Journal of Dentistry 153 (2025) 105469

Contents lists available at ScienceDirect

Journal of Dentistry

journal homepage: www.elsevier.com/locate/jdent

A personalized periodontitis risk based on nonimage electronic dental records by machine learning

Laura Swinckels [a,b,c,d,e,*], Ander de Keijzer [e,f], Bruno G. Loos [a], Reuben Joseph Applegate [g], Krishna Kumar Kookal [h], Elsbeth Kalenderian [i,j], Harmen Bijwaard [d,k], Josef Bruers [a,l]

[a] *Academic Centre for Dentistry Amsterdam (ACTA), University of Amsterdam and Vrije Universiteit Amsterdam, The Netherlands*
[b] *Department Oral Hygiene, Faculty of Health, Sports and Welfare, Inholland University of Applied Sciences, Amsterdam, The Netherlands*
[c] *Centre of Expertise Prevention in Health and Social Care, Inholland University of Applied Sciences, Haarlem, The Netherlands*
[d] *Medical Technology Research Group, Inholland University of Applied Sciences, The Netherlands*
[e] *Data Driven Smart Society Research Group, Inholland University of Applied Sciences, Alkmaar, The Netherlands*
[f] *Applied Responsible Artificial Intelligence Research Group, Avans University of Applied Sciences, Breda, The Netherlands*
[g] *Biomedical Informatics Group-Analytics Research Center, School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, USA*
[h] *Technology Services and Informatics, School of Dentistry, University of Texas Health Science Center at Houston, Houston, TX, USA*
[i] *Marquette University School of Dentistry, Milwaukee, WI, USA*
[j] *University of Pretoria School of Dentistry, Pretoria, South Africa*
[k] *National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands*
[l] *Royal Dutch Dental Association (KNMT), Utrecht, The Netherlands*

ARTICLE INFO

Keywords:
Artificial intelligence
Digital support
Dental data
Early detection
Predictive modeling
Prevention
Periodontal disease

ABSTRACT

Objective: This study aimed to develop a machine-learning (ML) model to predict the risk for Periodontal Disease (PD) based on nonimage electronic dental records (EDRs).

Methods: By using EDRs collected in the BigMouth repository, dental patients from the US were included. Patients were labeled as cases or controls, based on PD diagnosis, treatment and pocketing. By learning from their data, a model was trained. The ability of the developed model to predict PD was evaluated by the accuracy, sensitivity, specificity and area under the curve (AUROC) and the most important features were determined. The best-performing model was applied to the validation set.

Results: The final study population included 43,331 participants. Based on the development set, the Random Forest model performed with high sensitivity (81 %) and had an excellent AUROC (94 %), compared to four other ML and deep learning techniques. The most important predictors were bleeding proportion, age, the number of visits, prior preventive treatment, smoking and drugs usage. When the model was applied to the validation set, the model could detect almost all cases (91 %), but overestimated controls (specificity=0.54). When EDRs were retrieved 3 years before the PD diagnosis, the predictions for PD were still sensitive (89 %).

Conclusion: Based on consistent and complete EDR, ML has an excellent ability to assist with the early detection and prevention of PD cases. Further research is required to follow-up high-risk controls and improve the model's internal and external validation. Improved EDR documentation is an important first step.

Clinical significance: If such ML models become clinically applied, clinicians can be assisted with personalized risk predictions based on the individual. If the key riskcontributing factors for the individual are revealed/provided, ML can suggest targeted prevention interventions. These advancements can contribute to a reduced workload, sustainable EDRs, data-based dental care, and, ultimately, improved patient outcomes.

1. Introduction

Periodontal disease (PD) is a chronic multifactorial inflammatory disease of the tissues surrounding and supporting the teeth (periodontium) that is caused by lifestyle, microbiological, genetic and medical factors [1,2]. In response to the chronic persistence of dental plaque bacteria, the host induces a protective immune-inflammatory response in the periodontium [3], which can lead to irreversible

---

* Corresponding author.
E-mail address: L.Swinckels@acta.nl (L. Swinckels).

https://doi.org/10.1016/j.jdent.2024.105469
Received 13 July 2024; Received in revised form 13 November 2024; Accepted 16 November 2024

attachment and alveolar bone loss. However, PD can often be prevented with early screening and interventions [4,5].

To prevent PD, oral health professionals should regularly screen their patients for early signs [6,7]. This can be challenging due to individual variations in clinical expression [8–10] and the fact that many patients seem clinically asymptomatic until the disease has become severe [1]. Reducing the risk of PD may also further prevent the disease [5]. Factors associated with the risk for PD are age, tobacco and alcohol use, nutrition, stress, sex, an impaired immune response, health behavior and several systemic diseases (e.g. diabetes, obesity, metabolic syndrome, and cardiovascular diseases) [1,11]. Due to the large number of associated factors, it remains challenging to determine a specific risk for each patient individually. PD can be prevented if above-mentioned risk could be controlled, by assessing patients' disease risk [2].

Current risk assessment models used multiple features to calculate a concise risk or a risk class, offering great predictions [12]. A major limitation of these risk predictions is that they are based on data gathered by questionnaires and clinical examinations primarily for research purposes. Only a limited number of PD prediction models have extracted existing data. Prediction models that reused periodontal chart data or imaging data (including furcation and bone loss) capture past periodontal destruction or are incapable of incorporating different (non-) clinical features. Clinical applicability is therefore missing, but essential.

Studies show that early signs [13] and risk factors for PD can be retrieved from Electronic Dental Records (EDR) [14]. EDRs are electronic health records (EHRs) that document information considered to be integral to providing safe and high-quality dental care [15]. EDRs include dental and medical histories, oral examination findings, diagnoses, radiographs, treatment history, and clinical notes. Like all EHRs, EDRs are rich in detailed clinical outcomes [16] and have the benefit of containing longitudinal and multivariable information [17]. These data are assumed to be well-suited for the prevention of PD [18]. Specifically for patients who visit their dentist frequently, dental clinics possess a wealth of preventive data. However, the rising variety and volume of documented EDRs have outpaced clinicians' ability to use this information during oral examinations. Reviewing past documentation in EDRs is time-consuming [19] and many clinicians experience usability issues [20]. While the records in the EDR are intended to support dental care, they are barely used for prevention purposes.

To predict the risk for diseases, Machine Learning (ML) is frequently applied to non-image EHRs in other health domains [21–23]. ML is a subdomain of artificial intelligence (AI) where machines learn from experience with data, aiming to resolve issues without human input [24]. For longitudinal data such as EDRs, Deep Learning (DL) is a suitable component of ML because of its ability to include many variables (i.e. features), detect patterns over time, and incorporate complex interactions in a neural network with multiple layers. Such prediction models can provide dental clinicians with individual risk scores and suitable interventions based on existing data [25]. A scoping review identified studies that used ML and DL on EHRs for the early detection and prevention of diseases [26]. These ML and DL models had high diagnostic performances and could generate the most important features responsible for the predicted risk. The predicted risk scores can serve as a preliminary screening to prioritize patients on waiting lists for physical examination or before harmful or unnecessary examinations (e.g., radiographs). The use of these models also offers benefits such as a personalized care, new insights for healthcare policies, and a reduction of the clinical workload [26]. To date, there is only one ML prediction model using non-image EDRs to prevent PD [27]. Although this model seems promising, it is limited by its reliance on data from a single institution, and the records were not longitudinal. Most ML attempts in periodontology use imaging data to detect clinical attachment and alveolar bone loss, or focus on the severity and treatment of PD [28–31].

ML applied to non-image EDRs would support preventive screening by using existing data without additional impact on the patient and effort for clinicians. Therefore, this study aimed to develop a ML model to predict PD risk based on non-image, longitudinal EDRs. By applying this model, we investigated whether the early detection of PD can be supported.

