Prediction of Dental Caries in Pediatric Patients Using Machine Learning versus Traditional Statistical Models: Systematic Review and Meta-Analysis

---

A Thesis
Submitted to
The Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
MASTER'S OF ORAL HEALTH SCIENCES

---

by
Tanvi Shembekar
December 2023

Thesis Approvals:
Examining Committee Members:
Dr. Chukwuebuka Ogwo, Advisory Chair, Department of Oral Health Sciences
Dr. Marisol Tellez, Department of Oral Health Sciences
Dr. Louis DiPede, Restorative Dentistry

ABSTRACT

Objective: The objective of this systematic review was to determine the feasibility and accuracy of machine learning (ML) versus a traditional statistical model in predicting dental caries in children.

Methods: This is a systematic review comparing the feasibility and accuracy of ML versus traditional statistical models in predicting dental caries in children. The eligibility criteria were peer-reviewed studies published in English between January 1, 2007-December 31, 2022, that reported using both traditional statistical models and ML algorithms for caries prediction in primary dentition. The years 2007-2022 was chosen as the range because according to NIH.gov, in 2007, machine learning first started analyzing dental radiographs and images to look for tooth decay. Articles were extracted using search strategies from PubMed, Google Scholar, Embase, and Cochrane Library databases. Articles were screened using PRISMA guidelines, following a review for quality assessment using the JBI Critical Appraisal Checklists. Additionally, a meta-analysis was conducted comparing the studies that used traditional statistical models versus ML models using their pooled area under the curve (AUC) estimates. Only 3 of the 5 studies from each of the model types were analyzed based on the random effect estimate due to a limited number of studies. The meta-analysis was conducted using Med Calc software.

Results: ML-based models that were most successful in predicting dental caries in children were Multilayer Perceptron (MLP) and Random Forest (RF). These algorithms outperformed the traditional statistical model of Logistic Regression (LR) as confirmed in the meta-analysis. However, some LR models outperformed certain ML models such as STVM and SVM and did not have much difference in predictive performance scores compared to ML algorithms such as XGboost. Low-income, frequency of dental visits, and toothache were the most significant factors that predicted caries among many of the studies. Low fluoride exposure, consumption of sugary food/drinks, and tooth brushing frequency were additional significant factors contributing to caries prediction. Our meta analysis showed that the pooled AUC scores of ML and statistical models were 0.808 and 0.776 respectively. Heterogeneity assessment of the 3 studies that used traditional statistical models meta analysis showed no significant heterogeneity while the 3 articles that used ML models showed significant heterogeneity(Higgin's I^2 test = 28% and 91%, respectively). A forest plot showcased the pooled AUC scores and a funnel plot showcased publication bias for each model. The test for publication bias showed that both statistical and ML algorithms had low bias. However, this could be inaccurate due to the limited number of studies.

Conclusion: Machine learning is a highly plausible and successful method for caries prediction. Specifically, MLP and RF exceeded other ML algorithms and LR, in predictive performance. However, LR still outperformed or performed closely to some ML algorithms. Therefore, the best performing algorithms, MLP and RF, could be recommended as more robust and accurate analytical tools for caries predictions compared to LR, but LR also has predictive potential.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Caries, described as tooth decay, is one of the most common oral health conditions treated by dentists (Fleming et al., 2018). Caries can develop from factors such as bacteria, neglecting healthy oral habits like brushing, and frequent consumption of sugar (Fleming et al., 2018). Oral bacteria feed on sugar, producing acid that dissolves the surface of teeth. Teenage-aged pediatric patients in particular are prone to developing caries due to sugary consumption and a lesser understanding of dental hygiene and the repercussions of neglecting it. Additionally, children with caregivers who have low oral health literacy and less involvement in teaching oral hygiene such as tooth brushing become prone to caries development. However, caries can be preventable with early detection using advanced technology. The field of artificial intelligence (AI) is being researched as a tool that can used in the presence of adequate data to predict caries.

Artificial intelligence is a field that uses computers and datasets to interpret, solve, and predict problems based on data (International Business Machines (IBM, 2020)). With the dental field advancing towards digitalization, a method where images or text are processed by a computer, it is important to understand the role and impact of modern technologies such as AI. A branch of AI is machine learning which focuses on creating algorithms to imitate the way humans learn, resulting in accurate and efficient outcomes to help with decision making (IBM, 2020). In regard to the dental field, machine learning can potentially be used as a tool for predicting caries in children and pediatric patients, so that development of caries can be avoided. According to the literature, machine learning is a highly plausible method for caries predictions. In a study from University of North Carolina at Chapel Hill, the application of machine learning for early childhood caries classification was of value, and the algorithm used to compute actual probabilities of caries was accurate compared to observed probabilities (Karhade et al., 2021). Another study from the University of California used machine learning algorithms based on oral health surveys filled out by parents that helped dental providers identify key predictors of dental caries. The development of this algorithm proved to be highly useful for prediction, and therefore prevention, of dental caries among children (Ramos-Gomez et al., 2021).

