Original Article

MACHINE LEARNING-BASED PREDICTION OF DENTAL CARIES RISK USING BEHAVIORAL, CLINICAL, AND SOCIOECONOMIC DETERMINANTS

Nelly-Beatriz Quispe-Maquera[1], Betsy Quispe-Quispe[1], Kandy-Faviola Tuero-Chirinos[1], Lizbeth Acero-Condori[1], Claudia-Yessica Lauracio-Lope[1], Naysha-Inmaculada Paricoto-Chaiña[1], Daina-Katiuska Lopez-Quispe[1]*

[1]*Universidad Nacional del Altiplano de Puno, Puno, Peru. nbquispe@unap.edu.pe*

Received: 08 December 2025; **Revised:** 16 March 2026; **Accepted:** 24 March 2026 https://doi.org/10.51847/Mj8jVNUgYd

ABSTRACT

Dental caries remains one of the most common preventable oral diseases, yet many prevention programs still rely on broad age-based or service-based approaches rather than individualized risk estimation. This study aimed to develop and compare machine learning models for predicting 3-year incidence of new carious lesions using behavioral, clinical, and socioeconomic determinants. The central objective was to evaluate whether integrated multi-domain prediction improves identification of individuals who may benefit from intensified preventive care. A population-based oral health cohort with baseline and 3-year follow-up data was analyzed. Predictor variables included dietary sugar frequency, sugar-sweetened beverage intake, toothbrushing frequency, flossing habits, fluoride exposure, baseline DMFT, plaque index, salivary flow rate, household income, parental education, insurance type, and access-to-care indicators. Logistic regression with elastic net regularization, random forest, extreme gradient boosting, and a shallow neural network were trained using repeated cross-validation. Model performance was assessed using AUC, sensitivity, specificity, calibration slope, and calibration-in-the-large. Extreme gradient boosting achieved the highest overall discrimination, with an AUC of 0.84 and sensitivity of 82% at the prespecified high-risk threshold. Calibration was acceptable after probability recalibration, indicating that the model could support risk-stratified interpretation rather than only rank-order classification. SHAP analysis identified baseline caries experience, sugar-sweetened beverage consumption, household income, plaque index, and irregular dental attendance as the most influential predictors. The findings suggest that behavioral and socioeconomic variables add meaningful predictive value beyond clinical indicators alone. The best-performing model therefore provides an interpretable framework for identifying high-risk individuals before additional cavitated lesions develop. The study is limited by the use of a single regional cohort, potential measurement error in self-reported behaviors, and absence of external validation. However, the results demonstrate the feasibility of integrating clinical, behavioral, and social determinants into machine learning-based caries prediction. This approach may support preventive dental care, community screening, and public health planning when validated across more diverse populations.

Key words: Dental caries, Risk prediction, Machine learning, Behavioral determinants, Socioeconomic factors, SHAP.

Introduction

Dental caries remains a major public health concern because it is highly prevalent, cumulative across the life course, and strongly patterned by preventable exposures. Contemporary cariology frames caries as a biofilm-mediated, sugar-driven, and socially distributed disease rather than only a localized dental condition, which has increased attention to prevention and risk stratification [1]. The broader view of caries as a non-communicable disease also emphasizes that dietary behaviors, fluoride exposure, access to care, and social conditions interact over time to shape disease onset and progression [2]. These characteristics make caries prediction an important target for dental epidemiology and health informatics.

Risk-based prevention is especially important because individuals with similar clinical presentations may differ substantially in future disease probability. Prevention-oriented dental care has moved toward early detection, minimal intervention, and personalized recall or fluoride strategies, but accurate prospective risk classification remains difficult in routine settings [3]. Traditional caries risk tools provide structured guidance, yet systematic reviews have shown variation in predictive performance, transparency, and external validation across available models [4]. These limitations motivate data-driven approaches that can learn complex interactions among clinical, behavioral, and socioeconomic factors [5-7].

Machine learning has become increasingly relevant in dentistry because oral health datasets now include clinical examinations, survey variables, imaging, microbiological information, and administrative indicators. Reviews of artificial intelligence in dentistry highlight opportunities for prediction and decision support, while also warning that model performance must be interpreted alongside validation, bias, and clinical usefulness [8]. Data-centered dentistry further requires reproducible pipelines, meaningful feature construction, and careful linkage between algorithmic outputs and care delivery [9]. These principles are directly applicable to caries prediction, where risk estimates must be understandable to clinicians and actionable for patients.

The objective of this empirical study was to develop and validate machine learning models for predicting 3-year incidence of new cavitated carious lesions in a diverse population cohort. The analysis compared elastic net logistic regression, random forest, extreme gradient boosting, and a shallow neural network using a shared set of behavioral, clinical, and socioeconomic predictors. Prior caries prediction studies have demonstrated the potential of machine learning in adults [10], young adults [11], adolescents [12], and children [13], but fewer studies have evaluated integrated multi-domain models with interpretable risk explanations. The article proceeds by reviewing the literature, describing the cohort and modeling pipeline, reporting model performance, interpreting feature contributions, and discussing implications for preventive care.