2. Materials and methods

This study was approved by the Institutional Review Board of the Academic Centre for Dentistry Amsterdam (ACTA) and written according to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement for the development of prediction models in medicine [32]. The following paragraphs describe the methods conducted in sequential order (Result section 3.1).

2.1. *Source of data*

This case-control study used EDR data collected in the BigMouth dental repository [33]. This repository is an oral health database with EDR data from patients who received oral healthcare at academic dental institutions in the United States (US) at the time of the study [34]. These data are available for research purposes and contain data regarding demographics, dental diagnoses, medical and dental histories, caries risk assessment, periodontal data, medication, vital signs, allergies and dental treatment using axiUm® (Exan, Vancouver, BC, Canada). BigMouth members permitted the usage of de-identified and aggregated data (project BM-DR28), as such complying with the American Health Insurance Portability and Accountability Act (HIPAA) requirement [33], the European General Data Protection Regulation (GDPR) standards and the Data Use Agreement(DUA) in place between UTHealth Houston (data host) and participating BigMouth institutions. The data was extracted from BigMouth for patient visits between January 1, 2010 to December 31, 2022.

2.2. *Data preparation*

Two researchers (LS and AdK) had access to the data through a secured virtual machine (Linux). SQL Server was used to store and prepare the data. The raw data were loaded into the first database (LandingZone) and remained untouched. The data were assessed for completeness, correctness, consistency, and the extent of missingness. Completeness was evaluated by assessing whether features were recorded in all institutions and for both cases and controls, while correctness was checked by verifying valid, realistic and non-contradictory values. If issues with completeness and correctness relate to the labeling criteria, consequences occur in the labeling stage. In the second database (Staging), the inconsistent data were cleaned. Consistency was ensured through standardization of data formats, and missingness was evaluated by considering features with substantial records and predictive value. Features with too many varieties in the data (e.g. Prescriptions, Allergies, Dental Treatment, Medical Diagnoses and Dental Diagnoses) were grouped into categories. Next to this, BMI and the number of visits were calculated and converted into new features. The final data were loaded into the last database (Datawarehouse), where the data were ready for definite labeling. A final database consisted of a data hierarchy that was grouped per patient, sorted per visit and included all features registered at that visit. The data transformations that have been made can be found at https://github.com/keijzera/RiskForPDBasedOnEDR.

2.3. *Labeling of cases and controls*

Patients 18 years and older were included if they had data from at least three dental visits in a healthy condition (before their PD diagnosis or their last record in their EDR). Patients with prior PD treatment were excluded. To train a model based on patients who definitely suffered from PD and patients who are certainly PD-negative, we applied multiple criteria to label cases and controls. PD cases were defined when 1) a new diagnosis of PD was documented, followed by 2) PD treatment. This method is considered a valid way of detecting cases in the literature [35]. Eligible diagnoses included aggressive, chronic or necrotizing periodontitis or periodontitis associated with systemic diseases, according to the 1999 classification of periodontal diseases [36], or any form of periodontitis in stage I to IV, according to the classification of 2017 [37]. Treatment codes must relate to periodontitis (e.g. scaling and root planning). The absence of PD diagnosis terms and PD treatment codes in a patient's EDR, defined those patients as a control. However, by verifying the case-control labels, some seemed contradictory to their pocket depth data. Therefore, we applied an additional criterium to be sure that cases have at least one pocket >5mm and controls do not have any >5mm pocket recorded (i.e. ensuring data correctness). We decided to exclude patients who did not fulfill the third criterion to build a model based on data with assured case-control distinctions. Incomplete and inconsistent recording standards for the labeling features, lead to the exclusion of institutions as a whole. The final study population was composed (Result section 3.2) and ideally reflect the real-life prevalence (estimated between 11 and 23 % [38,39]) to be assumed as a representative [40].

2.4. Input features

This study used the following EDR features as input for the model, according to the data governance in BigMouth [34]. Input features were selected if they were assumed to be potential risk factors based on the literature. Input features were sex, age, number of visits, saliva test results, pain and problems, missing teeth, caries risk, dental diagnoses, prior dental treatment, prescribed and used medication, medical diseases (i.e., cancer, hepatitis, tuberculosis, blood, bronchial, cardiovascular, endocrine, kidney, neuro, viral, ear, nose and throat, and other diseases), medical conditions (i.e. allergies and health status), features related to lifestyle (i.e. smoking and drug use and oral hygiene), specific risk factors (diet, orthodontics, exposed roots, pits, recreational drugs, plaque), features related to vital signs (i.e. systolic blood pressure, diastolic blood pressure, BMI, length and weight) and early periodontal data (i.e. bleeding and plaque percentage) collected in the EDR. Looking at a patient's full EDR, some features are not single-registered (e.g. dental treatment codes can be registered every dental visit again). We used only the last record of each feature for non-temporal analyses, while the temporal analyses used all records over time to identify patterns over time. Therefore, each record in the longitudinal data was linked to the visit date. Due to possible privacy conflicts, disclosing an exact visit date was not allowed. Therefore, this was converted to the number of days before the diagnosis (cases) or the end of documentation in the EDR (controls). In other words, we set an 'index-date' at the end of each patient's EDRs to align varying EDR periods. Only patient-level features were used, rather than features on tooth- or even side-level. Input features that were indicative of the label were excluded. For example, caries risk was excluded since it was only recorded in institution 3. Any record of this feature would rather represent the prevalence of institution 3, than the specific result (low or high caries risk). Imaging data and unstructured clinical notes were beyond the scope of this research. The full list and descriptions of these input features are presented in Appendix A. The clinical data extraction was generated by the UTHealth team as they are the only ones who manage and have direct access to the repository.

2.5. Datasets

The final dataset in this study was split into a development set to build a model and a validation set to apply the model. Because the focus of this model is to detect cases (rather than controls), we wanted to prevent the model from being built on overrepresented control data. Therefore, a balanced development set was created (5000:5000). This development set included an 80 % training set to train and (4000:4000) a 20 % test set (1000:1000) to evaluate the developed model. The remaining patient data were used in the validation set to apply the model. For analyses with multifeatured predictions, larger amounts of data will improve the outcome. However, a minimum sample size was dependent on the amount of input features. Simple techniques require at least 10 cases per feature, but more complex techniques (e.g. neural networks) require hundreds per feature. Since we started with 50 input features, we aimed to include at least 5000 cases in the training set.

2.6. Model development

The model was developed based on the training set. Python, Keras, and TensorFlow were used to develop the model. The model was trained by learning from (patterns in) the data and built until the best performance metrics were reached. This was an iterative process of experiments with varying DL and ML techniques, selections of input features and decisions based on clinical relevance (i.e. threshold and sensitivity/specificity).

- Temporal techniques include Long Short Term Memory (LSTM), a Recurrent Neural Network (RNN), and Gated Recurrent Units (GRU) because these were assumed to be the most suitable for detecting development over time [26]. Non-temporal techniques included a Random Forest (RF), a basic Neural Network (NN) and Logistic Regression (LR), because these are most suitable for classification tasks [41,42]. While temporal techniques searched for patterns over time in longitudinal records (assuming temporal dependence within a feature), non-temporal techniques generated a static prediction based on only the last record of each feature.
- For each technique, we started with all features as input for the model. If Prescriptions, Allergies, Dental treatment, Medical diagnoses, and Dental Diagnoses were in the initial top 5 most important features, each 'category' was individually added to the model to identify specific disease or medication groups as predictors (e.g., diabetes). To reduce the number of input features, we selected the features accounting for 100 % of the total predictive value, which could vary per technique. Consequently, features having a feature importance of 0.0 were no longer selected in the next alteration of the model. (see Appendix A). While the selected features could differ per technique due to different contributions to the prediction, we aimed to select features in a consistent and unambiguous manner for all techniques.
- We used thresholds of 0.6 and 0.7 to evaluate and compare the performance of the models.
- Codes were optimized to enhance performance and speed up calculations. Model implementation details are available at https://github.com/keijzera/RiskForPDBasedOnEDR.

