Salivary Cystatin S in Children With Early Childhood Caries in Comparison With Caries-free Children; Statistical Analysis and Machine Learning

Maryam Koopaie

Tehran University of Medical Sciences

Mahsa Salamati

Tehran University of Medical Sciences

Roshanak Montazeri (✉ roshin_m65@yahoo.com)

Tehran University of Medical Sciences

Mansour Davoudi

Shiraz University

Sajad Kolahdooz

Tehran University of Medical Sciences

Research Article

Keywords: Early childhood caries (ECC), saliva, cystatin S, biomarker, machine learning

Posted Date: June 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-620613/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Early childhood caries is the most common infectious disease in childhood, with a high prevalence in developing countries. Recognition of the factors affecting early childhood caries and its pathophysiology, allows better control of disease. Cystatin S as one of the salivary proteins, has an important role in pellicle formation, tooth re-mineralization and protection. The aim of the present study is to assess salivary cystatin S levels and demographic data in early childhood caries children in comparison with caries-free ones using statistical analysis and machine learning methods.

Methods: A cross-sectional case-control study was undertaken on 20 cases of early childhood caries and 20 caries-free children as a control group. Unstimulated whole saliva samples by suction method was collected. Cystatin S concentrations were determined using human cystatin S ELISA kit. A checklist was collected for each participant about the demographic characteristics, oral health status and dietary habits by interviewing parents. The regression and receiver operating characteristic (ROC) curve analysis was done to evaluate the potential role of cystatin S salivary level and demographic using machine learning and statistical analysis.

Results: The mean value of salivary cystatin S concentration in early childhood caries group was 191.55 ± 81.90 (ng/ml) and in caries-free group was 370.06 ± 128.87 (ng/ml). T-test analysis showed that there is a statistically significant difference between early childhood caries and caries-free group in salivary cystatin S level ($p: 0.032$). Investigation of area under the curve and accuracy of ROC curve revealed that logistic regression model based on the salivary cystatin S levels and birth weight had most and acceptable potential for discriminating of early childhood caries from caries-free controls. After that the machine learning models and finally salivary cystatin S levels had more capability for differentiation of early childhood caries from caries-free controls.

Conclusion: Salivary cystatin S in caries-free children was higher than the children with early childhood caries. Therefore, cystatin S protein can be used as a biomarker for early prediction of early childhood caries, furthermore cystatin S is a protective factor against dental caries.

Background

Dental caries is one of the most common chronic infectious diseases in children and it is a major health problem in many countries. ECC is defined as the presence of one or more decayed (non-cavitated or cavitated lesions), missing teeth (because of caries), or filled tooth surfaces in any primary tooth in a child under 72 months of age [1, 2]. The first sign of ECC in children is demineralized areas in cervical and labial surfaces of maxillary incisors. Delay in treatment result in caries on occlusal surface of second primary molars and distal aspects of first primary molars [3]. Negative impacts of ECC on children's quality of life are numerous. ECC impair nutrition, speaking, occlusion, social behaviors and patient's self-esteem. According to World Health Organization (WHO) reports, 60 to 90 percent of students in the world are affected by dental caries. Dental caries is more common in Asian and Latin American students [4].

Dental caries is a multifactorial disease that results from multiple factors including immature immune systems; individual characteristics of saliva; cariogenic microorganisms; hypo salivation, enamel hypoplasia, enamel defects, cariogenic diet; night breast-feeding and oral hygiene care in early childhood [5]. Premature birth, low birth weight, Lack of access to nutritional supplements during pregnancy are other risk factors of ECC [6]. Socioeconomic factors such as parent's income and educational level, birth order and dental insurance coverage are the other effective factors on caries risk [7]. Recently salivary components and characteristics are introduced as one of the important factors, which have effect on ECC development[8]. ECC has several short and long-term side effects on children's quality of life. Untreated dental caries cause pain, difficulty in chewing, decrease appetite, sleep problems, behavioral changes (low self-esteem) and disturbance in school practice. ECC has adverse socioeconomic effects on the family [9]. ECC treatment generally does not have a long-term effect on the population of oral streptococcus mutant, and although these treatments help control the disease, caries recurrence is still common around or after repair, with a recurrence rate of approximately 40% reported during the first year [10, 11]. Hence, the management of caries in many countries redirects towards preventive measures. Biomarker-based diagnostic methods can play an important role in the early diagnosis of ECC. Based on prediction of child's caries risk, it is recommended that parents participate in oral health programs to prevent the development of dental caries by controlling the diet and oral health habits of the child [12].