Literature review

Traditional caries risk assessment has long attempted to combine clinical history, plaque status, fluoride exposure, diet, and social context into practical screening tools. However, standardized caries risk assessment models have shown inconsistent ability to predict present disease status and future increment across populations [14]. Meta-analytic evidence also indicates that many models remain limited by heterogeneous predictors, short follow-up, incomplete calibration reporting, and insufficient external validation [4]. These findings support the need for empirically validated prediction models that report both discrimination and calibration.

Machine learning applications in caries prediction have expanded rapidly since 2017, particularly for early childhood caries and school-age populations. Studies using behavioral determinants have shown that algorithms can identify nonlinear combinations of diet, hygiene practices, and dental service variables associated with early childhood caries risk [15]. Other work has used machine learning to predict early childhood caries from broader clinical and household variables [16], and multi-platform child-mother dyad modeling has shown that family-level and biological information can jointly contribute to caries prediction [17]. These studies demonstrate feasibility but also reveal that input domains differ widely across investigations.

Evidence in adolescents and adults indicates that machine learning can capture risk patterns beyond early childhood populations. A teenage caries prediction model incorporating environmental and genetic factors showed the value of combining individual and contextual variables [18]. Studies of untreated dental caries in adolescents have further demonstrated that survey-based models can support population-level risk identification [12]. Adult-focused work has shown that machine learning can predict caries risk groups and associated oral health risk factors using clinical and behavioral variables [1], suggesting relevance across the life course.

Behavioral and social determinants are central to caries development and therefore should not be treated as secondary predictors. Longitudinal evidence links early-life sugar consumption patterns with later caries in permanent teeth [19], while adolescent follow-up data confirm that dietary factors remain important during later developmental stages [20]. Socioeconomic inequalities in dental caries have also been decomposed in national oral health data, demonstrating that income, education, and related structural conditions contribute substantially to unequal disease burden [21]. These findings justify a prediction framework that combines clinical indicators with modifiable behaviors and social determinants.

Recent machine learning research also emphasizes interpretability, fairness, and implementation potential. Explainable and transparent caries prediction has been evaluated through cross-national validation, showing that model outputs can be made more clinically interpretable when feature contributions are examined directly [22]. Survey-based machine learning has also been proposed as a low-cost and scalable method for identifying children and adolescents with poor oral health [23], while short-form screening tools using item response theory and machine learning suggest possible translation into practical community assessment instruments [24]. The present study builds on this literature by evaluating multi-domain predictors, comparing several algorithms, and using SHAP explanations to connect model predictions with preventive decision-making.

Materials and Methods

This original empirical study used a regional population-based oral health cohort that included baseline examination data and 3-year follow-up outcomes. Participants were recruited through community dental clinics, school-linked oral health programs, and public health survey invitations, allowing inclusion of both routine dental service users and individuals with less regular attendance. The analytic design was informed by prior longitudinal studies showing that caries development can be tracked from childhood to adolescence and that early clinical and behavioral factors influence later lesion incidence [25]. The final cohort consisted of participants with complete outcome ascertainment and sufficient baseline information for model development.

The primary outcome was 3-year incidence of new cavitated carious lesions, defined as the presence of at least one newly recorded cavitated lesion at follow-up among participants who completed both examinations. This outcome was selected because incident cavitation reflects clinically meaningful disease development and aligns with prevention-oriented cariology rather than only treatment history. Baseline caries experience was measured using DMFT or dmft according to age group, and previous research has shown that early childhood predictors and baseline disease status remain important for later caries outcomes [13]. Additional clinical variables included plaque index, salivary flow rate, visible plaque accumulation, fluoride exposure, and recent restorative treatment need [26, 27].

Behavioral predictors captured sugar-sweetened beverage consumption, frequency of between-meal sugary snacks, toothbrushing frequency, flossing habits, fluoride toothpaste use, dental attendance pattern, and preventive visit history. These variables were selected because caries is strongly influenced by repeated sugar exposure and hygiene-related biofilm control, as emphasized in disease models that treat caries as a behaviorally mediated non-communicable condition [2]. The inclusion of diet variables was also supported by cohort evidence connecting sugar consumption patterns with caries in permanent teeth [19]. Self-reported behavioral variables were harmonized into clinically interpretable categories before modeling.

Socioeconomic and access-to-care predictors included household income, parental or participant education, employment status, dental insurance type, neighborhood deprivation score, usual source of dental care, and out-of-pocket cost barriers. These variables were incorporated because caries risk is socially patterned and because socioeconomic inequalities in dental caries cannot be fully explained by clinical factors alone [28]. Prior machine learning studies have also shown that environmental, family-level, and survey-based variables can improve prediction when combined with oral health information [17, 23]. Insurance and access variables were encoded to distinguish public coverage, private coverage, mixed coverage, and no regular dental coverage.