2.7. Evaluation of the model

By using the test set, we evaluated whether the developed model performs well in predicting PD. By providing a patient's full EDR, without disclosing the outcome (case/control label), the developed model predicted whether that patient would be a case or control. If the risk for PD was assessed higher than the threshold, it was assumed as a case and if the risk was lower, it was assumed as a control. The performance of various techniques was evaluated by a 5-fold cross-validation of the area under the receiver operating characteristic curve (AUROC), sensitivity, specificity and accuracy (Result section 3.3). The AUROC enabled the optimal threshold value, in which the detection of cases was more important than the detection of controls. To analyze the most important features (Result section 3.4), the Leaving One Feature Out technique was used for temporal techniques and the feature importance analysis was used for non-temporal techniques.

2.8. Validation of the model

The best-performing model was applied to a validation set and a risk per individual was predicted. Patients were classified into high risk (>60 %) and moderate/low risk (<60 %). A 5-fold cross-validation of the AUROC, sensitivity, specificity and accuracy was reported. The estimated risks of cases and controls were plotted per risk class (Result section 3.5). Last, we analyzed how much earlier the model was able to detect PD, compared to the moment of diagnosis (Result Section 3.6). By dividing the participants' EDRs into two periods: X years were observed (observation period) and based on these data, the risk of developing PD in the future was predicted (prediction period). An observation period was created by selecting those EDRs available at 3, 2 and 1 year before diagnosis. (i.e. EDRs recorded >1095 days, >730 days and >365 days before diagnosis). The prediction period included the remaining years (i.e., until they could have been diagnosed). This was carried out on the full dataset. These analyses explored whether the application of our model would be able to support an earlier detection of PD and prevention opportunities.

3. Results

3.1. The selection of EDRs

Fig. 1 shows the steps in selecting EDRs for this study (criteria outlined in the Method section). First, de-identified EDRs from 522,801 patients were extracted from the Bigmouth repository. After the data were cleaned, 401,951 unique records remained. These were labeled into cases and controls. By verifying the initial labels, we excluded the 'cases' who did not have any pocket >5 mm recorded and 'controls' who

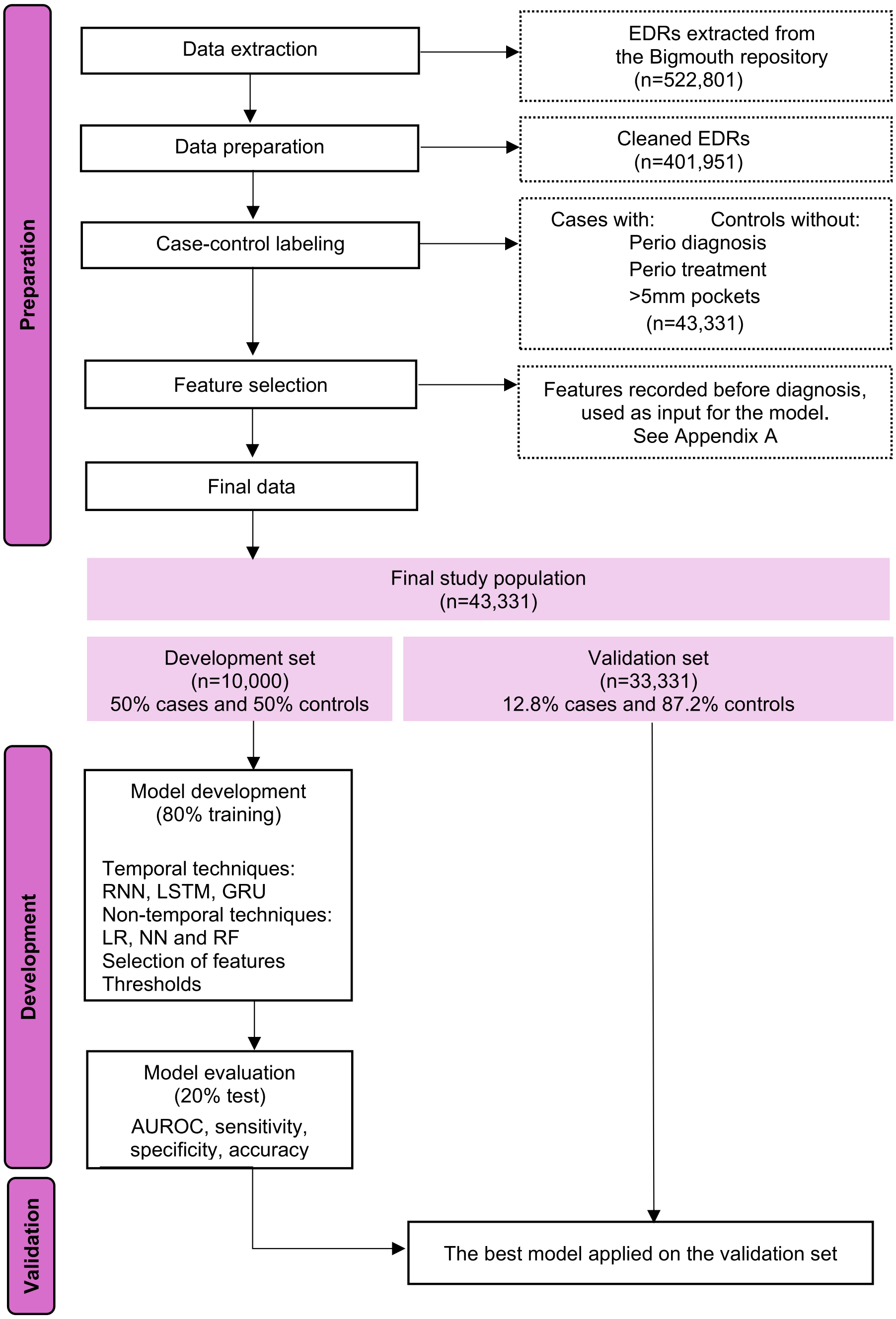

Fig. 1. Flowchart of the selection of EDR patient records used in this study.

had pockets >5 mm recorded in their EDRs, to include EDR data with assured case-control distinctions. EDRs from 43,331 patients were included in the final study population. From these patients, EDRs from 10,000 patients were included in the development set, and the remaining 33,331 patients were included in the validation set.

3.2. *Background characteristics of the study population*

Table 1 shows the background characteristics of the final study population (N = 43,331), grouped by cases (21 %) and controls (78.6 %). As per the design, the development set had an equal case-control ratio (50 %/50 %), while in the validation set, the case-control ratio was 21.8 % - 87.2 %, respectively. Patients from three institutions were included with consistent, complete and correct PD data. The age of the full study population was 46 years. PD cases were older (55 years) than controls (44 years). The case group included as many males as females, while the control group included more females. The EDR data from cases included fewer visits (9) and a shorter follow-up length (1.5 years) compared to the number of visits (11) and follow-up length (2.6 years) from controls.