Saliva is a unique body fluid which is easily accessible and contains complex components with great diagnostic value. Saliva collecting and storage is easy, noninvasive and cost-effective [13]. Salivary immune system composed of proteins with a significant effect on oral health. Many of the salivary proteins prevent adhesion and aggregation of cariogenic bacteria and some others has a role in defense against microorganisms [14]. Qualitative and quantitative changes in salivary proteome have a significant role in the initiation and intensity of oral disease [15]. Although proteomics component of saliva is only 30% of the bloods component, it is still known as a rich source of biomarkers [8]. Cystatins are a large family of proteins that control and reversibly inhibit the activity of extracellular cysteine proteinases in inflammatory conditions [16]. Cystatin S is a phosphorylated protein that is found in the tear and saliva. It is mainly secreted by submandibular gland and to a much lesser extent by the parotid and sublingual gland [17]. Unlike other proteins in this group, cystatin S has a lower cysteine proteinase inhibitory function [18].

Cystatin S has four sites for phosphorylation that bond with hydroxyapatite and has an important role in dental pellicle formation and calcium and phosphor equilibrium and cause enamel remineralization. It also prevents enamel demineralization by attachment to enamel surface. Cystatin S protects the oral mucosa by its antimicrobial and antiviral effects [19, 20]. Considering the role of cystatin S in caries prevention, the aim of the present study is a comparison of salivary level of cystatin S in ECC patients and caries-free (CF) children.

Methods

Ethical statement

This study was approved by Tehran University of Medical Science Ethical Committee (ethical code: IR.TUMS.DENTISTRY.REC.1398.099). After describing the study objectives, written informed consent was obtained from a parent or guardian for participants under 16 years old.

Samples

A cross-sectional case-control study was undertaken on 20 cases of ECC and 20 CF children as a control group. The participants enrolled in this study was selected randomly from children that referred to the dental clinic of Tehran University of Medical Sciences (Tehran, Iran) for routine oral examinations. The oral health status of each participant was determined by three professional dentists. The inclusion criteria were children aged between 48 and 72 months. CF controls were matched regarding age and gender.

After oral examination of participants, researcher completed the checklist by interviewing parents which collected information about the demographic characteristics of child and parents, dietary intakes [21], birth weight, oral hygiene behaviors, parental education level, medicine intake and night breastfeeding. Exclusion criteria include; children had chronic systemic diseases or syndromes, influenza or infection of respiratory system, taking medicine and received antibiotic therapy within three months, fluoride prophylaxis within 1 year and medical history of congenital diseases, parents refused to participate or refused to sign the informed consent and children who did not agree or cooperate with the participation. Children were assessed using Decayed, Missing, and Filled Teeth (dmft) index based on WHO Oral Health Surveys Basic Methods [22] and diagnosis of ECC was done based on the diagnostic criteria of American Academy of Pediatric Dentistry [23]. Those having a dmft index of zero were considered CF. After surveying and regarding the exclusion criteria finally, 20 children were selected as cases and 20 of them considered as control.

Saliva collection

Unstimulated whole saliva samples by suction method were collected. Before sampling children have been in rest position and they did not eat anything for 30 minutes. Saliva sampling was done between 9:00 am to 11:00 am to avoid circadian variations in case and control group. A protease inhibitor cocktail (Roche Diagnostics GmbH, Mannheim, Germany) was immediately added after the completion of saliva collection. Saliva samples centrifuged at 10,000×g for 15 min at 4°C, the supernatant was obtained and stored in - 80°C.

Determination of salivary cystatin S levels

Cystatin S concentrations were determined using human cystatin enzyme-linked immunosorbent assay (ELISA) kit (ZellBio GmbH, Ulm, Germany). The samples were thawed at 25°C and assayed in accordance with the manufacturer's instructions. The absorbance of samples at 450 nm was measured using Hyperion ELISA microplate reader. The concentrations of cystatin S were determined by spectrometer software based on standard curves.

Statistical analysis

Statistical analyses were performed with SPSS software (version 22; SPSS Inc., Chicago, IL, USA) and GraphPad Prism 8.2.1 for Windows (GraphPad Software, San Diego, California). Assuming normal distribution of data, in order to assess relationship between dental caries and age with cystatin S salivary level, T-test and Levene's test was used. Mean value and standard deviation of cystatin S level was reported. Regression analyses were performed for cystatin S level with backward stepwise method. Contribution of each variable (including ECC, demographic and clinical characteristics and nutrition habits) was expressed by the p-value (p) and standardized coefficients beta (β). The receiver operating characteristic (ROC) curve analysis was done to evaluate the potential role of cystatin S salivary level and combination of cystatin S salivary level with weight of birth to prediction of ECC. All results are presented as mean $\pm$ standard deviation (SD) and $p \leq 0.05$ was considered statistically significant.

Machine learning analysis

To evaluate the effectiveness of extracted features, including cystatin S, demographic and clinical characteristics, and nutrition habits, we used various supervised machine learning methods, including feed-forward neural network, XGBoost, Random Forest, Support vector machine (SVM), and Logistic regression. For implementing feed-forward neural network we used Python Software Foundation, Version 3.7, and the open-source deep learning package, namely Keras [24], which is a high-level neural network API.

