Wu et al. BMC Oral Health (2025) 25:389
https://doi.org/10.1186/s12903-025-05768-y

BMC Oral Health

RESEARCH

Open Access

Use machine learning to predict treatment outcome of early childhood caries

Yafei Wu[1,2], Maoni Jia[3], Ya Fang[1,4], Duangporn Duangthip[5], Chun Hung Chu[6] and Sherry Shiqian Gao[7*]

Abstract

Background Early childhood caries (ECC) is a major oral health problem among preschool children that can significantly influence children's quality of life. Machine learning can accurately predict the treatment outcome but its use in ECC management is limited. The aim of this study is to explore the application of machine learning in predicting the treatment outcome of ECC.

Methods This study was a secondary analysis of a recently published clinical trial that recruited 1,070 children aged 3- to 4-year-old with ECC. Machine learning algorithms including Naive Bayes, logistic regression, decision tree, random forest, support vector machine, and extreme gradient boosting were adopted to predict the caries-arresting outcome of ECC at 30-month follow-up after receiving fluoride and silver therapy. Candidate predictors included clinical parameters (caries experience and oral hygiene status), oral health-related behaviours (toothbrushing habits, feeding history and snacking preference) and socioeconomic backgrounds of the children. Model performance was evaluated using discrimination and calibration metrics including accuracy, recall, precision, F1 score, area under the receiver operating characteristic curve (AUROC) and Brier score. Shapley additive explanations were deployed to identify the important predictors.

Results All machine learning models showed good performance in predicting the treatment outcome of ECC. The accuracy, recall, precision, F1 score, AUROC, and Brier score of the six models ranged from 0.674 to 0.740, 0.731 to 0.809, 0.762 to 0.802, 0.741 to 0.804, 0.771 to 0.859, and 0.134 to 0.227, respectively. The important predictors of the caries-arresting outcome were the surface and tooth location of the carious lesions, newly developed caries during follow-ups, baseline caries experience, whether the children had assisted toothbrushing and oral hygiene status.

Conclusions Machine learning can provide promising predictions of the treatment outcome of ECC. The identified key predictors would be particularly informative for targeted management of ECC.

Keywords Machine learning, Early childhood caries, Predictor, Support vector machine, Extreme gradient boosting, SHAP

*Correspondence:
Sherry Shiqian Gao
sherrysgao@xmu.edu.cn
[1]School of Public Health, Xiamen University, Xiamen, China
[2]School of Nursing, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Hong Kong SAR, China
[3]Medical Department, School of Medicine, Xiang'an Hospital of Xiamen University, Xiamen University, Xiamen, China
[4]National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen, China
[5]College of Dentistry, The Ohio State University, Columbus, U.S.A.
[6]Faculty of Dentistry, The University of Hong Kong, Hong Kong SAR, China
[7]Department of Stomatology, School of Medicine, Xiamen University, Xiamen, China

Background

Precision medicine has recently gained a major consideration in health care. It is an emerging concept in clinical research and practice that integrates multi-modal or multi-dimensional data to personalise treatment and prevention plans for an individual [1]. The aim of precision medicine is to identify individual differences and utilise statistical strategies that can predict clinical outcomes [2]. Those clinical outcomes can include but are not limited to disease incidence after preventive strategies, effects after interventions, and consequences of diseases. It is noteworthy that many health issues are multi-factorial [3], and their clinical outcomes can be affected not only by the intervention itself but also by patients' sociodemographic backgrounds, health-related behaviours, general health conditions, etc [4]. These factors are considered confounders or effect modifiers in data analysis and interpretation. Confounders and effect modifiers are statistical concepts explaining how factors impact the relationship between variables [5]. Confounders are associated with both dependent and independent variables but may distort their correlation [5]. Effect modifiers are associated with the dependent variable but not the independent variable; they change the magnitude of the effect between variables [5]. Therefore, it is important to study confounders or effect modifiers to understand potential factors that may influence clinical outcomes. Although traditional statistical methods such as regression models were usually used for identifying the confounders or effect modifiers of treatment outcomes in clinical research [6], these methods often have difficulty in dealing with severe collinearity issues, require the specification of interaction terms in advance and usually have strict prior assumptions.

Prediction aims to estimate clinical outcomes using patients' data and identify the important confounders or effect modifiers (for prediction research, it refers to predictors). Prediction research is important to facilitate medical decision-making and achieve precise medicine to improve people's health outcomes [7]. Predictive models are different from traditional statistical methods and they utilise predictors that can forecast future development of health-related issues. It usually requires a large amount of data and an abundance of predictors to attain a higher accuracy of prediction. However, the predictive power of traditional statistical models was relatively weak due to the simple structure of models and limited predictors [8]. Therefore, the complex relationships between predictors and health outcomes for high-dimensional data are difficult to examine by traditional methods. Recently, new data management methods, such as machine learning, may allow for the development of algorithms that can accurately predict the clinical outcome. Machine learning can outperform traditional stepwise selection by capturing more relevant predictors simultaneously, with the great potential to efficiently address collinearity and automatically form flexible, empirically driven interactions without any specifications in advance, hence increasing the prediction accuracy and efficiency of prediction models [9]. Machine learning has been successfully used to predict the treatment responses of cancer, antidepressant medication and treatment outcomes in other medical issues [10, 11]. Meanwhile, with the unremitting efforts of computer and statistical scientists, lots of interpretable methods emerged as promising tools for uncovering the "black-box" of machine learning models, enabling the extraction of actionable insights for high-risk populations. Among them, the SHapley Additive exPlanations (SHAP) is the most popular one for assessing the contributions of each predictor by utilizing the unique Shapley values [12]. Overall, machine learning may provide more insightful views over traditional regression methods due to its complex structures and the emerging interpretable framework. However, no study has adopted machine learning in predicting the treatment outcome of early childhood caries (ECC) and examining the predictive values of diverse characteristics such as socioeconomic status, oral health-related behaviours, and clinical parameters.

