International Journal of Dental Research and Allied Sciences
2021, Volume 1, Issue 2, Page No: 57-65

Available online at: www.tsdp.net

ISSN: 3062-3502

Original Article

Machine Learning-Based Prediction of Dental Caries in Children: Insights from Parental Perceptions

Aksel Ibsen[1], Amalie Dahl[1]*

[1]Department of Clinical Dentistry, Faculty of Medicine, University of Bergen, Bergen, Norway.

***E-mail** ✉ amalie.dahl2025@gmail.com

Received: 01 August 2021; **Revised:** 23 October 2021; **Accepted:** 25 October 2021

ABSTRACT

Dental caries remains the most prevalent chronic infectious disease of childhood and a major public health concern in both low- and high-income nations, even though it is largely preventable. The present study assessed the feasibility of identifying dental caries in children through a machine learning model based on parental perceptions of their child's oral health collected via questionnaire. Data were gathered from 182 parents or caregivers and their children aged 2–7 years residing in Los Angeles County. The random forest algorithm was applied to determine which survey questions predicted active caries or overall caries experience. Three-fold cross-validation was implemented, and the cutoff point was defined by maximizing sensitivity and specificity with a minimum sensitivity of 70%. The predictive contribution of each survey item to the classification of active caries and caries experience was quantified using mean decreased Gini (MDG) and mean decreased accuracy (MDA). Strong predictors of active caries included parent's age (MDG = 0.84; MDA = 1.97), unmet healthcare needs (MDG = 0.71; MDA = 2.06), and being African American (MDG = 0.38; MDA = 1.92). For overall caries experience, key predictors were parent's age (MDG = 2.97; MDA = 4.74), reports of oral health problems within the past year (MDG = 2.20; MDA = 4.04), and the child having experienced dental pain (MDG = 1.65; MDA = 3.84). Findings support the viability of using parent-completed surveys analyzed through machine learning for caries risk screening among young children.

Keywords: Dental caries, Children, Oral health, Disparities, Machine learning, Random forest

How to Cite This Article: Ibsen A, Dahl A. Machine Learning-Based Prediction of Dental Caries in Children: Insights from Parental Perceptions. Int J Dent Res Allied Sci. 2021;1(2):57-65. https://doi.org/10.51847/oDc7gXlsl7

Introduction

Dental caries represents the leading chronic infection in childhood and continues to pose a serious global health issue, despite being largely avoidable [1–6]. In the United States (US), although caries prevalence has declined overall, cases among children under 5 years have increased [4, 7, 8]. Currently, early childhood caries (ECC) affects approximately 21% of American children between 2 and 5 years old [9]. Untreated lesions can lead to pain, infection, tooth loss, and hinder essential daily activities such as eating, sleeping, learning, and playing [7, 10, 11].

Multiple risk factors—diet, bacterial load, and oral hygiene—contribute to ECC, but preventive habits can mitigate most cases [4, 6–9]. While community-based measures like fluoridated water and public oral health programs play major roles, education remains a cornerstone of prevention [7, 12]. Dentists and public health professionals traditionally lead these efforts; however, equipping parents to recognize early indicators of ECC (e.g., enamel discoloration and visible plaque) is crucial for early detection and management.

The COVID-19 pandemic highlighted the necessity of parent involvement in oral care. Many non-urgent dental appointments were suspended, forcing reliance on parents' assessments of their children's oral conditions [13]. Even beyond the pandemic, growing use of teledentistry and virtual dental models means parental evaluations will increasingly supplement clinical screenings. Understanding caregivers' awareness and perceptions is therefore vital when face-to-face exams are limited.

Previous studies by the authors found significant correlations between parent- or child-reported survey responses and clinical oral health indices—particularly items related to tooth color satisfaction, alignment, and pain [14–16]. These results suggest that self- or proxy-reported surveys may help estimate oral health status when physical examinations are not feasible. While children over 8 years can often self-report accurately, those younger than 8 rely primarily on parental observation [15, 17–22].