The feed-forward neural network method has two hidden layers, and each hidden layer has 32 neurons (Additional file 1: Figure. S1, S2). We used the ReLU activation function for the hidden layers, and for the output layer, the sigmoid activation function was used. To evaluate the result of the feed forward neural network model, the binary cross-entropy loss function was used as the following form:

$$Loss = -\frac{1}{N} \sum_{j=0}^N [L_j \log \hat{L}_j + (1 - L_j) \log(1 - \hat{L}_j)] + \frac{\lambda}{2N} \sum_w W^2$$

Where N is the total number of training samples, L is the ground-truth label, $\hat{L}$ is the predicted label, and λ is the regularization parameter for the loss function. We use the 5-fold cross-validation and the Adam optimizer [25] for both datasets. The hyper-parameter settings for both datasets are 100 for batch size, 0.001 for the learning rate, 0.2 for the dropout probability, and $\lambda = 0.01$ for the regularization parameter of the loss function. Also, to implement other mentioned supervised machine learning methods, we used the scikit-learn library and ran all the classifiers using default parameter settings.

During the training phase of the constructed models, for each training data sample, the extracted features (including salivary cystatin S, demographic and clinical characteristics, oral health status, and dietary intakes) were entered into the models, and the output value indicating the label of the data sample was computed. 80% of data samples were used during the training and validation phase, and 20%

remaining were used during the test phase. The ROC curve analysis was done using demographic data and also a combination of demographic data and cystatin S salivary level by machine learning method.

Results

Patient characteristics

The demographic and clinical characteristics of forty sex- and age-matched subjects from the CF and ECC groups (20 subjects from each group) are depicted in Table 1.

Table 1
Participants demographic and clinical characteristics

	Control (CF) (n = 20)	Patients (ECC) (n = 20)
Age of participants, months (mean $\pm$ SD)	63.70 $\pm$ 8.32	63.25 $\pm$ 8.32
Sex		
Male	8 (40 %)	11(55%)
Female	12(60%)	9(45%)
Birth order		
First child	8 (40%)	11 (55%)
Second child	8 (40%)	5 (25%)
Third child	3 (15%)	2 (10%)
Forth child	1 (5%)	2(10%)
Father's education level (based on ISCED 2011[26])		
Post-secondary non-tertiary education and other lower Educational levels (ISCED 4 and lower levels)	1 (5%)	6 (30%)
Short-cycle tertiary education (ISCED 5)	2 (10%)	13 (65%)
Bachelor or equivalent (ISCED 6)	11 (55%)	0 (0.0%)
Master or equivalent (ISCED 7)	6 (30%)	1 (5%)
Mother's education level (based on ISCED 2011[26])		
Post-secondary non-tertiary education and other lower Educational levels (ISCED 4 and lower levels)	0 (5%)	4 (30%)
Short-cycle tertiary education (ISCED 5)	5 (10%)	15 (65%)
Bachelor or equivalent (ISCED 6)	2 (55%)	0 (0.0%)
Master or equivalent (ISCED 7)	13 (30%)	1 (5%)
Father age, years (mean $\pm$ SD)	35.10 $\pm$ 4.75	38.50 $\pm$ 4.74
Mother age, years (mean $\pm$ SD)	32.70 $\pm$ 3.93	33.75 $\pm$ 5.02
Diseases of parents		
Diabetes	1 (5%)	2 (10%)

	Control (CF) (n = 20)	Patients (ECC) (n = 20)
Rheumatism	1 (5%)	1 (5%)
Cardiovascular diseases	0 (0%)	1 (5%)
Dental visit regularly (at least every 6 months)		
Yes	17 (85%)	5 (25%)
No	3 (15%)	15 (75%)
Toothbrush		
Yes	19 (95%)	14 (70%)
No	1 (5%)	6 (30%)
Mouthwash		
Yes	7 (35%)	2 (10%)
No	13 (65%)	18 (90%)
Flossing		
Yes	13 (65%)	3 (15%)
No	7 (35%)	17 (85%)
Using fluoride Toothpaste		
Yes	18 (90%)	10 (50%)
No	1 (5%)	4 (20%)
Parental smoking daily		
Yes	6 (30%)	3 (15%)
No	14 (70%)	17 (85%)

In ECC group 75% of children did not have regular dental visits but in CF group 85% of children had regular dental visits. In ECC group 95% of children brush daily while 70% of ECC children brush daily. 65% of CF children had daily dental floss; this value was 15% for ECC group. Just 22.5% of children from both group used regular mouthwash. 75% of CF children brush twice a day, this value was 30% for ECC group. 70% of children in case and control group used fluoride toothpaste and 12.5% used toothpaste without fluoride. There was no difference in smoking by parents in case and control group.

Nutrition habits

The frequency of eating sweet snack in case and control group. Frequency of eating sweet snacks [21] in case and control group and frequency of eating dairy products in case and control group [27] are shown in Table 2. 97.5 % of children in case and control group consumed milk overnight in their infancy period. There was no difference in case and control group in this regard. There was no statistically significant difference in case and control group in using multivitamin and calcium supplements during pregnancy by mothers of participants.

