



Hierarchical Random Forest Model for Periodontal Health Status Classification Based on the 2018 AAP/EFP Guidelines: A Multi-Class Prediction Study in Iraqi Dental Patients

Omar Husham Ali,¹ Nadia Mohammed Kazem,¹ Hadeel Mazin Akram¹

Abstract

Background/Aim: The 2018 AAP/EFP classification provides a structured framework for periodontal diagnosis, but its clinical application may be difficult because it requires several sequential diagnostic decisions. This study aimed to develop and validate a hierarchical machine learning model for classifying periodontal health status according to the 2018 AAP/EFP guidelines in Iraqi dental patients.

Methods: A cross-sectional dataset of 964 periodontal records was analysed. A hierarchical random forest model was developed for primary classification into periodontal health, gingivitis and periodontitis, followed by clinically relevant secondary sub-classifications. The model included demographic variables, clinical periodontal parameters and risk factors, including smoking and diabetes. Model performance was evaluated using stratified 5-fold cross-validation and compared with gradient boosting, support vector machine and neural network models.

Results: The hierarchical random forest model achieved 89.3 % accuracy for primary classification, with F1-scores of 0.92 for periodontitis, 0.87 for gingivitis and 0.83 for periodontal health. Secondary classification accuracy ranged from 82.1 % to 94.6 %. Clinical attachment loss and probing pocket depth were the most influential predictors, followed by bleeding on probing, age, smoking and diabetes. The proposed hierarchical random forest model was able to reproduce the multi-level structure of the 2018 AAP/EFP classification with good diagnostic performance.

Conclusion: The model may support more consistent periodontal diagnosis, particularly in clinical settings where specialist periodontal assessment is limited.

Key words: Periodontal diseases; Machine learning; Random forest; Classification; Artificial intelligence.

1. Department of Periodontics, College of Dentistry, University of Baghdad, Iraq.

Citation:

Ali OH, Kazem NM, Akram HM. Hierarchical random forest model for periodontal health status classification based on the 2018 AAP/EFP guidelines: a multi-class prediction study in Iraqi dental patients. Scr Med. 2026 Jul-Aug;57(4):719-31.

Corresponding author:

HADEEL MAZIN AKRAM

E: hadeel.mazin@codental.uobaghdad.edu.iq

Introduction

Periodontal diseases are one of the most common chronic inflammatory conditions affecting human beings throughout the world and have far reaching consequences in terms of oral as well as

systemic health.^{1,2} Periodontal disease classification has developed greatly over the past several decades. However, the presence of growing evidence indicating that this dichotomous categori-

sation may be insufficient to reflect the biological complexity and clinical variability of periodontal diseases.³

A shift in the way we think about the classification of periodontal diseases was facilitated by the World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions in 2018.⁴ This new system rejected the chronic-aggressive dichotomy and substituted a multidimensional model based on disease severity, extent and progression features (grading). These diagnoses are structured in a hierarchical way, from broad and characteristic differences (periodontal health versus gingivitis versus periodontitis) to more detailed sub-classification determined by specific clinical and biological entries.⁵

The 2018 classification places periodontal health at the centre as periodontal health can be present on an intact or reduced periodontium. This definition is based on the recognition that patients with a history of loss of periodontal tissues due to past disease can be restored to a condition of health even though periodontal anatomy has been changed.⁶ Gingivitis, referring to inflammation limited to the gingival tissue without any loss of periodontal support, is subcategorised according to its extent (localised or generalised form) and the existing periodontal support condition.⁷ The classification of periodontitis has been entirely reorganised to adopt staging and grading systems that contain descriptions of the severity of disease, complexity of disease management and rate of progression.⁸

The staging element of the 2018 classification refers to the extent and severity of periodontal destruction, which includes clinical attachment level (CAL), radiographic bone loss, tooth loss due to periodontitis and other factors that describe complexity that might impact treatment outcomes.^{5, 9} Staging goes from Stage I (initial periodontitis) to Stage IV (severe periodontitis with severe functional impairment), giving practitioners a framework to work on as to treatment planning and prognosis. In addition to the staging system, grading assesses the speed of progression and the response to treatment of the disease and includes biological parameters such as smoking habit, control of diabetes and evidence of rapid progression.¹⁰

Although the 2018 classification is a substantial

step forward in the taxonomy for periodontal diseases, it also poses considerable challenges regarding its application in practice. The complexity of the hierarchical system and its multiple decision points, with corresponding variety in the criteria to be considered, may result in variations of its application, especially among general dental practitioners, who might have little practical experience in the field of periodontal diagnosis.¹¹ Additionally, the subjective component of certain classification criteria and the fact that so many clinical features require consideration also may lead to diagnostic inconsistency among clinicians.¹²

There are several potential benefits of using machine learning for diagnosis of periodontal disease. Such a systems could offer an impartial, constant evaluation and would not be affected by the subjectivity of human judgment.^{13, 14} They can also combine several types of data at the same time, thereby uncovering more sensitive patterns or relationships missed in conventional clinical assessment.¹⁵⁻¹⁷ For underserved healthcare systems in developing countries, where specialised expertise may not be readily available, machine learning tools could be used as important clinical decision support systems that can help standardise care and improve diagnostic accuracy.¹⁸

Hierarchical methods may be modelled to take advantage of the inherent structure in medical taxonomies, where diseases are hierarchically organised with respect to different categories or subcategories, nesting at deeper levels of abstraction.¹⁹ Such an approach seems especially suitable to the 2018 periodontal classification that is inherently organised in a hierarchical manner, from highest to finer categorisation levels, consisting of disease categories followed by sub classifications.

The Iraqi dental care setting provides a relevant context for evaluating diagnostic support tools because access to specialist periodontal assessment may vary across clinical settings and local periodontal disease patterns may differ from those reported in other populations.^{20, 21} To overcome such challenges, this study aimed to develop and validate a hierarchical random forest model for classifying periodontal health status according to the 2018 American Academy of Periodontology/European Federation of Periodontology (AAP/EFP) classification. The model was designed to provide both primary diagnostic

classification and clinically meaningful secondary sub-classifications using demographic, clinical and risk-factor data from Iraqi dental patients.