This investigation addresses that gap by employing a random forest (RF) algorithm to isolate the most predictive parental survey items for identifying active and past caries in children aged 2–7 years. Clinical dental examinations were also performed to confirm active lesions and caries experience, including counts of decayed, missing, and filled teeth (DMFT) [23].

Materials and Methods

Sample

This analysis included 182 parent–child pairs with children aged 2–7 years, drawn from a larger cohort of roughly 600 participants aged 2–<18 years, collected between August 2015 and October 2019 [15]. Surveys were completed using Audio Computer-Assisted Self-Interview (ACASI) software. Participants were recruited from 12 dental sites across Los Angeles County, including community clinics, health centers, and pediatric or general practices. Sites were selected to ensure socioeconomic and ethnic diversity—from low-income immigrant areas to affluent professional communities.

Eligibility was limited to one child per household, and those undergoing orthodontic treatment were excluded to avoid interference with dental evaluation. This subsample of 182 children represented approximately 30% of the total study population. A detailed methodological description is available in earlier publications [14, 15].

Ethical clearance was granted by the University of California, Los Angeles (UCLA) Institutional Review Board (Approval No. 13-001330). Parents provided written informed consent before participation. The study followed the STROBE guidelines for cross-sectional research [24, 25].

Data collection

Parents or caregivers completed a 34-item survey using the Audio Computer-Assisted Self-Interview (ACASI) system. The questionnaire included self-reported items related to oral health, addressing aspects of physical health, psychological well-being, and social functioning. The content was adapted from prior research examining multilevel influences on oral health within a life-course framework, encompassing genetic, biological, behavioral, social, and economic factors that evolve from childhood through adulthood [15, 16, 26].

For children aged 2–7 years, the survey covered topics such as:

- usage of fluoridated tap water and fluoridated toothpaste,
- access to dental care and fluoride varnish applications,
- oral health status, tooth and gum conditions,
- general health and well-being,
- preventive health behaviors carried out by parents,
- toothbrushing routines, oral hygiene practices, and socio-demographic details [15, 16, 26].

The majority of respondents were parents, although some were caregivers (e.g., grandparents). The analysis did not separate results based on respondent type. The questionnaire was available only in English, but Spanish and Chinese translators were present for assistance when necessary. On average, the survey required 20–45 minutes to complete. Participants received USD 55 in cash upon completing the questionnaire and having their child examined by a dentist.

Children underwent clinical dental examinations either before or after their parents filled out the survey. Examinations included inspection of all primary and permanent teeth, evaluation of the oral mucosa, detection of caries and decalcified white spots, assessment of plaque accumulation on central and molar teeth, and evaluation of gingival inflammation and bleeding on probing.

Two faculty dentists from the UCLA School of Dentistry conducted all examinations. They received training and calibration to ensure consistency, with Cohen's kappa used to assess inter- and intra-rater reliability. Duplicate dental exams were performed on the same children to measure consistency between and within examiners. During training at each site, 51 children were examined by both dentists to evaluate inter-rater reliability. Additionally, in every site, three randomly chosen children were re-examined by both pediatric dentists but recorded by the same assistant, with all clinical measurements repeated.

For active caries (DT), both examiners identified 2 children as having DT (DT > 0) and 44 as caries-free (DT = 0). The overall agreement between examiners for detecting active caries was 88%. However, due to the low prevalence (13%) of active caries, the Cohen's kappa (0.39) appeared low despite high agreement—an expected outcome in imbalanced datasets [27]. Therefore, percentage agreement was considered a more suitable reliability indicator [27].

For caries experience, both examiners identified 23 children with (DMFT > 0) and 23 without caries experience, achieving 90% agreement with a kappa value of 0.80. Dental findings were categorized by presence or absence of active caries (DT) and DMFT/dmft index scores to indicate total caries experience [23]. The DMFT/dmft index was coded as 0 for no caries and 1 for one or more decayed, missing, or filled teeth—covering both primary and permanent dentition. White spot lesions, crooked teeth, fractures, and positional anomalies were excluded from analysis.