Table 2
Participants nutrition habits characteristics

	Control (n = 20)	Patients (n = 20)
Consumption of sweet snacks		
Once a day	12 (60%)	9 (45%)
Twice a day	8 (40%)	6 (30%)
Three times a day and more	0 (0%)	5 (25%)
Consumption of sour snacks		
Once a day	12 (60%)	15 (75%)
Twice a day	8 (40%)	3 (15%)
Three times a day and more	0 (0%)	2 (10%)
Consumption of dairy units per day		
0 times a day (didn't consume)	0 (0%)	1 (5%)
Once a day	3 (15%)	7 (35%)
Twice a day	9 (45%)	7 (35%)
Three times a day and more	8 (40%)	5 (25%)
Consumption of milk at night during infancy		
Yes	19 (95%)	20 (100%)
No	1 (5%)	0 (0%)
Taking vitamin and calcium supplements by the mother during pregnancy		
Yes	18 (90%)	20 (100%)
No	2 (10%)	0 (0%)

Statistical analysis

The mean value of salivary cystatin concentration in ECC group was 191.55 ± 81.90 (ng/ml) and in CF group was 370.06 ± 128.87 (ng/ml). T-test analysis indicated that there is a statistically significant difference between ECC and CF group in salivary cystatin S level ($p: 0.032$). Considering 306.3 as cut off value, the sensitivity of salivary cystatin S in caries diagnosis was 95% and its specificity was 65% (Fig. 1-A).

Mean value of salivary cystatin S levels in CF and ECC groups based on the age of children, birth weight and birth order are shown in Figure 2.

In order to assess the effect of ECC and other risk factors on salivary cystatin S level, backward stepwise regression was performed and two models were concluded by including and excluding variables (criterion: Probability of F-to-remove ≥ 0.100). There was a statistically significant relationship between child's age ($p < 0.001$, $\beta: -0.413$), birth weight ($p: 0.002$, $\beta: 0.273$), birth order ($p: 0.021$, $\beta: -0.152$) and ECC ($p < 0.001$, $\beta: -0.658$) with salivary cystatin S level in regression model.

In addition, there was a statistically significant relationship between age of children ($p < 0.001$, $\beta: -0.533$), father education ($p: 0.007$, $\beta: 0.361$) and mother education ($p: 0.019$, $\beta: 0.308$) with salivary cystatin S level regression modeling of cystatin S level. In addition, logistic regression was used for modeling and prediction of ECC risk and based on the cystatin S and weight of birth ROC curve was presented for discriminating of ECC (Figure 1-B).

Machine learning results

To evaluate the effectiveness of cystatin S levels feature in the differentiation ECC from CF groups, we construct two different neural network models. The first model includes all extracted features except cystatin S levels and the second model, includes all features plus cystatin S levels feature. Tables 3 and 4 show the value of the accuracy, sensitivity, and specificity measures for the constructed models.

Table 3

the results of supervised machine learning methods, utilizing all features except cystatin S levels

All features except cystatin S	Feed-Forward neural network	XGBoost	Random Forest	SVM	Logistic Regression
Accuracy (%)	88.1	85.7	78.5	75	57.4
Sensitivity (%)	58.3	92.8	92.8	92.8	78.5
Specificity (%)	100	78.5	64.2	57.1	35.7

Table 4

the results of supervised machine learning methods, utilizing all features with cystatin S levels

All features	Feed-Forward neural network	XGBoost	Random Forest	SVM	Logistic Regression
Accuracy (%)	90.9	89.2	85.7	85.4	60.7
Sensitivity (%)	64.1	93.3	93.3	86.6	41.6
Specificity (%)	100	84.6	76.9	84.6	84.6

The results show that the feed-forward neural network model, including all features in addition to cystatin S levels achieves the most accuracy. ROC curve analysis of all feature except cystatin S related to the feed-forward neural network model is shown in Fig. 3-A (Area under the ROC curve (AUC): 0.87). In addition, ROC curve analysis of all features (including cystatin S, demographic and clinical characteristics and nutrition habits) related to the feed-forward neural network model revealed that the AUC: 0.93 (Fig. 3-B).

We also investigated the effectiveness of extracted features in the Random Forest classifier (Figure 4). Gini impurity [28] was utilized as a feature importance score. Figure 4 demonstrates the feature importance score of extracted features. The “cystatin S levels” was obtained the highest score and second highest score belongs to the “dental visit regularly” and the third rank belong to “mother’s education level”.

