*MEDINFO 2023 — The Future Is Accessible*
*J. Bichel-Findlay et al. (Eds.)*

*doi:10.3233/SHTI230958*

# Feasibility of Utilizing Electronic Dental Record Data and Periodontitis Case Definition to Automate Diagnosis

Jay PATEL[a,b], Daniel SHIN[c], Lisa WILLIS[a], Ahad ZAI[a] and Thankam Thyvalikakath[a,d,1]
[a]*Dental Informatics Division; Department of Cariology, Operative Dentistry, and Dental Public Health; Indiana University School of Dentistry; Indiana University Purdue University Indianapolis, Indianapolis, IN, USA*
[b]*Department of Health Services Administration and Policy, College of Public Health Temple University, Philadelphia, PA, USA*
[c]*Department of Periodontology, Indiana University School of Dentistry; Indiana University Purdue University Indianapolis, USA*
[d]*Center for Biomedical Informatics, Regenstrief Institute, Inc., Indianapolis, Indiana, USA*

**Abstract.** Periodontitis is an irreversible disease leading to tooth loss, and 42% U.S. population suffers from periodontitis. Hence, diagnosing, monitoring, and determining its prevalence is critical to develop preventive strategies. However, a nationwide epidemiological study estimating the prevalence reported a concern about the discontinuation of such studies due to cost and ethical reasons. Therefore, this study determined the feasibility of utilizing electronic dental record (EDR) data and periodontitis case definition to automate periodontitis diagnosis. We utilized EDR data from the Indiana University School of Dentistry of 28,908 unique patients. We developed and tested a computer algorithm to diagnose periodontitis using the case definition. We found 44%, 22%, and 1% of patients with moderate, severe, and mild periodontitis, respectively. The algorithm worked with 100% sensitivity, specificity, and accuracy because of the excellent quality of the EDR data. We concluded the feasibility of providing automated periodontitis diagnosis from EDR data to conduct epidemiological studies across the US.

**Keywords.** Periodontitis, electronic dental record, automation, 2012 case definition

## 1. Introduction

Periodontitis is an irreversible chronic condition that leads to tooth loss and poor quality of life. Nearly 42% of U.S. adults are suffering from periodontitis, and 7% of patients are suffering from severe periodontitis. Moreover, it is an irreversible condition; hence, diagnosing periodontitis at the right time is critical [1,2]. Periodontitis is typically diagnosed using patients' intraoral and soft tissue findings, radiographs, and periodontal charting findings such as clinical attachment loss (CAL), bleeding on probing (BOP), and periodontal pocket depth (PPD) [2]. For periodontitis, two major diagnostic criteria

[1] Corresponding Author: Thankam Thyvalikakath, email: tpt@iu.edu.

exist 1) the 2017 staging and grading of periodontitis [2] for clinical care and 2) the 2012 periodontitis case definition for surveillance and prevalence purposes in the U.S. [1,3,4].

The staging and grading periodontitis classification utilizes patients' periodontal charting findings, radiographic findings, medical histories, and social histories. However, the case definition is designed to be low cost with ethical compliance, such as only assessing periodontal findings on limited teeth and not using radiographs [1, 4]. The latest periodontitis prevalence study examining periodontitis in the U.S. utilized patient data collected through the Natural Health and Nutrition Examination Survey (NHANES) between 2009 and 2014 [1]. The study concluded that such a prevalence study is not anticipated in the future due to high costs associated with collecting data using prospective study designs. However, it is also critical to determine disease prevalence, and allocate appropriate resources to manage disease, and estimate healthcare costs [5].

The high use of electronic dental records (EDR) to document patient information could be a potential solution to this problem [6,7]. However, the diagnosis information in dental records is poor but could improve with time and education. Nevertheless, dental clinicians document patients' complete periodontal charting information. However, as per our best knowledge, no study has attempted to automate periodontitis diagnosis from periodontal charting data and using 2012 case definitions for epidemiological purposes [8]. Our long-term goal is to collaborate with other dental schools and private dental practices to examine the prevalence of periodontitis in the U.S. As the first step towards this goal, this study aims to determine the feasibility of generating periodontitis diagnosis automatically using the periodontal charting findings from the EDR data and using 2012 periodontitis case definition [1].

## 2. Methods

### *2.1. Data Source*

The IRB was approved (IRB#1909819686) by the Indiana University School of Dentistry (IUSD) and funded by IUSD, Dr. Thyvalikakath's start-up funds. We utilized EDR (axiUm®-EXAN) data from the pre-doctoral clinics to conduct this study. We included adult patients (>18 yeears) who visited IUSD clinics and received at least one comprehensive oral evaluation (COE) between Jan 1, 2009, to Dec 31, 2014. During the COE, dental students are mandated to document patients' complete periodontal charting information in the EDR. This period was selected because the NHANES epidemiological study estimated the prevalence of periodontitis in the U.S. using the same timeline. Selecting this similar timeline would allow an effective comparison between this study results and the results of the NHANES study [1].