In the current study, we collected available data from a recently published randomised controlled trial (RCT) that represents the ECC situation in preschool children in Hong Kong. Specifically, we compared the prediction power of six popular machine learning methods, aiming to (1) explore the application of machine learning in predicting the treatment outcome of ECC after receiving fluoride and silver therapies and (2) further identify key predictors of the treatment outcome by SHAP methods.

Methods

This study involved a secondary analysis of data from a recently published RCT that investigated the effectiveness of two fluoride and silver therapies, namely 25% $AgNO_3$ followed by 5% NaF and 38% SDF solution, for arresting ECC in preschool children in Hong Kong. The original RCT recruited 1,070 kindergarten children and followed for 30 months. Data including caries status, the participants' oral health-related behaviours and socioeconomic backgrounds were collected. This RCT was registered at ClinicalTrials.gov (No.: NCT02019160). The trial protocol was published in 2015 [13]. The main results were published in 2020 [14].

Treatment outcome

The treatment outcome included in the predictive model is the caries-arresting effect at the 30-month follow-up for each carious surface. The carious surface was examined by gentle probing using 0.5 mm ball-ended

Community Periodontal Index probes (Ash/Dentsply, Addlestone, UK). A binary classification (active vs. arrested) was used. If a carious lesion was hard upon probing, it was classified as arrested. A carious lesion was considered active if softness was detected upon probing [13, 15].

Candidate predictors

Based on the exploratory and data-driven approach to analysis, all available data were utilised in the predictive model to identify the important predictors. The candidate predictors include the following three domains. The detailed assignments of each predictor are shown in Appendix 1.

Domain 1 – Clinical parameters: tooth location and surface location of the carious lesions; decayed, missing (due to caries) and filled surfaces (dmfs) at baseline; newly developed dmfs during the study period (i.e. 30 months); oral hygiene status assessed by visible plaque index (VPI); fluoride and silver interventions.

Domain 2 – Oral health-related behaviours: age of stopping bottle-feeding, bottle-feeding behaviours before sleeping, age of starting toothbrushing, toothbrushing frequency, assisted toothbrushing, use of toothpaste, daily snacking habit, and history of regular and emergency dental visits.

Domain 3 – Socioeconomic status: birthplace, family status (both parents or single parent), monthly family income, father's education level, mother's education level, and main caretaker.

Machine learning algorithms and data preprocessing

Six machine learning algorithms were used to predict the caries-arresting outcome in this study, including four commonly used algorithms namely logistic regression (LR), Naive Bayes (NB), support vector machine (SVM) and decision tree (DT), and two representative ensemble learning methods namely random forest (RF) and extreme gradient boosting (XGBoost). LR is a kind of classical regression model that explores the relationship between a group of independent variables and dependent variables with a logit function, which has the advantage of intuitive explanation [16]. NB is a simple and powerful algorithm that can output the conditional probability on the basis of Bayes' theorem [17]. It is a probabilistic classifier that operates under the assumption that each feature in the model is independent of the existence of any other feature. In other words, each feature contributes to the predictions independently, with no interrelation between each other. The advantage of NB is its speed, making prediction for high-dimensional data simple. The SVM algorithm is widely used in machine learning because it can handle both linear and nonlinear classification tasks. When the data is not linearly separable, kernel functions are employed to transform the data into a higher-dimensional space, enabling linear separation. This technique is known as the 'kernel trick' [18]. DT is a flowchart-like structure where each internal node represents a 'test' on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label (the decision made after evaluating all attributes). The paths from the root to the leaves represent classification rules. DT performs better in decision-making because of its good model visualization. RF is an ensemble learning method, which consists of a multitude of decision trees and aggregates the final results in a parallel way based on the result of every single tree [19]. Based on its unique structure, RF is robust to noise and outliers, and can reduce overfitting issues compared to individual decision trees. XGBoost is also a decision-tree-based ensemble algorithm which can improve prediction performance by combining prediction results from weak prediction models [20]. Different from RF, XGBoost deploys the serial way to obtain the final predictions. Specifically, the prediction of the next decision tree in XGBoost depends on the output of the previous decision tree and certain weights were assigned to each tree according to their prediction accuracy.