3.3. *Performance of the model*

Table 2 shows the performance of various techniques evaluated on the test set (N = 2000), with a threshold of 0.6 and 0.7. The AUROC of the models varied between 0.50 and 0.94. Temporal techniques had an AUROC of 0.71 and 0.72, including a sufficient sensitivity for LSTM and RNN models (0.61–0.69) and a good sensitivity for the 0.6th GRU model (0.76). However, the AUROC did not change for any model when features were left out of the model. Since these temporal techniques seemed to perform insufficiently for these data, alternative techniques were employed to develop non-temporal prediction models. Basic NN, LR, and RF techniques were used to build a model based on the last record of each feature. The NN and LR showed a much lower AUROC (0.5 and 0.55 resp.), including unbalanced sensitivities and specificities (±1.0 and 0.0 resp.). An RF showed an excellent AUROC (0.94), including a good sensitivity (0.81) and specificity (0.91) when the threshold was set at 0.6. The accuracy was also the highest for RF (0.86) compared to other techniques (0.50–0.75). Based on the highest AUROC and sensitivity, the RF model was used for further analyses.

3.4. *The most important features for PD prediction*

The developed RF model is based on 38 features with predictive value. Table 3 lists the features responsible for ≥1.0 % of the prediction.

Table 1
Background characteristics of the study population.

	Study population	PD cases	Controls
N (%)	43,331 (100 %)	9271 (21.4 %)	34,060 (78.6 %)
Sub datasets			
Development set	10,000 (100 %)	5000 (50 %)	5000 (50 %)
Validation set	33,331 (100 %)	4271 (12.8 %)	29,060 (87.2 %)
n per institution			
2	11,137 (100 %)	1172 (10.5 %)	9965 (89.5 %)
3	22,669 (100 %)	6598 (29.1 %)	16,071 (70.9 %)
7	9525 (100 %)	1501 (15.8 %)	8024 (84.2 %)
Age* (avg)	46 yrs	55 yrs	44 yrs
Sex			
Female	25,749	4763	20,986
Male	17,404	4485	12,919
Other	7	1	6
Unknown	171	22	149
EDR length (avg) In days (years)	841 days (2.3 years)	547 days (1.5 years)	945 days (2.6 years)
# of visits	11	9	11

* at the last visits for controls, at the diagnosis date for cases.

Table 2
Performance results of PD detection applying several ML and DL techniques.

	th	AUROC	Sensitivity	Specificity	Accuracy
Temporal techniques*					
LSTM	0.6	0.71	0.68	0.55	0.68
	0.7	0.71	0.63	0.73	0.63
GRU	0.6	0.71	0.76	0.57	0.69
	0.7	0.71	0.42	0.81	0.75
RNN	0.6	0.71	0.69	0.56	0.69
	0.7	0.72	0.61	0.79	0.61
Non-temporal techniques**					
LR	0.6	0.5	0	1	0.50
NN	0.6	0.55	0.99	0.09	0.54
	0.7	0.71	0.13	>0.99	0.57
RF	0.6	0.94	0.81	0.91	0.86
	0.7	0.94	0.71	0.96	0.83

*Temporal techniques used all records documented over time.
**Non-temporal techniques used only the last record of each feature.
th = threshold.

Table 3
Most important features in the final RF model.

Features contributing for ≥1 %	Feature importance (%)
1. Bleeding proportion	18.5 %
2. Age	17.8 %
3. Number of visits	7.8 %
4. Preventive treatment	6.0 %
5. Smoking and drug use	5.3 %
6. Plaque proportion	3.6 %
7. Prescriptions	3.5 %
8. Health status	3.1 %
9. Pain and problems	3.1 %
10. Medical diseases	3.0 %
11. Allergy	2.7 %
12. Systolic blood pressure	2.6 %
13. Diastolic blood pressure	2.5 %
14. Medication (over-the-counter)	2.2 %
15. Restorative treatment	2.1 %
16. Oral and maxillofacial treatment	1.9 %
17. Weight	1.7 %
19. Caries	1.5 %
20. Endodontic diseases	1.3 %
21. Endodontic treatment	1.2 %
22–38	<1.0 %

The bleeding proportion, age, number of visits, preventive treatment, smoking and drug use were found to be the most important features for predicting PD. A full list of the most important features and their contribution to the prediction is displayed in Appendix A.

3.5. *Validation*

Table 4 shows our model's performance on the validation set (N = 33,331, including 4271 cases and 29,060 controls). When the developed model was applied to the validation set, the risk was estimated for each patient. The risk of 22,920 patients was 0.6 or higher, and the risk for 10,411 patients was lower than 0.6. The model had an AUROC of 0.85 on the validation set. By screening those patients in the high-risk class, almost all cases could be detected (sensitivity 91 %). However, only 54

Table 4
Performance of the final model applied on the validation set.

Estimated risk	High risk (0.6 – 1.0)	Low risk (0.0 – 0.6)
N	22,920	10,411
Cases	3907	364
Controls	19,013	10,047
Performance	Sensitivity = 0.91	Specificity = 0.54
AUROC	0.85	

% of all controls were found in the estimated low-risk class. Fig. 2 shows that the estimated risk for cases is ascending, but the risk for controls is equally distributed over the estimated risk classes.

3.6. Earlier detection

Fig. 3 shows whether the developed model could detect PD earlier in the patient's EDR than clinically diagnosed. When the model predicted the risk for PD one year before PD was diagnosed, the AUROC was 0.84 (sensitivity = 0.91 and specificity = 0.57). When the model observed the available EDRs two years before the PD diagnosis, the AUROC was 0.83 (sensitivity = 0.89 and specificity = 0.57). When the prediction was made three years earlier, the AUROC was 0.82 (sensitivity = 0.89 and specificity = 0.57). The performance of the predictions barely decreased when the prediction was made 1, 2, or 3 years before the diagnosis.

4. Discussion

This study aimed to develop a ML model to predict the risk for PD based on non-image, longitudinal EDRs, intending to support the early detection and prevention of PD.

With an AUROC of 0.94 (Result section 3.3), the performance of our model could be considered as 'excellent' according to Simundic's classification of diagnostic accuracy [43]. Compared to other AI models predicting PD [44], our AUROC is similar (AUC ≈ 0.90–0.95) and our accuracy (0.86) is slightly lower than others (Acc ≈ 0.90–0.95). Another systematic review of AI models for diagnosing PD, shows even lower accuracies (between 47 and 81 %) by using intraoral photographs [30]. AI models focusing on indirect PD outcomes (e.g. alveolar bone loss) report accuracies between 73 %−99 % [45,46], which is similar to our performance. However, imaging data capture the disease (i.e. by irreversible bone loss), while our non-image data are meant for prevention purposes (i.e. predicting a risk before the disease manifests). Additionally, imaging data perform better than non-imaging data in DL studies in periodontology [31]. Studies that used non-imaging data aimed to predict the onset of PD or its severity, alveolar bone loss, pain, oral malodor, or PD treatment outcomes. These models still reached accuracies between 0.73 and 0.98, but sample sizes were much lower (max. $N = 7500$) than in our study ($N = 43{,}331$).

While the overall performance seems promising, individual performance metrics must be evaluated as well to prevent over and underestimation. When our model was applied to the validation set, the sensitivity and specificity diverged significantly (Result section 3.5). The sensitivity increased from 81 % on the development set to 91 % on the validation set. In comparison, the sensitivity of clinical oral health examinations was evaluated much lower (42–56 %) [47]. However, the drop in our specificity from 91 % to 54 % raises concerns. The low specificity of our model is likely due to our model's overfit to cases in the training set, which reduced its ability to detect controls. By prioritizing ultimately prevention, we conceived the detection of cases as more important than the detection of controls. Therefore, we selected a higher proportion of cases (5000:5000) for training than the PD prevalence, while our imbalanced validation set included many more controls (4271:29,060). Therefore, the model needs further training on control data. Some controls may have a high risk for PD but are not yet diagnosed. Thus, while the model effectively detects cases, it seems to overestimate the controls in this study, reflecting its training focus and intentional overfitting to cases.

