

# Identification of Risk Group for Root Caries and Analysis of Associated Factors in Older Adults Using Unsupervised Machine Learning Clustering

Linxin Jiang<sup>1,2</sup>, Shaohong Huang<sup>1</sup>, Daniel R Reissmann<sup>2</sup>, Gerhard Schmalz<sup>3</sup>, Jianbo Li<sup>1</sup>

<sup>1</sup>Stomatological Hospital, School of Stomatology, Southern Medical University, Guangzhou, Guangdong, People's Republic of China; <sup>2</sup>Department of Prosthodontics, University of Leipzig, Leipzig, Germany; <sup>3</sup>Department of Conservative Dentistry and Periodontology, Brandenburg Medical School Theodor Fontane (MHB), Brandenburg/Havel, Germany

Correspondence: Jianbo Li, Stomatological Hospital, School of Stomatology, Southern Medical University, 366 Jiangnan South Avenue, Haizhu District, Guangzhou, Guangdong province, 510280, People's Republic of China, Email 435264691@qq.com; Gerhard Schmalz, Email Gerhard.Schmalz@mhb-fontane.de

**Purpose:** This study aimed to identify the high-risk group for root caries using unsupervised machine learning and to explore the associated factors.

**Patients and Methods:** This cross-sectional study included 423 adults aged 65 to 74 years, surveyed in 2021. Clusters representing root caries risk were identified using k-prototypes clustering, with the optimal number of clusters determined by the maximum silhouette index. The confusion matrix and alluvial diagram were used to visualize the predictive accuracy and composition of the clustering results. Binary logistic regression models further analyzed the associated factors, while ROC (receiver operating characteristic) curves and the random forest model visualized the predictive performance and the most important associated factors.

**Results:** Two clusters were identified: cluster 1, with low root caries risk (21.5% with and 78.5% without root caries), and cluster 2, with high root caries risk (83.7% with and 16.3% without root caries). The clustering results predicted root caries with an accuracy of 0.81, sensitivity of 0.79, and specificity of 0.83. Overlapping results from binary logistic regression and the random forest model indicated that older age, more periodontal pockets, more attachment loss, female, a history of systemic diseases, presence of xerostomia, and presence of unrestored tooth loss were positively associated with cluster 2. Brushing tooth  $\geq 2$  times per day and a high level of oral health knowledge were negatively associated with cluster 2. The ROC curve for the binary logistic regression model showed an AUC (area under the curve) of 0.84.

**Conclusion:** Individuals who are older, female, with poorer oral and systemic health status, suboptimal oral hygiene behaviors, and lower oral health knowledge levels are more likely to be identified as high-risk group. The identified factors, revealed through unsupervised machine learning, can facilitate personalized prevention and management strategies for root caries in older adults.

**Keywords:** older adults, oral health, risk analysis, machine learning

## Introduction

Root caries is a destructive lesion of the hard tissues that occurs on the root surface.<sup>1</sup> Due to the lower mineral content of root surfaces, they are more prone to demineralization.<sup>2</sup> The pathogenesis of root caries is closely associated with root surface exposure caused by gingival recession. Aging is accompanied by unique oral physiological and behavioral characteristics, such as gingival recession due to periodontitis, reduced salivary secretion, and changes in oral hygiene habits.<sup>3,4</sup> These factors create an environment conducive to microbial colonization and biofilm formation,<sup>5</sup> thus increasing the risk of root caries.<sup>6,7</sup> With global aging, improved tooth retention has been achieved, yet the burden of root caries among older adults remains substantial, posing a major public health challenge in dentistry. This highlights the importance of root caries prevention and management.<sup>2,8</sup> However, early-stage changes in root caries are often subtle, and the primary diagnostic method—visual-tactile examination—is subject to the subjectivity of examiner.<sup>2</sup> Adjunctive diagnostic tools, such as X-ray and fluorescence-based methods, lack strong evidence to support their efficacy.<sup>9</sup> Thus, the

diagnosis of root caries, particularly in its early stages, remains controversial.<sup>9,10</sup> Those show that the critical need for risk prediction in root caries. Identifying high-risk group can facilitate targeted prevention and management.

The k-prototypes clustering algorithm is an unsupervised machine learning approach for analyzing mixed-type data. This algorithm enables simultaneous clustering of both quantitative and qualitative variables, making it a novel application in the medical field in recent years.<sup>11</sup> By identifying latent grouping patterns within the data, it allows direct recognition and classification of population characteristics without the need for predefined outcome indicators.<sup>12</sup> This method is particularly advantageous for exploring complex factors associated with high-risk group and identifying associated factors.<sup>11</sup> Unsupervised machine learning has been studied for identifying high-risk group in various diseases, such as acute exacerbation of chronic obstructive pulmonary disease,<sup>13</sup> nonspecific neck pain,<sup>14</sup> heart failure,<sup>15</sup> and immune-related liver injury.<sup>16</sup> However, there is no application of this method for identifying risk factors and high-risk group for root caries.

The objective of this study is to apply the k-prototypes clustering algorithm to identify high-risk group for root caries among older adults. By integrating binary logistic regression model, the study further investigates associated factors, providing a scientific basis for the precise identification of high-risk group and the optimization of oral health intervention strategies for older adults. The study hypothesizes that: (1) The k-prototypes clustering results are stable and effective in identifying high-risk group for root caries. (2) Multidimensional factors (such as periodontal and coronal conditions, oral hygiene behaviors) play significant roles in the risk of root caries.

## Methods

### Source of Data and Participants

The data for this study were derived from the 2021 oral health surveillance project for key populations in Guangdong Province, China, a representative cross-sectional study.<sup>17</sup> Prior to participation, all individuals were informed about the study details and signed written informed consent forms. This study strictly was approved by the Stomatological Ethics Committee of the Chinese Stomatological Association (Permit Number: 2014-003), and complied with the Declaration of Helsinki.