Derivation and evaluation of the predictive models

In this study, the prediction unit was the treatment outcome in each carious surface. Specifically, data on 880 participants with 4157 carious surfaces were randomly assigned to a training set and a test set with a ratio of 7:3. For continuous variables, we used the normalization method to scale the value of each predictor to the same range (0–1), which would be helpful for model training and result interpretation. For categorical variables, one-hot encoding was applied to convert categorical data into a numerical format suitable for the training process of machine learning algorithms. The model was fitted on the training set and evaluated on the test set with a 1000-time bootstrap technique. Because the ratio of active and arrested surfaces was about 1:3, representing a class imbalance issue, the synthetic minority oversampling technique (SMOTE) was applied to generate a balanced outcome in the training data. The SMOTE is designed to address class imbalance in datasets, particularly in the context of machine learning predictions. It works by generating synthetic examples of the minority class (arrested class in our study) to increase its representation in the dataset. SMOTE creates these synthetic samples by selecting a data point from the minority class and interpolating between it and one of its nearest neighbours. This approach helps to provide a more balanced class distribution, which can improve the performance of classifiers by preventing them from being biased towards the majority class. Hyperparameters were tuned by traversing multiple combinations using a technique that combines cross-validation and grid search. Model performance was evaluated by assessing the discrimination and calibration of the predictive models on the test set. Specific metrics used in this study include: (1) balanced accuracy – balanced accuracy gives equal weight to each class, making it particularly useful when the classes are imbalanced, and it is calculated by the average of sensitivity and specificity; (2) weighted recall – weighted recall is the average recall across all classes, weighted by the number of true instances for each class (this ensures that the recall for each class contributes proportionally to its size in the dataset, providing a more balanced evaluation of the model's performance across all classes); (3) weighted precision – similarly, weighted precision is the average precision across all classes, weighted by the number of true instances for each class; (4) weighted F1 score – It combines both precision and recall into a single metric, providing a balanced measure of a model's accuracy across all classes while accounting for the size of each class; (5) area under the ROC curve (AUROC) – the area under the receiver operating characteristic curve (ROC), measuring the overall classification performance; (6) Brier score – it examines the differences between predicted probabilities and true labels. All the aforementioned metrics range from 0 to 1. Except for the Brier score, higher values for the other metrics indicate better performance. In this study, AUROC was used as the primary metric for selecting the optimal model. SHAP was used for interpreting the machine learning models. SHAP is a unified framework for interpreting predictions made by complicated machine learning models. It is based on the cooperative game theory, specifically the Shapley value, which fairly distributes the 'payout' among players based on their contributions to the total 'game'. In the context of machine learning, Shapley values quantify the contribution of each predictor to a particular prediction, offering insights into how different predictors impact the model's output. The whole process of machine learning analysis is listed in Fig. 1.

For statistical analysis, continuous variables were presented by means (standard deviation) for normal distribution or by median (interquartile) for skewed distribution. Categorical variables were presented by numbers (percentage). For comparisons of characteristics, a t-test or Wilcoxon test was conducted for continuous variables and a chi-square test or Fisher exact test was used for categorical variables. The two-tailed value of $p < 0.05$ was regarded as statistically significant. The statistical analysis was performed using SPSS 25.0. The machine learning procedures were performed with the scikit-learn toolkit in Python 3.7.6.

Results

Characteristics of the data set

The RCT recruited 1,070 kindergarten children at baseline with 535 children in each intervention group. At the 30-month follow-up, 880 children with 4,157 carious surfaces remained in this study and the data was used for