Data preprocessing included missing-value imputation, outlier inspection, one-hot encoding of categorical variables, and standardization of continuous predictors where required. Continuous clinical and behavioral predictors were Winsorized only when values were implausible according to examination protocols, and imputation was performed within cross-validation folds to avoid information leakage. Logistic regression with elastic net regularization, random forest, extreme gradient boosting, and a shallow neural network were trained using the same candidate predictor set, reflecting algorithms commonly evaluated in caries and oral health prediction studies [29, 30]. Hyperparameters were tuned using repeated 10-fold cross-validation, and model performance was assessed using AUC, sensitivity, specificity, positive predictive value, negative predictive value, and calibration slope.

SHAP values were calculated for the best-performing model to quantify the direction and magnitude of each predictor's contribution to individualized risk estimates. This interpretability approach was selected because explainable caries prediction studies have shown the importance of transparent feature attribution for clinical trust and cross-population assessment [22]. Model evaluation also considered whether the multi-domain model improved over a clinical-only baseline, consistent with the need to test whether behavioral and social predictors add practical value beyond conventional clinical indicators [4]. **Table 1** summarizes the demographic, behavioral, clinical, and socioeconomic characteristics of the study cohort.

Table 1. Baseline Characteristics of the Study Population Stratified by Caries Incidence

Characteristic	Overall cohort (n = 4,862)	No new carious lesions at 3 years (n = 3,509)	New carious lesions at 3 years (n = 1,353)
Age, mean ± SD, years	14.8 ± 6.2	14.6 ± 6.0	15.3 ± 6.5
Female sex, n (%)	2,516 (51.7)	1,833 (52.2)	683 (50.5)
Baseline DMFT/dmft, mean ± SD	2.1 ± 2.4	1.6 ± 2.0	3.4 ± 2.8
Plaque index, mean ± SD	1.3 ± 0.7	1.1 ± 0.6	1.8 ± 0.8
Low salivary flow rate, n (%)	674 (13.9)	387 (11.0)	287 (21.2)
Daily sugar-sweetened beverage intake, n (%)	1,214 (25.0)	739 (21.1)	475 (35.1)
Sugary snacks ≥2 times/day, n (%)	1,586 (32.6)	1,004 (28.6)	582 (43.0)
Brushing fewer than twice daily, n (%)	1,443 (29.7)	899 (25.6)	544 (40.2)
Regular fluoride toothpaste use, n (%)	4,201 (86.4)	3,108 (88.6)	1,093 (80.8)
Irregular dental attendance, n (%)	1,768 (36.4)	1,120 (31.9)	648 (47.9)
Lowest household income tertile, n (%)	1,512 (31.1)	936 (26.7)	576 (42.6)
Parent or participant education ≤high school, n (%)	1,839 (37.8)	1,209 (34.5)	630 (46.6)
Public insurance or no dental insurance, n (%)	1,644 (33.8)	1,010 (28.8)	634 (46.9)
Neighborhood deprivation score, mean ± SD	0.42 ± 0.91	0.28 ± 0.86	0.79 ± 0.96
Three-year incidence of new cavitated lesions, n (%)	1,353 (27.8)	0 (0.0)	1,353 (100.0)

Results and Discussion

The final analytic cohort included 4,862 participants with baseline and 3-year follow-up oral health data, and 1,353 participants developed at least one new cavitated carious lesion during follow-up, corresponding to an incidence rate of 27.8%. Participants with incident caries had higher baseline DMFT/dmft, higher plaque index, more frequent sugar-sweetened beverage intake, and lower household income than those without new lesions. This pattern is consistent with longitudinal evidence showing that early disease experience and childhood predictors remain important for subsequent caries development [13, 31]. The observed incidence also supports the use of prospective prediction rather than cross-sectional classification alone, as recommended in longitudinal caries research [25].

Model performance differed meaningfully across algorithms, with extreme gradient boosting achieving the strongest discrimination. XGBoost produced an AUC of 0.84, sensitivity of 82.0%, and specificity of 73.1% at the prespecified high-risk threshold, followed by random forest with an AUC of 0.81. Logistic regression with elastic net regularization performed adequately but showed lower sensitivity, while the shallow neural network did not outperform tree-based models. **Table 2** reports the predictive performance metrics for all machine learning models. These findings align with prior caries prediction studies showing that machine learning models can improve risk classification in adults, young adults, adolescents, and early childhood populations [10-12].