423 participants were recruited from community populations. The inclusion criteria were: adults aged 65 to 74 years; residents of the communities covered by the study who had not been away from the community for more than six months; voluntary participation with signed informed consent; ability to undergo oral health examinations and complete the questionnaires; and absence of severe mental or cognitive impairments.

Exclusion criteria included: individuals with severe systemic diseases that prevented them from completing the study (such as advanced cancer, end-stage renal disease); those who had undergone severe oral surgeries or treatments affecting oral health within the past six months; individuals with severe cognitive impairments preventing comprehension of the study or completion of the questionnaires; and those who refused to sign the informed consent or withdrew from the study during the research process.

### Study Design

The sample size was calculated using the formula:  $N = \text{deff} \frac{v^2 \alpha \rho (1-\rho)}{\delta^2}$ . In the formula,  $N$  represents the sample size, and the design effect ( $\text{deff}$ ) was set at 4.5. The significance level  $\alpha$  was set at 0.05, corresponding to a  $u$  value of 1.96 when the cumulative probability in the standard normal distribution is  $\alpha/2$ . The relative allowable error ( $\delta$ ) for the expected prevalence ( $p$ ) was controlled at 10%. Based on this calculation, the minimum required sample size for the survey was initially determined to be 390 participants. To account for potential non-responses, the non-response rate was estimated at 10%, resulting in a final sample size requirement of 429 participants.

A stratified random sampling method was employed to select the study samples. The stratification was based on geographic regions and population distribution, with 13 disease surveillance points designated as independent strata. Within each surveillance point, residential communities were used as the basic sampling units, and three communities were randomly selected using simple random sampling. In each selected community, 11 individuals aged 65 to 74 years were randomly chosen, resulting in a total of 33 individuals per surveillance point.

During the survey, six individuals could not complete the study due to reasons such as declining to participate, or health issues. Ultimately, 423 participants were included in the analysis. This exceeded the minimum sample size requirement.

## Variables Design

The variables in this study were defined based on the results of two components: questionnaire survey and oral health examination. The questionnaire survey was conducted face-to-face on-site by two trained professional surveyors. The survey primarily gathered information on demographic characteristics, oral health behaviors, and health conditions. Demographic variables obtained included age (years) and gender. Additionally, qualitative variables potentially associated with root caries were extracted, including education level (low, medium, high), history of systemic diseases (yes, no), presence of xerostomia (yes, no), presence of non-standard denture restoration, presence of unrestored tooth loss, use of fluoride toothpaste (yes, no), toothbrushing frequency ( $\leq 1/\text{day}$ ,  $\geq 2/\text{day}$ ), and oral health knowledge level (low, medium, high).

The oral health examination was performed under artificial light using a combination of visual inspection and probing. The instruments included a plane mouth mirror and a CPI probe. The examination followed a systematic sequence, beginning with the second permanent molar in the upper right quadrant and progressing tooth by tooth to the second permanent molar in the lower left quadrant, covering a total of 28 teeth. Third molars were not included in the examination. Diagnostic criteria were based on the basic methods for oral health surveys (5th edition) published by the World Health Organization.<sup>18</sup>

The clinical examination provided quantitative variables potentially related to root caries, including periodontal pockets (probing depth  $\geq 4$  mm), attachment loss (attachment loss  $\geq 4$  mm, examined at six specified sites: 17/16, 11, 26/27, 36/37, 31, 46/47), DMFT of coronal caries (including teeth with coronal caries, filled teeth with caries, filled teeth without caries, and teeth missing due to caries), and DMFT of root caries (including teeth with root caries, filled root caries, filled root caries without active lesions, and residual roots).