### *2.2. Periodontal Charting Data Pre-Processing*

We developed a computer algorithm that converts and stores each patient's periodontal charting information in individual text files. Each converted text file contained patients' I.D., charting date, periodontal findings (CAL, PPD, or BOP), tooth number, and tooth sites (mesiolingual, mesial, mesiofacial, distal, distolingual, distofacial). The text file format helped in tracking patients' P.D. diagnosis during each visit. By the end of this step, we created individual text files by unique patient I.D. & visit dates for further processing and analysis [9].

### 2.3. Case Definition and Logic Development for Automated Diagnosis

The case definition developed by the American Academy of Periodontology as described in the NHANES study was used to classify periodontitis status into mild, moderate, or severe periodontitis cases [1,8]. Severe periodontitis: Presence of >=2 interproximal sites with ≥ 6 mm CAL (not on the same tooth) and >=1 interproximal site(s) with ≥ 5 mm PPD. Moderate periodontitis: >=2 interproximal sites with ≥ 4 mm clinical CAL (not on the same tooth) or >=2 interproximal sites with PPD ≥ 5 mm, also not on the same tooth. Mild periodontitis: >=2 interproximal sites with ≥ 3 mm CAL and >=2 interproximal sites with ≥ 4 mm PPD (not on the same tooth) or 1 site with ≥ 5 mm (Figure 1).

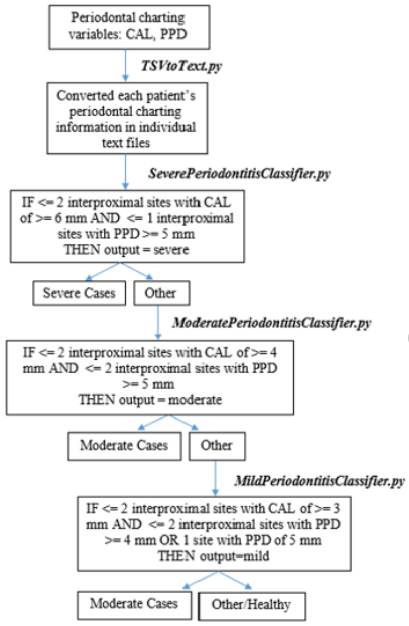


**Figure 1.** Workflow to automate diagnosis.

### 2.4. Rule-Based Algorithm to Classify Periodontitis Automatically

We developed three step rule-based computational algorithms that automatically classified patients' periodontitis status using their CAL and PPD information. As demonstrated in Figure 1, the algorithms utilize filtering approach and examines the maximum CAL and PPD in the patient's periodontal charting. For example, if the patient had two or more interproximal sites with CAL of more than 6 mm, and at least one interproximal site with a PPD of 5 mm or more, the program would classify this patient into the severe periodontitis category.

### 2.5. Manual Review & Examining Performance of Algorithm

Two clinical faculty members reviewed 50 common records and diagnosed patients' periodontitis status based on the rules described in the NHANES study. The inter-rater agreement was 0.9 (Cohen's Kappa value), which indicated excellent agreement. Next, each reviewer reviewed 150 records independently, resulting in an overall dataset of 350

cases. The algorithm's output was then compared with the reviewers' diagnoses. Based on the computer algorithm's ability to correctly diagnose periodontitis cases, true positives, false positives, and false negatives were calculated. Using these measures, we calculated the performance of the automated algorithm's precision, recall, and f-measure.

## 3. Results

### *3.1. Study Cohort*

This study cohort consisted of 28,908 unique patients who received at least one COE between January 1, 2009, and December 31, 2014. The mean age of the patient population was 46 years old (standard error=0.09, standard deviation=16.74). The patient population consisted of more female patients (54%) than male (46%) patients. The majority of the patient population were Caucasians (49%), followed by African Americans (13%) and others (41%).

### *3.2. Automated Periodontitis Diagnosis*

We found 3,708 (16%) patients were healthy and did not have periodontitis. Out of the remaining 25,200 patients, 182 (0.78%) had mild, 12,635 had moderate (55%), and 6,317 (27%) had severe periodontitis when the periodontitis case definitions were used. For 6,028 patients, the charting information was missing.

### *3.3. Evaluation of Algorithm's Performance*

While comparing patients' P.D. diagnoses generated from the computer algorithms against the gold standard, we found 100% sensitivity, specificity, and accuracy. All records belonged to the true positive case, which resulted in excellent accuracy. This is because the data is structured, and the information in each text file is written consistently. For example, the indexing method was used to locate a tooth number in the text file. In each text file, each tooth was present on the 10th element. As a result, computer algorithms were able to identify all patient cases correctly.