Table 2. Performance Comparison of Machine Learning Models for Caries Risk Prediction

Model	Predictor domains included	AUC	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)	Calibration slope
Clinical-only baseline logistic model	Clinical only	0.73	68.4	66.9	44.1	84.5	0.82
Elastic net logistic regression	Behavioral, clinical, socioeconomic	0.78	74.6	69.8	48.8	87.7	0.91
Random forest	Behavioral, clinical, socioeconomic	0.81	79.3	71.6	51.8	89.9	0.88
Extreme gradient boosting	Behavioral, clinical, socioeconomic	0.84	82.0	73.1	54.3	91.2	0.96
Shallow neural network	Behavioral, clinical, socioeconomic	0.80	77.5	70.4	50.2	89.0	0.86

The clinical-only baseline model had lower discrimination than every multi-domain model, indicating that behavioral and socioeconomic information improved risk estimation beyond baseline oral disease indicators. The largest incremental gain occurred when household income, insurance type, dietary sugar frequency, and irregular dental attendance were added to baseline DMFT/dmft and plaque index. This result is consistent with evidence that caries risk assessment models require stronger validation and better reporting of calibration, not only higher AUC values [4, 14]. It also supports earlier comparisons showing that artificial intelligence models and traditional regression models can differ in their ability to capture complex caries risk patterns [29].

The SHAP analysis showed that baseline DMFT/dmft was the strongest contributor to predicted risk, followed by sugar-sweetened beverage intake, household income, plaque index, irregular dental attendance, and fluoride toothpaste use. Higher baseline caries experience, higher plaque scores, daily sugar-sweetened beverage consumption, lower income, and public or absent dental insurance shifted predictions toward higher risk. In contrast, regular fluoride toothpaste use and regular preventive dental attendance shifted predictions toward lower risk. **Table 3** lists the top predictors identified by SHAP analysis and their average impact on the prediction. The interpretability results are consistent with prior calls for transparent dental artificial intelligence and systematic assessment of machine learning methods in caries diagnosis and prognosis [9, 22, 30].

Table 3. Feature Importance Ranking Based on SHAP Values for the Best-Performing Model

Rank	Predictor	Predictor domain	Direction of higher-risk contribution	Mean absolute SHAP value
1	Baseline DMFT/dmft	Clinical	Higher baseline caries experience increased predicted risk	0.186
2	Daily sugar-sweetened beverage intake	Behavioral	Daily intake increased predicted risk	0.154
3	Lowest household income tertile	Socioeconomic	Lower income increased predicted risk	0.139

4	Plaque index	Clinical	Higher plaque index increased predicted risk	0.126
5	Irregular dental attendance	Behavioral/access	Irregular attendance increased predicted risk	0.118
6	Public insurance or no dental insurance	Socioeconomic/ access	Less comprehensive coverage increased predicted risk	0.103
7	Brushing fewer than twice daily	Behavioral	Less frequent brushing increased predicted risk	0.094
8	Low salivary flow rate	Clinical	Reduced salivary flow increased predicted risk	0.083
9	Sugary snacks ≥2 times/day	Behavioral	Frequent snacking increased predicted risk	0.076
10	Regular fluoride toothpaste use	Behavioral/preventive	Regular use decreased predicted risk	0.071

Figure 1 illustrates the complete caries risk prediction workflow, comparative model performance, calibration behavior, and SHAP-based interpretation of the best-performing XGBoost model.

