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

doi:10.3233/SHTI231179

Utilizing Electronic Dental Records to Predict Neuro-Degenerative Diseases in a Dental Setting: A Pilot Study

Jay PATEL[a,1] and Huanmei WU[a]
[a]*Department of Health Services Administration and Policy, College of Public Health Temple University, Philadelphia, PA, USA*

Abstract. Limited research demonstrates the possible correlations between dental diseases and neurodegenerative diseases like Alzheimer's disease (AD) and Parkinson's disease (PD). Nevertheless, dental diseases are often overlooked while assessing the risk of AD and PD in clinical settings. It is unknown whether AD/PD risk can be predicted using electronic dental record (EDR) data collected in a routine dental setting. This pilot study determined the feasibility of predicting AD/PD using 84 features routinely captured in the EDR. We utilized the Temple University School of Dentistry clinic data of 27,138 patients. Using a natural language processing (NLP) approach (accuracy=97%), we identified patients with AD/PD and their matched controls (matched by age and gender). XGBoost machine learning model with 10-fold cross-validation was applied for prediction. With 77% accuracy, we found 53 features significantly associated with AD/PD that could be utilized to predict the risk of AD/PD. Further studies are warned to confirm these findings.

Keywords. Alzheimer's disease, Parkinson's disease, electronic dental records

1. Introduction

Neurodegenerative diseases (NDs) like Alzheimer's disease (AD) and Parkinson's disease (PD) affect millions of people worldwide in which irreversible damages occur to cells and nervous system connections essential for coordination, strength, mobility, sensation, and cognition [1,2]. AD is characterized by progressive cognitive decline and memory loss, while PD is a movement disorder that causes serious impairment and significantly reduces the quality of life. Mounting evidence suggests that oral health is a significant part of systemic health especially due to strong mouth-to-body connection such as associations between diabetes and periodontitis [3]. Limited evidence also demonstrates a possible correlation between neurodegenerative diseases like AD/PD, [1,2,4,5]. However, it is not well understood whether oral diseases, habits, and other oral social determinants of health factors are associated with an increased risk for AD/PD. Determining bidirectional relationships between Oral diseases and NDs would help clinicians, researchers, and policymakers to make critical decisions, especially with early diagnosis, and develop preventive strategies and policies to cover more dental treatments in ND populations. A longitudinal study examining the association between periodontitis and the risk of AD concluded a significantly greater incidence of developing AD in perio

[1] Corresponding Author: Jay Patel, email: patel.jay@temple.edu.

group versus the healthy control [6]. Verhoeff et al. found that PD patients reported significantly more often bruxism during sleep and wakefulness and reported more serious temporomandibular joint problems [5]. In addition to bi-directional relationships between oral diseases and NDs, some dental treatments may negatively affect systemic health [5] like dental amalgam filling, which is associated with a higher risk of PD [7].

Despite this anecdotal evidence, the inclusion of oral diseases in AD/PD risk assessment is limited. One of the reasons for this could be very few studies have reported these connections in the conjugation of various oral diseases such as dental caries, periodontal disease, and temporomandibular joint problems [2,5]. Moreover, as per our best knowledge, no study has attempted to include the oral disease status of AD/PD, their oral habits, and oral health-related social determinants of health (SDoH) in a single model. Understanding the associations between AD and PD with oral health habits and SDoH beyond just specific disease diagnosis would allow clinicians to provide precision preventive approaches to maintain good oral health that may slow the initiation and progression of ND disease. Therefore, this pilot study aims to determine the feasibility of predicting AD/PD using 84 features routinely captured in the EDRs.

2. Methods

This study was approved by the Temple University Institutional Review Board (#28321). We first obtained EDRs from a large academic dental hospital. We then developed and tested a natural language processing (NLP) application to extract patient information who self-reported AD/PD diagnosis. We then created a master dataset of these patients' oral health habits, dental history, and SDoH factors for analysis. We then ran the XGBoost machine learning (ML) model to develop a prediction model using 84 features.

2.1. Patient Cohort

We obtained the patient data from the Temple University Kornberg School of Dentistry dataset between 01/2017-12/2021. Our inclusion criteria consisted of adult patients who received at least one comprehensive oral evaluation. We utilized EDR sections, including patient demographic, medical history, dental history, social habits, behavioral factors, and social determinants of health.