This study showed that no disease development was detectable over time. Up front, we assumed that PD develops slowly, so we explored whether it was detectable in longitudinal EDR [26]. However, our analyses showed that temporal techniques performed insufficiently in detecting cases (Result section 3.3). A systemic review of the use of generative AI on EHRs similarly found that the history of a patient's illness becomes omitted by reliance on static data [48].The development of some diseases might be detectable in data that are continuously measured in time-series, e.g., an ECG [49]. The EHRs in our study were not as regularly timed at specific time intervals as time-series, and are therefore assumed as event data [50]. Compared to patterns in an ECG or continuous glucose monitoring, it is not surprising that we could not detect any 'progression' in EDRs. However, it appears that PD can be predicted more accurately based on only the last record of each feature, rather than based on all records over time. Since the timing of forecasting barely affected the performance, we conclude that a simple cross-sectional prediction is preferable to a temporal prediction. We could have expected this because other models in dentistry could also

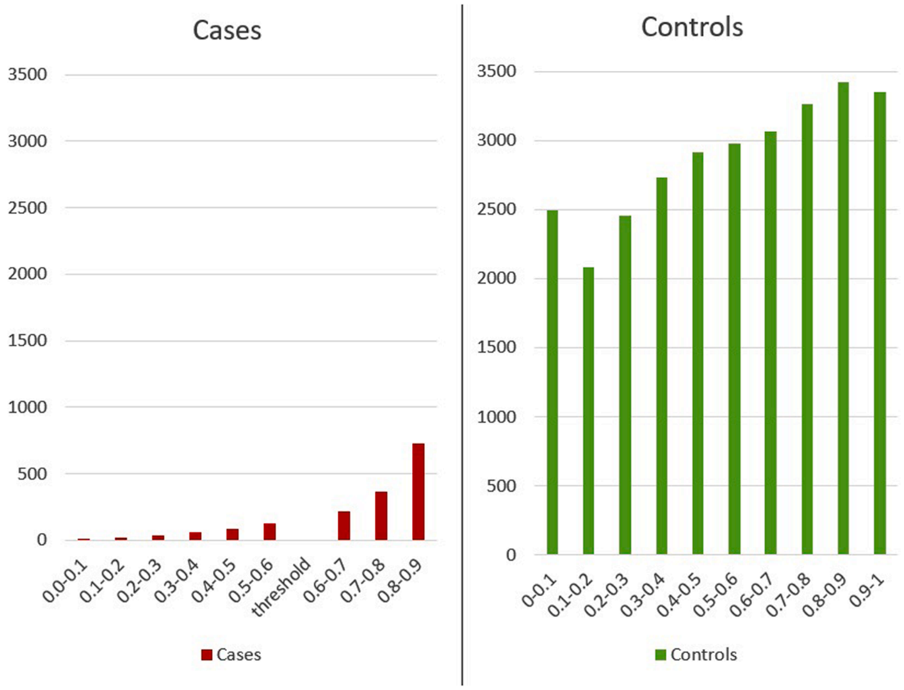

Fig. 2. Graphical presentation of the number (y-axis) of cases (panel A) and controls (panel B) per risk class (x-axis).

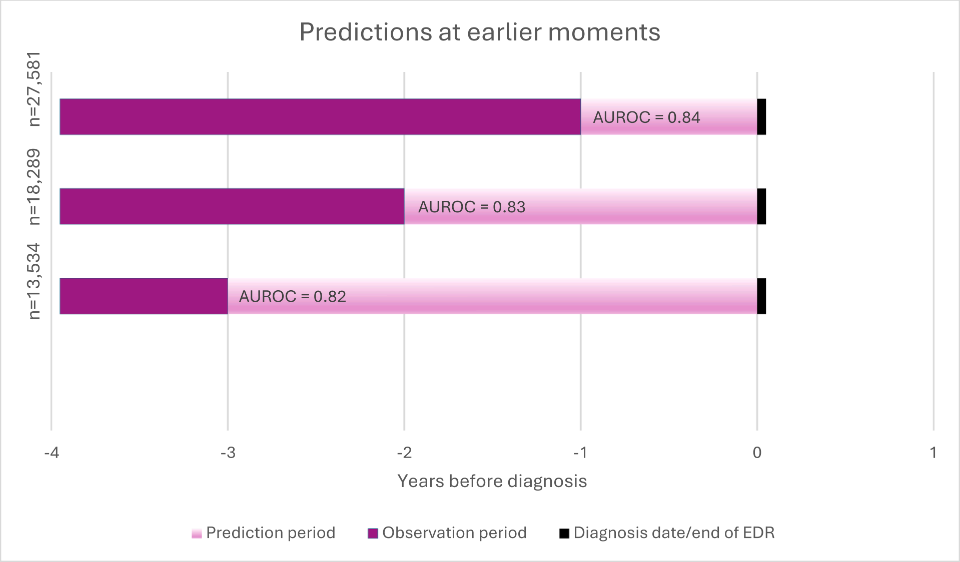

Fig. 3. The performance of predictions made at 1, 2 and 3 years before the diagnosis (x-axis).

successfully predict dental diseases with basic LR, RF, SVM, and neural network techniques [51]. A Random Forest is indeed well-known for its high performance in basic classification, irrespective of the number of features [42,52].

The most important features found in our study (Results section d) are well-known for PD, such as bleeding on probing, increased age, the amount of plaque, certain medical diseases and hypertension [2,11]. The fact that the prediction is based upon 38 features confirms that PD is a disease with a multicausal etiology, with most of them found in lifestyle, systemic conditions, and dentition-related clusters [53]. While age and chronic medical diseases are often fairly well-documented in the patient's EDR, blood pressure is less common to document in EDR but was still found as an important feature. In contrast, brushing, saliva test results and diet are well-known to be associated with PD risk [54–56], but interestingly, they were not found to be a relevant feature in our study. It is possible that factors may not be identified as the most important features because they are scarcely registered, thus being buried by the sheer volume of data and loss of data content. Therefore, we can only confirm the significance of the most important features that are found in this study, but not the irrelevance of the features that were not found. As the interaction between predictors plays a role, their individual contribution might be skewed. This study provides new insights into the relevance of the dental history, since prior preventive treatment, the number of dental visits, and earlier restorative treatment belong to the main predictors. We conclude that the identified features could significantly bolster preventive care, provided they are consistently and thoroughly documented. After all, the features found in this study depend on the documentation by clinicians.

The results of this study are based on a selection of EDRs that was consistently and completely documented. On one hand, this study affirms the 'imperfection' of non-imaging EDRs. We have to highlight that a large proportion of the overall EDRs were excluded because of different documentation standards (Results section a). A recently published framework might help with analyses of heterogeneous EHR data [57]. The EDRs that were consistently and completely recorded were primarily focused on billing and scheduling, while EDRs focused on prevention included many missings (e.g., oral health behavior and specific risk factors). The less these features are recorded, the less predictive power could be discovered. On the other hand, this study shows that when these EDRs are consistently recorded, they appear to be capable of succeeding in an accurate risk prediction, at least to detect cases. It is imperative to recognize that these 'imperfect' data are also currently leading our clinical practice. Dental clinicians make decisions based on the available information, even though it may be incomplete or inaccurate. It will be interesting to explore how the model performs with lower-quality input data in further research. We demonstrate that with optimal record keeping, AI techniques can help clinicians in the prediction of PD.