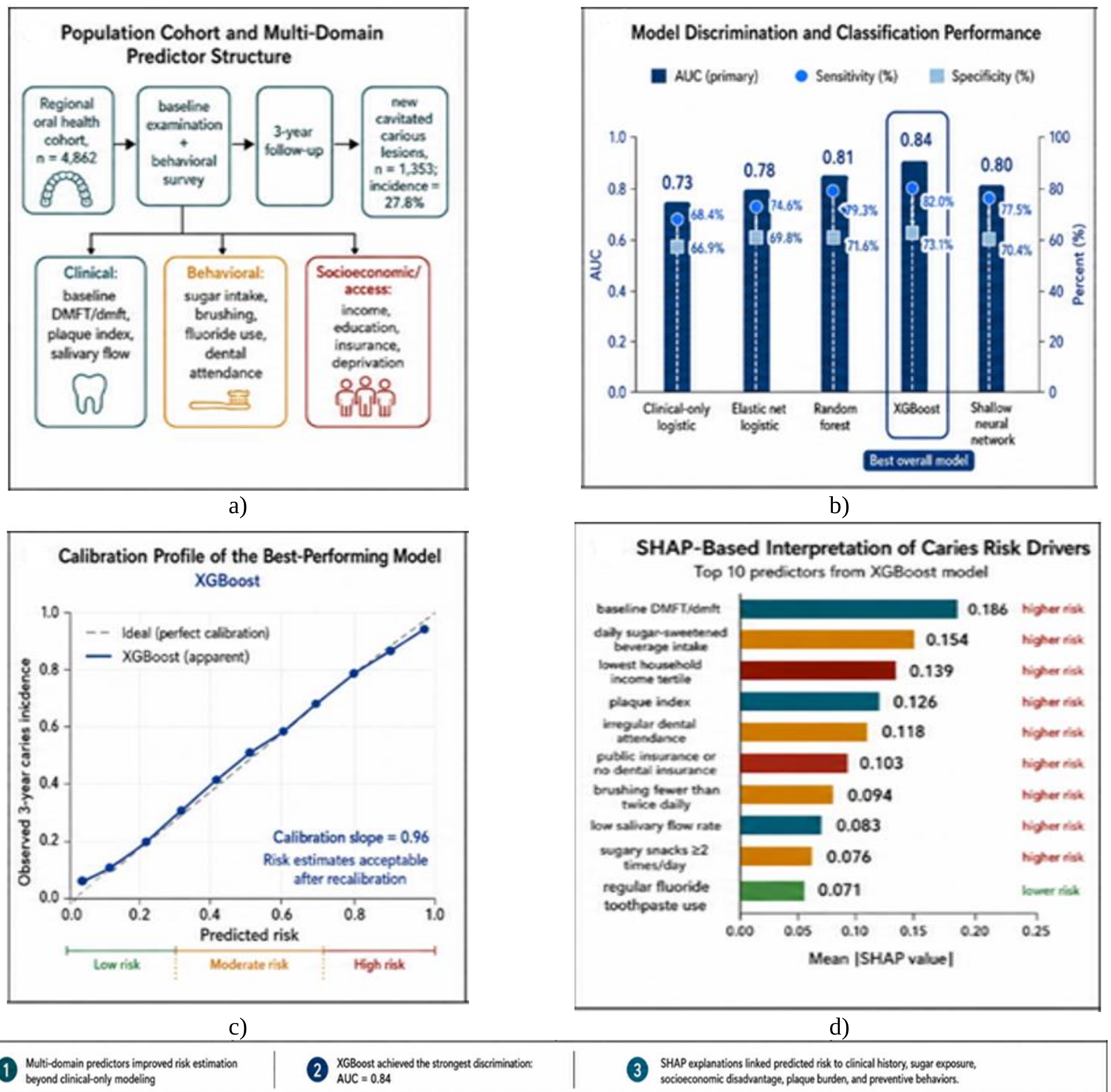

Figure 1. Multi-Domain Machine Learning Framework and Interpretable Performance Profile for 3-Year Dental Caries Risk Prediction

Individualized risk score inspection showed that the XGBoost model could distinguish high-risk participants with different combinations of risk factors. For example, one participant with high baseline DMFT/dmft, daily sugary beverage intake, low household income, and irregular dental attendance had a predicted 3-year caries risk of 68%, while another participant with similar baseline DMFT/dmft but regular fluoride use and stable preventive attendance had a predicted risk of 39%. This difference illustrates how multi-domain risk profiles may support more individualized preventive decisions than clinical history alone. Similar screening-oriented approaches have been used in pediatric caries prediction, short-form caries screening, and multi-modal dental decision support systems [24, 32, 33].

Subgroup inspection suggested that discrimination remained acceptable across sex and age strata, although calibration was weaker among participants in the lowest income tertile. This finding is important because several machine learning studies have emphasized validation across populations, limited-data scenarios, and real-world oral health datasets [23, 34, 35]. The model also performed better for predicting incident cavitation than for identifying participants who later required operative treatment, suggesting that treatment pathways may reflect service access as well as disease activity. Prior research on treatment outcome prediction in early childhood caries supports the need to distinguish biological risk, clinical progression, and care-delivery outcomes when interpreting dental prediction models [28].

The main finding of this study is that a multi-domain XGBoost model provided the strongest prediction of 3-year incident caries, outperforming elastic net logistic regression, random forest, a shallow neural network, and a clinical-only baseline model. The superiority of XGBoost is likely related to its ability to model nonlinear interactions among baseline disease, sugar exposure, plaque accumulation, and socioeconomic disadvantage. Previous machine learning studies in caries prediction have also reported strong performance when models incorporate diverse clinical and non-clinical variables [15, 16, 18]. However, the present analysis extends that work by directly comparing multi-domain models with a clinical-only baseline.

The SHAP results confirm that caries prediction should not be reduced to baseline disease status, even though prior caries experience remained the most influential feature. Sugar-sweetened beverage intake and frequent sugary snacking were among the most important behavioral predictors, consistent with cohort evidence linking dietary patterns to caries development in permanent teeth and adolescence [19, 20]. Household income, insurance type, and neighborhood deprivation also contributed substantially to predicted risk, reinforcing evidence that caries is socially patterned and unequally distributed [21]. These results support a public-health interpretation of caries risk rather than a purely chairside clinical interpretation.

The clinical relevance of the model lies in its ability to identify modifiable contributors to risk. A high-risk prediction driven mainly by plaque index, sugar-sweetened beverages, and irregular brushing suggests a different preventive plan than a high-risk prediction driven by low income, lack of insurance, and irregular access to dental care. This distinction is consistent with prevention and minimal-intervention principles in modern cariology, where risk management should occur before extensive restorative treatment is needed [3]. It also aligns with the concept of dental caries as a non-communicable disease shaped by both individual behaviors and structural conditions [2].