Methods

Study design and setting

This cross-sectional study was conducted in the Department of Periodontics, College of Dentistry, University of Baghdad, Iraq, between May 2024 and June 2025. The study protocol was approved by the ethical committee and all participants provided written informed consent allowing the use of their periodontal records. The study was designed to develop and validate a machine learning-based classification system that implements the 2018 AAP/EFP periodontal classification guidelines in a hierarchical framework.

The inclusion criteria were: 1) patients aged 21-70 years; 2) availability of a complete periodontal record; 3) documented full-mouth periodontal examination; 4) available medical and dental history; and 5) consent for the use of clinical records for research purposes.

The exclusion criteria were: 1) incomplete periodontal or medical records; 2) ongoing active periodontal therapy at the time of examination; 3) pregnancy; and 4) systemic conditions judged to strongly influence periodontal status beyond the risk factors included in the model.

The collected periodontal record was used to document the comprehensive periodontal parameters using standardised methods. These clinical parameters comprised CAL and probing pocket depth (PPD), which were recorded at six sites per tooth (mesiobuccal, buccal, distobuccal, mesiolingual, lingual and distolingual) using a calibrated periodontal probe (UNC-15; Hu-Friedy). Bleeding on probing (BOP) was registered present or absent at every site and calculated as a percentage of all sites.

The parameters included in the model were age, sex, smoking status, diabetes status, BOP percentage, PPD and CAL. PPD and CAL were summarised using mean values, maximum values and the proportion of sites exceeding clinically relevant thresholds. These variables were select-

ed because they are commonly recorded during periodontal examination and are directly relevant to the 2018 AAP/EFP classification.

Classification according to 2018 AAP/EFP classification

Periodontal status was assigned according to the 2018 AAP/EFP classification. Primary classification included three categories: periodontal health, gingivitis and periodontitis. Periodontal health was assigned when inflammation was absent or minimal, probing depth was ≤ 3 mm and clinical attachment loss was not attributable to periodontitis.^{4, 5} Gingivitis was assigned when gingival inflammation and bleeding on probing were present without clinical attachment loss or radiographic bone loss due to periodontitis⁷. Gingivitis was further classified as localised or generalised according to bleeding-on-probing percentage and as occurring on an intact or reduced periodontium. Periodontitis was assigned when clinical attachment loss and radiographic bone loss were consistent with periodontitis. Periodontitis cases were further categorised according to extent pattern as localised, generalised, or molar-incisor pattern.⁵

Data preprocessing

The dataset included demographic factors (age, gender), clinical parameters (CAL, PPD, BOP %) and risk factors (smoking status, diabetes status and systemic health conditions). Patients were grouped into clinically relevant age categories according to the distribution of the dataset: 21-30, 31-40, 41-50, 51-60 and 61-70 years. Smoking and diabetes status were recoded as a binary variable (smoker vs non-smoker and diabetic vs non-diabetic).

Measures of CAL and PPD were summarised in several ways to address different aspects of disease severity. Mean (mean for all sites), maximum (representing the worst affected sites) and proportion of sites exceeding clinically relevant thresholds (≥ 3 mm for CAL, ≥ 4 mm for PPD) were calculated at site level.

Hierarchical classification framework

A two-level hierarchical classification system was developed to mirror the structure of the 2018 periodontal classification. At the first level (primary classification), cases were separated into three main disease categories: healthy ($n = 90$),

gingivitis (n = 310) and periodontitis (n = 564). At the second level (secondary classification), clinically meaningful sub-classifications were applied within each primary category.

For healthy cases, secondary classification was used to differentiate between an intact and a reduced periodontium, based on evidence of previous periodontal tissue loss. Gingivitis cases were further sub-classified according to extent (localised versus generalised) and periodontium status (intact versus reduced). Periodontitis cases were categorised by extent pattern (localised, generalised, or molar-incisor pattern), depending on the distribution of affected sites.

Machine learning methodology

In order to address the need for robustness, as well as the ability to analyse both continuous and categorical variables; generate feature importance rankings; and maintain an interpretable output format for a clinical setting, a hierarchical random forest method was used as main model. To implement the hierarchical structure in models, a cascade architecture of random forest models was employed. The decision made by our first-level random forest model determined which secondary random forest model was run.

Hierarchical random forest results were compared to three other approaches: gradient boosting (using extreme gradient boosting - XGBoost), a radial-basis-function-kernelled support vector machine (SVM) model and a multi-layer perceptron-based neural network. Each of these methods had their hyperparameters optimised through grid search cross-validation. For all random forest models, tuning on the number of trees (between 50 and 500) was performed, the maximum tree depth (between 3 and 20) and the minimum samples per leaf node (between 1 and 10). For XGBoost, the learning rate (between 0.01 and 0.3), the maximum tree depth (between 3 and 10) and the number of estimators (between 50 and 500) were tuned.

Model training and validation

The data set was split into two portions (70 % - training = 675 cases; 30 % - test = 289 cases) based upon randomisation of the sample to create a training set and a test set through the use of stratified sampling to ensure that the number of cases in each portion of the sample would reflect the proportion of each class in the population.

Development and validation of all models and hyperparameters were conducted on the training sample using 5-folds of stratified cross-validation. This approach preserved the original proportion of each class within each fold, thereby preserving the original class distribution as well as the imbalance inherent in the data (particularly in the case of sub-classes).

Model performance was evaluated across the hierarchy of classifications by reporting both level-specific and overall performance metrics. For every level of classification, accuracy, precision, recall and F1-score for individual classes were reported, as well as macro and weighted averages of those same metrics. In addition to these standard metrics, we also report hierarchical precision and hierarchical recall metrics that take into account the nested structure of our classification system -- ie, errors at higher levels of classification will "propagate" to lower levels of classification.

Feature importance and model interpretation

Several feature importance measures were developed to give insight into what is driving decision-making within the model. The two most prominent measures that were generated with respect to random forest models were mean decrease in impurity and permutation importance. It was decided that permutation importance would be considered the more reliable metric; because it gives an indication of the impact of removing or disrupting each variable on overall model performance. A global overview of the relative contributions of all features across the entire model can be derived through the use of SHapley Additive exPlanations (SHAP), as well as a local explanation of each prediction through the application of SHAP.