The classical statistical model (Logistic regression (LR) and ML algorithms were used in a study conducted in Chengdu, China and both showed favorable performances in predicting early childhood caries, identifying high-risk groups, and using preventive treatments (Qu et al., 2022). So far, studies have shown machine learning as a potentially viable tool for predicting dental caries with varying degrees of accuracy. Thus, the objective of this systematic review was to determine the feasibility and accuracy of machine learning versus traditional statistical models in predicting dental caries in children.

CHAPTER 2

METHODS

This was a systematic review, a summary of available primary literature based on a certain topic, of the prediction of dental caries in children using machine learning. The PRISMA flow chart was used to display the process of choosing studies. The following electronic databases were utilized: PubMed, Google Scholar, Embase, and Cochrane Library. The search was limited to studies in English and published between the years 2007-2022 due to the recent nature of machine learning.

MeSH terms used during the search were “caries” OR “dental caries” OR “carious dentin” OR “dental white spot” AND “prediction” AND “machine learning” OR “artificial intelligence” OR “deep learning” OR “supervised learning” OR “neural network” OR “natural language processing” AND “children” or “pediatric patients”. Only studies that used machine learning algorithms to create prediction models using data from children will be considered for this review. The search will be filtered to exclude systematic reviews, case reports, meta-analyses, ex vivo studies, and in vitro studies. Additionally, studies will be excluded if it includes children over the age of 17, and if they are written in a language other than English. Inclusion criteria for this systematic review is studies must be in English, is not limited to any geographic location, and include pediatric patients aged 1-17 as the sample. All literature must include use of a machine learning algorithm to evaluate caries prediction. All literature must consider key factors and predictors of development of caries.

To evaluate the evidence, a PICO (Population, Intervention, Comparison, Outcome) strategy was defined:

Table 1. Population, Intervention, Comparison, Outcome (PICO) Strategy

Population (P)	In pediatric patients without caries
Intervention (I)	how feasible are machine learning algorithms
Comparison (C)	versus traditional statistical models
Outcome (O)	in predicting dental caries?