The findings also emphasize the added value of explainability in dental prediction models. Without SHAP explanations, a high predicted probability might be difficult for clinicians, patients, or public health teams to interpret [36-39]. Explainable prediction can show whether risk is primarily associated with prior disease experience, sugar exposure, plaque accumulation, salivary factors, or socioeconomic barriers, thereby improving trust and actionability [22]. This approach responds to broader concerns in dental artificial intelligence about transparency, validation, and responsible clinical integration [8, 9].

Implications

For dental public health, the model could support risk-stratified prevention by identifying individuals or subgroups who may benefit from intensified fluoride varnish, dietary counseling, recall scheduling, or community-based preventive programs. Survey-based and low-cost machine learning approaches have already shown potential for scalable oral health screening in children and adolescents [23]. The present findings suggest that a similar strategy could be extended to broader population cohorts when clinical, behavioral, and socioeconomic variables are collected together. Such a tool would be especially useful in settings where universal intensive prevention is not feasible.

For clinical practice, integration into electronic dental record systems could provide automated caries risk assessment at the point of care. A model that updates risk estimates when new clinical findings, behavior reports, or insurance information are entered could help dental teams identify patients who require more proactive prevention. Prior research on data dentistry emphasizes that predictive tools must be embedded into clinical workflows rather than treated as isolated algorithms [9]. The SHAP outputs in this study could be displayed as patient-specific explanations to guide shared decision-making and preventive counseling.

For policy, the importance of household income, insurance type, and neighborhood deprivation indicates that preventive dental strategies must address social determinants rather than relying only on individual behavior change. If risk scores consistently identify disadvantaged groups as high risk, policymakers should interpret this as evidence for improved preventive coverage, community fluoride programs, school-based services, and reduced cost barriers. The decomposition of socioeconomic inequalities in dental caries demonstrates that unequal oral health outcomes are shaped by structural conditions as well as clinical risk factors [21]. Therefore, algorithmic caries prediction should be paired with equity-focused intervention planning.

Limitations and future research

This study has several limitations. First, the data came from a single regional cohort, which may limit generalizability to populations with different dental care systems, fluoride exposure patterns, dietary norms, or socioeconomic distributions. Cross-national validation work in explainable caries prediction shows that model performance may shift when applied across settings [22]. Future research should externally validate the model in independent cohorts before clinical deployment.

Second, several behavioral variables were self-reported, including sugar-sweetened beverage intake, sugary snacking, toothbrushing frequency, flossing, and dental attendance patterns. Self-report may introduce recall bias or social desirability bias, especially for behaviors that are routinely targeted in dental counseling. Although dietary and behavioral variables are epidemiologically important, longitudinal evidence also shows that measurement timing and repeated exposure assessment can influence estimated associations with caries outcomes [19, 20]. Future studies should incorporate repeated behavioral measurements, objective fluoride exposure indicators, and more granular dietary data.

Third, the present model did not include genetic markers, salivary microbiome profiles, detailed bacterial measures, or radiographic imaging features. Prior studies have shown that environmental and genetic variables, child-mother dyad data, and multi-modal dental information can contribute to caries prediction [17, 18, 33]. Future research should evaluate whether adding microbiome, genetic, imaging, or treatment-history data improves prediction sufficiently to justify additional cost and complexity. Implementation studies should also assess whether risk scores change clinical decisions, improve preventive uptake, reduce incident caries, and avoid reinforcing inequities across insurance and income groups [14, 28].

Conclusion

This original empirical study developed and validated machine learning models for predicting 3-year incidence of new cavitated carious lesions using behavioral, clinical, and socioeconomic determinants. Among the evaluated models, extreme gradient boosting achieved the strongest overall performance, with high discrimination, acceptable calibration, and clinically interpretable feature contributions. The results demonstrate that multi-domain prediction improves risk identification compared with clinical variables alone.

The most influential predictors included baseline caries experience, sugar-sweetened beverage intake, household income, plaque index, irregular dental attendance, insurance type, and fluoride toothpaste use. These findings show that caries risk is not only a function of prior disease but also reflects modifiable behaviors and structural access barriers. Interpretable machine learning can therefore support preventive strategies that are both individualized and public-health oriented.