Data analysis

A Random Forest (RF) algorithm was implemented for data modeling, which operates by combining multiple decision trees [28, 29]. Each tree is generated by recursively splitting variables within the training dataset to create two optimal subsets, forming branches and leaf nodes based on a splitting criterion. At every node, data points are classified by conditional thresholds applied to predictor variables. During prediction, each observation travels from the root node to a terminal leaf, and results are aggregated across multiple trees to enhance accuracy [28, 29]. RF is suitable for complex datasets and can perform both classification and regression analyses [28].

For model development, 70% of participants were randomly allocated to a training set and 30% to a testing set, stratified by outcome variables (active caries and caries experience) [28, 29]. Given the limited sample size and low caries prevalence, a three-fold cross-validation procedure was applied within the training data to tune model parameters [28–30].

The training set was randomly divided into three equal parts. In each cross-validation cycle, two folds were used for model training and the remaining one for validation. This process was repeated three times, ensuring each fold functioned once as the validation set. Results from all three iterations were aggregated to form the final model [28–30].

Key parameters derived for each decision tree included:

- mtry (number of variables considered per tree),
- ntree (total number of trees grown).

The optimal model was selected based on the highest aggregated area under the ROC curve (AUC) during cross-validation. A classification threshold was defined by maximizing the sum of sensitivity and specificity, while ensuring minimum sensitivity ≥ 70%. Model performance was then assessed using the testing dataset [28–30].

Table 1 presents RF model parameters and validation performance for both active caries and caries experience outcomes.

Table 1. Random Forest performance metrics for cross-validation and testing datasets regarding active caries and caries experience.

Condition	Tuning Parameters	Threshold	3-Fold Cross-Validation Accuracy	3-Fold Cross-Validation Sensitivity	3-Fold Cross-Validation Specificity	Test Accuracy	Test Sensitivity	Test Specificity
Active Caries	mtry = 41; ntree = 100	0.08	0.71	0.94	0.68	0.62	0.57	0.63
Caries Experience	mtry = 2; ntree = 100	0.36	0.71	0.78	0.64	0.73	0.92	0.55

Notes: (1) mtry denotes the number of variables used per tree; (2) ntree is the total number of trees generated; (3) threshold indicates the cutoff for binary classification.

The relative influence of each questionnaire variable on the classification of oral health categories was assessed using two indices: the mean reduction in accuracy (MDA) and the mean reduction in Gini impurity (MDG) [29, 30]. The MDA quantifies how much predictive accuracy declines when a specific item is randomly permuted, representing the loss in model precision due to its exclusion. In contrast, MDG reflects how consistently a variable contributes to cleaner data partitions within the trees, calculated from the average decline in Gini impurity (pi(i − pi)) whenever that feature is used to create a split. Elevated MDG values occasionally occur because the R statistical package computes impurity using raw counts rather than normalized proportions. Both MDA and MDG are dimension-free. MDA is expressed as the proportion of correctly classified records, while MDG measures the quality of node separation. Since no standard thresholds exist for either index [29, 30], variables with comparatively higher MDA and MDG scores were interpreted as more predictive for the study outcomes (active caries and caries experience). Integrating both indices resulted in more stable and precise classifications. All random forest analyses were conducted using the R platform.

Figures 1 and 2 display the items (y-axis) and their importance coefficients (x-axis), where longer bars signify greater predictive weight.