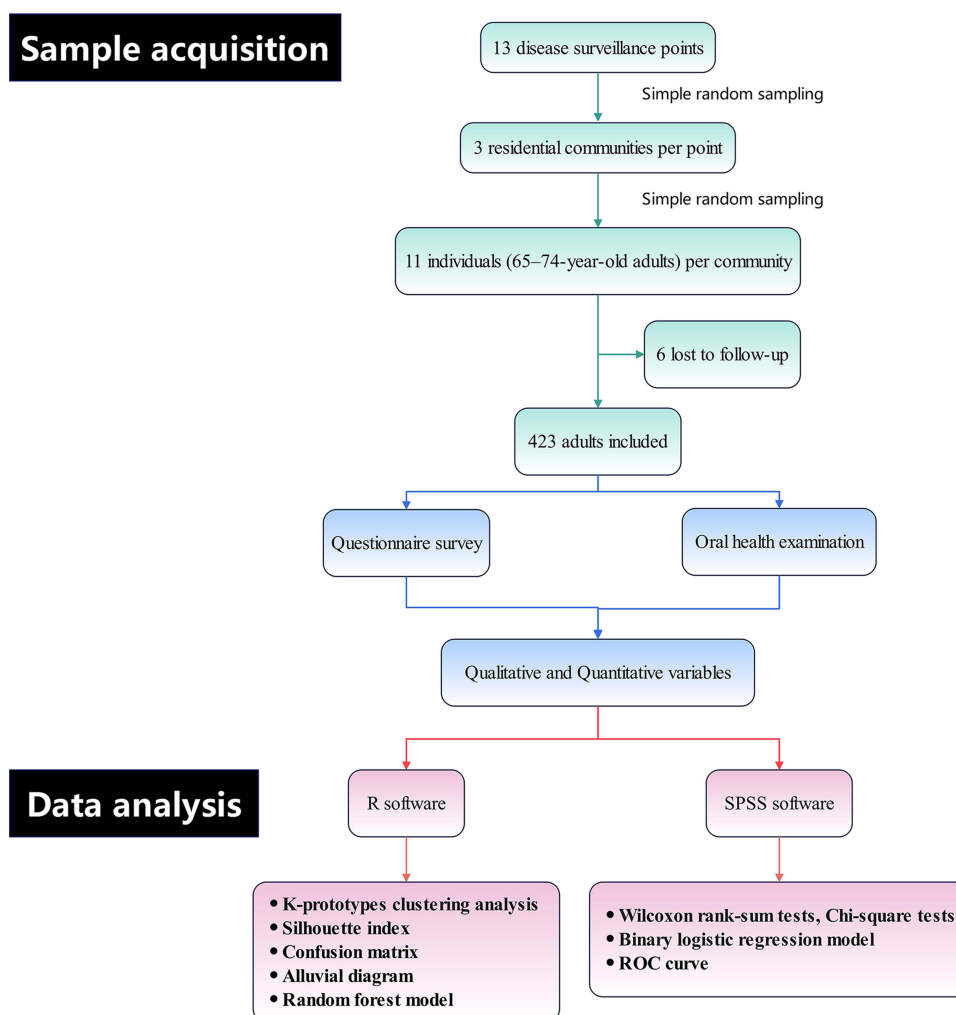
## Data Analysis

Cluster analysis in this study was conducted using the ClustMixType package (version 0.4–2) in R (version 4.4.2), employing the k-proto function for k-prototypes clustering.<sup>19</sup> K-prototypes is an unsupervised machine learning method that extends the classical k-means clustering approach, enabling the analysis of mixed data containing both qualitative and quantitative variables. Its core principle integrates euclidean distance and variable mismatch dissimilarity to comprehensively evaluate the similarity between samples. By iterative optimization, each sample is assigned to the most suitable cluster, ultimately achieving a rational grouping of samples. To ensure the robustness of the clustering results, the Silhouette index was calculated to determine the optimal number of clusters. This index evaluates clustering quality by assessing the distance of a sample within its own cluster and relative to other clusters.<sup>20</sup> The Silhouette index (S) ranges from  $-1$  to  $1$ , where values close to  $1$  indicate well-clustered samples, values near  $0$  suggest samples on cluster boundaries, and negative values imply potential misclassification. The optimal number of clusters (K) was determined by identifying the K value that maximized the average Silhouette index.

The consistency between clustering results (predicted labels) and root caries distribution (true labels) was validated using a confusion matrix,<sup>21</sup> and the differences in composition between clustering results and root caries distribution were visualized with an alluvial diagram.<sup>22</sup>

Associated factors were analyzed using SPSS software (version 25.0), with clustering results as the dependent variable and the quantitative and qualitative variables used for clustering as independent variables. For univariate analysis, the Wilcoxon rank-sum test was employed for non-normally distributed quantitative variables, while the Chi-square test was used for qualitative variables. Variables with a p-value  $< 0.05$  were included in a binary logistic regression model, and the predictive performance of the binary logistic regression model was assessed using ROC (receiver operating characteristic) curves.<sup>23</sup> To enhance the robustness of the findings, a random forest model was constructed using R. The most significant associated factors were identified by comparing and overlapping results from the logistic regression and Random Forest models.<sup>24</sup>

Figure 1 summarizes the research workflow.



**Figure 1** Research workflow for sample acquisition and data analysis.

## Results

### Machine Learning

#### Clustering Characteristics

This study included a total of 423 adults. After determining the optimal number of clusters as 2 based on the Silhouette index, k-prototypes clustering analysis categorized the samples into cluster 1 and cluster 2. Univariate analyses of both quantitative and qualitative variables revealed statistically significant differences ( $P < 0.05$ ) between the clusters for all variables except for the presence of non-standard denture restoration and use of fluoride toothpaste (Table 1). Compared to cluster 2, cluster 1 was characterized by a younger age, fewer numbers of periodontal pockets (teeth), attachment loss (teeth), and DMFT of coronal caries. Cluster 1 also had higher proportions of males, individuals with higher educational level, no history of systemic diseases, no presence of xerostomia, no unrestored of tooth loss, a toothbrushing frequency of  $\geq 2$  times/day, and a higher level of oral health knowledge.

#### Clustering Model Evaluation

Based on the clustering results and the distribution of root caries, the confusion matrix (Figure 2A) showed that among 221 samples with root caries, 175 were correctly predicted, while among 202 samples without root caries, 168 were correctly predicted. The accuracy of classification was 0.81 (95% CI: 0.77 to 0.85). The sensitivity and specificity of the predictions was 0.79 and 0.83.

**Table 1** Comparison of Demographic, Clinical, and Behavioral Characteristics Between Cluster 1 and Cluster 2

Variables	Category	Cluster (mean $\pm$ SD or n, %)		Statistical values	P
		1	2		
<b>Total</b>		214(50.6)	209(49.4)		
<b>Age (years)</b>		68.70 $\pm$ 2.62	69.44 $\pm$ 2.66	-2.811	0.005
<b>Periodontal Pockets (teeth)</b>		2.55 $\pm$ 3.86	5.82 $\pm$ 6.63	-6.033	<0.001
<b>Attachment Loss (teeth)</b>		0.97 $\pm$ 1.44	1.99 $\pm$ 1.86	-6.070	<0.001
<b>DMFT of coronal caries</b>		4.36 $\pm$ 5.96	4.84 $\pm$ 4.30	-3.743	<0.001
<b>Gender</b>	Male	121(59.6)	82(40.4)	12.690	<0.001
	Female	93(42.3)	127(57.7)		
<b>Education Level</b>	Low	39(48.8)	41(51.2)	8.935	0.011
	Medium	73(42.9)	97(57.1)		
	High	102(59.0)	71(41.0)		
<b>History of Systemic Diseases</b>	Yes	80(40.6)	117(59.4)	14.698	<0.001
	No	134(59.3)	92(40.7)		
<b>Presence of Xerostomia</b>	Yes	37(38.9)	58(61.1)	6.645	0.011
	No	177(54.0)	151(46.0)		
<b>Presence of non-standard denture restoration</b>	Yes	18(52.9)	16(47.1)	0.082	0.859
	No	196(50.4)	193(49.6)		
<b>Presence of Unrestored Tooth Loss</b>	Yes	71(34.1)	137(65.9)	44.336	<0.001
	No	143(66.5)	72(33.5)		
<b>Use of fluoride toothpaste</b>	Yes	150(51.0)	144(49.0)	0.071	0.833
	No	64(49.6)	65(50.4)		
<b>Toothbrushing Frequency</b>	$\leq$ 1 time/day	93(40.3)	138(59.7)	21.731	<0.001
	$\geq$ 2 times/day	121(63.0)	71(37.0)		
<b>Oral Health Knowledge Level</b>	Low	62(44.0)	79(56.0)	11.263	0.004
	Medium	65(45.8)	77(54.2)		
	High	87(62.1)	53(37.9)		