## 4. Discussion

We demonstrated the feasibility of diagnosing periodontitis status using the 2012 case definitions and periodontal charting data automatically from the EDR data. This method can be applied to other institutions and private practices' EDR data to determine periodontitis diagnosis and estimate prevalence. The rationale is that due to the high cost of conducting prospective epidemiological studies and the ethical purposes of exposing patients to radiographs, further NHANES studies will not be conducted.

Our approach also has several advantages over the epidemiological study [1,4]. While calculating the prevalence of periodontitis, Eke et al. used measurements from four interproximal sites (mesiobuccal, distobuccal, mesiolingual, distolingual) with an assumption that those sites are most affected by the disease and excluded midbuccal and midlingual sites. Measurements from the mid-buccal and the mid-lingual sites that

potentially could indicate furcation involvement were not included in the study. In addition, they also excluded the involvement of people for medical reasons and people who are institutionalized such as nursing home residents who may have introduced selection bias. Because of time constraints, the examiners did not assess bleeding on probing sites which could provide information to estimate gingivitis prevalence which is a precursor of periodontitis. The study authors acknowledged that they might have underestimated the disease prevalence [1]. In contrast, this study used all six sites per tooth, which helped in estimating the health of the entire tooth. Moreover, no patients were excluded based on their medical conditions or institutionalization, which represented a real-world patient population.

Finally, compared to Eke et al. [1], we found a significantly higher number of patients in severe periodontitis cases (8% to 27%). There could be multiple reasons for these discrepancies. First, we have only used one institute's dataset compared to the entire U.S. sample, as demonstrated in Eke et al. Also, in academic institutions, there is a high proportion of patients who belong to low socio-economical status and visit dental clinic when the disease is in the severe stages.

## 5. Conclusions

We utilized automated approaches to diagnose periodontitis cases using the 2012 case definitions [1,8]. Once the accuracy of the computer algorithm is confirmed by further studies, this approach can then be utilized to estimate prevalence in the U.S. by collaborating with other dental academic institutions and private dental practitioners.

## References

[1] Eke PI, Thornton-Evans GO, Wei L, Borgnakke WS, Dye BA, Genco RJ. Periodontitis in US adults: national health and nutrition examination survey 2009-2014. J Am Dental Assoc. 2018 Jul;149(7):576-88, doi: 10.1016/j.adaj.2018.04.023.

[2] Tonetti MS, Greenwell H, Kornman KS. Staging and grading of periodontitis: framework and proposal of a new classification and case definition. J Periodontol. 2018 Jun;89:S159-72, doi: 10.1002/JPER.18-0006.

[3] Eke PI, Dye BA, Wei L, Thornton-Evans GO, Genco RJ. Prevalence of periodontitis in adults in the United States: 2009 and 2010. J Dent Res. 2012 Oct;91(10):914-20, doi: 10.1177/0022034512457373.

[4] Eke PI, Dye BA, Wei L, Slade GD, Thornton-Evans GO, Borgnakke WS, Taylor GW, Page RC, Beck JD, Genco RJ. Update on prevalence of periodontitis in adults in the United States: NHANES 2009 to 2012. J Periodontol. 2015 May;86(5):611-22, doi: 10.1177/0022034512457373.

[5] Holtfreter B, Albandar JM, Dietrich T, Dye BA, Eaton KA, Eke PI, Papapanou PN, Kocher T. Standards for reporting chronic periodontitis prevalence and severity in epidemiologic studies: Proposed standards from the Joint EU/USA Periodontal Epidemiology Working Group. J Clin Periodontol. 2015 May;42(5):407-12, doi: 10.1111/JCPE.12392.

[6] Song M, Liu K, Abromitis R, Schleyer TL. Reusing electronic patient data for dental clinical research: a review of current status. J Dent. 2013 Dec;41(12):1148-63, doi: 10.1016/j.jdent.2013.04.006.

[7] Thyvalikakath TP, Duncan WD, Siddiqui Z, LaPradd M, Eckert G, Schleyer T, Rindal DB, Jurkovich M, Shea T, Gilbert GH, National Dental PBRN Collaborative Group. Leveraging electronic dental record data for clinical research in the national dental PBRN practices. Applied Clinical Informatics. 2020 Mar;11(02):305-14, doi: 10.1055/s-0040-1709506.

[8] 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, doi: 10.1902/JOP.2012.110664.

[9] Patel JS. Utilizing electronic dental record data to track periodontal disease change. Indiana University-Purdue University Indianapolis; 2020.