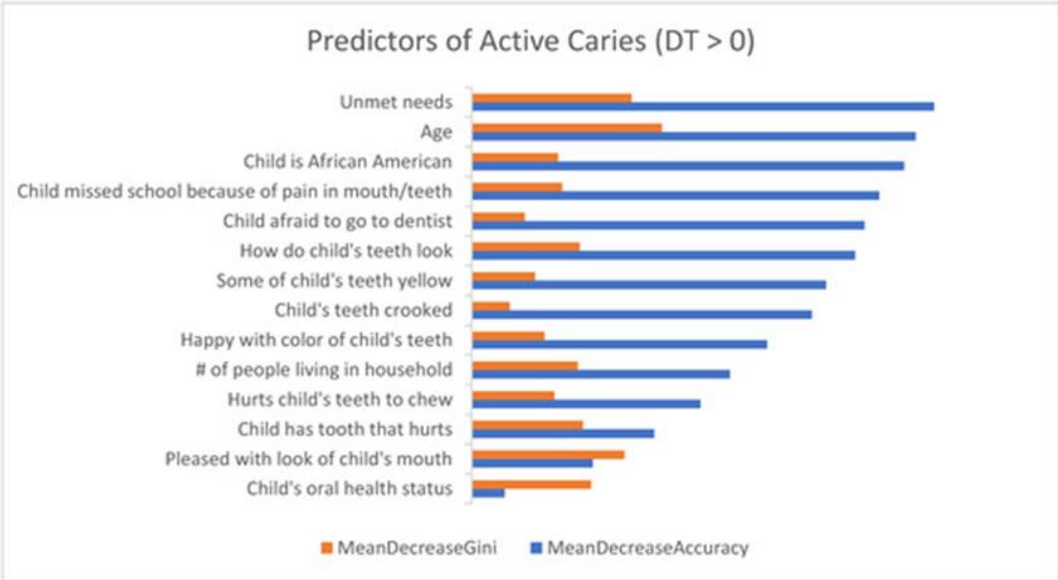

Figure 1. Demographic and oral-health attributes predicting active caries (DT > 0) based on MDA and MDG indices.

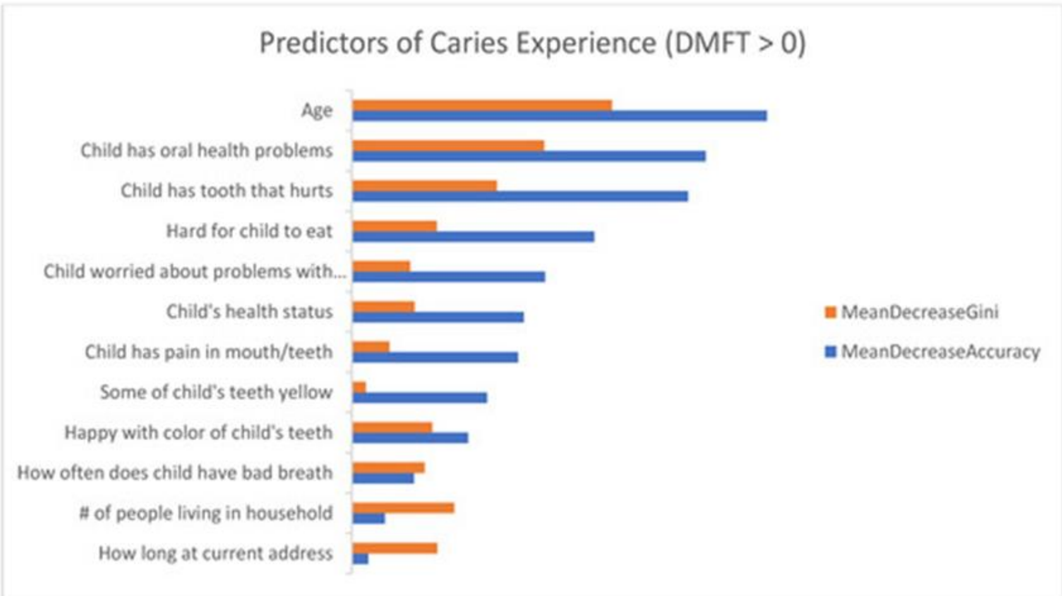

Figure 2. Demographic and oral-health attributes predicting caries experience (DMFT > 0) derived from the same measures.