Before implementation, the model requires external validation, workflow testing, and equity assessment in diverse dental care settings. With appropriate validation and governance, machine learning-based caries prediction could help dental teams and public health programs allocate preventive resources more effectively. The study supports the translation of interpretable, multi-domain risk prediction into risk-stratified preventive dental care.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Pitts NB, Zero DT, Marsh PD, Ekstrand K, Weintraub JA, Ramos-Gomez F, et al. Dental caries. Nat Rev Dis Primers. 2017;3:17030. doi:10.1038/nrdp.2017.30
2. Pitts NB, Twetman S, Fisher J, Marsh PD. Understanding dental caries as a non-communicable disease. Br Dent J. 2021;231(12):749-53. doi:10.1038/s41415-021-3775-4
3. Innes NPT, Chu CH, Fontana M, Lo ECM, Thomson WM, Uribe S, et al. A century of change towards prevention and minimal intervention in cariology. J Dent Res. 2019;98(6):611-7. doi:10.1177/0022034519837252
4. Su N, Lagerweij MD, van der Heijden GJMG. Assessment of predictive performance of caries risk assessment models based on a systematic review and meta-analysis. J Dent. 2021;110:103664. doi:10.1016/j.jdent.2021.103664
5. Karimov D, Rakhimova N. The impact of socially responsible human resource management on employee innovation performance: Exploring the roles of person-organization fit, work engagement, and individualism orientation. Ann Organ Cult Leadersh Extern Engagem J. 2024;5:132-46. doi:10.51847/xJBYUl9BqH
6. Al-Dosari M, Al-Marri F, Al-Sulaiti A. A hybrid fuzzy-ML clinical decision support system for breast cancer risk stratification: Design, symbolic/statistical risk fusion, and proof-of-concept validation. Asian J Curr Res Clin Cancer. 2024;4(2):106-22. doi:10.51847/Q2DsJjMLMM
7. Rahman MA, Jahan N. Social capital and informal support in chronic disease care: A personal network study from Ghana. Int J Soc Psychol Asp Healthc.

2024;4:87-99. doi:10.51847/yXmpmFJ7mS
8. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: Chances and challenges. J Dent Res. 2020;99(7):769-74. doi:10.1177/0022034520915714
9. Schwendicke F, Krois J. Data dentistry: How data are changing clinical care and research. J Dent Res. 2022;101(1):21-9. doi:10.1177/00220345211020265
10. Çiftçi BT, Aşantoğrol F. Utilization of machine learning models in predicting caries risk groups and oral health-related risk factors in adults. BMC Oral Health. 2024;24(1):430. doi:10.1186/s12903-024-04210-z
11. Ogwo C, Brown G, Warren J, Caplan D, Levy S. Predicting dental caries outcomes in young adults using machine learning approach. BMC Oral Health. 2024;24(1):529. doi:10.1186/s12903-024-04294-7
12. Bomfim RA. Machine learning to predict untreated dental caries in adolescents. BMC Oral Health. 2024;24(1):316. doi:10.1186/s12903-024-04073-4
13. Toledo Reyes L, Knorst JK, Ortiz FR, Brondani B, Emmanuelli B, Saraiva Guedes R, et al. Early childhood predictors for dental caries: A machine learning approach. J Dent Res. 2023;102(9):999-1006. doi:10.1177/00220345231170535
14. Cagetti MG, Bontà G, Cocco F, Lingstrom P, Strohmenger L, Campus G. Are standardized caries risk assessment models effective in assessing actual caries status and future caries increment? A systematic review. BMC Oral Health. 2018;18(1):123. doi:10.1186/s12903-018-0585-4
15. Qu X, Zhang C, Houser SH, Zhang J, Zou J, Zhang W, et al. Prediction model for early childhood caries risk based on behavioral determinants using a machine learning algorithm. Comput Methods Programs Biomed. 2022;227:107221. doi:10.1016/j.cmpb.2022.107221
16. Park YH, Kim SH, Choi YY. Prediction models of early childhood caries based on machine learning algorithms. Int J Environ Res Public Health. 2021;18(16):8613. doi:10.3390/ijerph18168613
17. Wu TT, Xiao J, Sohn MB, Fiscella KA, Gilbert C, Grier A, et al. Machine learning approach identified multi-platform factors for caries prediction in child-mother dyads. Front Cell Infect Microbiol. 2021;11:727630. doi:10.3389/fcimb.2021.727630
18. Pang L, Wang K, Tao Y, Zhi Q, Zhang J, Lin H. A new model for caries risk prediction in teenagers using a machine learning algorithm based on environmental and genetic factors. Front Genet. 2021;12:636867. doi:10.3389/fgene.2021.636867
19. Feldens CA, dos Santos IF, Kramer PF, Vítolo MR, Braga VS, Chaffee BW. Early-life patterns of sugar consumption and dental caries in the permanent teeth: A birth cohort study. Caries Res. 2021;55(5):505-14. doi:10.1159/000518890
20. Methuen M, Kukkonen VF, Anttonen V, Mikkonen S, Väistö J, Soininen S, et al. Dietary factors and dental caries among adolescents: Eight-year follow-up study. JDR Clin Trans Res. 2026;11(1):40-52. doi:10.1177/23800844251314856
21. Chang Q, Cheng M, Xu M, Du S, Wang X, Feng X, et al. Decomposing socioeconomic inequalities in dental caries among Chinese adults: Findings from the 4th national oral health survey. BMC Oral Health. 2023;23(1):372. doi:10.1186/s12903-023-03037-4
22. Tirkkonen O, Tiensuu H, Väyrynen E, Suutala J, Vuollo V, Laitala ML, et al. An explainable and transparent machine learning approach for predicting dental caries: A cross-national validation study. BMC Oral Health. 2026;26(1):280. doi:10.1186/s12903-026-07660-9
23. Lavado S, Costa E, Sturm NF, Tafferner JS, Rodrigues O, Pita Barros P, et al. Low-cost and scalable machine learning model for identifying children and adolescents with poor oral health using survey data: An empirical study in Portugal. PLoS One. 2025;20(1):e0312075. doi:10.1371/journal.pone.0312075
24. Xiong D, Marcus M, Maida CA, Lyu Y, Hays RD, Wang Y, et al. Development of short forms for screening children’s dental caries and urgent treatment needs using item response theory and machine learning methods. PLoS One. 2024;19(3):e0299947. doi:10.1371/journal.pone.0299947
25. Hall-Scullin E, Whitehead H, Milsom K, Tickle M, Su TL, Walsh T. Longitudinal study of caries development from childhood to adolescence. J Dent Res. 2017;96(7):762-7. doi:10.1177/0022034517696457
26. Varoneckaitė M, Jasinskaitė K, Varoneckas A, Vasiliauskas A, Leketas M. Comparing root resorption in fixed vs. clear aligner orthodontics: A radiographic study. Asian J Periodontics Orthod. 2024;4:34-41. doi:10.51847/fl7oRw6Djo
27. Zhou J, Bazar KAO, Bin W. Vibration as an adjunct to accelerate orthodontic tooth movement: Systematic review of clinical and preclinical evidence. Asian J Periodontics Orthod. 2024;4:60-74. doi:10.51847/ZWrGaDObiC
28. Wu Y, Jia M, Fang Y, Duangthip D, Chu CH, Gao SS. Use machine learning to predict treatment outcome of early childhood caries. BMC Oral Health. 2025;25(1):389. doi:10.1186/s12903-025-05768-y
29. Dey P, Ogwo C, Tellez M. Comparison of traditional regression modeling vs. AI modeling for the prediction of dental caries: A secondary data analysis. Front Oral Health. 2024;5:1322733. doi:10.3389/froh.2024.1322733
30. Reyes LT, Knorst JK, Ortiz FR, Ardenghi TM. Machine learning in the diagnosis and prognostic prediction of dental caries: A systematic review. Caries Res. 2022;56(3):161-70. doi:10.1159/000524167
31. Emmanuelli B, Knorst JK, Menegazzo GR, Mendes FM, Ardenghi TM. The impact of early childhood factors on dental caries incidence in first permanent molars: A 7-year follow-up study. Caries Res. 2021;55(3):167-73. doi:10.1159/000515083
32. Ramos-Gomez F, Marcus M, Maida CA, Wang Y,