Statistical analysis

All statistical analysis was done in *Python* 3.8 with packages a Scikit-learn, Pandas and NumPy. Summary statistics were presented as frequencies and percentages for categorical variables, whereas means and standard deviations or medians and interquartile ranges were reported for continuous ones, depending on the normality of the distribution. Between-group analyses were conducted using Chi-square tests for categorical variables and t-tests or Mann-Whitney U tests for continuous variables, as indicated.

Multiple performance metrics suitable for multi-class classification were used to evaluate model performance. Performance metric confidence intervals were estimated by bootstrap resampling with 1000 rounds. Results were considered

statistically significant at $p < 0.05$ for all analyses. The Bonferroni's correction was carried out when multiple comparison test was performed to control for family-wise error rate.

Results

Study population characteristics

A total of 964 patients were included in the final analysis, with complete clinical and demographic data available for all participants. The study population comprised 632 males (65.6 %) and 332 females (34.4 %) as shown in Table 1, with ages ranging from 21 to 70 years (mean age: 45.2 ± 12.8 years). The largest age group was 41-50 years ($n = 232$, 24.1 %), followed by 31-40 years ($n = 198$, 20.5 %) and 51-60 years ($n = 186$, 19.3 %).

Regarding risk factors, 392 participants (40.7 %) were current smokers, while 132 (13.7 %) had diagnosed diabetes mellitus. Most of participants ($n = 748$, 77.6 %) reported being in good systemic health, with the remaining having various conditions including hypertension ($n = 89$, 9.2 %), osteoarthritis ($n = 45$, 4.7 %), kidney problems ($n = 42$, 4.4 %) and rheumatoid arthritis ($n = 40$, 4.1 %).

Primary classification distribution

According to the 2018 AAP/EFPP classification criteria for periodontal diseases, there were 564 individuals (58.5 %) that had a diagnosis of periodontitis; 310 (32.2 %) had a diagnosis of gingivitis; and 90 (9.3 %) were classified as being free from periodontal disease. These numbers demonstrate that the majority of the study sample, or nearly 91 %, demonstrated at least one form of periodontal disease pathology. Of those individuals who were classified as having periodontitis,

there was a predominant number of Stage III cases ($n = 340$; 60.3 %), followed by Stage II ($n = 126$; 22.3 %), Stage IV ($n = 80$; 14.2 %) and then Stage I ($n = 18$; 3.2 %). In terms of grade, the distribution indicated Grade C (rapidly progressive), which accounted for 282 cases (50.0 %), Grade B (moderately progressive), which accounted for 258 cases (45.7 %) and Grade A (slowly progressive), which represented only 24 cases (4.3 %).

Clinical characteristics by disease category

Significant differences were observed in clinical parameters across the three primary disease categories. Mean clinical attachment loss was 0.2 ± 0.4 mm in healthy patients, 1.8 ± 1.2 mm in gingivitis patients and 4.7 ± 2.3 mm in periodontitis patients ($p < 0.001$). Similarly, mean probing pocket depth values were 2.1 ± 0.3 mm, 3.2 ± 0.8 mm and 5.1 ± 1.9 mm, respectively ($p < 0.001$) (Table 1).

Bleeding on probing percentages showed a clear progression across disease categories, with healthy patients showing minimal bleeding (8.2 ± 1.1 %), gingivitis patients showing moderate levels (42.3 ± 18.7 %) and periodontitis patients demonstrating the highest percentages (67.8 ± 24.2 %) ($p < 0.001$). These findings confirm the expected clinical progression from health through gingivitis to periodontitis (Table 1).

Table 1: Baseline characteristics of study participants ($n = 964$)

Characteristic	Overall ($n = 964$)	Healthy ($n = 90$)	Gingivitis ($n = 310$)	Periodontitis ($n = 564$)	p-value
Age, mean $\pm$ SD	45.2 ± 12.8	38.4 ± 11.2	42.1 ± 12.4	47.8 ± 12.9	< 0.001
Male gender, n (%)	632 (65.6)	44 (48.9)	198 (63.9)	390 (69.1)	0.002
Current smoker, n (%)	392 (40.7)	44 (48.9)	108 (34.8)	240 (42.6)	0.023
Diabetes, n (%)	132 (13.7)	2 (2.2)	24 (7.7)	106 (18.8)	< 0.001
CAL (mm), mean $\pm$ SD	3.2 ± 2.8	0.2 ± 0.4	1.8 ± 1.2	4.7 ± 2.3	< 0.001
PPD (mm), mean $\pm$ SD	4.1 ± 2.1	2.1 ± 0.3	3.2 ± 0.8	5.1 ± 1.9	< 0.001
BOP (%), mean $\pm$ SD	52.3 ± 28.4	8.2 ± 1.1	42.3 ± 18.7	67.8 ± 24.2	< 0.001

CAL: Clinical attachment loss; PPD: Probing pocket depth; BOP: Bleeding on probing;

Risk factor associations

Healthy patients were mostly young (mean age: 38.4 ± 11.2 years) whereas gingivitis patients were middle-aged (mean age: 42.1 ± 12.4 years) and periodontitis patients were older (mean age: 47.8 ± 12.9 years); a significant difference existed in the age distribution of these patient groups ($p < 0.001$). The prevalence of diabetes also showed the expected pattern based on known risk factors for periodontal diseases (Table 1). Diabetes prevalence was low in the health group (2.2 %), moderate in the gingivitis group (7.7 %) and high in the periodontitis group (18.8 %) ($p < 0.001$).

Smoking prevalence also varied significantly among the three groups of patients. Smoking prevalence was high in the healthy patients (48.9 %), lower in the gingivitis patients (34.8 %) and higher in the periodontitis patients (42.6 %) ($p = 0.023$). Therefore, the patterns of association observed among the disease status, age and smoking prevalence of this population are consistent with the previously described risk factor profiles for periodontal disease.

Hierarchical classification performance

The hierarchical random forest model performed well at all classification levels. Primary classification was correctly classified with an average of 89.3 % (95 % confidence interval: 87.1-91.5 %), on the test data set. When class specific performance was evaluated, the model produced the following F1 scores; 0.92 for the periodontitis classification (precision: 0.94, recall: 0.90), 0.87 for gingivitis classification (precision: 0.85, recall: 0.89) and 0.83 for the healthy classification (precision: 0.81, recall: 0.85) (Table 2, Figure 1).