Results and Discussion

Sample description

Table 2 summarizes the main demographic and outcome characteristics. Most parent or caregiver respondents were women (n = 126; 69%) between 33 and 44 years old (n = 115; 63%). Ethnic distribution included Hispanic/Latino (n = 71; 39%), White (n = 45; 24%), Asian (n = 21; 11%), and African American (n = 14; 8%). Nearly 42% of the children (n = 76) were 2–4 years old. Reported child ethnicities were

Hispanic/Latino (n = 71; 39%), White (n = 43; 24%), Asian (n = 21; 12%), African American (n = 14; 7%), and multiracial (n = 21; 12%).

Table 2. Overview of primary outcomes, demographics, and oral-health variables (n = 182).

Category	Variable	n (%)
Primary Outcome Measures		
Active Caries	Present	23 (13)
	Absent	159 (87)
Caries Experience (DMFT)	Present	86 (47)
	Absent	96 (53)
Parent Characteristics		
Age, years	Under 30	42 (23)
	30–44	115 (63)
	45 or older	25 (14)
Gender	Male	56 (31)
	Female	126 (69)
Race/Ethnicity	White/Caucasian	45 (24)
	African American	14 (8)
	Hispanic/Latino	71 (39)
	Asian	21 (11)
	Other	31 (18)
Household Size	3 or fewer	31 (17)
	4–5	101 (56)
	6 or more	50 (27)
Duration at Current Residence	1 year or less	33 (18)
	More than 1 to 5 years	82 (45)
	More than 5 to 10 years	36 (20)
	More than 10 years	31 (17)
Child Characteristics		
Age, years	2	21 (12)
	3	25 (14)
	4	30 (16)
	5	22 (12)
	6	44 (24)
	7	40 (22)
Gender	Male	93 (51)
	Female	89 (49)
Race/Ethnicity	White/Caucasian	43 (24)
	African American	14 (7)
	Hispanic/Latino	71 (39)
	Asian	21 (12)
	Multi-racial	21 (12)

	Other	12 (6)
Oral Health Survey Responses		
Child's General Oral Health Status	Excellent or Very Good	120 (66)
	Good	50 (27)
	Fair or Poor	12 (7)
Oral Health Problem in Past 12 Months	Yes	55 (30)
	No	127 (70)
Satisfaction with Child's Mouth Appearance (Past 4 Weeks)	Always or Almost Always	149 (82)
	Often or Sometimes	30 (16)
	Almost Never or Never	3 (2)
Child's Pain/Discomfort in Mouth (Past 4 Weeks)	Always or Almost Always	0 (0)
	Often or Sometimes	14 (8)
	Almost Never or Never	168 (92)
Frequency of Child's Bad Breath	Always or Almost Always	10 (6)
	Often or Sometimes	91 (50)
	Almost Never or Never	81 (44)
Appearance of Child's Teeth	Looks Fine	119 (66)
	Could Look Slightly Better	48 (26)
	Could Look Much Better	15 (8)
Child's Concern About Oral Problems (Past 4 Weeks)	Always or Almost Always	1 (1)
	Often or Sometimes	7 (4)
	Almost Never or Never	174 (95)
Child's Mouth Pain	Always or Almost Always	0 (0)
	Often or Sometimes	9 (5)
	Almost Never or Never	173 (95)
Child Has a Painful Tooth	Always or Almost Always	1 (1)
	Often or Sometimes	15 (8)
	Almost Never or Never	166 (91)
Pain When Chewing	Always or Almost Always	0 (0)
	Often or Sometimes	7 (4)
	Almost Never or Never	175 (96)
Difficulty Eating Due to Mouth Pain	Always or Almost Always	0 (0)
	Often or Sometimes	9 (5)
	Almost Never or Never	173 (95)
Satisfaction with Child's Tooth Color	Very Much or Quite a Bit	141 (78)
	Somewhat	24 (13)
	A Little Bit or Not at All	17 (9)
Child's Teeth Are Yellow	Yes	21 (12)
	No	161 (88)
Child's Teeth Are Crooked	Yes	15 (8)
	No	167 (92)
Unmet Dental Care Needs (Past 12 Months)	Yes	8 (4)

	No	174 (96)
Child's Fear of Dentist	Not at All	97 (53)
	A Little Bit or Somewhat	76 (42)
	A Great Deal	9 (5)
School Days Missed Due to Oral Pain (Past School Year, if Applicable)	Never	168 (92)
	1 to 3 Days	13 (7)
	4 Days or More	1 (1)

Clinical assessment showed 13% (n = 23) of children had active caries, and 47% (n = 86) exhibited caries experience as defined by the DMFT index.