Discussion

ECC is premature dental caries in deciduous teeth that mainly affects smooth surfaces and is one of the most common chronic infectious diseases in children [29]. ECC is a preventable disease and its early diagnosis is very important [30]. Besides diagnosed risk factors of ECC, changing in salivary composition is one of the risk factors with significant importance in early disease diagnosis [31]. Biomarker-based diagnostic methods can play an important role in early diagnosis of this disease. Studies have shown that some salivary compounds have different levels of biological activity in active caries individuals in comparison with CF ones. However, there are limited studies on the difference between salivary protein components in ECC and CF children. Cystatin S is a cystatin protease inhibitor, which is mainly present in submandibular saliva. Phosphorylated form of cystatin S has a significant role in regulation of salivary calcium level [32],[33]. In our study, assessment of salivary cystatin S by ELISA, revealed that salivary cystatin S was significantly higher in CF children than ECC group ($p < 0.005$). Furthermore, the sensitivity of cystatin S in caries diagnosis was 95% and its specificity was 65%. Considering the AUC, we concluded that cystatin S was a proper salivary biomarker for early diagnosis of ECC and prevention of the disease.

These results are in line with Vittrino et al. study. They assessed salivary proteins involved in dental pellicle formation and concluded that the level of cystatin S in dental pellicle in CF individuals was higher than dental caries ones and there was a direct relationship between salivary cystatin S and lower values of dmft [34]. Wang et al. in a study with the aim of salivary proteomics investigation in children with and without caries in the age range of 10-12 years concluded that the level of salivary cystatin S in the group without caries is significantly higher than the group with high dental caries [35]. In a study by Siqueira et al., compares the protein composition of dental pellicle in patients with and without dental caries. They revealed that cystatin S level in CF group was higher than dental caries group [36].

Odanaka et al. revealed that cystatin S in dental pellicle originated from saliva [37]. Age is one of the important factors that cause significant changes in saliva composition and plays an important role in determining the proteins of healthy and cariogenic saliva. Cabras et al. stated that the level of salivary cystatin S changes at different ages and their results indicated that the salivary cystatin level increase with aging [38]. Furthermore, assessment of salivary proteome in the first 48 month of life, revealed that cystatin isn't present in saliva in first 6 months of life. It can be concluded that salivary cystatin concentration is age related and increase in age and physiologic condition have great impact on it [39].

The level of cystatin S in elderly is different from childhood. Preza et al. in a comparison between elderly patients with root caries and elderly patients without root caries as control group concluded that the level of parotid cystatin S in elderly patients with root caries was significantly higher than control group [40]. These results are contrary to results of our study, which can be attributed to the role of age on concentration of salivary cystatin S. Regarding relatively higher expression of salivary cystatin S in CF group, we anticipate a significant relationship between salivary cystatin S and caries risk factors. Considering high level of cystatin S in CF children a statistically significant relationship between cystatin S and caries risk factors is anticipated. Julihn suggested that birth order is associated with caries increment in young children [41].

Compared with first-born children, the highest risk of caries increment occurred in fifth- or later-born children. In our study, there was a statistically significant inverse relationship between salivary cystatin S and birth order. The level of salivary cystatin S in first-born children of the family was higher. These results are consistent with the study by Julihn which stated that second-, third-, fourth-, and fifth- or later-born children have a significantly increased risk of developing new caries lesions between age 3 to 7 years compared with first-born children [41].

The other risk factor of dental caries is low birth weight. Some studies demonstrated a significant relationship between dental caries and preterm low birth weight status. Bernabé et al. concluded that low birth weight and maternal smoking cause increase in caries rate. The results of this study were in line with ours. Nicolau et al. in a study concluded that low birth weight children which were second-born child in the family had higher dmft index [42]. Hallett et al. in their study reported that prevalence of ECC in forth born children of the family was significantly higher ($p:0.001$) [43].

These differences between siblings can be for a variety of reasons, including the time parents spend with children or changes in the family as a result of the presence of children of different ages. The first child usually gets all the attention of the parents, at least in the first period of life. Although younger children can benefit from the education of their older siblings, this is a mutual benefit, and it has even been said that older children benefit more from teaching their younger siblings. It is also not hard to imagine that many things that children learn from their older siblings may have a negative effect [44].

Birth problems, such as low birth weight, can lead to problems such as decreased immune function or enamel hypoplasia and also early establishment of cariogenic bacteria in the oral cavity which result in increases the risk of dental caries [40]. In our study there was a statistically significant relationship between birth weight and salivary levels of cystatin S which was similar to Rajshekar et al.'s study. They concluded that prevalence of dental caries has a significant relationship with birth weight and low birth weight children had higher dental caries than normal weight children [45].

Bernabé et al. also found that maternal weight and maternal smoking were associated with changes in dmft and that children with low birth weight and smoking mothers showed a higher rate of caries than children with normal birth weight and non-smoking mothers [46]. These results were in line with ours but Athamneh et al. assessed 30 to 48 months children and concluded that there isn't a significant relationship between birth weight and dental caries [47]. Some studies have shown that the children of parents with higher education have better oral health. In the present study a statistically significant positive correlation between salivary cystatin S and parent, education was found. Highly educated families have a better economic situation, which in turn leads to more benefits from health services and improving their children's oral health.

Another aspect of the impact of parental education is increasing parental awareness of oral hygiene and caries prevention [48]. The present study did not show a significant relationship between salivary cystatin S levels and oral health and nutritional variables. Given that these variables are all part of the risk factors for dental caries and various studies have found a significant relationship between them and dental caries, so the result of our study can be due to the dishonesty of parents in responding to questions about oral health and nutritional habits because parents tend to show their child's health and eating habits better than they already are.