Table 2. Search Strategy for Selection of Scientific Evidence

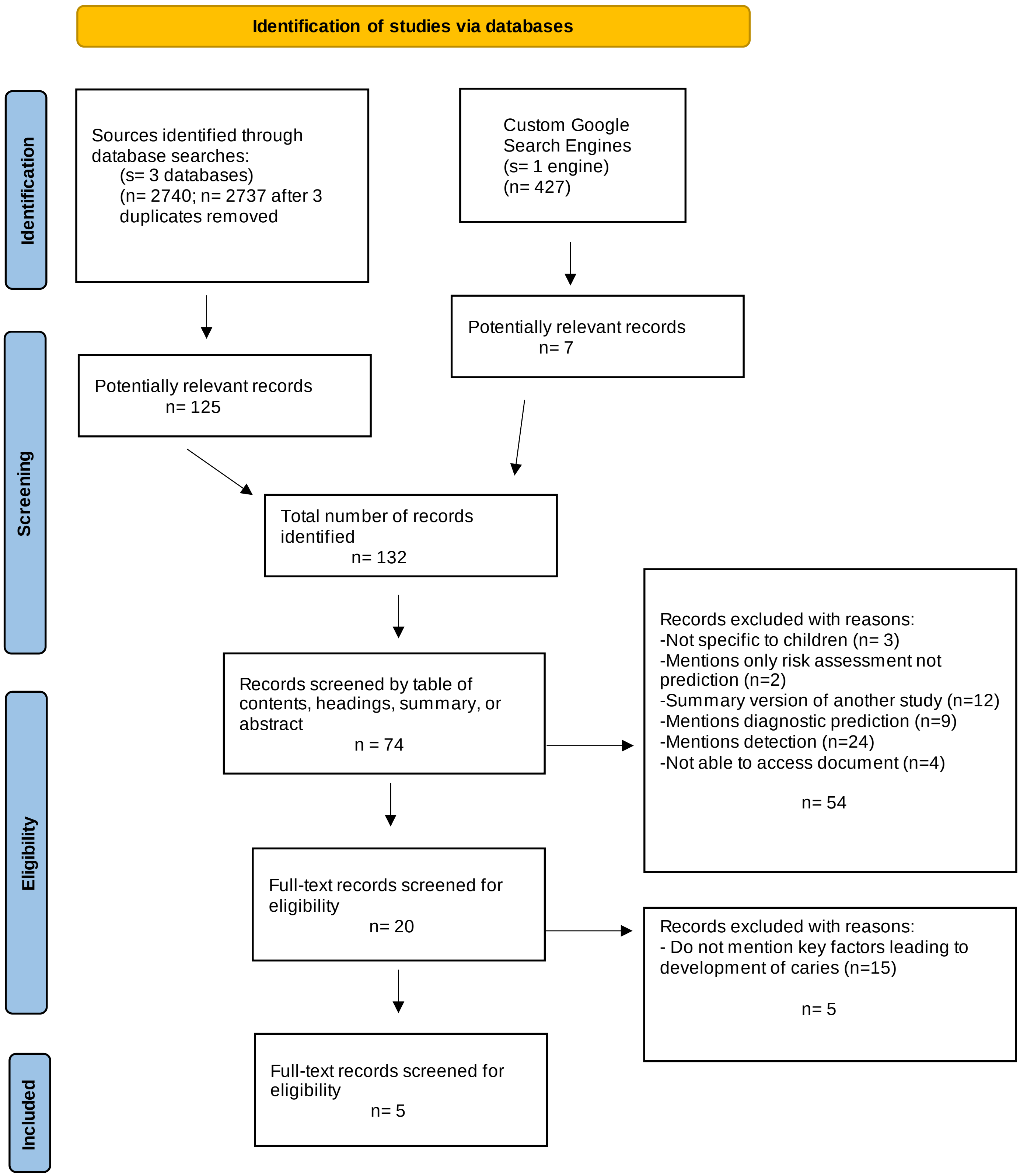
Database and Search Strategy
PubMed
((((((((((((((((((cari*[tiab]) OR (dental cari*[tiab])) OR (dental caries [Mesh])) OR (tooth cari*[tiab])) OR (dental white spot*[tiab])) OR (carious dentin*[tiab])) AND (Machine learn*[tiab])) OR (ML[tiab])) OR (AI[tiab])) OR (artificial intelligence*[tiab])) OR (active learning*[tiab])) OR (neural network*[tiab])) OR (natural language processing*[tiab])) OR (unsupervised learning*[tiab])) OR (supervised learning*[tiab])) OR (deep learning*[tiab])) OR (machine intelligence*[tiab])) AND (predict*[tiab])) AND (children*[tiab])
Embase
((‘dental caries’/exp OR ‘caries’ OR ‘tooth caries’ OR ‘dental white spot’ OR ‘carious dentin’) AND ‘machine learning’/exp OR ‘artificial intelligence’ OR ‘natural language processing’ OR ‘machine intelligence’) AND ‘prediction’/exp AND ‘children’
Cochrane Library
#1 MeSH descriptor: [Dental Caries] explode all trees

Table 2. (continued)

#2	("dental caries"):ti,ab,kw OR ("carious dentin"):ti,ab,kw OR ("dental white spot"):ti,ab,kw OR ("tooth caries"):ti,ab,kw OR ("caries"):ti,ab,kw (Word variations have been searched)
#3	MeSH descriptor: [Machine Learning] explode all trees
#4	("machine learning"):ti,ab,kw OR ("ML"):ti,ab,kw OR ("artificial intelligence"):ti,ab,kw OR ("AI"):ti,ab,kw OR ("deep learning"):ti,ab,kw (Word variations have been searched)
#5	#1 OR #2
#6	"prediction"
#7	("children"):ti,ab,kw OR ("pediatric patients"):ti,ab,kw (Word variations have been searched)
#8	#5 AND #6 AND #7
Google Scholar	
dental caries OR caries OR carious dentin OR dental white spot OR caries AND machine learning OR ML OR AI OR artificial intelligence OR deep learning OR machine intelligence OR natural language processing OR supervised learning AND prediction AND children OR pediatric children.	

Joanna Briggs Institute (JBI) critical appraisal tool was used to assess the quality of each study and extent to which the study addressed possible bias. The risk of bias and external validity of each study was assessed. An extraction table was created to include relevant information such as study characteristics and results from each study. A meta-analysis was conducted to provide a quantitative assessment of the results of the studies.