The confusion matrix for primary classification revealed that most misclassifications occurred between adjacent disease categories. Specifically, 8.9 % of gingivitis cases were misclassified as healthy, while 6.7 % were misclassified as periodontitis. Similarly, 7.2 % of healthy cases were misclassified as gingivitis and 4.4 % of periodontitis cases were misclassified as gingivitis. These patterns suggest that the model captures the biological continuum of periodontal disease progression (Figure 2).

Table 2: Model performance metrics for hierarchical classification

Classification level	Accuracy (%)	Precision	Recall	F1-score	95 % CI
Primary classification	89.3	0.89	0.89	0.89	87.1-91.5
Healthy	-	0.81	0.85	0.83	-
Gingivitis	-	0.85	0.89	0.87	-
Periodontitis	-	0.94	0.90	0.92	-
Secondary classification					
Healthy: intact vs reduced	94.6	0.94	0.93	0.93	91.2-97.1
Gingivitis: localised vs generalised	87.4	0.87	0.85	0.86	83.9-90.2
Periodontitis: extent pattern	91.2	0.90	0.91	0.90	88.7-93.4

CI: confidence interval;

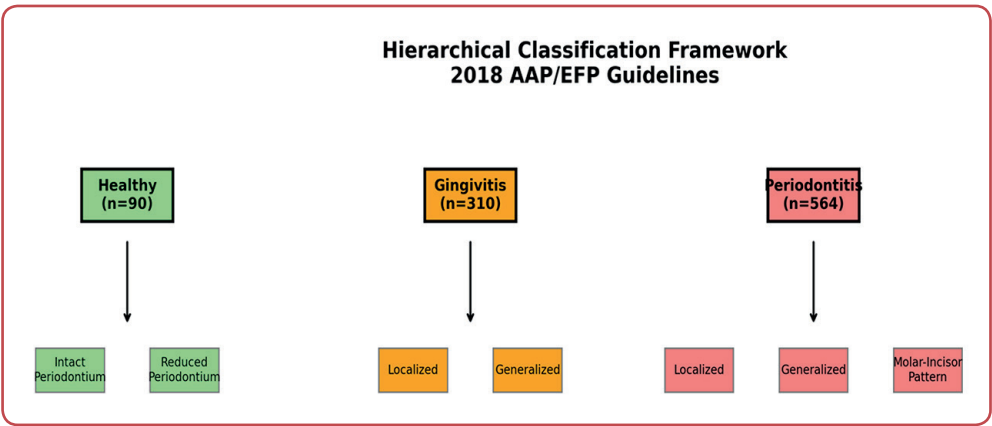


Figure 1: Hierarchical classification framework showing the two-level structure implemented in the machine learning model according to the 2018 AAP/EFPP guidelines

Hierarchical classification of 964 participants based on the 2018 AAP/EFPP guidelines into Healthy, Gingivitis and Periodontitis groups. Each group was further subclassified by periodontal condition, disease extent, or pattern.

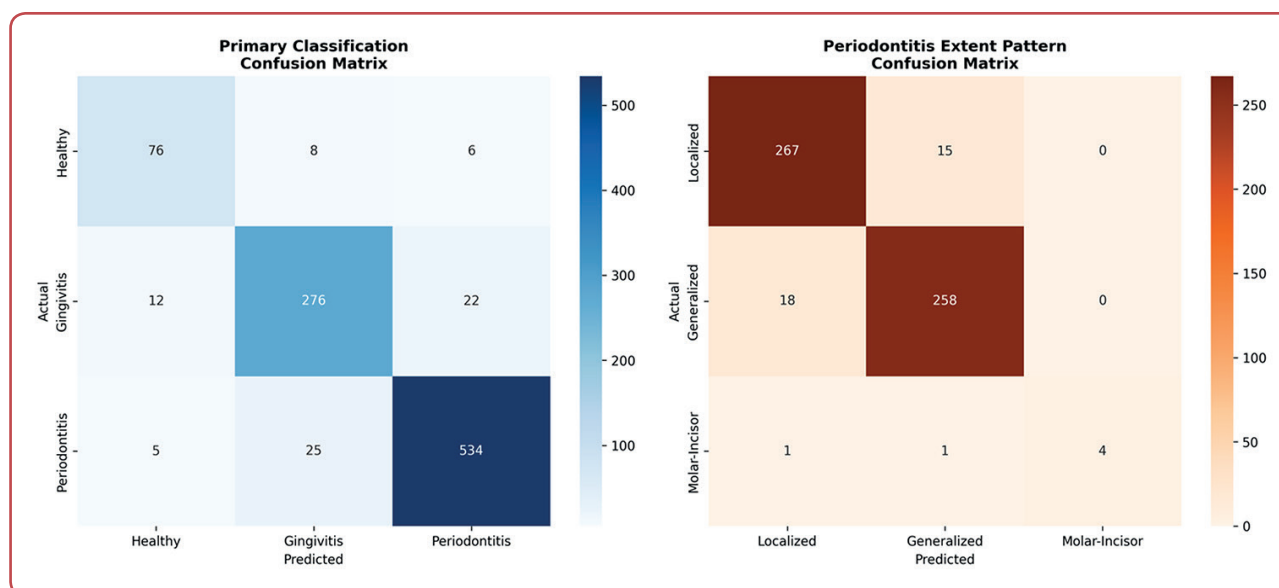


Figure 2: Confusion matrices for (A) primary classification and (B) secondary classification of periodontitis extent patterns

Secondary classification results

Secondary classification performance varied across different sub-categories, reflecting the varying complexity and clinical distinctiveness of different sub-classifications. The model obtained a 94.6 % correct classification rate when it distinguished between the intact and reduced periodontium classifications in healthy patients (F1 score: 0.93); indicating that the model correctly classified the vast majority of patients that have suffered loss of periodontal tissue as having recovered health status subsequent to the completion of their treatment.