2.2. Patient Assignments in Case and Healthy Control Groups

We developed an NLP application to extract patients' information in a structured format because medical history in the EDR is documented in free-text format. For the NLP, we first created a semantic schema that includes variations of writing AD/PD and medications in the EDR. To create the semantic schema, we reviewed several self-reported histories using a bottom-up approach and also conducted a literature review using the top-down approach. We then used this schema to develop a named entity recognition (NER) program described in [8] to identify patients with AD/PD. This application [8] was also utilized to extract other medical histories (e.g., cardiovascular diseases, diabetes, etc.) from EDRs. We manually checked 200 records identified as AD/PD positive and 200 AD/PD negative. We then calculated true positive (TP), false positive (FP), true negative (TN), and false negative (FN) to measure sensitivity,

specificity, and accuracy. To prepare the healthy control group (*3*N of case group*), we matched patient records in the AD/PD negative group by age (*+/-5*) and gender.

2.3. ML Model Training, Testing & Determine Feature Importance Through SHAP Values

We utilized XGBoost, a powerful ML model which combines multiple decision trees to construct a strong model [9]. Since AD/PD are not common medical conditions, model overfitting is possible (correct prediction of healthy controls only). Therefore, to avoid overfitting and to find correct optimal hyperparameters, a 10-fold cross-validation strategy was used. We used 70% dataset for training and 30% dataset for testing the ML model. Specifically, the training set was split into ten folds with equal size. Each time, a fold was used as the hold-out validation set to calculate model performance, while the remaining nine folds were used to train the model.

2.4. Determine Feature Importance through SHAP Values

We utilized SHapley Additive exPlanation (SHAP) to determine the feature importance and to enhance model interpretability [10]. It is a well-designed tool that can interpret each predictor's output (e.g., decision) and values to assess whether their contributions differ in AD/PD versus healthy controls.

3. Results

3.1. Patient Demographics

There were 27,138 patients who received at least one COE during our study period. Two hundred and three patients either had a self-reported diagnosis of AD/PD or were taking at least one medication. The matched healthy control group consisted of 537 patients. Therefore, 740 patients were used for prediction modeling with their 84 distinct features. The majority of the patient cohort [502 (68%)] belonged to the age group 50-65 years, followed by > 65 years [153 (21%)] and <50 [85 (11%)]. There were more female patients [421 (57%)] than male patients [319 (43%)].

3.2. Performance of the NLP and ML Model

Our NLP model performed with 95% sensitivity, 99% specificity, and 97% accuracy (TP=189, FP=1, TN=200, FN=10). We achieved this excellent accuracy because of two reasons. First, a separate section is provided for patients to document their neurological conditions. Therefore, the clinical notes were not too lengthy, eliminating FP cases. Second, the medication records were obtained from patients' prescription histories, and as a result, there were little to no spelling errors in obtaining this information. Our ML model performed with 77% accuracy after balancing the data. As demonstrated in Figure 1, we found 110 TN cases, 61 TP cases, 39 FN cases, and 12 FP cases.

3.3 Factors Associated with AD/PD

Out of 84 features, we found 53 features/variables associated with AD/PD. **Figure 2** shows the top 20 clinically significant factors. Three features related to periodontitis include bone loss, periodontal diagnosis, and bleeding on probing. Decay missing filled teeth was also associated with AD/PD, demonstrating that the patients had a significantly higher incidence of decayed teeth. SDOH and behavioral factors such as high-stress levels, anxiety, and inadequate patient compliance are also predictors for AD/PD. Further, we found teeth level factors such as higher levels of calculus, plaque, inadequate brushing and flossing, and the number of teeth is also associated with AD/PD. Finally, we observed associations between other systemic diseases, such as cardiovascular diseases, diabetes, pulmonary diseases, and AD/PD.

4. Discussion

In this pilot study, we determined the feasibility of utilizing EDR data to predict AD/PD in dental settings. This was the first study that utilized 84 features from EDRs to predict the risk of AD/PD. We found 53 prediction features from the EDR that could be utilized to predict the risk of AD/PD onset. As demonstrated in the results, we observed the two most prevalent dental diseases, such as periodontal disease and dental caries, highly associated with AD/PD. This could be because the literature shows that numerous microorganisms are suspected in AD brains ranging from bacteria, viruses, and cysts. For instance, bacteria of oral and non-oral Treponema species, viruses such as herpes simplex type I, and some cysts are found in both periodontium and AD brains. This causal relationship has been explained between periodontal pathogens and non-oral Trepedoma species via the amyloid-beta and inflammatory links [4]. Therefore, managing these patients' good oral health is crucial to delay the onset and prevent disease progression. Oral health factors should also be incorporated in AD/PD risk assessment.

Figure 1. Prediction Performance.