Figure 1. PRISMA Flow Diagram



Meta-Analyses

The meta-analysis was conducted on 3 out of 5 studies due to their consistency in using the AUC as their performance parameter. The pooled effect estimation, tests for heterogeneity, and publication bias tests were performed for the ML algorithms and LR models and their output tables, forest plot, funnel plots generated. The output table was used to provide a summary of the pooled AUC, fixed and random effects, standard error, confidence interval, and p-value for the ML algorithms and LR models each. The forest plots were used to visualize of the pooled AUC estimates of fixed and random effects. The test for heterogeneity was conducted using the Cochrane's Q, and the Higgins (I^2) statistic was used to validate the result of the Q test. A significant Q test indicated the presence of heterogeneity Publication bias was assessed using two publication bias tests - Egger's and Begg's tests. The funnel plot was then used to show a visual representation of publication bias for both the ML algorithm and logistic regression. All Meta-analyses was conducted using MedCalc software and p-value set at 0.05.

CHAPTER 3

RESULTS

Table 3. Extraction Table

Study Title and Author	Location of Study	Objective	Study Design and Duration	Age of Subjects	Sample Size	Variables	Machine Learning Prediction Models	Outcomes
A Comparative Study of Machine Learning Techniques for Caries Prediction Montenegro et al., (2008)	Department of Computing Systems, Polytechnic School, Pernambuco State University, Brazil	To build models to predict the presence of caries in preschool children and display the factors associated to caries prediction	Cross-sectional	>5 years old	998 preschool-aged children	<u>Independent variables:</u> 1. Gender: male/female 2. Age in months 3. Parent's opinion about oral health of child (excellent, good, regular, bad, very bad) 4. Has the child already had a toothache? (yes/no) 5. Family income (1-7, or more) in minimum wages (yes/no) 6. Child has already gone to the dentist and a caries was diagnosed (yes/no) 7. Child has never gone to the dentist for another reason (yes/no) 8. Child has already gone	1. 4. k-nearest neighbors (kNN) classifier (k= 19) 2. Decision tree C4.5 (C= 0.25) 3. Decision tree C4.5 (C= 0.001) 4. Multilayer perceptron (MLP) neural networks (hidden layer units= 2, learning rate= 0.01, epochs= 500) 5. Support vector machines (SVM) (C= 1, σ= 0.1) 6. LR	**Caries Prediction Results (15 input variables)** <u>10-fold cross-validation error:</u> 1. kNN (k=19): 26.75% 2. C4.5 (C= 0.25): 25.95% 3. C4.5 (C= 0.001): 30.96% 4. MLP: 22.75% 5. SVM (C=1): 23.65% 6. LR= 24.43% <u>Area under the curve (AUC) scores:</u> 1. kNN: 0.8178 2. C4.5 (C= 0.25): 0.7985

Table 3. (continued)

						to the dentist (yes/no) 9. Child has already visited the dentist for having a toothache (yes/no) 10. Presence of failure in the enamel (yes/no) 11. Presence of fistula (yes/no) 12. Political-administrative region (1 to 6) 13. Child has never gone to the dentist for access reason (yes/no) 14. Child has gone to the dentist for prevention reason (yes/no) 15. Child has never gone to the dentist for financial questions (yes.no) <u>Outcome variable:</u> 1. Presence of caries (yes/no)		3. C4.5 (C= 0.001): 0.7193 4. MLP: 0.8452 5. SVM: 0.7635 6. LR= 0.7592 Children aged 23 months were more prone to caries. Family income, children with presence of toothache, and if children already had caries diagnoses were significant influences on development of caries. Overall, in terms of performance criteria, classification error, and AUC value, MLP method followed by kNN, was most successful.
DCP (deep closest point)	Department of Computer and	To identify DMFT and its	Cross-sectional	12 years old	22,287 children	<u>Independent Variables</u>	1.Random forest (RF)	<u>Accuracy of models:</u> 1. RF: 92%

Table 3. (continued)

Prediction of Dental Caries Using Machine Learning in Personalized Medicine Kang et al., (2022)	Electronics Convergence Engineering, Sun Moon University, Korea	related factors and predict dental caries which can be al basis for establishing individual oral prevention strategy			1.1695 (children with presence of caries) 2. 20,593 (children with absence of caries)	1. Presence of dental caries 2.Area of residence of subject 3. Region of residence of subject 4. Gender 5. Previous experience with caries 6. Awareness of oral health 7. Dental treatment experience in past year 8. Teeth brushed before and after lunch, before and after dinner, and after snack 9. Teeth brushed before going to bed 10. Teeth not being brushed 11. Flossing frequency 12. Mouthwash usage 13. Electric toothbrush usage 14. Use of fluoride toothpaste 15. Sticky snacks consumption	2. Artificial neural network (ANN) 3. Convolutional neural network (CNN) 4. Gradient-boosted decision trees (GBDT) 5. Support vector machine (SVM) 6. LR 7. Long and short-term memory (LSTM)	2. ANN: 88% 3. CNN: 87% 4. GBDT: 85% 5. SVM: 83% 6. LR: 82% 7. LSTM: 75% F1 scores (measurement of the model's accuracy on the dataset): 1. RF: 91% 2. ANN: 87% 3. CNN: 87% 4. GBDT: 81% 5. SVM: 79% 6. LR: 78% 7. LSTM: 74% Precision scores (model's ability to identify relevant data points): 1. RF: 94% 2. ANN: 87% 3. CNN: 87% 4. GBDT: 83%