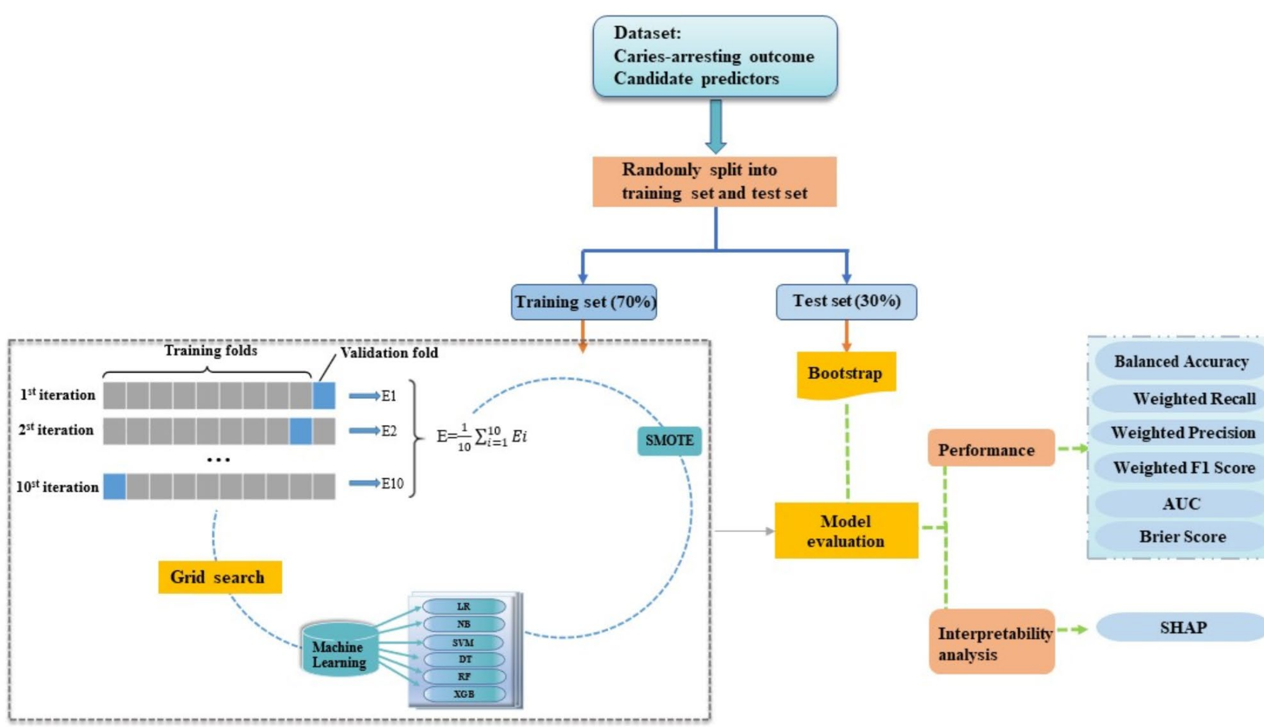

Fig. 1 Process of machine learning development

developing predictive models. The items, i.e., carious surfaces, were randomly allocated to a training set and test set with a ratio of 7:3 (2,909:1,248). The T-test and chi-square test showed that there was no difference between the training set and test set regarding the distributions of all candidate predictors (Appendix 1). The comparison between treatment outcome groups (active vs. arrested) was presented in Appendix 2.

Performance of the predictive models

Six machine learning algorithms were used for constructing the predictive models. The results of the model performance are shown in Table 1. The ROC of all models is presented in Fig. 2. All six models showed desirable performance in general. The accuracy values of all models were higher than 0.7 except SVM with the lowest value of 0.674. All models showed remarkable recall and precision values that were higher than 0.7, among which RF and XGBoost showed better recall and precision that were higher than 0.8. The AUROC of all models is around 0.8, among which RF and XGBoost showed relatively higher values (> 0.85). The results of the Brier score showed that all models have good calibration results.

Interpretability analysis of the predictive models

Because RF and XGBoost models showed the highest AUROC, and the other metrics were comparable among models, these two models were further utilized for interpretability analysis by SHAP. The results are presented in Figs. 3 and 4. The top five predictors with the greatest impact on the output of the RF model were the surface location and tooth location of the carious lesions, newly developed dmfs during the follow-up period, dmfs at baseline, and whether the children received assisted toothbrushing (Fig. 3A). Figure 3B shows the direction of how the predictors influenced the treatment outcome. Specifically, carious lesions on the buccal surface and anterior tooth were expected to have better treatment outcomes. Participants with more newly developed dmfs during the follow-up period and higher dmfs scores at baseline showed worse caries-arresting effects. In addition, assisted toothbrushing contributed to a better treatment outcome. According to the XGBoost model, the top five predictors with the greatest impact on the predictive model were surface and tooth location of the carious lesions, dmfs at baseline, newly developed dmfs during the follow-up period and oral hygiene status, which are quite similar to the results of the RF model (Fig. 4A and B). Beeswarm plot of the XGBoost model showed that a higher VPI sore would result in a worse caries arresting effect (Fig. 4B). A summary of the important predictors identified by machine learning models and the effect modifiers reported by the original trial is presented in Table 2.

Discussion

This study is a secondary analysis of a recently published RCT. It is the first study that uses machine learning algorithms to build predictive models to predict the treatment outcome of ECC. This study demonstrated that machine learning can be a promising strategy to predict caries-arresting outcomes after receiving fluoride and silver therapies in preschool children. In addition, based on the prediction results, we found that the important predictors were not the same as the effect modifiers reported by the original study. Dental researchers and clinicians can refer to the specific predictors that impact the caries-arresting effect and develop tailor-made caries management plans to achieve a more desirable treatment outcome.

Traditional regression models usually suppose the data fits a specific equation and calculate the equation parameters that represent the correlation between independent and dependent variables. Therefore, traditional

Table 1 Performance of machine learning predictive models (mean and 95% confidence interval)

	Accuracy	Recall	Precision
NB	0.713 (0.685, 0.741)	0.731 (0.706, 0.756)	0.762 (0.738, 0.785)
LR	0.735 (0.707, 0.764)	0.775 (0.751, 0.799)	0.783 (0.759, 0.807)
DT	0.719 (0.690, 0.747)	0.790 (0.767, 0.812)	0.783 (0.759, 0.807)
RF	0.737 (0.709, 0.764)	0.809 (0.787, 0.831)	0.802 (0.779, 0.825)
SVM	0.674 (0.646, 0.703)	0.790 (0.768, 0.813)	0.779 (0.754, 0.805)
XGBoost	0.740 (0.712, 0.767)	0.807 (0.786, 0.828)	0.801 (0.779, 0.823)
	F1 Score	**AUROC**	**Brier Score**
NB	0.741 (0.717, 0.764)	0.771 (0.741, 0.802)	0.227 (0.207, 0.248)
LR	0.778 (0.755, 0.802)	0.800 (0.772, 0.829)	0.170 (0.158, 0.182)
DT	0.785 (0.762, 0.809)	0.777 (0.747, 0.806)	0.167 (0.152, 0.182)
RF	0.804 (0.781, 0.826)	0.858 (0.835, 0.880)	0.134 (0.121, 0.146)
SVM	0.771 (0.744, 0.797)	0.837 (0.812, 0.862)	0.158 (0.145, 0.171)
XGBoost	0.803 (0.781, 0.825)	0.859 (0.837, 0.881)	0.142 (0.128, 0.156)

Note: precision, recall, and accuracy were weighted results. NB: Naive Bayes; LR: logistic regression; DT: decision tree; RF: random forest; SVM: support vector machine; XGBoost: extreme gradient boosting