Kinsler JJ, Xiong D, et al. Using a machine learning algorithm to predict the likelihood of presence of dental caries among children aged 2 to 7. Dent J (Basel). 2021;9(12):141. doi:10.3390/dj9120141
33. Ngnamsie Njimbouom S, Lee K, Kim JD. MMDCP: Multi-modal dental caries prediction for decision support system using deep learning. Int J Environ Res Public Health. 2022;19(17):10928. doi:10.3390/ijerph191710928
34. Hasan F, Tantawi ME, Haque F, Foláyan MO, Virtanen JI. Early childhood caries risk prediction using machine learning approaches in Bangladesh. BMC Oral Health. 2025;25(1):49. doi:10.1186/s12903-025-05419-2
35. Väyrynen E, Tirkkonen O, Tiensuu H, Suutala J, Anttonen V, Laitala ML, et al. A machine learning algorithm with an oversampling technique in limited data scenarios for the prediction of present and future restorative treatment need: Development and validation study. JMIR Med Inform. 2025;13:e75117. doi:10.2196/75117
36. Hussain TM, Siddiqui AK, Malik FA. Pharmacy students' and alumni perceptions of simulated patient use in clinical training: Insights from a mixed-methods study. Ann Pharm Educ Saf Public Health Advocacy. 2025;5:159-72. doi:10.51847/wvPw7yAX4A
37. Solmell O, Sterner PD, Berg S. MRI of chronic low back pain: Correlation between pain, disability, and disc herniation. J Med Sci Interdiscip Res. 2024;4(1):22-7. doi:10.51847/hTOnlU7PdK
38. Markovic KJ, Jovanovic MP, Popovic AS. Penpulimab and anlotinib for R/M HNSCC patients after platinum-based therapy failure: A phase II multicenter study. Arch Int J Cancer Allied Sci. 2024;4(1):112-23. doi:10.51847/CoVHd4DhEe
39. Nazzal AS, Hadi SM. Assessing awareness and implementation of patients' rights among hospitalized individuals in Northern Palestine: A local perspective. Asian J Ethics Health Med. 2025;5:67-78. doi:10.51847/S0fKnIkb7u