An alluvial diagram (Figure 2B) illustrated the correspondence between clustering results and the distribution of root caries. Cluster 2 consisted of 83.7% individuals with root caries and 16.3% without root caries, identifying it as a high-risk group for root caries. In contrast, cluster 1 consisted of 21.5% individuals with root caries and 78.5% without root caries, indicating it as a low-risk group for root caries.

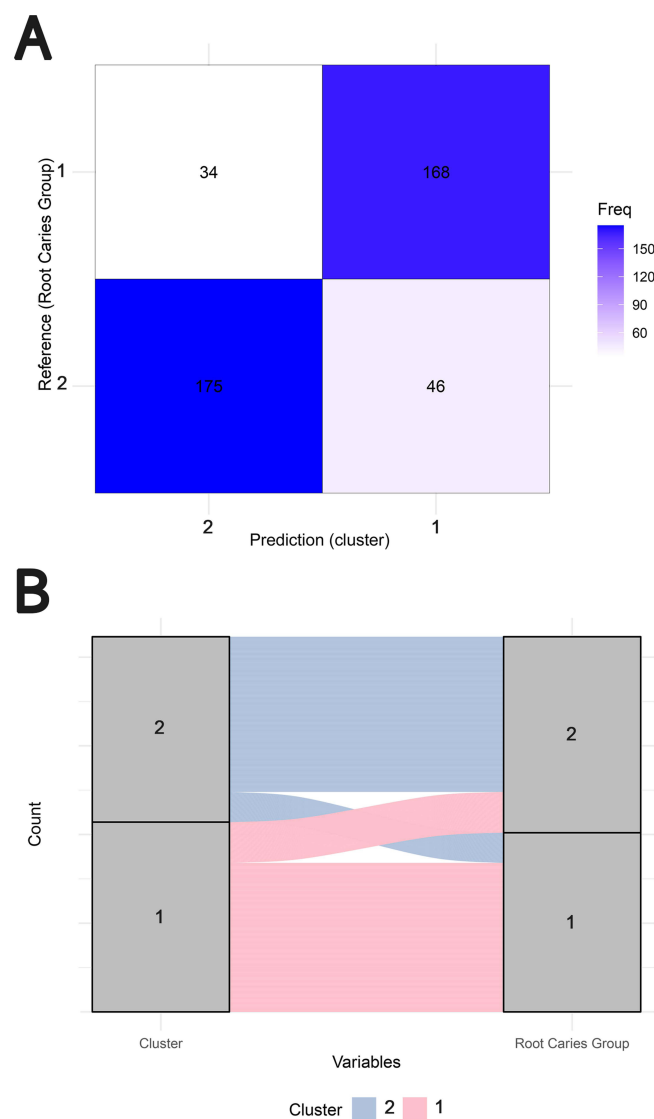
## Associated Analysis

### Logistic Model

A binary logistic regression model (Table 2) was applied to further analyze four quantitative variables and seven qualitative variables that demonstrated statistical significance ( $P < 0.05$ ) in the univariate analysis. The results indicated that the risk of being categorized into cluster 2 increased with age, the number of periodontal pockets and attachment losses. Additionally, female, having a history of systemic diseases, the presence of xerostomia, and the presence of unrestored tooth loss were associated with a higher risk of being included in cluster 2 ( $OR > 1$ ). Conversely, brushing teeth  $\geq 2$  times/day and having a high level of oral health knowledge were associated with a lower risk of being included in cluster 2 ( $OR < 1$ ).

### Logistic Model Evaluation

The ROC curve of the logistic regression model in Figure 3 showed an AUC (area under the curve) of 0.84 (95% CI: 0.80 to 0.88), indicating good predictive performance and reliable discrimination between samples.



**Figure 2** Evaluation of the clustering results and root caries distribution. **(A)** The confusion matrix illustrating the prediction accuracy of the clustering model. **(B)** The alluvial diagram demonstrating the correspondence between clustering results and the distribution of root caries. 1 in root caries group represents the absence of root caries, and 2 represents the presence of root caries.

The results of the random forest model in [Figure 4](#) highlighted the most important associated factors, all of which were included among the variables with statistically significant differences in the binary logistic regression model. This consistency reinforces the robustness of the identified factors in predicting cluster membership.

**Table 2** Variables Associated with Clusters Identified by the Binary Logistic Regression Model

Variable	Category	OR (95% CI)	P
Age (years)		1.11 (1.02~1.22)	0.024
Periodontal Pockets (teeth)		1.13 (1.07~1.19)	0.000
Attachment Loss (teeth)		1.26 (1.07~1.49)	0.007
DMFT of coronal caries		1.01 (0.96~1.06)	0.634

(Continued)

Table 2 (Continued).