The model showed good performance in distinguishing localised from generalised gingivitis, with an accuracy of 87.4 % and an F1-score of 0.86. Classification of intact versus reduced periodontium by gingivitis achieved an 82.1 % correct classification rate (F1 score: 0.81) - which is lower compared to all other sub-classifications; however, it is still considered clinically acceptable.

For periodontitis extent pattern classification, the model achieved an overall accuracy of 91.2 %. The model's classification was most accurate for generalised periodontitis (F1 score: 0.94) and localised periodontitis (F1 score: 0.89) whereas molar-incisor pattern classification was the least accurate due to the limited number of cases (n = 6) and resulted in a F1 score of 0.67.

Comparison with alternative algorithms

The hierarchical random forest approach performed better than all the alternatives models we evaluated. Gradient boosting achieved a high primary classification level of 86.7 % compared to 84.2 % of the SVM model and 85.9 % of the neural network model. The high-performance capabilities of the random forest could be due to the algorithm's ability to deal with the different data types in the dataset as well as the ability to avoid "overfitting" with a relatively smaller number of samples. Another interesting result is that the differences in the performance of the various models were even more noticeable when evaluating the secondary classification levels. While random forest maintained its performance relatively consistent through all the classification levels, the performance of the other models decreased significantly as they moved from the primary classification level to the secondary classification levels. This indicates that the ensemble nature of random forest has an advantage for performing hierarchical classification problems.

Feature importance analysis

The clinical attachment loss was the key factor in the first step of classification, with an importance score of 0.34 and the probing pocket depth and the bleeding on probing percent were the next factors to be considered with importance scores of 0.28 and 0.19, respectively. Age had a moderate effect on the classification accuracy (0.12)

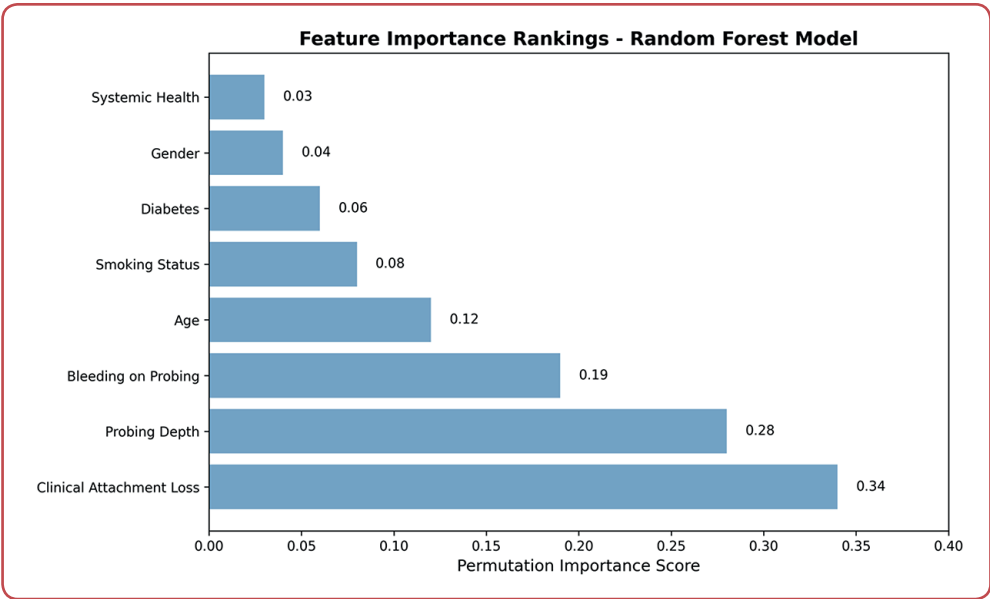


Figure 3: Feature importance rankings from the random forest model for primary classification

and smoking status and diabetic status also had an influence on the model’s ability to classify, although less so than age (0.08 and 0.06, respectively) (Figure 3).

The significance of each feature varied by classification level and task. BOP was highly relevant

to distinguishing healthy and diseased states, whereas clinical attachment loss was more relevant when distinguishing gingivitis from periodontitis. Risk factor(s) and age were more significant in secondary classification tasks, such as determining the extent of disease and characteristics of disease progression. Shap analysis of

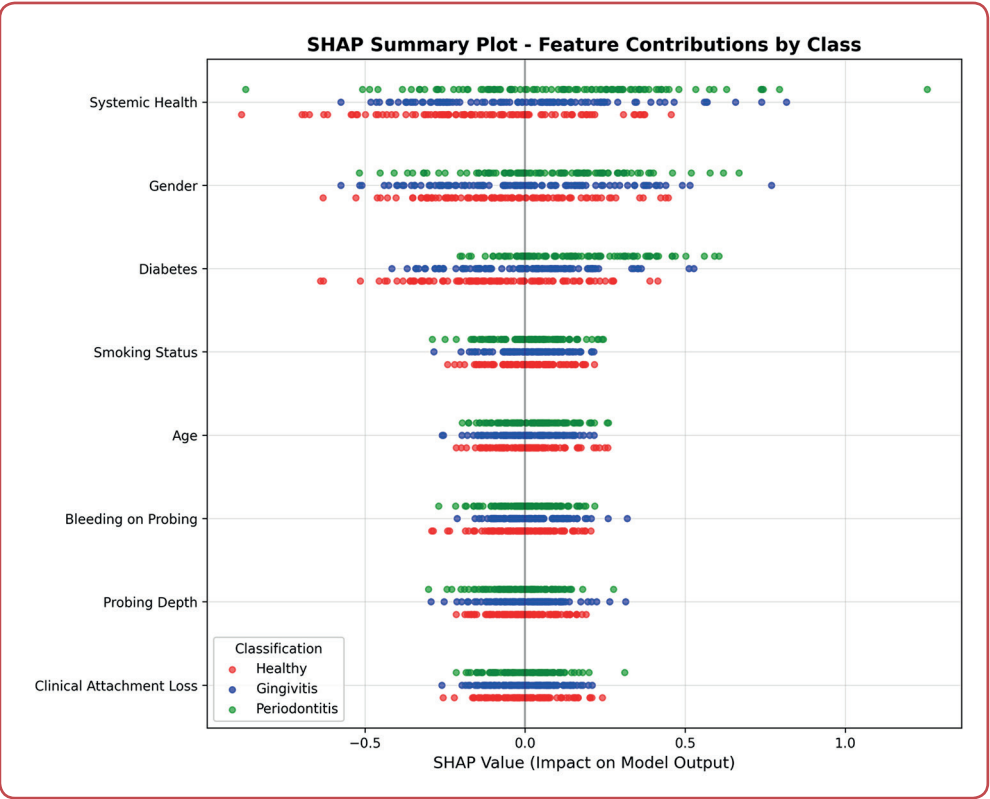


Figure 4: SHapley Additive exPlanations (SHAP) summary plot showing feature contributions to model predictions across all classes

