periodontal diseases on health, nutrition and wellbeing of mankind: A call for global action. J Clin Periodontol. 2017 May;44(5):456-62. PMID: 28419559. doi: 10.1111/jcpe.12732.
4. Linden GJ, Lyons A, Scannapieco FA. Periodontal systemic associations: review of the evidence. J Clin Periodontol. 2013 Apr;40 Suppl 14:S8-19. PMID: 23627336. doi: 10.1111/jcpe.12064.
5. Mattila KJ, Nieminen MS, Valtonen VV, Rasi VP, Kesäniemi YA, Syrjälä SL, et al. Association between dental health and acute myocardial infarction. Bmj. 1989 Mar 25;298(6676):779-81. PMID: 2496855. doi: 10.1136/bmj.298.6676.779.
6. Lertpimonchai A, Rattanasiri S, Tamsailom S, Champaiboon C, Ingsathit A, Kitiyakara C, et al. Periodontitis as the risk factor of chronic kidney disease: Mediation analysis. J Clin Periodontol. 2019 Jun;46(6):631-9. PMID: 30993705. doi: 10.1111/jcpe.13114.
7. Monsarrat P, Blaizot A, Kémoun P, Ravaud P, Nabet C, Sixou M, et al. Clinical research activity in periodontal medicine: a systematic mapping of trial registers. J Clin Periodontol. 2016 May;43(5):390-400. PMID: 26881700. doi: 10.1111/jcpe.12534.
8. Cyrino RM, Miranda Cota LO, Pereira Lages EJ, Bastos Lages EM, Costa FO. Evaluation of self-reported measures for prediction of periodontitis in a sample of Brazilians. J Periodontol. 2011 Dec;82(12):1693-704. PMID: 21563951. doi: 10.1902/jop.2011.110015.
9. Eke PI, Dye BA, Wei L, Slade GD, Thornton-Evans GO, Beck JD, et al. Self-reported measures for surveillance of periodontitis. J Dent Res. 2013 Nov;92(11):1041-7. PMID: 24065636. doi: 10.1177/0022034513505621.
10. Lai H, Su CW, Yen AM, Chiu SY, Fann JC, Wu WY, et al. A prediction model for periodontal disease: modelling and validation from a National Survey of 4061 Taiwanese adults. J Clin Periodontol. 2015 May;42(5):413-21. PMID: 25817519. doi: 10.1111/jcpe.12389.
11. Verhulst MJL, Teeuw WJ, Bizzarro S, Muris J, Su N, Nicu EA, et al. A rapid, non-invasive tool for periodontitis screening in a medical care setting. BMC Oral Health. 2019 May 23;19(1):87. PMID: 31122214. doi: 10.1186/s12903-019-0784-7.
12. Wu X, Weng H, Lin X. Self-reported questionnaire for surveillance of periodontitis in Chinese patients from a prosthodontic clinic: a validation study. J Clin Periodontol. 2013 Jun;40(6):616-23. PMID: 23557490. doi: 10.1111/jcpe.12103.
13. Zhan Y, Holtfreter B, Meisel P, Hoffmann T, Micheelis W, Dietrich T, et al. Prediction of periodontal disease: modelling and validation in different general German populations. J Clin Periodontol. 2014 Mar;41(3):224-31. PMID: 24313816. doi: 10.1111/jcpe.12208.
14. Hsiao C. Panel data analysis—advantages and challenges. TEST. 2007 2007/05/01;16(1):1-22. doi: 10.1007/s11749-007-0046-x.
15. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. Nat Methods. 2018 Apr;15(4):233-4. PMID: 30100822. doi: 10.1038/nmeth.4642.
16. Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and Machine Learning forecasting methods: Concerns and ways forward. PLoS One. 2018;13(3):e0194889. PMID:

29584784. doi: 10.1371/journal.pone.0194889.
17. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol. 2019 Mar 19;19(1):64. PMID: 30890124. doi: 10.1186/s12874-019-0681-4.
18. Farhadian M, Shokouhi P, Torkzaban P. A decision support system based on support vector machine for diagnosis of periodontal disease. BMC Res Notes. 2020 Jul 13;13(1):337. PMID: 32660549. doi: 10.1186/s13104-020-05180-5.
19. Ozden FO, Özgönenel O, Özden B, Aydogdu A. Diagnosis of periodontal diseases using different classification algorithms: a preliminary study. Niger J Clin Pract. 2015 May-Jun;18(3):416-21. PMID: 25772929. doi: 10.4103/1119-3077.151785.
20. Shankarapillai R, Mathur LK, Nair M, Rai N, Mathur A. Periodontitis Risk Assessment using two artificial Neural Networks-A Pilot Study. International Journal of Dental Clinics. 2010;2:36-40.
21. Vathesatogkit P, Woodward M, Tanomsup S, Ratanachaiwong W, Vanavanan S, Yamwong S, et al. Cohort profile: the electricity generating authority of Thailand study. Int J Epidemiol. 2012 Apr;41(2):359-65. PMID: 21216741. doi: 10.1093/ije/dyq218.
22. O'Leary TJ, Drake RB, Naylor JE. The plaque control record. J Periodontol. 1972 Jan;43(1):38. PMID: 4500182. doi: 10.1902/jop.1972.43.1.38.
23. Eke PI, Page RC, Wei L, Thornton-Evans G, Genco RJ. Update of the case definitions for population-based surveillance of periodontitis. J Periodontol. 2012 Dec;83(12):1449-54. PMID: 22420873. doi: 10.1902/jop.2012.110664.
24. Riley RD, Ensor J, Snell KIE, Harrell FE, Jr., Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. Bmj. 2020 Mar 18;368:m441. PMID: 32188600. doi: 10.1136/bmj.m441.
25. Bureau of Dental Health DoH, Ministry of Public Health Thailand. The Eighth National Oral Health Survey of Thailand 2017. Nonthaburi: Samcharoen Panich (Bangkok) Co., Ltd.; 2018. ISBN: 9786161137519.
26. Wong GYC, Mason WM. The Hierarchical Logistic Regression Model for Multilevel Analysis. Journal of the American Statistical Association. 1985;80:513-24.
27. Sturdivant; DWHJSLRX. Model-Building Strategies and Methods for Logistic Regression. Applied Logistic Regression: John Wiley & Sons, Inc.; 2013. p. 89-151.
28. Al-Askar H, Radi N, MacDermott Á. Chapter 7 - Recurrent Neural Networks in Medical Data Analysis and Classifications. In: Al-Jumeily D, Hussain A, Mallucci C, Oliver C, editors. Applied Computing in Medicine and Health. Boston: Morgan Kaufmann; 2016. p. 147-65.
29. Hajjem A, Bellavance F, Larocque D. Mixed-effects random forest for clustered data. Journal of Statistical Computation and Simulation. 2010;84:1313 - 28.
30. Schölkopf BB, Peter L.; Smola, Alex; Williamson, Robert C., editor. Shrinking the Tube: A New Support Vector Regression Algorithm. NIPS; 1998.
31. Bookstein A. Informetric distributions, part I: Unified overview. J Am Soc Inf Sci. 1990;41:368-75.
32. Ling CXS, Victor S., editor. Cost-Sensitive Learning and the Class Imbalance Problem. 2008.
33. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. Journal of the American Statistical Association. 1993;88:9-25.
34. Hajjem A, Larocque D, Bellavance F. Generalized mixed effects regression trees. Statistics & Probability Letters. 2010;126:114-8.
35. Chen JJ, Tsai CA, Moon H, Ahn H, Young JJ, Chen CH. Decision threshold adjustment in class prediction. SAR QSAR Environ Res. 2006 Jun;17(3):337-52. PMID: 16815772. doi: 10.1080/10659360600787700.

Supplementary Files

Figures

Magnitude of associations in term of ln(odds ratios) between predictors and severe periodontitis: A mixed-effect logistic regression.



Multimedia Appendixes

Supplementary Document.
URL: http://asset.jmir.pub/assets/189a958006df9aae16bb3656a0fd5d8a.docx