Table 3. (continued)

						16. Pain in gums or bleeding when brushing <u>Outcome Variable</u> 1. Dental caries 17. Pain or discomfort in teeth past year 18. Smoking experience 19. Living with grandfather, grandmother, father, mother, stepfather, stepmother, brother, or sister 20. Socioeconomic status 21. Weekly allowance 22. Tartar build up 23. Tooth speckle		5. SVM: 82% 6. LR: 80% 7. LSTM: 74% <u>Recall scores (measures how correctly a model has identified positive classes in the dataset):</u> 1. RF: 88% 2. ANN: 87% 3. CNN: 87% 4. GBDT: 78% 5. SVM: 76% 6. LR: 76% 7. LSTM: 74% RF was the most successful model in terms of accuracy, F1 score, precision score, and recall score. Significant caries predictors were region of subject's residence, teeth brushed after breakfast,

Table 3. (continued)

								not being able to receive dental treatment, previous experience with caries, and household income.
A New Model for Caries Risk Prediction in Teenagers Using a Machine Learning Algorithm Based on Environmental and Genetic Factors Pang et al., (2021)	Guangdong Provincial Key Laboratory of Stomatology, Department of Preventive Dentistry, Guanghua School of Stomatology, Hospital of Stomatology, Sun Yat-sen University, Guangzhou, China	To construct a new caries risk prediction model (CRPM) that includes genetic and environmental risk factors for caries development in teenagers	Cross-sectional	Teenagers aged 13-14 years	1055 teenagers (cohort 1: 710, cohort 2: 345)	<u>Independent variables</u> 1. Pit and fissure sealant (no, yes) 2. Sex (female, male) 3. Frequency of tooth brushing (<1 times/day, 1 times/day, or 2 times/day) 4. Toothpaste (no, yes) 5. Mouthwash (no, yes) 6. Dental flossing (no, yes) 7. Professional application of fluoride (no, yes) 8. Dental attendance in the past 6 months (no, yes) 9. One-child family (no, yes) 10. Activity (no, yes)	1. CRPM model via random forest (RF) 2. Logistic Regression	<u>Logistic Regression Model AUC Scores</u> Training cohort: 0.70 Testing cohort: 0.74 Average: 0.72 <u>Random Forest Model AUC Scores</u> Training cohort: 0.78 Testing cohort: 0.81 Average: 0.795 <u>Environmental factors significantly associated with caries presence:</u> 1. Gender: p-value: 0.041 2. Dental attendance in past 6 months: p-value: 0.009

Table 3. (continued)

						11. Cariostatic score (low, medium, high) 12. Plaque index (low, medium, high) 13. Residence (urban, rural) 14. Toothpaste (non-fluoride, fluoride) 15. Saliva buffering capability (pH<3.5, pH 3.5-4.24, pH 4.25-4.75, pH>4.75) 16. Dental insurance (no, yes) 17. Caregiver (mother, father, grandparents, nursemaid, no regular caregiver) 18. Education of caregiver (<9 years, greater than or equal to 9 years) 19. Household monthly income in CNY (<3000, 3000-7000, greater than or equal to 7000) 20. Frequency of snacks consuming (<1 per day, greater than or equal to 1 per day)		3. Cariostat score: p-value: <0.001 4. Past caries experience: p-value: <0.001 Genetic factors significantly associated with caries presence: 1.rs3790506 (SNP of TUFTI1 associated with enamel development, mineralization, and interaction with *Streptococcus mutans*): p-value: 0.024 2. rs1996315 (SNP of AQP5 that codes for water channel protein in salivary glands): p-value: 0.0042

Table 3. (continued)