Conclusion

In sum, this study showed that Cystatin S protein levels were significantly lower in children with early childhood caries than in CF ones. This study confirms the association between cystatin S protein and dental caries in children aged 48-71 months. In addition, a significant relationship was found with some risk factors of ECC and salivary cystatin S levels by statistical analysis and machine learning. Cystatin S is involved in the process of enamel remineralization and dental pellicle formation and plays a protective role in tooth structure. Therefore, reducing the amount of this protective protein in saliva in people who have risk factors for caries makes teeth prone to decay. Salivary levels of cystatin S can be introduced as

a prognostic biomarker for dental caries risk assessment. Considering the acceptable sensitivity and specificity of salivary cystatin S in ECC diagnosis, we suggest it as a biomarker for prediction of early childhood caries. Considering the ROC curve analysis, logistic regression model based on the salivary cystatin S levels and birth weight, machine learning models and finally salivary cystatin S levels had more capability for discriminating of early childhood caries from caries-free controls, respectively.

Abbreviations

Area under the ROC curve (AUC)

Caries-free (CF)

Decayed, Missing, and Filled Teeth (dmft)

Early childhood caries (ECC)

Enzyme-linked immunosorbent assay (ELISA)

Receiver operating characteristic (ROC)

Standard deviation (SD)

Support vector machine (SVM)

World Health Organization (WHO)

Declarations

Ethics approval and consent to participate

This study was approved by Tehran University of Medical Science Ethical Committee (ethical code: IR.TUMS.DENTISTRY.REC.1398.099). After describing the study objectives, written informed consent was obtained from a parent or guardian for participants under 16 years old.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. Machine learning code that was used is accessible via the following address:

https://codeberg.org/mansur/Salivary_cystatin_S_in_early_childhood_caries

Competing interests

The authors declare that they have no competing interests.

Funding

There is no source of financial support or funding.

Acknowledgements

Not applicable

Authors' contributions

MK and MS conceived the study idea and led data collection. MK, MS, and RM created the study protocol and wrote the original draft. MK, MS and SK contributed to data analysis / interpretation and preparation of the manuscript. MK, MS and RM led the writing- review & editing. MD and SK performed machine learning. MK and MD interpreted the results. All authors read and approved the final manuscript.

References

1. Anil S, Anand PS: **Early childhood caries: prevalence, risk factors, and prevention.** *Frontiers in pediatrics* 2017, **5**:157.
2. Augustin JG, Lepine C, Morini A, Brunet A, Veyer D, Brochard C, Mirghani H, Péré H, Badoual C: **HPV detection in head and neck squamous cell carcinomas: what is the issue?** *Frontiers in Oncology* 2020, **10**.
3. Boscolo-Rizzo P, Zorzi M, Del Mistro A, Da Mosto MC, Tirelli G, Buzzoni C, Rugge M, Polesel J, Guzzinati S, Group AW: **The evolution of the epidemiological landscape of head and neck cancer in Italy: Is there evidence for an increase in the incidence of potentially HPV-related carcinomas?** *PloS one* 2018, **13**(2):e0192621.
4. Hu J, Jiang W, Lin X, Zhu H, Zhou N, Chen Y, Wu W, Zhang D, Chen H: **Dental caries status and caries risk factors in students ages 12–14 years in Zhejiang, China.** *Medical science monitor: international medical journal of experimental and clinical research* 2018, **24**:3670.
5. Tinanoff N, Baez RJ, Diaz Guillory C, Donly KJ, Feldens CA, McGrath C, Phantumvanit P, Pitts NB, Seow WK, Sharkov N: **Early childhood caries epidemiology, aetiology, risk assessment, societal burden, management, education, and policy: Global perspective.** *International journal of paediatric dentistry* 2019, **29**(3):238-248.
6. Kirthiga M, Murugan M, Saikia A, Kirubakaran R: **Risk factors for early childhood caries: a systematic review and meta-analysis of case control and cohort studies.** *Pediatric dentistry* 2019, **41**(2):95-112.
7. dos Santos Junior VE, de Sousa RMB, Oliveira MC, de Caldas Junior AF, Rosenblatt A: **Early childhood caries and its relationship with perinatal, socioeconomic and nutritional risks: a cross-**