A limitation of this study was the case-control labeling. Due to the extensive sample size disclosed, we labeled based on codes instead of manual labeling. By employing three criteria (PD diagnosis, PD treatment, and pocket depths > 5 mm), we ensured accurate labeling of both groups, but a major limitation of these criteria is that the sample size was reduced significantly. As a result of strict labeling criteria, many 'grey area' patients who met only 1 or 2 out of the 3 criteria, were excluded and may go underdetected, such as those with early stages PD or those opting to delay treatment. First, the initial case labeling was affected by the absence of the diagnosis documentation in some EDRs. As it turned out that some academic institutions did not document diagnoses in their EDR, we could not include participants from these institutions. In retrospect, we could have foreseen this, since the completeness of PD diagnosis documentation varied between 8 % and 99 % across the participating US institutions [58]. Automated diagnosis documentation might be a promising solution for future research [59]. By exploring the pocket depth data, we found that some 'controls' had pockets > 5 mm and 'cases' did not have pockets >5mm, at least according to their EDR. Therefore, we extended the labeling criteria by adding pocket depth. By sharpening the case-control labeling, the sample size was again reduced considerably, but the validity of the labels improved. As a result, the model was reliably developed and still based on an adequate amount of data. These differences in recording standards of the PD criteria across the participating institutions led to unfavorable decisions for the research. For the clinical application of the model to new EDR, these PD criteria do not need to be documented.

The large sample size ($N = 43.331$), the use of real-world EDRs (assumed as representative [40] and confirmed by a similar prevalence), and the inclusion of multiple institutions in our dataset represent strengths of this study. However, the fact that our model was developed and validated within the same population presents a limitation and limits the model's generalizability. It seems that the performance of predictive models for PD might be excellent during internal testing, but it dropped when validated externally [60]. This was already seen when our model was applied to an internal validation set, and it might drop even further in populations outside the U.S., non-academic settings, or

EDRs recorded in other software. In addition, we included a subset of EDRs that were extensively and well-documented, which does not represent the EDRs in typical clinical settings, which limits clinical applicability. For clinical use, we suggest that a minimum set of recently updated input features should be required to make a prediction valid. Short questionnaires can be offered digitally to dental patients, asking for the most important predictors or updating these if something has changed. Nevertheless, future clinical implications require improved generalizability, applicability, and integration into EDR software.

5. Conclusions

Based on multicenter consistent and complete EDR, ML showed an excellent ability to assist with the early detection of PD. 91 % of all PD cases could be detected based on their EDRs with Random Forest techniques. Further research is required to explore controls and further improve the specificity. If such ML models become accurate enough in classifying both cases and controls, future clinicians can be assisted with personalized risk predictions and suggestions for targeted prevention interventions based on the individual patient's circumstances. Improved documentation of EDR records is an important first step.

CRediT authorship contribution statement

Laura Swinckels: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Project administration, Validation. **Ander de Keijzer:** Conceptualization, Methodology, Software, Formal analysis, Writing – review & editing, Supervision. **Bruno G. Loos:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Reuben Joseph Applegate:** Resources, Data curation. **Krishna Kumar Kookal:** Resources, Writing – review & editing, Project administration. **Elsbeth Kalenderian:** Conceptualization, Resources, Writing – review & editing, Supervision. **Harmen Bijwaard:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Josef Bruers:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Declaration of competing interest

We hereby declare that this manuscript is original and has not been submitted or published elsewhere. All authors have contributed to and approved the submission of this manuscript. Furthermore, there are no conflicts of interest to disclose.

Declaration of generative AI in the writing process

During the preparation of this work the authors used ChatGPT in order to improve the readability of this manuscript. After using this tool, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the published article.

Acknowledgment

We extend our sincerest gratitude to Erik Ellinger for his invaluable assistance in data management within the SQL server environment and for streamlining the intricate data transformations necessary to ensure the structure and quality of the data. Furthermore, we would like to express our appreciation to the member institutions of the BigMouth repository, who made the extraction of their data for this study possible. This collaboration was essential in facilitating our research and advancing the understanding of ML models using electronic dental records.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jdent.2024.105469.

References

[1] B.L. Pihlstrom, B.S. Michalowicz, N.W. Johnson, Periodontal diseases, Lancet 19 (2005) 1809–1820, https://doi.org/10.1016/s0140-6736(05)67728-8.

[2] N.P. Lang, J.E. Suvan, M.S. Tonetti, Risk factor assessment tools for the prevention of periodontitis progression a systematic review, J. Clin. Periodontol. 42 (2015) 59–70, https://doi.org/10.1111/jcpe.12350.

[3] A. Cekici, A. Kantarci, H. Hasturk, T.E. Van Dyke, Inflammatory and immune pathways in the pathogenesis of periodontal disease, Periodontology 64 (2000) 57–80, https://doi.org/10.1111/prd.12002, 2014.

[4] I.L. Chapple, F. Van der Weijden, C. Doerfer, D. Herrera, L. Shapira, D. Polak, P. Madianos, A. Louropoulou, Machtei E, N. Donos, H. Greenwell, A.J. van Winkelhoff, B.E. Kuru, N. Arweiler, W. Teughels, M. Aimetti, A. Molina, E. Montero, F. Graziani, Primary prevention of periodontitis: managing gingivitis, J. Clin. Periodontol. 42 (2015) 71–76, https://doi.org/10.1111/jcpe.12366.

[5] D.C. Matthews, Prevention and treatment of periodontal diseases in primary care, Evid. Based Dent. 15 (2014) 68–69, https://doi.org/10.1038/sj.ebd.6401036.

[6] Jr A.C. Tanner, R. Kent, T. Van Dyke, S.T. Sonis, L.A. Murray, Clinical and other risk indicators for early periodontitis in adults, J. Periodontol. 76 (2005) 573–581, https://doi.org/10.1902/jop.2005.76.4.573.

[7] R. Palmer, P. Floyd, Periodontal examination and screening, Br. Dent. J. 235 (2023) 707–713, https://doi.org/10.1038/s41415-023-6410-8.

[8] K.S. Kornman, M.G. Newman, R. Alvarado, T.F. Flemmig, S. Nachnani, J. Tumbusch, Clinical and microbiological patterns of adults with periodontitis, J. Periodontol. 62 (1991) 634–642, https://doi.org/10.1902/jop.1991.62.10.634.

[9] L. Trombelli, C. Scapoli, G. Calura, D.N. Tatakis, Time as a factor in the identification of subjects with different susceptibility to plaque-induced gingivitis, J. Clin. Periodontol. 33 (2006) 324–328, https://doi.org/10.1111/j.1600-051x.2006.00914.x.

[10] D.N. Tatakis, L. Trombelli, Modulation of clinical expression of plaque-induced gingivitis, I. Background review and rationale, J. Clin. Periodontol. 31 (2004) 229–238, https://doi.org/10.1111/j.1600-051x.2004.00477.x.

[11] B.L. Pihlstrom, Periodontal risk assessment, diagnosis and treatment planning, Periodontology 25 (2000) 37–58, https://doi.org/10.1034/j.1600-0757.2001.22250104.x, 2001.

[12] E. Raittio, R. Lopez, V. Baelum, Contesting the conventional wisdom of periodontal risk assessment, Comm. Dent. Oral Epidemiol. 52 (2024) 487–498, https://doi.org/10.1111/cdoe.12942.

[13] J.S. Patel, D. Shin, L. Willis, A. Zai, K. Kumar, T.P. Thyvalikakath, Comparing gingivitis diagnoses by bleeding on probing (BOP) exclusively versus BOP combined with visual signs using large electronic dental records, Sci. Rep 13 (2023) 17065, https://doi.org/10.1038/s41598-023-44307-z.

[14] T.P. Thyvalikakath, R. Padman, K. Vyawahare, P. Darade, R. Paranjape, Utilizing dental electronic health records data to predict risk for periodontal disease, Stud Health Technol. Inform 216 (2015) 1081, https://doi.org/10.3233/978-1-61499-564-7-1081.