Predictive items for active caries

Figure 1 illustrates the top 10 survey variables most strongly associated with active caries, based on the random forest algorithm. These included parent age (MDG = 0.84; MDA = 1.97), unmet needs (MDG = 0.71; MDA = 2.06), parental satisfaction with the child's dental appearance (MDG = 0.68; MDA = 0.54), overall perceived oral health (MDG = 0.53; MDA = 0.14), toothache presence (MDG = 0.50; MDA = 0.81), general dental aesthetics (MDG = 0.48; MDA = 1.70), household size (MDG = 0.47; MDA = 1.15), school absence due to mouth pain (MDG = 0.40; MDA = 1.81), African American child (MDG = 0.38; MDA = 1.92), and pain when chewing (MDG = 0.37; MDA = 1.02).

Predictive items for caries experience (DMFT Index)

As seen in **Figure 2**, the ten most predictive factors for caries experience were: parent age (MDG = 2.97; MDA = 4.74), child with an oral issue in the prior 12 months (MDG = 2.20; MDA = 4.04), child's tooth pain (MDG = 1.65; MDA = 3.84), number of household residents (MDG = 1.17; MDA = 0.38), length of residence (MDG = 0.98; MDA = 0.19), difficulty eating due to oral pain (MDG = 0.97; MDA = 2.77), parental satisfaction with tooth color (MDG = 0.92; MDA = 1.33), frequency of halitosis (MDG = 0.83; MDA = 0.71), overall child health (MDG = 0.72; MDA = 1.97), and child's concern about oral problems (MDG = 0.66; MDA = 2.21).

The 3-fold cross-validation yielded an accuracy of 0.71, sensitivity of 0.94, and specificity of 0.68 for active caries; and accuracy of 0.71, sensitivity of 0.78, and specificity of 0.64 for caries experience **(Table 1)**.

In this investigation, Random Forest (RF) modeling was applied to determine which questionnaire variables could predict active dental caries and caries experience—measured through the DMFT index—among children aged 2 to 7 years.

This paper adds to a growing body of research developing algorithmic approaches and diagnostic frameworks that can assist clinicians, oral health researchers, and policymakers in areas such as screening, monitoring, and health planning [14–18, 26]. Previous works focusing on children aged 8–17 years (where both parents and children responded to the survey) [14–18] identified predictors of oral health similar to those observed in the present study of younger children (2–7 years), whose data were parent-reported. For instance, this study confirmed that aesthetic perceptions—such as satisfaction with the child's dental appearance, mouth, and gum condition, or the color of the teeth—served as strong indicators of caries risk. Tooth color (e.g., white, yellow, or brown), commonly reflecting oral cleanliness and self-care behavior [17], was among the consistent predictors across both age groups. Moreover, functional indicators such as dental pain or difficulty eating due to oral discomfort also strongly predicted caries presence and overall oral health in this age range, again mirroring findings in older children [14–18]. These outcomes highlight the potential of oral health questionnaires as efficient predictive tools for pediatric oral assessments. Furthermore, identifying critical determinants of dental caries through machine learning offers clinicians an opportunity to guide families toward better hygiene practices and dietary habits to prevent early childhood caries, as many survey questions address preventive behaviors [31, 32].

The analysis also revealed that unmet needs and demographic profiles—such as the parent's age, the child's race (specifically African American), household size, and duration of residence at the current address—were significant predictors of both active caries and caries experience in the 2–7-year group. These findings reinforce the influence of socio-demographic determinants on oral health, aligning with existing evidence that characteristics like race, income, and household composition contribute to inequalities in dental health and treatment accessibility [2, 5, 6, 8]. Access to this type of predictive information allows dental professionals to perform more comprehensive caries risk evaluations encompassing biological, social, and cultural determinants, as well as indicators of disease and protective factors, and when necessary, to connect families with community or social support services [32].