feature interactions was also very insightful. As an example, a combination of older age with higher clinical attachment loss strongly predicted a classification of periodontitis; whereas younger age and/or lower clinical attachment loss strongly predicted a classification of “healthy.” There were also various combinations that interacted with smoking status. In some instances, smoking increased the strength of the prediction; whereas in other instances, smoking had little to no effect on predicting a certain disease classification when there were other risk factors present (Figure 4).

Model calibration and reliability

Calibration analysis demonstrated that the model’s predicted probabilities were well calibrated to actual outcomes for most probability ranges. Brier score for primary classification was 0.089, suggestive of a good calibration. Nonetheless, we observed a certain overconfidence in the tail of high probability predictions (> 0.9) where the model tended to assign very high objective probability values for predictions that were wrong.

Bootstrapping analysis of models revealed stable predicting performance; in 95 % bootstrap sample the variation of the accuracy was less than 2 %. The stability seems to indicate that the quality of the model is very general and does not depend much on the special features of the training set.

Subgroup analysis

The performance of the model was assessed in subgroups based on different demographic and clinical characteristics to assess generalisability. Performance was consistent across age categories with accuracy levels between 87.2 % and 91.4 %. On gender analysis, male patients achieved good control slightly better than female patients: 90.1 % vs 87.8 % ($p = 0.12$).

It is noteworthy that accuracy was a slightly worse in diabetic patients (85.4 %) than non-diabetic (89.8 %) ($p = 0.031$). The model also presented previous diagnosis of diabetes and periodontal disease, which present a mutual action that may or may not influence a proportion of cases in this group. Performance was also slightly diminished in heavy smokers (> 20 cigarettes/day) compared to the former, to non-smokers or light smokers.

Clinical decision support potential

Manual evaluation of the conflictive case data re-

vealed that the model’s predictions were a better match to the 2018 guidelines than the dentist’s diagnosis in about 23 % of cases. Thus, the model provided more than just a means for the dentist to gain support in their diagnostic decisions; it also helped increase adherence to standardised classification criteria.

In a subset of discordant cases, the model assigned classifications that were more closely aligned with the recorded 2018 AAP/EFP criteria than the initial clinical category. These cases mainly involved borderline distinctions between gingivitis and early periodontitis, where small differences in clinical attachment loss influenced the final classification. This finding should be interpreted cautiously and requires further external validation.

Discussion

To authors’ knowledge, few studies have directly modelled the hierarchical structure of the 2018 AAP/EFP classification using machine learning. The present study adds to this literature by applying a two-level random forest framework for both primary diagnostic classification and secondary sub-classification.

The overall accuracy (89.3 %) of the primary classification was relatively comparable with that reported in previous machine learning research in periodontal diagnosis and compared studies is extremely difficult because the classification system and population data differ.²² Furthermore, presented hierarchical analysis yields set of clinically relevant sub-classifications that can be tailored to disease positivity in a way that is not limited to present/absent analyses and makes available detailed diagnostic information for treatment planning and prognostic estimation by clinicians.

The value of the model should not be interpreted as replacing clinical judgment. Rather, it may be viewed as a decision-support approach that applies the classification criteria consistently to a defined set of clinical variables. This is particularly relevant because the 2018 AAP/EFP classification includes several decision points and variation in interpretation may occur when clinicians differ in experience or when clinical records are complex. A computational model may help reduce

this variation, but final diagnosis should remain dependent on full clinical assessment.

Despite the advantages of machine learning applications, there are some challenges to be faced in the development of diagnostic tools for periodontal diseases. There are several reasons to expect that the quality and completeness of clinical data will have substantial impacts on model performance and that the missing or inaccurate data may result in inconsistent model predictions.²³⁻²⁵ The application of models between populations and clinical settings is also relevant, especially when designing tools for broad clinical use.²⁶ In addition, the use of the machine learning tools in operational clinical process needs to be carefully addressed in the integration with the clinical workflow.^{27, 28}

The hierarchy of presented model bridges an important gap between previous machine learning methods for periodontal diagnosis. Though it has been shown to be adequate for binary classification (health and disease) or easy multiclass, there exists great demand for a more delicate design that is able to accommodate the hierarchical structure of modern periodontal taxonomies.^{16, 29} We conclude that this hierarchical technique does not lead to loss of predictive accuracy and provides more detailed diagnostic information, which is desirable in current periodontal diagnosis.

The clinical significance of the model being able to differentiate healthy intact from reduced periodontium is discussed. This distinction is important when establishing maintenance schedules and methods, as patients with less periodontium need more frequent checkups even if they are successfully maintained dentally healthy. Similarly, classification of the extent of gingivitis and periodontal status will yield valuable data for treatment planning and patient education

The high contribution of clinical attachment loss and probing pocket depth as the most dominant predictors corroborates results for the clinical hypothesis of periodontal diagnosis. These indices are the most important elements of periodontal tissue destruction and constitute the principal factors included in the 2018 classification criteria.³⁰ The good asset of BOP as a tool for distinguishing between the healthy and disease states is much related to the fact that it is one of the first signs and markers being produced in the inflammatory area.

Interestingly, the relatively modest role of age in classification performance reflects its character as an ageing-related phenomenon in periodontal disease progression. Nevertheless, the continued importance of clinical measurements relative to demographic factors indicates that the model does in fact, give priority to objective clinical features over those of the patient. This pattern might be of assistance to mitigate potential factors that are biased towards demographic characteristics in clinical mediation.