[15] O. Tokede, R.B. Ramoni, M. Patton, J.D. Da Silva, E. Kalenderian, Clinical documentation of dental care in an era of electronic health record use, J. Evid. Based Dental Pract. 16 (2016) 154–160, https://doi.org/10.1016/j.jebdp.2016.07.001.

[16] M. Song, K. Liu, R. Abromitis, T.L. Schleyer, Reusing electronic patient data for dental clinical research: a review of current status, J. Dent 41 (2013) 1148–1163, https://doi.org/10.1016/j.jdent.2013.04.006.

[17] W.Q. Wei, J.C. Denny, Extracting research-quality phenotypes from electronic health records to support precision medicine, Genome Med. 7 (2015) 41, https://doi.org/10.1186/s13073-015-0166-y.

[18] B. Tokede, A. Yansane, J. White, S. Bangar, J. Mullins, R. Brandon, S. Gantela, K. Kookal, D. Rindal, C.T. Lee, G.H. Lin, H. Spallek, E. Kalenderian, M. Walji, Translating periodontal data to knowledge in a learning health system, J. Am. Dental Asso. 153 (2022) 996–1004, https://doi.org/10.1016/j.adaj.2022.06.007.

[19] B.G. Arndt, J.W. Beasley, M.D. Watkinson, J…L. Temte, W.J. Tuan, C.A. Sinsky, V. J. Gilchrist, Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations, Ann. Fam. Med. 15 (2017) 419–426, https://doi.org/10.1370/afm.2121.

[20] J.L. Howe, K.T. Adams, A.Z. Hettinger, R.M. Ratwani, Electronic Health Record Usability Issues and Potential Contribution to Patient Harm, JAMa 319 (2018) 1276–1278, https://doi.org/10.1001/jama.2018.1171.

[21] C.A. Stevens, A.R. Lyons, K.I. Dharmayat, A. Mahani, K.K. Ray, A.J. Vallejo-Vaz, M. T.A. Sharabiani, Ensemble machine learning methods in screening electronic health records: a scoping review, Digit. Health 9 (2023) 1–17, https://doi.org/10.1177/20552076231173225.

[22] S.N. Payrovnaziri, Z. Chen, P. Rengifo-Moreno, T. Miller, J. Bian, J.H. Chen, X. Liu, Z. He, Explainable artificial intelligence models using real-world electronic health record data: a systematic scoping review, J. Am. Med. Inform. Ass. 27 (2020) 1173–1185, https://doi.org/10.1093/jamia/ocaa053.

[23] S. Yang, P. Varghese, E. Stephenson, K. Tu, J. Gronsbell, Machine learning approaches for electronic health records phenotyping: a methodical review, J. Am. Med. Inform. Assoc. 30 (2023) 367–381, https://doi.org/10.1093/jamia/ocac216.

[24] C. Janiesch, P. Zschech, K. Heinrich, Machine learning and deep learning, Electr. Markets 31 (2021) 685–695, https://doi.org/10.1007/s12525-021-00475-2.

[25] N.R. Panda, S.S. Satapathy, S.K. Bhuyan, R. Bhuyan, Impact of machine learning and prediction models in the diagnosis of oral health conditions, Int. J. Stat. Med. Res. 12 (2023) 51–57, https://doi.org/10.6000/1929-6029.2023.12.07.

[26] L. Swinckels, F.C. Bennis, K.A. Ziesemer, J.F.M. Scheerman, H. Bijwaard, A. de Keijzer, J.J. Bruers, The use of deep learning and machine learning on longitudinal electronic health records for the early detection and prevention of diseases: scoping review, J. Med. Internet Res. 26 (2022) 48320, https://doi.org/10.2196/48320.

[27] J.S. Patel, C. Su, M. Tellez, J.M. Albandar, R. Rao, V. Iyer, E. Shi, H. Wu, Developing and testing a prediction model for periodontal disease using machine learning and big electronic dental record data, Front. Artific. Intellig. 5 (2022) 979525, https://doi.org/10.3389/frai.2022.979525.

[28] V.P. Kearney, A.M. Yansane, R.G. Brandon, R. Vaderhobli, G.H. Lin, H. Hekmatian, W. Deng, N. Joshi, H. Bhandari, A.S. Sadat, J.M. White, A generative adversarial inpainting network to enhance prediction of periodontal clinical attachment level, J. Dent. 123 (2022) 104211, https://doi.org/10.1016/j.jdent.2022.104211.

[29] C.T. Lee, T. Kabir, J. Nelson, S. Sheng, H.W. Meng, T.E. van Dyke, M.F. Walji, X. Jiang, S. Shams, Use of the deep learning approach to measure alveolar bone level, J. Clin. Periodontol. 49 (2022) 260–269, https://doi.org/10.1111/jcpe.13574.

[30] M. Revilla-León, M. Gómez-Polo, A.B. Barmak, W. Inam, J.Y.K. Kan, J.C. Krois, O. Akal, Artificial intelligence models for diagnosing gingivitis and periodontal disease: a systematic review, J. Prosthet. Dent 130 (2022) 816–824, https://doi.org/10.1016/j.prosdent.2022.01.026.

[31] H. Mohammad-Rahimi, S.R. Motamedian, Z. Pirayesh, A. Haiat, S. Zahedrozegar, E. Mahmoudinia, M.H. Rohban, J. Krois, J.H. Lee, F. Schwendicke, Deep learning in periodontology and oral implantology: a scoping review, J. Periodont. Res. 57 (2022) 942–951, https://doi.org/10.1111/jre.13037.

[32] G.S. Collins, J.B. Reitsma, D.G. Altman, K.G.M. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement, The BMJ 350 (2015) g7594, https://doi.org/10.1136/bmj.g7594.

[33] M.F. Walji, E. Kalenderian, P.C. Stark, J.M. White, K.K. Kookal, D. Phan, D. Tran, E. V. Bernstam, R. Ramoni, BigMouth: a multi-institutional dental data repository, J. Am. Med. Inform. Ass. 21 (2014) 1136–1140, https://doi.org/10.1136/amiajnl-2013-002230.

[34] M.F. Walji, H. Spallek, K.K. Kookal, J. Barrow, B. Magnuson, T. Tiwari, U. Oyoyo, M. Brandt, B.J. Howe, G.C. Anderson, J.M. White, E. Kalenderian, BigMouth: development and maintenance of a successful dental data repository, J. Am. Inform. Assoc. 29 (2022) 701–706, https://doi.org/10.1093/jamia/ocac001.

[35] J.M. White, E. Kalenderian, P.C. Stark, R.L. Ramoni, R. Vaderhobli, M.F. Walji, Evaluating a dental diagnostic terminology in an electronic health record, J. Dent. Educ. 75 (2011) 605–615, https://doi.org/10.1002/j.0022-0337.2011.75.5.tb05086.x.

[36] G.C. Armitage, Development of a classification system for periodontal diseases and conditions, Ann. Periodont. 4 (1999) 1–6, https://doi.org/10.1902/annals.1999.4.1.1.

[37] P.N. Papapanou, M. Sanz, N. Buduneli, T. Dietrich, M. Feres, D.H. Fine, T. F. Flemmig, R. Garcia, W.V. Giannobile, F. Graziani, H. Greenwell, D. Herrera, R. T. Kao, M. Kebschull, D.F. Kinane, K.L. Kirkwood, T. Kocher, K.S. Kornman, P. S. Kumar, B.G. Loos, E. Machtei, H. Meng, A. Mombelli, I. Needleman, S. Offenbacher, G.J. Seymour, R. Teles, M.S. Tonetti, Periodontitis: consensus report of workgroup 2 of the 2017 world workshop on the classification of periodontal and Peri-implant diseases and conditions, J. Periodontol. 89 (2018) S173–SS82, https://doi.org/10.1002/jper.17-0721.