- sectional study.** *BMC oral health* 2014, **14**(1):1-5.
8. Peltzer K, Mongkolkeha A: **Severe early childhood caries and social determinants in three-year-old children from Northern Thailand: a birth cohort study.** *BMC Oral Health* 2015, **15**(1):1-7.
 9. Kobayashi K, Hisamatsu K, Suzui N, Hara A, Tomita H, Miyazaki T: **A review of HPV-related head and neck cancer.** *Journal of clinical medicine* 2018, **7**(9):241.
 10. Kofler B, Laimer J, Bruckmoser E, Steinbichler TB, Runge A, Scharfetter JH, von Laer D, Borena W: **The role of HPV and non-HPV sexually transmitted infections in patients with oropharyngeal carcinoma: A case control study.** *Cancers* 2020, **12**(5):1192.
 11. McDermott JD, Bowles DW: **Epidemiology of head and neck squamous cell carcinomas: impact on staging and prevention strategies.** *Current treatment options in oncology* 2019, **20**(5):1-13.
 12. Rodriguez J, Herrero A, Li S, Rauch N, Quintanilla A, Wynne K, Krstic A, Acosta JC, Taylor C, Schlisio S: **PHD3 regulates p53 protein stability by hydroxylating proline 359.** *Cell reports* 2018, **24**(5):1316-1329.
 13. Yoshizawa JM, Schafer CA, Schafer JJ, Farrell JJ, Paster BJ, Wong DT: **Salivary biomarkers: toward future clinical and diagnostic utilities.** *Clinical microbiology reviews* 2013, **26**(4):781-791.
 14. Si Y, Ao S, Wang W, Chen F, Zheng S: **Magnetic bead-based salivary peptidome profiling analysis for severe early childhood caries.** *Caries research* 2015, **49**(1):63-69.
 15. Kaczor-Urbanowicz KE, Martin Carreras-Presas C, Aro K, Tu M, Garcia-Godoy F, Wong DT: **Saliva diagnostics—Current views and directions.** *Experimental Biology and Medicine* 2017, **242**(5):459-472.
 16. Ochieng J, Chaudhuri G: **Cystatin superfamily.** *Journal of health care for the poor and underserved* 2010, **21**(1 Suppl):51.
 17. Prunk M, Perišić Nanut M, Sabotić J, Kos J: **Cystatins, cysteine peptidase inhibitors, as regulators of immune cell cytotoxicity.** *Periodicum biologorum* 2016, **118**(4):353-362.
 18. Dickinson D: **Salivary (SD-type) cystatins: over one billion years in the making—but to what purpose?** *Critical Reviews in Oral Biology & Medicine* 2002, **13**(6):485-508.
 19. Laputková G, Schwartzová V, Bánovčin J, Alexovič M, Sabo J: **Salivary protein roles in oral health and as predictors of caries risk.** *Open life sciences* 2018, **13**(1):174-200.
 20. Hemadi AS, Huang R, Zhou Y, Zou J: **Salivary proteins and microbiota as biomarkers for early childhood caries risk assessment.** *International journal of oral science* 2017, **9**(11):e1-e1.
 21. Naidu R, Nunn J, Kelly A: **Socio-behavioural factors and early childhood caries: a cross-sectional study of preschool children in central Trinidad.** *BMC oral health* 2013, **13**(1):1-7.
 22. Organization WH: **Oral health surveys: basic methods:** World Health Organization; 2013.
 23. Pediatrics AAo: **Policy on early childhood caries (ECC): classifications, consequences, and preventive strategies.** *Pediatric dentistry* 2008, **30**(7 Suppl):40-43.
 24. Chollet F: **keras.** In.; 2015.
 25. Kingman D, Ba J: **Adam: A Method for Stochastic Optimization. Conference paper.** In: *3rd International Conference for Learning Representations: 2015*; 2015.