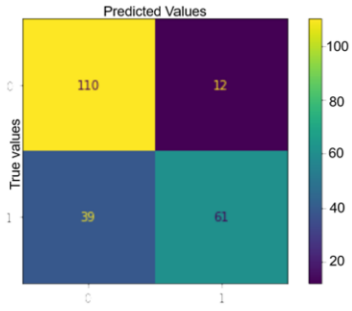
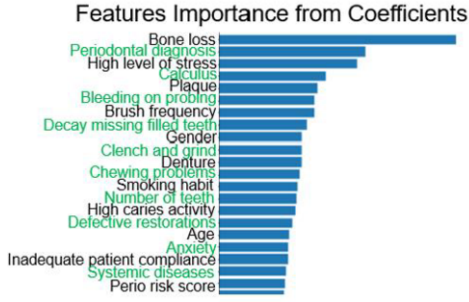


Figure 2. Feature Importance.



Several oral behavioral and SDoH factors, such as high calculus, plaque, and inadequate compliance levels, are associated with AD/PD. These patients often experience dementia and may not comply with the instructions provided by dental clinicians associated with poor oral health outcomes. Therefore, interventions such as mobile applications or sensors should be developed to improve patient compliance.

Our NLP application performed excellently in identifying patients who self-reported AD/PD or related medications which can be applied to other institutions' datasets. On the other hand, the ML model performed moderately with 77% accuracy. However, this

accuracy can be considered good especially when we solely use the EDR data for prediction. Matching patients' dental data with their medical records may help improve the accuracy of the prediction. This study has several limitations. First, the prevalence of AD/PD is higher than what we found in our population. We found <1% of patients with AD/PD because they often do not self-report their medical histories in their dental records. Self-reported measures may not provide the most accurate information about their health. In the future, we will match our dental patient records with their medical to obtain their up-to-date and complete medical histories. We also did not validate the algorithm's performance on any external data source which we will do in our next study. Finally, we grouped AD/PD patients into one category, and these two distinct diseases may have different risk factors which we will address in our future studies using matched medical-dental datasets.

5. Conclusions

We demonstrated the feasibility of developing and implementing a prediction model to assess the risk of AD/PD in dental settings. This model was developed using 84 distinct features (53 significant factors) from the EDR. Future studies are warranted to confirm the results of this pilot study and the validity of the model's performance on external data.

References

[1] Jeong E, Park JB, Park YG. Evaluation of the association between periodontitis and risk of Parkinson's disease: a nationwide retrospective cohort study. Sci Rep. 2021 Aug;11:16594(2021), doi: doi:10.1038/s41598-021-96147-4.

[2] Teixeira FB, Saito MT, Matheus FC, Prediger RD, Yamada ES, Maia CS, Lima RR. Periodontitis and Alzheimer's disease: a possible comorbidity between oral chronic inflammatory condition and neuroinflammation. Front Aging Neurosci. 2017 Oct 10;9:327, doi: 10.3389/FNAGI.2017.00327.

[3] Alpert PT. Oral health: the oral-systemic health connection. Home Health Care Manag Pract. 2017 Feb;29(1):56-9, doi: 10.1177/1084822316651658.

[4] Olsen I, Singhrao SK. Can oral infection be a risk factor for Alzheimer's disease?. J Oral Microbiol. 2015 Jan;7(1):29143, doi: 10.3402/JOM.V7.29143.

[5] Verhoeff MC, Lobbezoo F, Wetselaar P, Aarab G, Koutris M. Parkinson's disease, temporomandibular disorders and bruxism: a pilot study. J Oral Rehabil. 2018 Nov;45(11):854-63, doi: 10.1111/JOOR.12697.

[6] Chen CK, Wu YT, Chang YC. Association between chronic periodontitis and the risk of Alzheimer's disease: a retrospective, population-based, matched-cohort study. Alzheimers Res Ther. 2017 Dec;9:56, doi: 10.1186/S13195-017-0282-6.

[7] Hsu YC, Chang CW, Lee HL, Chuang CC, Chiu HC, Li WY, Horng JT, Fu E. Association between history of dental amalgam fillings and risk of Parkinson's disease: a population-based retrospective cohort study in Taiwan. PloS One. 2016 Dec;11(12):e0166552, doi: 10.1371/JOURNAL.PONE.0166552.

[8] Sureshbhai Patel J, Rao R, Brandon R, Iyer V, Albandar JM, Tellez M, Krois J, Wu H. Develop a natural language processing pipeline to automate extraction of periodontal disease information from electronic dental clinical notes. In: Proceedings of the 6th International Conference on Medical and Health Informatics; 2022 May 13. p. 61-8, doi: 10.1145/3545729.3545744.

[9] T. Chen, and C. Guestrin, XGBoost: A Scalable Tree Boosting System, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. (n.d.). doi:10.1145/2939672.

[10] Chen T, Guestrin C. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 Aug 13. p. 785-94.

[11] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30. https://github.com/slundberg/shap (accessed Dec 2, 2022).