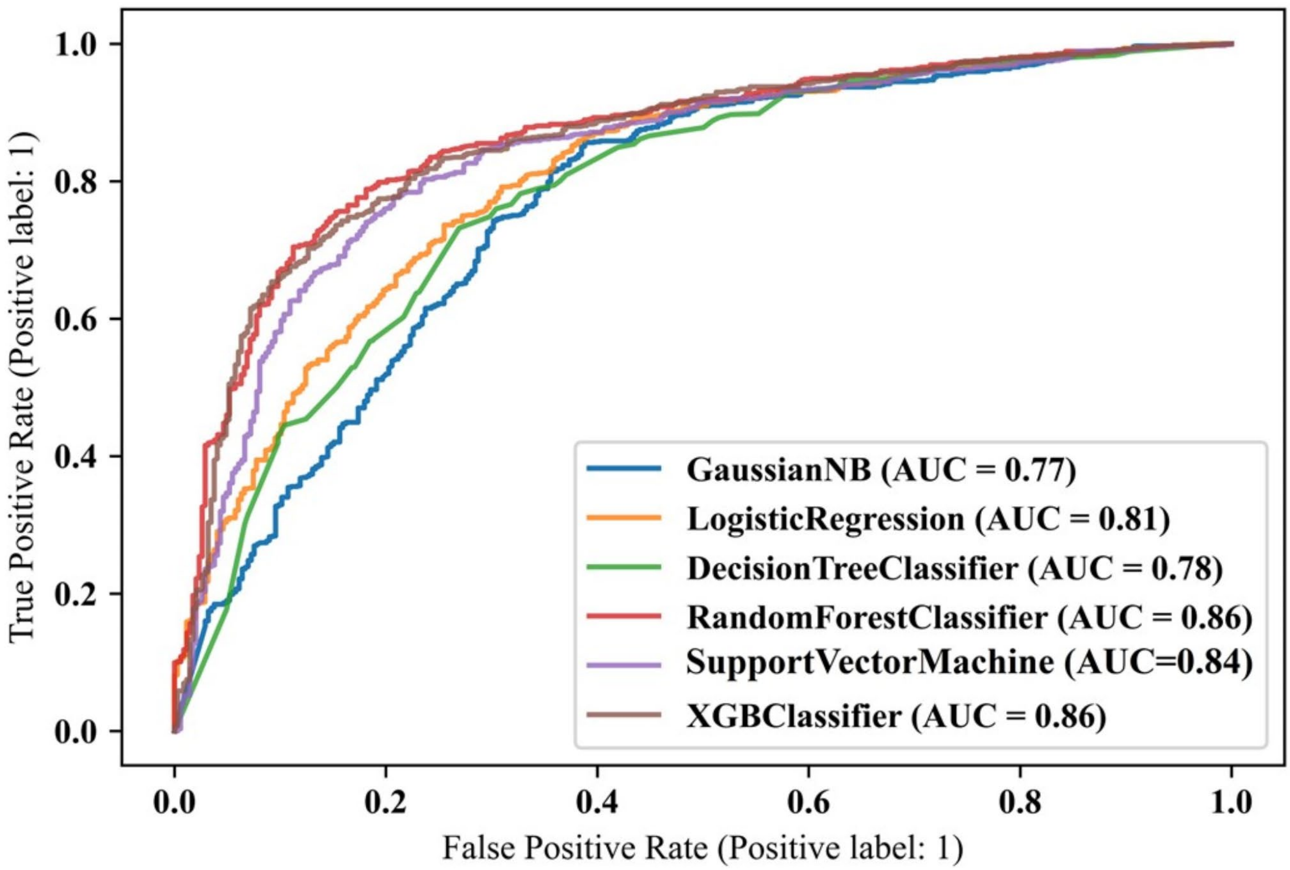

Fig. 2 The receiver operating characteristic curve (ROC) of all models

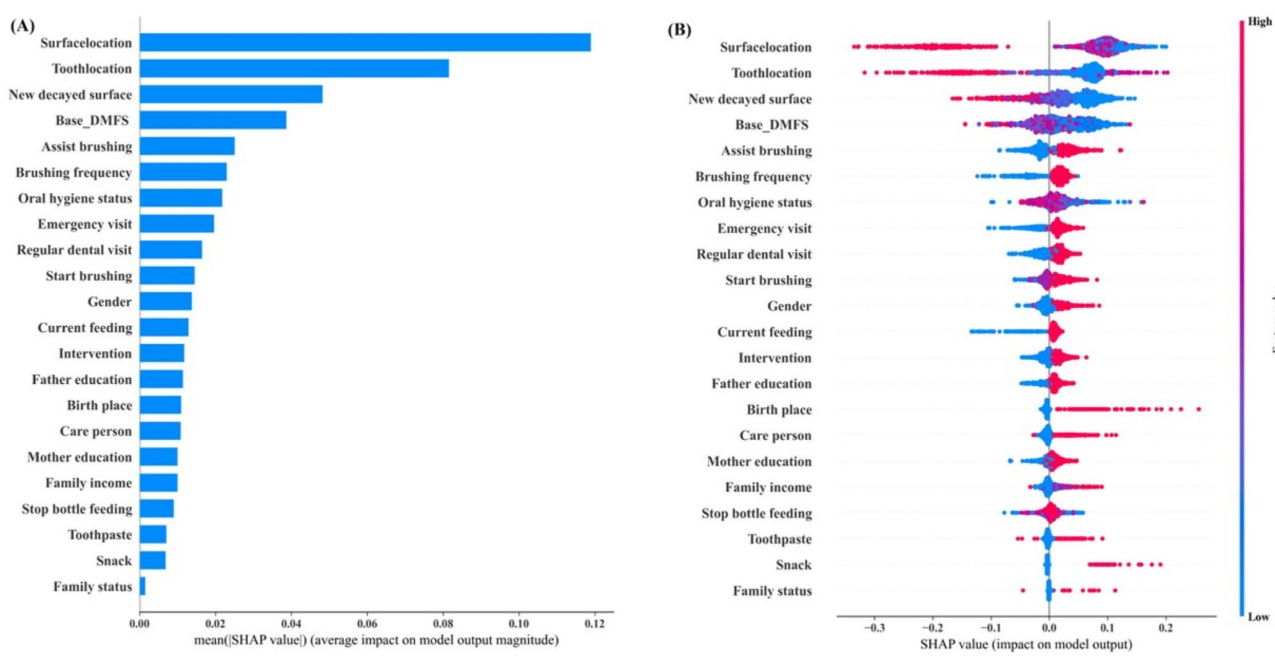

Fig. 3 Feature importance of RF model (**A**: typical bar chart in SHAP, **B**: beeswarm plot in SHAP. The Y-axis indicates the feature importance ranking of all predictors [descending order], and the X-axis indicates the Shapley values, with higher values indicating larger contributions to the predicted outcome. The colour bar represents the magnitude of the values, with red indicating the highest value and blue indicating the lowest value.)