26. Statistics Ulf: **International standard classification of education: ISCED 2011**: UNESCO Institute for Statistics Montreal; 2012.
27. Badruddin IA, Zhafarina AR, Rahardjo A, Prabawanti C, Adiatman M: **The Correlation between a Mother's Behavior Regarding Her Child's Dietary Habits and Early Childhood Caries (Based on the Theory of Planned Behavior)**. *Journal of International Dental and Medical Research* 2017, **10**:574-582.
28. Nembrini S, König IR, Wright MN: **The revival of the Gini importance?** *Bioinformatics* 2018, **34**(21):3711-3718.
29. Sun X, Huang X, Tan X, Si Y, Wang X, Chen F, Zheng S: **Salivary peptidome profiling for diagnosis of severe early childhood caries**. *Journal of translational medicine* 2016, **14**(1):1-11.
30. Martins C, Buczynski AK, Maia LC, Siqueira WL, de Araujo Castro GFB: **Salivary proteins as a biomarker for dental caries—a systematic review**. *Journal of dentistry* 2013, **41**(1):2-8.
31. Vieira AR, Vieira NM, Limesand K, Modesto A: **Differences in Proteomic Profiles Between Caries Free and Caries Affected Children**. *Pesquisa Brasileira em Odontopediatria e Clínica Integrada* 2020, **20**.
32. Rudney J, Staikov R, Johnson J: **Potential biomarkers of human salivary function: a modified proteomic approach**. *Archives of oral biology* 2009, **54**(1):91-100.
33. Devarajan H, Somasundaram S: **Salivary proteins and its effects on dental caries—A review**. *Drug Invention Today* 2019, **11**(6).
34. Vitorino R, de Moraes Guedes S, Ferreira R, Lobo MJC, Duarte J, Ferrer-Correia AJ, Tomer KB, Domingues PM, Amado FM: **Two-dimensional electrophoresis study of in vitro pellicle formation and dental caries susceptibility**. *European journal of oral sciences* 2006, **114**(2):147-153.
35. Wang K, Wang Y, Wang X, Ren Q, Han S, Ding L, Li Z, Zhou X, Li W, Zhang L: **Comparative salivary proteomics analysis of children with and without dental caries using the iTRAQ/MRM approach**. *Journal of translational medicine* 2018, **16**(1):1-13.
36. Siqueira W, Custodio W, McDonald E: **New insights into the composition and functions of the acquired enamel pellicle**. *Journal of dental research* 2012, **91**(12):1110-1118.
37. Odanaka H, Obama T, Sawada N, Sugano M, Itabe H, Yamamoto M: **Comparison of protein profiles of the pellicle, gingival crevicular fluid, and saliva: possible origin of pellicle proteins**. *Biological research* 2020, **53**.
38. Cabras T, Pisano E, Boi R, Olinas A, Manconi B, Inzitari R, Fanali C, Giardina B, Castagnola M, Messina I: **Age-dependent modifications of the human salivary secretory protein complex**. *Journal of proteome research* 2009, **8**(8):4126-4134.
39. Manconi B, Cabras T, Pisano E, Sanna MT, Olinas A, Fanos V, Faa G, Nemolato S, Iavarone F, Castagnola M: **Modifications of the acidic soluble salivary proteome in human children from birth to the age of 48 months investigated by a top-down HPLC–ESI–MS platform**. *Journal of proteomics* 2013, **91**:536-543.
40. Preza D, Thiede B, Olsen I, Grinde B: **The proteome of the human parotid gland secretion in elderly with and without root caries**. *Acta odontologica Scandinavica* 2009, **67**(3):161-169.

41. Julihn A, Soares FC, Hammarfjord U, Hjern A, Dahllöf G: **Birth order is associated with caries development in young children: a register-based cohort study.** *BMC public health* 2020, **20**(1):1-8.
42. Nicolau B, Marcenes W, Bartley M, Sheiham A: **A life course approach to assessing causes of dental caries experience: the relationship between biological, behavioural, socio-economic and psychological conditions and caries in adolescents.** *Caries research* 2003, **37**(5):319-326.
43. Hallett K, O'Rourke P: **Social and behavioural determinants of early childhood caries.** *Australian dental journal* 2003, **48**(1):27-33.
44. Price J: **Parent-child quality time does birth order matter?** *Journal of human resources* 2008, **43**(1):240-265.
45. Rajshekar SA, Laxminarayan N: **Comparison of primary dentition caries experience in pre-term low birth-weight and full-term normal birth-weight children aged one to six years.** *Journal of Indian Society of Pedodontics and Preventive Dentistry* 2011, **29**(2):128.
46. Bernabé E, MacRitchie H, Longbottom C, Pitts NB, Sabbah W: **Birth weight, breastfeeding, maternal smoking and caries trajectories.** *Journal of dental research* 2017, **96**(2):171-178.
47. Athamneh T, Al-Shraideh Z, Malkawi TF, Ersan MY, Ersheidat A: **The association between low birth weight and dental caries of primary teeth among Jordanian children.** *Middle East Journal of Family Medicine* 2014, **12**(3).
48. Rajab LD, Hamdan M: **Early childhood caries and risk factors in Jordan.** *Community dental health* 2002, **19**(4):224-229.

Figures

AUC	0.885	0.990
95% CI	0.7856 to 0.9844	0.9694 to 1.000
p-value	<0.0001	<0.0001
Accuracy (%)	72.5	95.0

Figure 1

The receiver operating characteristic (ROC) curve analysis by statistical analysis using A): Cystatin S, B): Cystatin S & birth weight for discriminating of ECC from CF controls.

Figure 2

Left) Mean value of salivary cystatin S levels in ECC and CF groups based on the age of subjects. Middle) Mean value of salivary cystatin S levels in ECC and CF groups based on the birth weight of subjects. Right) Mean value of salivary cystatin S levels in ECC and CF groups based on the birth order of subjects.

Figure 3

ROC curve by machine learning using A): All features except Cystatin S levels, B): All features.

Figure 4

Feature importance score of extracted features in incidence of ECC using random forest classifier. Features description: A): Cystatin S levels B): Age C): Gender D): Weight of birth E): Birth order F): Father's education level G): Mother's education level H): Father age I): Mother age J): Using Toothbrush K): Flossing L): Using Mouthwash M): Systemic diseases of parents N): Dental visit regularly O): Oral health instrument using P): Number of tooth brushing daily Q): Consumption of sweet snacks R): Consumption of sour snacks S): Consumption of dairy units per day T): Consumption of milk at night during infancy U): Taking vitamin and calcium supplements by the mother during pregnancy V): Parental smoking daily

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)