Several limitations must be acknowledged. The relatively small sample (n = 182) and low prevalence of active caries (13%) constrained the power to detect predictive variables. Nevertheless, the RF approach was well-suited for handling the nonlinear, high-dimensional dataset, demonstrating strong sensitivity and moderate specificity. Another limitation arises from reliance on parent- or caregiver-completed questionnaires, which may introduce biases such as social desirability or response bias, particularly if non-parental guardians participated. While self-reported measures from children are ideal, proxy reports remain appropriate when participants are too young to provide reliable responses [15, 22, 33]. Additionally, potential selection bias exists because participants received financial incentives and were recruited from dental clinics already providing care, which may limit the generalizability to populations without regular dental access.

Conclusion

This study illustrates that applying machine learning techniques to oral health surveys can help clinicians pinpoint essential risk factors for caries in infants and preschool-aged children. Once these predictors are recognized, they can be incorporated into standard caries risk assessments and used as educational opportunities for families to improve oral hygiene practices. Moreover, as teledentistry and virtual home-based dental evaluations expand, parents will increasingly assess their child's oral health using guided questionnaires targeting high-probability indicators of active caries and caries experience. Consequently, developing algorithm-driven "toolkits" for dental professionals could substantially enhance early prevention strategies and management of pediatric dental caries.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Colak H, Dulgergil CT, Dalli M, Hamidi MM. Early childhood caries update: a review of causes, diagnoses, and treatments. J Nat Sci Biol Med. 2013;4(1):29–38.
2. Anil S, Anand PS. Early childhood caries: prevalence, risk factors, and prevention. Front Pediatr. 2017;5:157.
3. Pitts NB, Zero DT, Marsh PD, Elkstrand K, Weintraub JA, Ramos-Gomez F, et al. Dental caries. Nat Rev Dis Primers. 2017;3(1):17030.
4. Hoeft KS, Barker JC, Shiboski S, Pantoja-Guzman E, Hiatt RA. Effectiveness evaluation of contra caries oral health education program for improving Spanish-speaking parents' preventive oral health knowledge and behaviors for their young children. Community Dent Oral Epidemiol. 2016;44(6):564–76.
5. World Health Organization. Dental caries: Key facts. 2020. Available from: https://www.who.int/news-room/fact-sheets/detail/oral-health (accessed 13 Nov 2021).
6. Peres MA, Macpherson L, Weyant RJ, Daly B, Venturelli R, Mathur MR, et al. Oral diseases: a global public health challenge. Lancet. 2019;394(10194):249–60.
7. Hoeft K. Using community participation to assess acceptability of "Contra Caries", a theory-based, promotora-led oral health education program for rural Latino parents: a mixed methods study. BMC Oral Health. 2015;15(1):103–14.
8. Centers for Disease Control and Prevention. Disparities in Oral Health. 2021. Available from: https://www.cdc.gov/oralhealth/oral_health_disparities/index.htm (accessed 13 Nov 2021).
9. Fleming E, Afful J. Prevalence of total and untreated dental caries among youth: United States, 2015–2016. 2017. Available from: https://www.cdc.gov/nchs/data/databriefs/db307.pdf (accessed 13 Nov 2021).
10. Naavaal S, Kelekar U. School hours lost due to acute/unplanned dental care. Health Behav Policy Rev. 2018;5(1):66–73.
11. Barker JC, Horton SB. An ethnographic study of Latino preschool children's oral health in rural California: intersections among family, community, provider and regulatory sectors. BMC Oral Health. 2008;8(1):31.
12. Quock RL. The evidence supporting fluoride varnish. 2017. Available from: https://decisionsindentistry.com/article/evidence-supporting-fluoride-varnish/ (accessed 13 Nov 2021).
13. Bhanushali P, Katge F, Deshpande S, Chimata VK, Shetty S, Pradhan D. COVID-19: changing