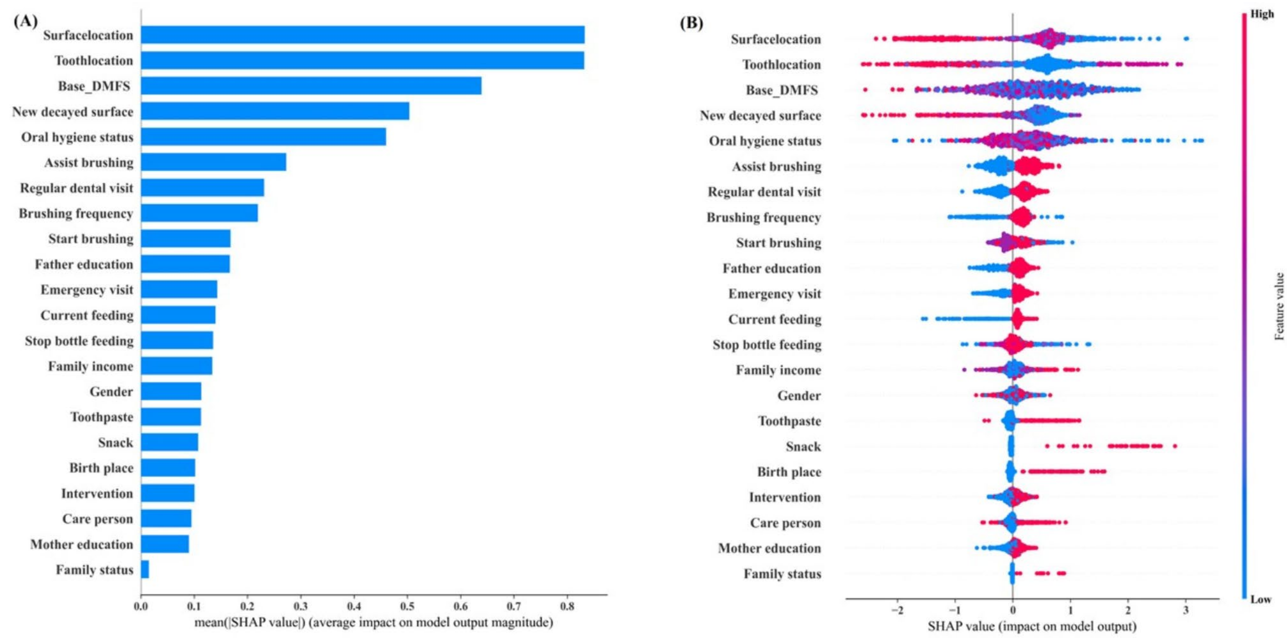

Fig. 4 Feature importance of XGBoost model (**A**: typical bar chart in SHAP, **B**: beeswarm plot in SHAP. The Y-axis indicates the feature importance ranking of all predictors [descending order], and the X-axis indicates the Shapley values, with higher values indicating larger contributions to the predicted outcome. The colour bar represents the magnitude of the values, with red indicating the highest value and blue indicating the lowest value.)