Variable	Category	OR (95% CI)	P
<b>Gender</b>	Male <sup>†</sup>	–	–
	Female	3.12 (1.83~5.32)	<0.001
<b>Education Level</b>	Low <sup>†</sup>	–	–
	Medium	1.54 (0.78~3.03)	0.213
	High	1.05 (0.51~2.18)	0.886
<b>History of Systemic Diseases</b>	No <sup>†</sup>	–	–
	Yes	2.14 (1.33~3.45)	0.002
<b>Presence of Xerostomia</b>	No <sup>†</sup>	–	–
	Yes	1.92 (1.08~3.44)	0.028
<b>Presence of Unrestored Tooth Loss</b>	No <sup>†</sup>	–	–
	Yes	3.03 (1.87~4.91)	<0.001
<b>Toothbrushing Frequency</b>	≤1 time/day <sup>†</sup>	–	–
	≥2 times/day	0.45 (0.28~0.74)	0.002
<b>Oral Health Knowledge Level</b>	Low <sup>†</sup>	–	–
	Medium	1.03 (0.57~1.87)	0.911
	High	0.54 (0.30~0.97)	0.040

Note: <sup>†</sup>Reference category.

Abbreviations: OR, odds ratio; CI, confidence interval.

## Discussion

Unsupervised clustering has demonstrated robust capabilities in handling complex datasets and plays a crucial role in uncovering intricate relationships between factors.<sup>11</sup> In this study, the use of k-prototype clustering reduced manual intervention through automated feature extraction, allowing researchers to more intuitively interpret the characteristics of different clusters. Additionally, the algorithm enhances robustness to outliers by assigning them to marginal or independent clusters. Compared to traditional multifactor logistic regression models, k-prototype clustering provides a novel approach for classification that avoids subjective assumptions. It addresses the limitations of logistic regression models stemming from linearity assumptions, offering more comprehensive and structured inputs that mitigate the

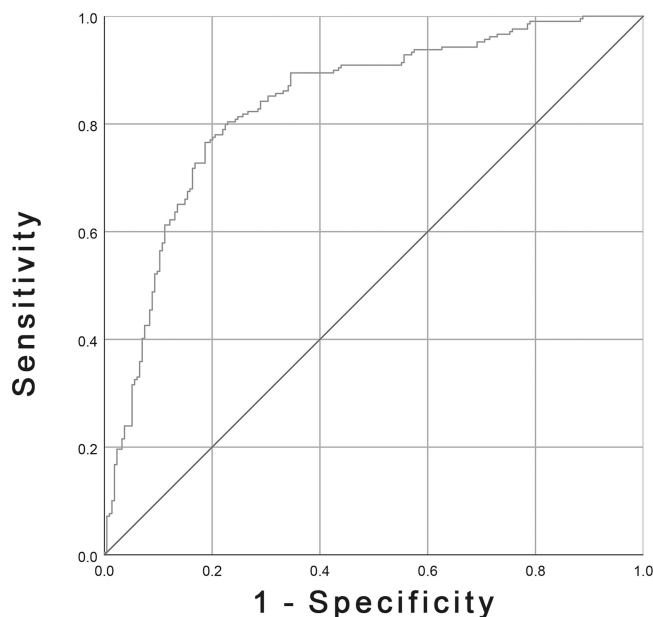
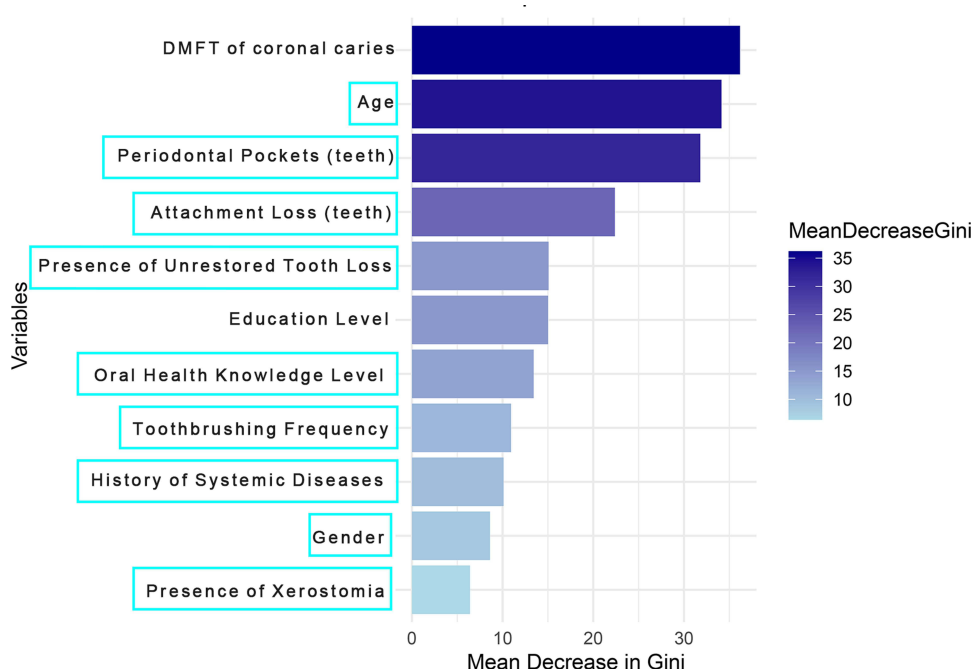


Figure 3 ROC curve of the binary logistic regression model showing predictive performance.





**Figure 4** Key variables of clusters identified by the random forest model.

negative impact of outliers on overall model performance and enhance the interpretability of predictive results. This analytical method shows considerable potential for application in risk prediction fields.

In this study, high-risk group for root caries among the older adults were effectively identified, and important associated factors were determined. The clustering analysis revealed significant differences across various indicators between high-risk individuals (cluster 2) and low-risk individuals (cluster 1). The characteristics of cluster 2 in this study were generally consistent with previous literature,<sup>6,25–27</sup> demonstrating a demographic profile of older age, a higher proportion of females, and lower educational level. Additionally, cluster 2 exhibited poorer oral and systemic health status, suboptimal oral hygiene behaviors, and lower levels of oral health knowledge. In contrast, cluster 1 showed distinct advantages across these variables. By integrating the results from the random forest model and the multivariable logistic regression analysis, key factors associated with the high-risk group were identified, including age, gender, the number of periodontal pockets and attachment loss, a history of systemic diseases, presence of xerostomia, presence of unrestored tooth loss, toothbrushing frequency, and the level of oral health knowledge.