trends and its impact on future of dentistry. Int J Dent. 2020;2020(1):1–6.
14. Marcus M, Wang Y, Xiong D. Child and parent demographic characteristics and oral health perceptions associated with clinically measured oral health. JDR Clin Transl Res. 2018;3(3):302–13.
15. Liu H, Hays RD, Marcus M, Coulter I, Maida C, Ramos-Gomez F, et al. Patient-reported oral health outcome measurement for children and adolescents. BMC Oral Health. 2016;16:1–9.
16. Wang Y, Hays RD, Marcus M, Maida CA, Shen J, Xiong D, et al. Developing children's oral health assessment tool kits using machine learning algorithm. JDR Clin Transl Res. 2019;4(3):233–43.
17. Maida CA, Marcus M, Hays RD, Coulter ID, Ramos-Gomez F, Lee SY, et al. Child and adolescent perceptions of oral health over the life course. Qual Life Res. 2015;24(11):2739–51.
18. Maida CA, Marcus M, Hays RD, Coulter ID, Ramos-Gomez F, Lee SY, et al. Qualitative methods in the development of a parent survey of children's oral health status. J Patient Rep Outcomes. 2018;2(1):7.
19. Walsh TR, Irwin DE, Meier A, Varni JW, DeWalt DA. The use of focus groups in the development of the PROMIS pediatrics item bank. Qual Life Res. 2008;17(5):725–35.
20. Varni JW, Seid M, Rode CA. The PedsQL™: Measurement model for the pediatric quality of life inventory. Med Care. 1999;37(2):126–39.
21. Matza LS, Patrick DL, Riley AW, Alexander JJ, Rajmil L, Pleil AM, et al. Pediatric patient-reported outcome instruments for research to support medical product labeling: report of the ISPOR PRO good research practices for the assessment of children and adolescents task force. Value Health. 2013;16(3):461–79.
22. Irwin DE, Varni JW, Yeatts K, DeWalt DA. Cognitive interviewing methodology in the development of a pediatric item bank: a patient-reported outcomes measurement information system (PROMIS) study. Health Qual Life Outcomes. 2009;7(1):3.
23. Koch AL, Gershen JA, Marcus M. Children's oral health status index based on dentists' judgment. JADA. 1985;110(1):36–42.
24. Von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP; STROBE Initiative. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61(4):344–9.
25. STROBE: Strengthening the reporting of observational studies in epidemiology. Available from: https://www.strobe-statement.org/ (accessed 14 Nov 2021).
26. Marcus M, Xiong D, Wang Y, Maida CA, Hays RD, Coulter ID, et al. Development of toolkits for detecting dental caries and caries experience among children using self-report and parent report. Community Dent Oral Epidemiol. 2019;47(6):520–7.
27. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol. 1990;43(6):551–85.
28. Calle ML, Urrea V. Letter to the editor: Stability of random forest importance measures. Brief Bioinform. 2011;12(1):86–9.
29. Kelleher JD, Mac Namee B, D'arcy A. Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and CAE studies. Cambridge, MA: MIT Press; 2015.
30. Yap BW, Rani KA, Rahman HAA, Fong S, Khairudin Z, Abdullah N. An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets. In: Proceedings of the First International Conference on Advanced Data and Information Engineering. Singapore: Springer; 2014. p. 13–22.
31. Pitts NB, Baez RJ, Diaz-Guillory C, Donly KJ, Feldens CA, McGrath C, et al. Early childhood caries: IAPD bangkok declaration. J Dent Child. 2019;86(2):72.
32. Hurlbutt M. CAMBRA: best practices in dental caries management. 2011. Available from: https://pdfs.semanticscholar.org/f6ac/6833549fe10821f8baa6bb927e19bcfa8591.pdf (accessed 21 Oct 2021).
33. Wang Y, Hays R, Marcus M, Maida C, Shen J, Xiong D, et al. Development of a parent's short form survey of their children's oral health. Int J Paediatr Dent. 2019;29(3):332–44.