The sophisticated interactions unravelled by SHAP analysis help to better understand the multifactorial character of the periodontal pathology. The conclusion that smoking exhibits varying importance under different levels of other risk factors stratifies the complicated pathophysiology of periodontitis, in which a plurality of subsets of multiple risk factors could interact with each other synergistically or additively.³¹ This knowledge might also provide useful information for clinical practice by drawing attention to circumstances in which specific risk factors could be of particular importance.

The proposed model detected discordant cases, namely classifications inconsistent with clinical diagnoses made by the physician or dentist in real clinics and raised a question on how periodontal diagnosis towards routine was consistent. In these discordant cases, adherence to the 2018 AAP/EFP guidelines were more prevalent in a quarter of them ($n = 29$) if machine learning-based diagnostic classification was applied, which may help towards making diagnosis more standardised and limiting the inter-examiner variability.

Such improvements in diagnostic concordance are potentially salient, particularly with the latest classification revision in 2018 that represented significant deviations from the previous system. This new criterion is a bit complex to apply, which may result in inconsistent usage, particularly among non-expert reviewers who might have limited experience applying the new criteria.³² In this setting, machine learning approaches could be applied as aiding tools for encouraging the less subjective application of classification criteria in different clinical situations.

Limitations and considerations

There are several limitations that must be ac-

knowledge with respect to this study. The fact that this was a single-centre study and the particular focus on an Iraqi population could reduce the generalisability to other populations and health-care systems. Patterns of periodontal disease, presentation of risk factors and clinical practice may vary greatly from one geographical and cultural setting to another.³³ To validate generalisation across diverse populations, it will be mandatory for widespread clinical application.

Because of the cross-sectional research design, how well the model tracks progression of disease over time could not be assessed. The ROR-RI in reference to the 2018 classification system has great emphasis on progression features, especially in the grading parameter and its full value might be achieved under longitudinal information conditions.³⁴ In future work, we will look at incorporating temporal aspects to improve model performance for progression-related classifications.

Although presented dataset was completed for the parameters included, a number of clinically relevant variables which could be important were not included (eg more detailed radiographic measurements, microbiological factors and genetic profile). The model performance might be further improved by considering more clinical factors, especially in difficult classification tasks such as grade prediction.³⁵ However, the utilisation of such modes in practice needs to weigh their benefit against increased data acquisition.

Future research directions

There are several opportunities in future research that arise from these results. Future research may use additional data sources (eg radiographs) to increase the sensitivity and specificity of classifications; further, advances in machine learning will likely allow for rapid implementation of models that automatically analyse radiographic images to provide additional information in conjunction with clinical parameters. Another promising area of future research is the development of dynamic models that can be used with longitudinal data to model disease progression over time; this type of model would allow researchers to capture and understand how diseases progress, a key consideration in the grading component of the 2018 classification. Dynamic models may also help clinicians predict patient disease trajectories and monitor patient responses to treatments.

Conclusion

The results show that the 2018 AAP/EFP periodontal classification could be used to make periodontal diagnoses using a hierarchical machine learning process. In addition to high accuracy in the identification of each primary disease category, the study found very high accuracy in identifying clinically significant sub-classifications of the diseases identified by the system. These results provide evidence to support the use of computerised diagnostic tools in routine dental practice, especially where they can support clinical decision making. Consistent with current principles of periodontal diagnosis, CAL and PPD were determined to be the two most important variables in the models developed in this study. Further studies are needed to validate these results on a large scale and across a wide range of populations before machine learning will be ready to support standardisation of periodontal diagnosis.

Ethics

All study procedures were conducted in accordance with the Declaration of Helsinki. Study was approved by the Ethical Committee of the College of Dentistry, decision No: 1246, dated 5 January 2025.

Acknowledgement

None.

Conflicts of interest

The authors declare that there is no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data access

The data that support the findings of this study are available from the corresponding author upon reasonable individual request.

Author ORCID numbers

Omar Husham Ali (OHA):
0000-0002-8548-2349
Nadia Mohammed Kazem (NMK):
0009-0009-9786-2267
Hadeel Mazin Akram (HMA):
0000-0001-7620-5131

Author contributions

Conceptualisation: OHA, NMK, HMA
Methodology: OHA
Software: OHA
Validation: OHA, NMK, HMA
Formal analysis: OHA
Investigation: NMK
Data curation: OHA
Writing - original draft: OHA
Writing - review and editing: NMK, HMA
Supervision: HMA

References

- Kassebaum NJ, Bernabé E, Dahiya M, Bhandari B, Murray CJ, Marcenes W. Global burden of severe periodontitis in 1990-2010: a systematic review and meta-regression. *J Dent Res*. 2014 Nov;93(11):1045-53. doi: 10.1177/0022034514552491.
- 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. doi: 10.1111/jcpe.12732.
- Albandar JM. Aggressive periodontitis: case definition and diagnostic criteria. *Periodontol* 2000. 2014 Jun;65(1):13-26. doi: 10.1111/prd.12014.
- Caton JG, Armitage G, Berglundh T, Chapple IL, Jepsen S, Kornman KS, et al. A new classification scheme for periodontal and peri-implant diseases and conditions - Introduction and key changes from the 1999 classification. *J Clin Periodontol*. 2018 Jun;45 Suppl 20:S1-S8. doi: 10.1111/jcpe.12935.
- 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 Suppl 1:S159-S172. doi: 10.1002/JPER.18-0006.
- Chapple IL, Mealey BL, Van Dyke TE, Bartold PM, Dommisch H, Eickholz P, et al. Periodontal health and gingival diseases and conditions on an intact and a reduced periodontium: Consensus report of workgroup 1 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. *J Periodontol*. 2018 Jun;89 Suppl 1:S74-S84. doi: 10.1002/JPER.17-0719.
- Trombelli L, Farina R, Silva CO, Tatakis DN. Plaque-induced gingivitis: Case definition and diagnostic considerations. *J Clin Periodontol*. 2018 Jun;45 Suppl 20:S44-S67. doi: 10.1111/jcpe.12939.
- Papapanou PN, Sanz M, Buduneli N, Dietrich T, Feres M, Fine DH, et al. Periodontitis: Consensus report of workgroup 2 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. *J Periodontol*. 2018 Jun;89 Suppl 1:S173-S182. doi: 10.1002/JPER.17-0721.
- Needleman I, Garcia R, Gkranias N, Kirkwood KL, Kocher T, Iorio AD, et al. Mean annual attachment, bone level, and tooth loss: A systematic review. *J Clin Periodontol*. 2018 Jun;45 Suppl 20:S112-S129. doi: 10.1111/jcpe.12943.
- Kornman KS, Papapanou PN. Clinical application of the new classification of periodontal diseases: Ground rules, clarifications and "gray zones". *J Periodontol*. 2020 Mar;91(3):352-60. doi: 10.1002/JPER.19-0557.
- Ertas K, Pence I, Casmeli MS, Ay ZY. Determination of the stage and grade of periodontitis according to the current classification of periodontal and peri-implant diseases and conditions (2018) using machine learning algorithms. *J Periodontal Implant Sci*. 2023 Feb;53(1):38-53. doi: 10.5051/jpis.2201060053.
- Klinger A. Periodontal Classification Revisited: The classification system in the test of time. *Compend Contin Educ Dent*. 2022 Mar;43(3):184-8. PMID: 35917589.
- Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017 Feb 2;542(7639):115-8. doi: 10.1038/nature21056.
- Cholan P, Ramachandran L, Umesh SG, Tadepalli A, Palanisamy S. The impetus of artificial intelligence on periodontal diagnosis: a brief synopsis. *Cureus*. 2023 Aug 16;15(8):e43583. doi: 10.7759/cureus.43583.
- Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016 Dec 13;316(22):2402-10. doi: 10.1001/jama.2016.17216.
- Vinayaka AM, Laxmi SK. Artificial intelligence in periodontics: Transforming the future of periodontal care. *IP*