						21. Saliva secretion in ml/min (0.1-0.25, >0.25) 22. Frequency of sweet drinks consuming (<1 per day, greater than or equal to 1 per day) 23. Past caries experience (no, yes) <u>Outcome variables</u> 1. DMFT increment (ΔDMFT) over 21 months of follow-up		
Prediction Models of Early Childhood Caries Based on Machine Learning Algorithms Park et al., (2021)	Department of Biostatistics, Yonsei University Wonju College of Medicine, Korea	To develop prediction models for early childhood caries (ECC) via national survey data and evaluate performance by comparing the ML-based models with a regression model	Cross-sectional	Children aged 1-5 years	4195 children	<u>Variables used from Korea National Health and Nutrition Examination Survey (KHANES) Dataset</u> <u>Independent variables</u> 1. Age of children (1-5) 2. Gender 3. Children with siblings (only child, 1, or greater than or equal to 2) 4. Household 5. Tooth brushing frequency	<u>Variable selection by logistic regression using backward elimination and permutation importance</u> 1. Logistic regression 2. Extreme gradient (XG) boost 3. Random Forest (RF) 5. Light gradient	**Predictive Performance of Prediction Models** <u>AUROC</u> 1. LR: 0.784 2. XG boost: 0.785 3. RF: 0.780 4. Light GBM: 0.774 <u>Accuracy</u> 1. LR: 0.765 2. XG boost: 0.766 3. RF: 0.755

Table 3. (continued)

						6. Education level of the mother 7. Age of the mother at the time of giving birth (<35 or greater than or equal to 35) 8. Use of dental floss or interdental toothbrush of mother 9. Tooth brushing frequency of the mother 10. DMFT of the mother Outcome variables 1. Early childhood caries (ECC) 2. ECC high-risk groups	boosting machine (GBM)	4. Light GBM: 0.764 Sensitivity 1. LR: 0.799 2. XG boost: 0.769 3. RF: 0.759 4. Light GBM: 0.782 Specificity 1. LR: 0.531 2. XG boost: 0.581 3. RF: 0.400 4. Light GBM: 0.546 Misclassification Rate (MCC) 1. LR: 0.258 2. XG boost: 0.148 3. RF: 0.040 4. Light GBM: 0.204 Risk factors of ECC: 1. Children's age ($p<0.001$)) 2. Teeth brushing frequency ($p<0.001$)

Table 3. (continued)

								3. Mother's birthing age (p<0.001) 4. Mother's DMFT (p=0.001)) XGboost outperformed LR in terms of AUC value with a score of 0.785, compared to LR's 0.784. XG boost also performed best in accuracy, with a score of 0.766, compared to LR's 0.765.
Dental Caries Risk Assessment in Children 5 Years Old and Under via Machine Learning Sadegh-Zadeh et al., (2022)	Department of Computing, Staffordshire University, UK	To investigate dental caries risk among children under 5 and find the potential approach(es) to lower the risk of dental caries in high-risk individuals using ML and personal prescriptions	Cross-sectional	Children aged 0-5 years	780 parents and their child	<u>Independent variables</u> 1. Fluoride exposure 2. Sugary food/drinks 3. Regular dental visits 4. Special needs) 5. Chemo/radiotherapy 6. Eating disorders 7. Medications reducing salivary flow 8. Cavitated/non-cavitated 9. Carious lesion (visual,	1.Decision tree (DT) 2. Extreme gradient boosting (EGB) 3. K-nearest neighbours (kNN) 4. Logistic regression (LR) 5. Multilayer perception (MLP) 6. Random forest (RF) 7. Support vector	<u>Accuracy (%) of models:</u> 1. DT: 96.2 2. EGB: 97.4 3. kNN: 96.2 4. LR: 94.9 5. MLP: 97.4 6. RF: 97.4 7. SVM linear: 93.6 8. SVM rbf: 94.9 9. SVM poly: 96.2 10. SVM sigmoid: 96.2

Table 3. (continued)

						radiographically) 10. Teeth extracted due to caries within past 36 months 11. Visible plaque 12. Unusual tooth morphology that causes plaque retention 13. Proximal restorations 14. Dental/orthodontic appliances 15. Parents'/carers' education 16. Parents'/carers' monthly income <u>Outcome variables</u> 1. Dental caries	machine (SVM) kernel= 'linear' 8. Support vector machine (SVM) kernel= 'rbf' 9. Support vector machine (SVM) kernel= 'poly' 10. Support vector machine (SVM) kernel= 'sigmoid'	MLP and RF models had most accurate results, and SVM linear had least accurate results. Consumption of sugary foods/drinks, not attending dental appointments, low socioeconomic status, and low fluoride exposure were significant contributing factors in predicting caries.