[38] N.J. Kassebaum, E. Bernabé, M. Dahiya, B. Bhandari, C.J. Murray, W. Marcenes, Global burden of severe periodontitis in 1990-2010: a systematic review and meta-regression, J. Dent. Res. 93 (2014) 1045–1053, https://doi.org/10.1177/0022034514552491.

[39] D. Trindade, R. Carvalho, V. Machado, L. Chambrone, J.J. Mendes, J. Botelho, Prevalence of periodontitis in dentate people between 2011 and 2020: a systematic review and meta-analysis of epidemiological studies, J. Clin. Periodontol. 50 (2023) 604–626, https://doi.org/10.1111/jcpe.13769.

[40] M. Qi, O. Cahan, M.A. Foreman, D.M. Gruen, A.K. Das, K.P. Bennett, Quantifying representativeness in randomized clinical trials using machine learning fairness metrics, JAMIa Open. 4 (2021) ooab077, https://doi.org/10.1093/jamiaopen/ooab077.

[41] S. Dreiseitl, L. Ohno-Machado, Logistic regression and artificial neural network classification models: a methodology review, J. Biomed. Inform 35 (2002) 352–359, https://doi.org/10.1016/s1532-0464(03)00034-0.

[42] A.B. Shaik, S. Srinivasan, A brief survey on random forest ensembles in classification model, in: S. Bhattacharyya, A. Hassanien, D. Gupta, A. Khanna, I. Pan (Eds.), International conference on innovative computing and communications, Lecture Notes in Networks and Systems, Springer, Singapore, 2019, pp. 253–260.

[43] A.M. Šimundić, Measures of diagnostic accuracy: basic definitions, J. Internat. Federat. Clin. Chem. Labor. Med. 19 (2009) 203–211.

[44] A. Polizzi, V. Quinzi, A. Lo Giudice, G. Marzo, R. Leonardi, G. Isola, Accuracy of artificial intelligence models in the prediction of periodontitis: a systematic review, JDR Clin. Translat. Res. 9 (2024), https://doi.org/10.1177/23800844241232318, 23800844241232318.

[45] J. Issa, M. Jaber, I. Rifai, P. Mozdziak, B. Kempisty, M. Dyszkiewicz-Konwinska, Diagnostic test accuracy of artificial intelligence in detecting periapical periodontitis on two-dimensional radiographs: a retrospective study and literature review, Medicina (B Aires) 59 (2023) 768, https://doi.org/10.3390/medicina59040768.

[46] M. Revilla-Leon, M. Gomez-Polo, A.B. Barmak, W. Inam, J.Y.K. Kan, J.C. Krois, O. Akal, Artificial intelligence models for diagnosing gingivitis and periodontal disease: a systematic review, J. Prosthet. Dent 130 (2023) 816–824, https://doi.org/10.1016/j.prosdent.2022.01.026.

[47] R. Chou, S.S. Selph, C. Bougatsos, C. Nix, A. Ahmed, J. Griffin, E. Schwarz, Referral Screening, Behavioral counseling, and preventive interventions for oral health in adults: a systematic review for the US preventive services task force, JAMa 330 (2023) 1780–1790, https://doi.org/10.1001/jama.2023.20685.

[48] X. Du, Z. Zhou, Y. Wang, Y.W. Chuang, R. Yang, W. Zhang, X. Wang, R. Zhang, P. Hong, D.W. Bates, L. Zhou, Generative large language models in electronic health records for patient care since 2023: a systematic review (preprint), medRxiv. (2024). https://doi.org/10.1101/2024.08.11.24311828.

[49] Q.J. Lv, H.Y. Chen, W.B. Zhong, Y.Y. Wang, J.Y. Song, S.D. Guo, L.X. Qi, C.Y. C. Chen, A multi-task group Bi-LSTM networks application on electrocardiogram classification, IEEE J. Transl. Eng. Health Med 8 (2020) 1–11, https://doi.org/10.1109/JTEHM.2019.2952610.

[50] R. AlSaad, A. Abd-Alrazaq, S. Boughorbel, A. Ahmed, M.A. Renault, R. Damseh, J. Sheikh, Multimodal large language models in health care: applications, challenges, and future outlook, J. Med. Internet Res. 26 (2024) e59505, https://doi.org/10.2196/59505.

[51] W. Beak, J. Park, S. Ji, Data-driven prediction model for periodontal disease based on correlational feature analysis and clinical validation, Heliyon. 10 (2024) e32496, https://doi.org/10.1016/j.heliyon.2024.e32496.

[52] J. Park, J.W. Kim, B. Ryu, E. Heo, S.Y. Jung, S. Yoo, Patient-level prediction of cardio-cerebrovascular events in hypertension using nationwide claims data, J. Med. Internet Res. 21 (2019) e11757, https://doi.org/10.2196/11757.

[53] B.G. Loos, T.E. Van Dyke, The role of inflammation and genetics in periodontal disease, Periodontology 83 (2020) 26–39, https://doi.org/10.1111/prd.12297, 2000.

[54] G. Mainas, P. Santamaria, M. Ide, V. Longo, M. Vinciguerra, J. Nart, L. Nibali, Could dietary restrictions affect periodontal disease? A systematic review, Clin. Oral Investig 27 (2023) 4107–4116, https://doi.org/10.1007/s00784-023-05052-9.

[55] H. Zimmermann, N. Zimmermann, D. Hagenfeld, A. Veile, T.S. Kim, H. Becher, Is frequency of tooth brushing a risk factor for periodontitis? A systematic review and meta-analysis, Comm. Dent Oral Epidemiol 43 (2015) 116–127, https://doi.org/10.1111/cdoe.12126.

[56] J. Jeong, H.S. Kim, D. Lee, K. Kim, Y.H. Kim, Association between four dietary patterns and the risk of periodontal diseases: a systematic review and meta-analysis, Nutrients. 14 (2022) 4363, https://doi.org/10.3390/nu14204362.

[57] L. Heumos, P. Ehmele, T. Treis, J. Upmeier zu Belzen, E. Roellin, L. May, A. Namsaraeva, N. Horlava, V.A. Shitov, X. Zhang, L. Zappia, R. Knoll, N.J. Lang, L. Hetzel, I. Virshup, L. Sikkema, F. Curion, R. Eils, H.B. Schiller, A. Hilgendorff, F. J. Theis, An open-source framework for end-to-end analysis of electronic health record data, Nat. Med. (2024), https://doi.org/10.1038/s41591-024-03214-0.

[58] J. Mullins, A. Yansane, S.V. Kumar, S. Bangar, A. Neumann, T.R. Johnson, G. E. Olson, K.K. Kookal, E. Sedlock, A. Kim, E. Mertz, R. Brandon, K. Simmons, J. M. White, El. Kalenderian, M.F. Walji, Assessing the completeness of periodontal disease documentation in the EHR: a first step in measuring the quality of care, BMC. Oral Health 21 (2021) 282, https://doi.org/10.1186/s12903-021-01633-w.

[59] J. Patel, D. Shin, L. Willis, A. Zai, T. Thyvalikakath, Feasibility of utilizing electronic dental record data and periodontitis case definition to automate diagnosis, Stud. Health Technol. Inform. 310 (2024) 214–218, https://doi.org/10.3233/shti230958.

[60] N.Z. Bashir, Z. Rahman, S.L. Chen, Systematic comparison of machine learning algorithms to develop and validate predictive models for periodontitis, J. Clin. Periodontol. 49 (2022) 958–969, https://doi.org/10.1111/jcpe.13692.