- Int J Maxillofac Imag. 2024;10(3):129-31. doi: 10.18231/j.ijmi.2024.028.
17. Akram HM. Artificial intelligence in dentistry: advancements in periodontology and other specialties, diagnosis, treatment planning, and ethical considerations. *Dentistry Rev*. 5(2):100157. doi: 10.1016/j.dentre.2025.100157.
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. doi: 10.1186/s13104-020-05180-5.
19. Forman GH, Suermondt HJ. Hierarchical categorization method and system with automatic local selection of classifiers. Google Patents; 2008.
20. Sharma N, Kaushik P. Integration of AI in healthcare systems—A discussion of the challenges and opportunities of integrating AI in healthcare systems for disease detection and diagnosis. *AI in Disease Detection: Advancements and Applications*. 2025:239-63. doi: 10.1002/9781394278695.ch11
21. Vasylyevkyi V, Stepanov I, Koval R, Soputnyak M, Liutianska N, Sheyko V, et al. Machine learning algorithms implementation in the healthcare system as a prospective area for science, healthcare, and business. *Med Sci Ukraine (MSU)*. 2021;17(3):98-109. doi: 10.32345/2664-4738.3.2021.11
22. apantonopoulos G, Takahashi K, Bountis T, Loos BG. Artificial neural networks for the diagnosis of aggressive periodontitis trained by immunologic parameters. *PLoS One*. 2014 Mar 6;9(3):e89757. doi: 10.1371/journal.pone.0089757.
23. Farhadian M, Torkashvand Z, Shokouhi P, Torkzaban P. Developing a Prediction Model for Periodontal Disease Using Ordinal Random Forest. *Avicenna J Dental Res*. 2024;16(3):177-83. doi: 10.34172/ajdr.1766.
24. Patel JS, Su C, Tellez M, Albandar JM, Rao R, Iyer V, et al. Developing and testing a prediction model for periodontal disease using machine learning and big electronic dental record data. *Front Artif Intell*. 2022 Oct 13;5:979525. doi: 10.3389/frai.2022.979525.
25. Chen JH, Asch SM. Machine learning and prediction in medicine—beyond the peak of inflated expectations. *N Engl J Med*. 2017 Jun 29;376(26):2507-9. doi: 10.1056/NEJMp1702071.
26. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019 Apr 4;380(14):1347-58. doi: 10.1056/NEJMr1814259.
27. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. *Ann Intern Med*. 2018 Dec 18;169(12):866-72. doi: 10.7326/M18-1990.
28. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. *JAMA*. 2018 Dec 4;320(21):2199-200. doi: 10.1001/jama.2018.17163.
29. Urenda J, Avila N, Gordillo N, Kreinovich V. Hierarchical multiclass classification works better than direct classification: an explanation of the empirical fact. 2019. Department Tech Rep. 2019;5:1324.
30. Jepsen S, Caton JG, Albandar JM, Bissada NF, Bouchard P, Cortellini P, et al. Periodontal manifestations of systemic diseases and developmental and acquired conditions: Consensus report of workgroup 3 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. *J Periodontol*. 2018 Jun;89 Suppl 1:S237-S248. doi: 10.1002/JPER.17-0733.
31. Genco RJ, Borgnakke WS. Risk factors for periodontal disease. *Periodontol 2000*. 2013 Jun;62(1):59-94. doi: 10.1111/j.1600-0757.2012.00457.x.
32. Dietrich T, Ower P, Tank M, West N, Walter C, Needleman I, et al; British Society of Periodontology. Periodontal diagnosis in the context of the 2017 classification system of periodontal diseases and conditions - implementation in clinical practice. *Br Dent J*. 2019 Jan 11;226(1):16-22. doi: 10.1038/sj.bdj.2019.3.
33. Eke PI, Dye BA, Wei L, Slade GD, Thornton-Evans GO, Borgnakke WS, et al. Update on Prevalence of Periodontitis in Adults in the United States: NHANES 2009 to 2012. *J Periodontol*. 2015 May;86(5):611-22. doi: 10.1902/jop.2015.140520.
34. Fine DH, Patil AG, Loos BG. Classification and diagnosis of aggressive periodontitis. *J Clin Periodontol*. 2018 Jun;45 Suppl 20:S95-S111. doi: 10.1111/jcpe.12942.
35. Loos BG, Van Dyke TE. The role of inflammation and genetics in periodontal disease. *Periodontol 2000*. 2020 Jun;83(1):26-39. doi: 10.1111/prd.12297.