Table 4. Critical Appraisal: Cross-Sectional Studies

Study	Montenegro et al., (2008)	Kang et al., (2022)	Pang et al., (2021)	Park et al., (2021)	Sadegh-Zadeh et al., (2022)
Were the criteria for inclusion in the sample clearly defined?	Y	Y	Y	Y	Y
Were the study subjects and the setting described in detail?	Y	N	Y	Y	Y
Was the exposure measured in a valid and reliable way?	Y	Y	Y	Y	Y
Were objective, standard criteria used for measurement of the condition?	Y	Y	Y	Y	Y
Were confounding factors identified?	N	N	Y	Y	Y

Table 4. (continued)

Were strategies to deal with confounding factors stated?	N	N	Y	Y	Y
Were the outcomes measured in a valid and reliable way?	Y	Y	Y	Y	Y
Was appropriate statistical analysis used?	Y	Y	Y	Y	Y

As seen in Table 4, the overall risk of bias for all studies was low. All studies contained the desired inclusion criteria, and reliable and valid ways of measuring outcomes via sensitivity, specificity, AUC/AUROC scores, and cross validation.

The Montenegro et al. (2008)'s study was cross sectional with sample size of 3,864 Brazilian preschool children under the age of 5 years. A total of 4 machine learning algorithms (kNN, decision trees, MLP, SVM), and 1 traditional statistical model (logistic regression) were used in the study. Fifteen independent variables considered as caries contributors, were inputted into the algorithms and traditional model. These variables were answered by the children's parents in a survey, and were related to child's oral health, socioeconomic status, and dental experience. The study used classification error via 10-fold cross-validation error and area under the curve (AUC) scores to determine performance of the algorithms. Results showed that MLP had the best performance with a 10-fold cross validation error of 22.75% and AUC score of 0.8452, compared with LR at 24.43% error and AUC score of 0.7592. The study also mentions that socioeconomic status was the most significant contributor to caries development, making it an important factor in predicting caries.

Kang et al was a cross-sectional study with sample size 22,287 children. The data used in the study was from Korean Center for Disease Control and Prevention, and produced by medical doctors through consultations with their patients. The study used 6 ML algorithms (RF, SVM, GBDT, ANN, LSTM, CNN) and traditional model logistic regression. Forty-five variables related to caries history, calculus information, and oral health will asked in a survey, and results were entered into the algorithms and traditional model. This study utilized precision, recall, and F1 (combination of recall and precision) scores, as well as accuracy via the equation:

$$Accuracy = \frac{TP + FP}{(TN + TP + FN + FP)}$$

to analyze performance of each algorithm. Random forest exhibited an accuracy of 92%, F1 score of 91%, precision score of 94%, and recall score of 88%. Logistic regression exhibited an accuracy of 82%, F1 score of 78%, precision score of 80%, and recall score of 76%. Significant caries predictors were found to be region of residence of subject, teeth brushed after breakfast, not being able to receive treatment, previous experience with caries, and household income. So, Kang et al concluded that random forest (RF) outperformed LR and was the best performing ML algorithm.

Pang et al was a cross-sectional study with 1,055 teenagers aged 13-14 years from urban and rural schools in Foshan, China. Random forest was the only algorithm used in the study, in addition to the traditional statistical model - LR. The survey consisted of 3 categories:

demographic, socioeconomic information, and oral habits. In addition to a questionnaire, the study also utilized single nucleotide polymorphism sequencing analysis to determine presence of caries-related pathway genes in the subjects. The results showed that genetic factors significantly associated with caries presence: were rs3790506 (SNP of TUFTI1 associated with enamel development, mineralization, and interaction with *Streptococcus mutans*) and rs1996315 (SNP of AQP5 that codes for water channel protein in salivary glands). Significant predictive environmental factors were gender, dental attendance in past 6 months, cariostat score, past caries experience. In terms of performance, the AUC score of RF was 0.78, and 0.74 for LR, concluding that RF outperformed LR in predictive performance.

Park et al was a cross-sectional study with 4,195 children aged 1-5. Data was taken from Korea National Health and Nutrition Examination (KNHANES), a surveillance system that produces statistics about health behaviors, and food and nutrient consumption of the Korean population. The survey included demographic variables, oral hygiene, maternal details such as education, birthing age, and oral hygiene, and DMFT and DFT scores conducted by dentists. The study utilized 3 ML algorithms (XGBoost, RF, LightGBM), and LR. The study analyzed performance by calculating area under the receiving operating characteristics (AUROC), 1-accuracy, sensitivity, specificity, and misclassification scores. The AUROC scores of XGboost and LR was 0.785 and 0.784 respectively. Additionally, XGboost had an accuracy of 76.6%, while LR had an accuracy of 76.5%. The study concluded that children's age ($p<0.001$), toothbrushing frequency ($p<0.001$), mother's birthing age ($p<0.001$), and mother's DMFT score ($p=0.001$) were the most significant risk factors to developing ECC. Park et al concluded that XGboost was the most successful ML algorithm, slightly outcompeting LR.