Table 2 Summary of the important predictors identified by machine learning models and comparison with the effect modifiers reported by the original trial

	Important predictor	**Effect modifier**
Clinical parameters	Tooth location of the caries	Tooth location of the caries
	Surface location of the caries	Surface location of the caries
	Newly developed dmfs	Newly developed dmfs
	Baseline dmfs	
	Visible plaque index	
Oral health-related behaviours		Daily snack intake
	Assisted toothbrushing	

Note: Effect modifiers were extracted from the original study (Gao SS et al. 2020). dmfs – decayed, missing (due to caries) and filled surfaces

regression analysis is often used to study the confounders or effect modifiers that are related to the treatment outcome; their predictive power is usually weak [6]. Machine learning, as the main branch of artificial intelligence, uses different algorithms that can learn the statistical patterns and structures of data to build predictive models. It can include more factors and generate models with more accurate predictions [21]. Machine learning is currently gaining popularity in dental research. It is usually applied in disease diagnosis and prediction of prognosis. In cariology, machine learning has been adopted for the detection of dental caries [22] and the prediction of caries risk [23]. Because no study used machine learning in predicting the treatment outcome of dental caries, we tried six typical algorithms, namely LR, NB, SVM, DT, RF and XGBoost, to build the predictive models in this study. We used six metrics, namely accuracy, recall, precision, F1 score, AUROC and Brier score, to comprehensively evaluate the performance of each predictive model. The results supported that all six models showed good performance when being used in predicting the caries-arresting outcome of ECC, especially for the RF and XGBoost. Dental researchers can consider using these machine learning algorithms when generating their own predictive models for predicting caries treatment outcomes. Most sophisticated algorithms present only the risk probability in the output and are usually absent of visual interpretability. SHAP is a tool that can help to improve the interpretability of complex machine learning models. SHAP can describe the importance of a particular factor in a model when making a prediction, with positive or negative values representing the direction of the effect [24]. In this study, we used SHAP bar charts and beeswarm plots to provide a visual interpretation of RF and XGBoost models. It helps the readers quickly identify the importance of predictors and the directions of how the predictors impact the treatment outcome in the predictive models.

In the original study, we used logistic regression with generalized estimating equations to investigate the effects of the intervention, clinical parameters, oral health-related behaviours, and socioeconomic backgrounds on the caries-arresting results. We found that tooth and surface location of the carious lesions, newly developed dmfs during the follow-up period and daily snack intake behaviours were significantly associated with the treatment outcome of ECC [14]. In the current study, tooth and surface location of the carious lesions and newly developed dmfs during the follow-up period were included in the top five predictors of caries arresting results. Carious lesions on buccal surfaces or anterior teeth are expected to have more favourable treatment outcomes after fluoride and silver therapy; dental clinicians can consider the topical application of 38% SDF solution or 25% $AgNO_3$ solution followed by 5% NaF varnish for treating those cases. In the predictive models, we identified several new important predictors that were not reported in the original analysis, namely the dmfs score at baseline, whether the children received assisted toothbrushing and the oral hygiene status. Children with higher baseline dmfs scores will have a worse treatment outcome of ECC. Therefore, for children with high caries risk (i.e., higher dmfs scores and higher numbers of newly developed dmfs), dental professionals should consider reducing the risk to achieve a better treatment outcome of fluoride and silver therapies. In the XGBoost model, oral hygiene status is the fifth predictor for the caries-arresting result. Therefore, plaque control is critical for improving the treatment outcome of ECC. Toothbrushing is the most effective way to remove visible plaque. For young children who are not able to brush their teeth well, assisted toothbrushing is often required. We found that whether the children had assisted toothbrushing was listed as the fifth predictor in the RF model and the sixth predictor in the XGBoost model. Hence, if parents can assist their young children to perform better toothbrushing practice, the caries-arresting effect of fluoride and silver therapies is expected to be more favourable.

There are several strengths of this study. First, the data utilised in this analysis is from a well-designed RCT. Data with high quality provides a good foundation to build reliable predictive models. Second, this study pioneered machine learning in predicting the caries-arresting outcome of ECC and demonstrated that machine learning can be a promising strategy for predicting treatment outcomes in cariology. Dental professionals can adopt a similar methodology in conducting predictions for other disease and treatment combinations. Third, the predictive models identified several key predictors that can impact the caries-arresting outcome. Dental clinicians can refer to the key predictors and develop tailor-made strategies to manage ECC in individuals. Some limitations should also be noticed. First, because there is no previous research investigating a similar topic, we could only choose the commonly used six machine learning algorithms to generate the predictive models. There may be other algorithms that fit the data better. Future studies can focus on testing other potential algorithms and developing applications of predictive models in dental practices. Second, we should be cautious that the identified key predictors only represent associations instead of causality. Recently, a systematic review pointed out that relevance and importance do not guarantee biological significance, because the machine learning studies lack considerations of causation needed to correlate their predictions with biological relevance [25]. However, the top predictors may help us discover and understand the potential causal relationships. Also, the data analyzed in our study was derived from an RCT design, which enhanced the understanding of causality between our key predictors and ECC to some extent. Finally, the prediction performance of our machine learning algorithms was only slightly better than the logistic regression model, which might be attributed to the relatively small sample size. In further studies, we should validate the models with large population-based cohorts, and maximize the advantages of machine learning algorithms.

Conclusion

In this study, we demonstrated that machine learning algorithms can provide promising predictions of the treatment outcome of ECC. The location of carious lesions, caries experience and toothbrushing behaviours are the important predictors of the caries-arresting outcome. These important predictors identified by machine learning models would be particularly informative for targeted management of ECC.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12903-025-05768-y.

Supplementary Material 1

Acknowledgements

The authors thank the kindergarten teachers, children, and parents for their support in the original trial.