Several studies have demonstrated that increasing age is associated with a higher risk of root caries, which is in line with the findings of this current study.<sup>28–30</sup> Additionally, multiple studies have reported that gender influences the prevalence of root caries, although there is controversy as to whether the effect is greater in males or females.<sup>31–34</sup> In this study, females were more likely to be clustered into high-risk groups. Analysis of gender differences in oral health knowledge level revealed a significantly higher proportion of females with lower level, which may partly explain this classification disparity. However, this could also be attributed to gender differences in factors such as socioeconomic and educational levels, smoking habits, and access to healthcare services in different regions. Further investigation is warranted.

The increased risk of root caries in older adults, influenced by periodontal pockets and attachment loss, which contribute to gingival recession, root surface exposure, and an increase in pathogenic bacteria, has been extensively studied and validated over the years.<sup>5,35–37</sup> In this study, the high-risk group also exhibited characteristics such as lower rates of tooth restoration, poorer oral health knowledge, and lower toothbrushing frequency. These factors may contribute to an increased periodontal burden on the remaining teeth,<sup>38</sup> plaque retention,<sup>39</sup> and other adverse effects on periodontal health. Additionally, some studies have explored the negative impact of coronal caries on root caries.<sup>34,40,41</sup> In this study, the number of DMFT for coronal caries was significantly higher in cluster 2 compared to cluster 1, with a statistically significant difference. However, in the



multivariable analysis, the contribution of this difference to identifying groups at varying levels of risk was not significant. Future research could be designed to investigate the relationship between coronal caries and root caries in greater detail.

The relationship between systemic health and oral health has garnered significant research attention in recent years.<sup>42–44</sup> Systemic diseases, particularly diabetes, are among the most common conditions that may exacerbate the prevalence and progression of periodontitis, which negatively impact the development of root caries.<sup>37</sup> Patients with systemic diseases often take multiple medications, which can lead to a reduction in salivary flow, resulting in xerostomia. This condition further impairs oral hygiene, promoting plaque accumulation and acid erosion, thereby increasing the risk of root caries.<sup>45</sup> Consequently, oral health management for patients with systemic diseases is crucial. Strategies such as educating older adults, caregivers, and family members to improve oral hygiene practices,<sup>46</sup> along with interventions like the use of artificial saliva, immunomodulators to alleviate xerostomia symptoms,<sup>47</sup> can provide benefits in reducing the risk of root caries.

Overall, the findings of this study provide valuable guidance for public health initiatives and clinical practices targeting root caries. The results of the cluster analysis offer a basis for optimizing resource allocation, enabling public health services to more precisely reach high-risk groups, thus conserving public resources. For high-risk groups, personalized intervention strategies are recommended, such as regular periodontal treatment, oral health support and education. Additionally, the study highlights the interplay between systemic health and oral health, underscoring the need to incorporate oral health interventions into comprehensive health management for older adults.

The study has some limitations. First, the cross-sectional design only reveals associations between variables and cannot establish causal relationships. Second, the variables included in the cluster analysis were selected based on expert opinions and literature review, a process that may cause subjectivity. Third, the dichotomous decision-making (root caries: yes or no) did not account for the severity, number, or activity of carious lesions, which are critical clinical considerations that might have influenced the findings. Additionally, some data were derived from questionnaires, which could be influenced by subjective perceptions, recall bias, or social desirability effects, introducing potential bias. Lastly, as the study sample was drawn from a homogenous Chinese ethnic group, the generalizability of the findings to diverse racial and ethnic populations worldwide is limited.

In light of the findings and limitations, several recommendations could be proposed to guide future research. Longitudinal studies are essential to establish causal relationships between the identified risk factors and the development of root caries. Future investigations may benefit from incorporating more refined clinical indicators, such as the severity and activity of root carious lesions, and examining the potential influence of salivary biomarkers. Expanding the study population through multicenter collaborations will enable the inclusion of more diverse demographic groups and enhance the generalizability of the results. Furthermore, integrating unsupervised machine learning approaches with additional data sources, such as radiographic imaging and salivary diagnostics, holds promise for developing more clinically relevant and targeted strategies for root caries prevention.

## Conclusion

The findings of this study indicate that unsupervised machine learning clustering can effectively identify populations at varying risk of root caries. Factors such as age, gender, periodontal health status, systemic health, xerostomia, dental restorations, oral hygiene behaviors, and oral health awareness play critical roles in increasing the risk of root caries and warrant targeted interventions. Moreover, further optimization of machine learning models and the expansion of datasets in future research may uncover additional potential factors and interaction mechanisms.

## Data Sharing Statement

The data cannot be made publicly available because it contains the private personal data of participants.

## Ethics Approval and Informed Consent

This study was approved by the Stomatological Ethics Committee of the Chinese Stomatological Association (Permit Number: 2014-003). All participants provided written informed consent prior to participation.

## Author Contributions

All authors made significant contributions to the work reported, whether in the conception, study design, execution, data acquisition, analysis and interpretation, or in all these areas. They drafted or written, or substantially revised or critically reviewed the article. They agreed on the journal to which the article will be submitted. They reviewed and agreed on all versions of the article before submission, during revision, the final version accepted for publication, and any significant changes introduced at the proofing stage. They all agree to take responsibility and be accountable for the contents of the article.