Author contributions

YW performed the acquisition, analysis and interpretation of data, and drafted the manuscript. MJ performed acquisition and analysis. YF contributed to the conception. DD performed acquisition and analysis. CHC contributed to the conception and design of the work. SSG contributed to the conception and design of the work and drafted the manuscript. All authors read and approved the final manuscript.

Funding
This study was supported by Natural Science Foundation of Xiamen, China (No. 3502Z20227001).

Data availability
The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
Ethical approval for this study was obtained from the Institutional Review Board of the University and the Hospital Authority (No.: UW 13–569). Written informed consent was collected from the parents of the participating children.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 26 March 2024 / Accepted: 7 March 2025
Published online: 15 March 2025

References

1. MacEachern SJ, Forkert ND. Machine learning for precision medicine. Genome. 2021;64(4):416–25.
2. Hamburg HA, Collins FS. The path to personalized medicine. N Engl J Med. 2010;363:301–4.
3. Zeng X, Wang A, Zhang Y, Zhang H. Factors influencing oral health service utilization for autistic children. Int Dent J. 2024;74:S413.
4. Brondani B, Knorst JK, Tomazoni F, Cósta MD, Vargas AW, Noronha TG, Mendes FM, Ardenghi TM. Effect of the COVID-19 pandemic on behavioural and psychosocial factors related to oral health in adolescents: A cohort study. Int J Paediatr Dent. 2021;31(4):539–46.
5. Correia KF, Dodge LE, Farland LV, Hacker MR, Ginsburg E, Whitcomb BW, Wise LA, Missmer SA. Confounding and effect measure modification in reproductive medicine research. Hum Reprod. 2020;35(5):1013–8.
6. Pearson R, Pisner D, Meyer B, Shumake J, Beevers CG. A machine learning ensemble to predict treatment outcomes following an internet intervention for depression. Psychol Med. 2019;49:2330–41.
7. Efthimiou O, Seo M, Chalkou K, Debray T, Egger M, Salanti G. Developing clinical prediction models: a step-by-step guide. BMJ. 2024;386.
8. Rajula HS, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. Medicina. 2020;56(9):455.
9. Chen X, He L, Shi K, Wu Y, Lin S, Fang Y. Interpretable machine learning for fall prediction among older adults in China. Am J Prev Med. 2023;65(4):579–86.
10. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, Cannon TD, Krystal JH, Corlett PR. Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatry. 2016;3:243–50.
11. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Inf. 2007;2:59–77.
12. Qiu W, Chen H, Dincer AB, Lundberg S, Kaeberlein M, Lee SI. Interpretable machine learning prediction of all-cause mortality. Commun Med. 2022;2(1):125.
13. Chu CH, Gao SS, Li SK, Wong MCM, Lo ECM. The effectiveness of the biannual application of silver nitrate solution followed by sodium fluoride varnish in arresting early childhood caries in preschool children: study protocol for a randomised controlled trial. Trials. 2015;16:1–7.
14. Gao SS, Chen KJ, Duangthip D, Wong MCM, Lo ECM, Chu CH. Arresting early childhood caries using silver and fluoride products–a randomised trial. J Dent. 2020;103:103522.
15. Fung MHT, Duangthip D, Wong MCM, Lo ECM, Chu CH. Randomized clinical trial of 12% and 38% silver Diamine fluoride treatment. J Dent Res. 2018;97:171–8.
16. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019;110:12–22.
17. Golpour P, Ghayour-Mobarhan M, Saki A, Esmaily H, Taghipour A, Tajfard M, Ghazizadeh H, Moohebati M, Ferns GA. Comparison of support vector machine, Naïve Bayes and logistic regression for assessing the necessity for coronary angiography. Int J Environ Res Public Health. 2020;17:6449.
18. Hu C, Anjur V, Saboo K, Reddy KR, O'Leary J, Tandon P, Wong F, Garcia-Tsao G, Kamath PS, Lai JC, Biggins SW. Low predictability of readmissions and death using machine learning in cirrhosis. Am J Gastroenterol. 2021;116:336–46.
19. Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. Front Aging Neurosci. 2017;9:329.
20. Hsu YC, Weng HH, Kuo CY, Chu TP, Tsai YH. Prediction of fall events during admission using eXtreme gradient boosting: a comparative validation study. Sci Rep. 2020;10:1–7.
21. Park YH, Kim SH, Choi YY. Prediction models of early childhood cries based on machine learning algorithms. Int J Environ Res Public Health. 2021;18:8613.
22. Casalegno F, Newton T, Daher R, Abdelaziz M, Lodi-Rizzini A, Schürmann F, Krejci I, Markram H. Caries detection with nearinfrared transillumination using deep learning. J Dent Res. 2019;98:1227–33.
23. Hung M, Voss MW, Rosales MN, Li W, Su W, Xu J, Bounsanga J, Negrón BR, Lauren E, Licari FW. Application of machine learning for diagnostic prediction of root caries. Gerodontology. 2019;36:395–404.
24. Li R, Shinde A, Liu A, Glaser S, Lyou Y, Yuh B, Wong J, Amini A. Machine learning–based interpretation and visualization of nonlinear interactions in prostate cancer survival. JCO Clin Cancer Inf. 2020;4:637–46.
25. Martin SA, Townend FJ, Barkhof F, Cole JH. Interpretable machine learning for dementia: a systematic review. Alzheimers Dement. 2023;19(5):2135–49.

Publisher's note