## Funding

This research was funded by Guangdong General Higher Education Key Area Special Project, Guangdong Provincial Department of Education (2024ZDZX2031) and Science research cultivation program of stomatological hospital, Southern medical university (PY2023021). Linxin Jiang received doctoral study support from the China Scholarship Council (CSC) (CSC No. 202408440099) at University of Leipzig.

## Disclosure

The authors report no conflicts of interests in this work.

## References

1. Hellyer P. Root caries. *Br Dent J*. 2021;231(1):32. doi:10.1038/s41415-021-3240-4
2. AlQranei MS, Balhaddad AA, Melo MAS. The burden of root caries: updated perspectives and advances on management strategies. *Gerodontology*. 2020;38(2):136–153. doi:10.1111/ger.12511
3. Chavda RM, Patel E, Shah S, Shah S, Sheth S, Munia V. A study to assess the unique oral health challenges faced by elderly individuals, including denture use, dry mouth, and periodontal diseases. *Inter J Oral Care Res*. 2023;11(44):–47. doi:10.4103/INJO.INJO\_20\_23
4. Hopcraft MS, Morgan MV, Satur JG, Wright FAC, Darby IB. Oral hygiene and periodontal disease in Victorian nursing homes. *Gerodontology*. 2012;29(2):e220–228. doi:10.1111/j.1741-2358.2010.00448.x
5. Do T, Damé-Teixeira N, Naginyte M, Marsh PD. Root surface biofilms and caries. *Monographs Oral Sci*. 2017;26:26–34.
6. Zhang J, Leung KCM, Sardana D, Wong MCM, Lo EC-M. Risk predictors of dental root caries: a systematic review. *Journal of Dentistry*. 2019;89:103166. doi:10.1016/j.jdent.2019.07.004
7. Gil-Montoya JA, Sánchez-Lara I, Carnero Pardo C, et al. Oral hygiene in the elderly with different degrees of cognitive impairment and dementia. *J Am Geriatr Soc*. 2017;65(3):642–647. doi:10.1111/jgs.14697
8. Hayes M, Burke F, Allen PF. Incidence, prevalence and global distribution of root caries. *Monogr Oral Sci*. 2017;26:1–8.
9. Fee PA, Macey R, Walsh T, Clarkson JE, Ricketts D. Tests to detect and inform the diagnosis of root caries. *Cochrane Database Syst Rev*. 2020;12(12):Cd013806. doi:10.1002/14651858.CD013806
10. Spielman R, Ameh G, Brandes I, Berkowitz L, Elson N, Blum IR. Challenges in differential diagnosis and treatment of cervical root resorption vs. root caries. *Primary Dental J*. 2024;13(2):65–70. doi:10.1177/20501684241256468
11. Eckhardt CM, Madjarova SJ, Williams RJ, et al. Unsupervised machine learning methods and emerging applications in healthcare. *Knee Surge Sports Traumatol Arthro*. 2023;31(2):376–381. doi:10.1007/s00167-022-07233-7
12. Friedman J. The elements of statistical learning: data mining, inference, and prediction. (No Title). 2009.
13. Chen Z, Wang J, Wang H, et al. Machine learning reveals sex differences in clinical features of acute exacerbation of chronic obstructive pulmonary disease: a multicenter cross-sectional study. *Front Med*. 2023;10:1105854. doi:10.3389/fmed.2023.1105854
14. Hwang UJ, Kwon OY, Kim JH. Unsupervised machine learning for clustering forward head posture, protraction and retraction movement patterns based on craniocervical angle data in individuals with nonspecific neck pain. *BMC Musculoskeletal Disor*. 2024;25(1):376. doi:10.1186/s12891-024-07485-z
15. Simonsen J, Modin D, Skaarup K, et al. Utilizing echocardiography and unsupervised machine learning for heart failure risk identification. *Int J Cardiol*. 2025;418:132636. doi:10.1016/j.ijcard.2024.132636
16. Yamamoto T, Morooka H, Ito T, Ishigami M. Clustering using unsupervised machine learning to stratify the risk of immune-related liver injury. *J Gastroenterol Hepatol*. 2023;38(2):251–258. doi:10.1111/jgh.16038
17. Tian T, Yang Z, Li S, et al. Cross-sectional survey and analysis of factors influencing the prevalence of dental caries among older individuals aged 65–74 in Guangdong Province in 2021. *BMC Oral Health*. 2024;24(1):24. doi:10.1186/s12903-023-03776-4
18. Organization WH. *Oral Health Surveys: Basic Methods*. World Health Organization; 2013.
19. Huang Z. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data min Knowl Discov*. 1998;2(3):283–304. doi:10.1023/A:1009769707641
20. Mur A, Dormido R, Duro N, Dormido-Canto S, Vega J. Determination of the optimal number of clusters using a spectral clustering optimization. *Expert Syst Appl*. 2016;65:304–314. doi:10.1016/j.eswa.2016.08.059
21. Susmaga R. Confusion matrix visualization. In: Kłopotek MA, Wierchoń ST, Trojanowski K, editors. *Intelligent Information Processing and Web Mining*. Berlin, Heidelberg: Springer; 2004:107–116.
22. Yen PT, Chien TW. Using the alluvial diagram to display variable characteristics for COVID-19 patients and research achievements on the topic of COVID-19, epidemiology, pathogenesis, and vaccine (CEPV). *Bibliometric Analysis*. 2023;102:e33873.

23. Grey DR, Morgan BJT. Some aspects of ROC curve-fitting: normal and logistic models. *J math psychol.* 1972;9(1):128–139. doi:10.1016/0022-2496(72)90009-0
24. Fawagreh K, Gaber MM, Elyan E. Random forests: from early developments to recent advancements. *Syst Sci Control Eng.* 2014;2(1):602–609. doi:10.1080/21642583.2014.956265
25. Tokumoto K, Kimura-Ono A, Mino T, et al. Risk factors for root caries annual incidence and progression among older people requiring nursing care: a one-year prospective cohort study. *J Prostho Res.* 2022;66(2):250–257. doi:10.2186/jpr.JPR\_D\_20\_00272
26. Zhang J, Leung KCM, Chu CH, Lo ECM. Risk indicators for root caries in older adults using long-term social care facilities in Hong Kong. *Commun Dent Oral Epidemiol.* 2020;48(1):14–20. doi:10.1111/cdoe.12495
27. Hayes M, Da Mata C, Cole M, McKenna G, Burke F, Allen PF. Risk indicators associated with root caries in independently living older adults. *J Dent.* 2016;51:8–14. doi:10.1016/j.jdent.2016.05.006
28. Souza JGS, Costa Oliveira BE, Sampaio AA, Lages VA, Romão DA, Martins A. Contextual and individual determinants of root caries in older people. *Caries Res.* 2018;52(3):253–261. doi:10.1159/000479800
29. Sen S, Kumar S, Chakraborty R, Srivastava S, Mishra G, Choudhary A. Prevalence and risk factors of root caries in the geriatric population in the rural sector. *J Family Med Primary Care.* 2020;9(2):771–776. doi:10.4103/jfmpe.jfmpe\_1053\_19
30. Chen W, Zhu T, Zhang D. The prevalence and common risk indicators of root caries and oral health service utilization pattern among adults, a cross-sectional study. *PeerJ.* 2023;11:e16458. doi:10.7717/peerj.16458
31. Badr F, Sabbah W. Inequalities in untreated root caries and affordability of dental services among older american adults. *Int J Environ Res Public Health.* 2020;18(1):17. doi:10.3390/ijerph18010017
32. Gao YB, Hu T, Zhou XD, et al. How root caries differs between middle-aged people and the elderly: findings from the 4th national oral health survey of China. *Chin J Dental Res.* 2018;21(3):221–229. doi:10.3290/j.cjdr.a41078
33. Mamai-Homata E, Topitsoglou V, Oulis C, Margaritis V, Polychronopoulou A. Risk indicators of coronal and root caries in Greek middle aged adults and senior citizens. *BMC Public Health.* 2012;12(1):484. doi:10.1186/1471-2458-12-484
34. Tepox-Puga SM, Rocha-Ortiz JA, Medina-Solís CE, Castrejón-Pérez RC, Sánchez-García S, Borges-Yáñez SA. Prevalence, and risk indicators of coronal and root caries in mexican older adults in nursing homes. *Clin Cosmetic Invest Dentis.* 2023;15:333–347. doi:10.2147/CCIDE.S439342
35. Li Y, Xiang Y, Ren H, et al. Association between periodontitis and dental caries: a systematic review and meta-analysis. *Clin Oral Invest.* 2024;28(6):306. doi:10.1007/s00784-024-05687-2
36. Vehkalahti M, Paunio I. Association between root caries occurrence and periodontal state. *Caries Res.* 1994;28(4):301–306. doi:10.1159/000261990
37. Gavriilidou NN, Belibasakis GN. Root caries: the intersection between periodontal disease and dental caries in the course of ageing. *Br Dent J.* 2019;227(12):1063–1067. doi:10.1038/s41415-019-0973-4
38. Guarnieri R, Ippoliti S. Restoration of periodontally compromised dentitions using telescopic full-arch retrievable prosthesis supported by tooth-implant combination: a long-term retrospective study. *Inter J Periodon Restor Den.* 2018;38(2):217–224. doi:10.11607/prd.3102
39. Ruiz Núñez MDR, da Luz Raulino M, Goulart Castro R, Schaefer Ferreira de Mello AL. Dental plaque control strategies for the elderly population: a scoping review. *Inter J Dental Hygiene.* 2022;20(1):167–181. doi:10.1111/idh.12497
40. Vehkalahti MM. Relationship between root caries and coronal decay. *J Dent Res.* 1987;66(10):1608–1610. doi:10.1177/00220345870660102001
41. Sugihara N, Maki Y, Kurokawa A, Matsukubo T. Cohort study on incidence of coronal and root caries in Japanese adults. *The Bulletin of Tokyo Dental College.* 2014;55(3):125–130. doi:10.2209/tdcpub.55.125
42. Coll PP, Lindsay A, Meng J, et al. The prevention of infections in older adults: oral health. *J Am Geriatr Soc.* 2020;68(2):411–416. doi:10.1111/jgs.16154
43. Weintraub JA, Kaeberlein M, Perissinotto C, Atchison KA, Chen X. Geroscience. *Aging Oral Health Res.* 2023;31:2–15.
44. Kapila YL. Oral health's inextricable connection to systemic health: special populations bring to bear multimodal relationships and factors connecting periodontal disease to systemic diseases and conditions. *Periodontology 2000.* 2021;87(1):11–16. doi:10.1111/prd.12398
45. Chan AKY, Tamrakar M, Jiang CM, Ecm L, Leung KCM, Chu CH. Common medical and dental problems of older adults: a narrative review. 2021;6.
46. Todayama N, Hara R, Tabata T, Hatanaka Y. Systemic and oral characteristics of convalescent inpatients requiring oral-health management by a dental specialist during hospitalization. *Geriatrics.* 2024;9(3). doi:10.3390/geriatrics9030082
47. Baer AN, Walitt B. Update on Sjögren syndrome and other causes of sicca in older adults. *Rheum Dis Clin North Am.* 2018;44(3):419–436. doi:10.1016/j.rdc.2018.03.002

## Clinical Interventions in Aging

## Publish your work in this journal

Clinical Interventions in Aging is an international, peer-reviewed journal focusing on evidence-based reports on the value or lack thereof of treatments intended to prevent or delay the onset of maladaptive correlates of aging in human beings. This journal is indexed on PubMed Central, MedLine, CAS, Scopus and the Elsevier Bibliographic databases. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/clinical-interventions-in-aging-journal>

**Dovepress**  
Taylor & Francis Group