Lastly, Sadegh-Zadeh et al was a cross-sectional study with 780 parents and their child aged 2-5 years. The survey data was collected from a dental clinic that runs frequent examination sessions for their pediatric patients. Seventeen variables were used to assess prediction of caries, and the main survey categories were past caries experience, socioeconomic factors, and biological factors such as diet, sugar intake, and fluoride use. ML models used in the study were decision tree, EGB, XGBoost, KNN, RF, MLP and SVM, and the traditional model LR. In a survey meant to determine high risk factors for caries development, consumption of sugary food/drinks received a (“yes”) response rate of 70.5%. Only half of the responders said they went for regular dental visits. Half of the responders said they had a medium monthly income, while the other 50% were in the high-income category. Lastly, 76.9% of responders answered they do not have fluoride exposure. Both random forest and MLP exhibited an accuracy of 97.4% and LR exhibited accuracy of 94.9%. Sadegh-Zadeh et al concluded that multilayer perceptron (MLP) and random forest had the highest accuracy among all ML algorithms used, and outperformed LR.

All five studies were cross-sectional and assessed caries prediction in children, using various ML algorithms. Pang et al was the only study to also utilize the longitudinal study design to assess caries risk prediction over a 21-month period. The sample size among all five studies different significantly, ranging from 780 to 22,287 children. Pang et al was the only study to use teenage subjects aged 13-14, and Kang et al was the only study to use subjects 12 years old. All studies utilized a questionnaire/survey to assess key predictors of caries prevalence. All studies differed in geographic location in terms of where the study was conducted. The Montenegro et al study was conducted in Brazil, Kang et al and Park et al in Korea, Pang et al in China, and Sadegh-Zadeh et al in the U.K. All studies except for Montenegro et al used RF as one of their

ML algorithms. All studies used LR as a traditional statistical model to be compared against the ML algorithms. Gender was used as an independent variable in all studies except Sadegh-Zadeh et al. Race was used as an independent variable in only one study: Sadegh-Zadeh et al. Montenegro et al, and Park et al, used parent's perception of their child's oral health status as factors. All studies recommend ML-based algorithms for prediction of caries. Methods of analyzing performance of various ML algorithms differed across the studies.

Multilayer perceptron (MLP) neural network performed the best in the Montenegro et al (Area Under the Receiver Operating Characteristics (AUROC) score of 0.8452 and 10-fold cross-validation error of 22.75%) and Sadegh-Zadeh et al (97.4% accuracy)) studies. The RF model performed the best in the Kang et al (92% accuracy), Sadegh-Zadeh et al (97.4% accuracy), and Pang studies (average AUC of 0.76). Study conducted by Pang et al was the only to use genetic markers as a potential contributing factor to caries prediction; this method was analyzed via single nucleotide polymorphism (SNP) sequencing analysis. Park et al was the only study that reported XGBoost as the highest performing ML algorithm with AUROC value of 0.785.

To summarize LR vs ML in predictive performance, the Montenegro et al study concluded that their most successful ML model MLP outperformed LR in 10-cross validation with a score of 22.75%, with LR being 24.43%. This ML model also outperformed LR in terms of AUC scores with 0.8452 for MLP and 0.7592 for LR. Kang et al concluded that their most successful ML model, RF, also outperformed LR in terms of accuracy (RF: 92%, LR: 82%), F1 (RF: 91%, LR: 78%), precision (RF: 94%, LR: 80%), and recall (RF: 88%, LR: 76%) scores. In the Pang et al study, RF outperformed LR in terms of AUC scores, 0.795 and 0.72 respectively. Park et al concluded that their most successful ML model, XGboost, outperformed LR through

AUROC (XGboost: 0.785, LR: 0.784), accuracy (XGboost: 0.766, LR: 0.765), sensitivity (XGboost: 0.769, LR: 0.799), specificity (XGboost: 0.581, LR: 0.531), and misclassification (XGboost: 0.148, LR: 0.258) scores. Lastly, Sadegh-Zadeh et al concluded that both MLP and RF had the same accuracy scores and outperformed LR with scores of 97.4% for MLP and RF, and 94.9% for LR.

Although most ML models and especially the best performing ML model outperformed LR, LR still outperformed other ML models such as LSTM and SVM. In Pang et al, the best performing ML model, XG boost, had an AUROC score of 0.785, whereas LR exhibited a score of 0.784. In terms of accuracy, XG boost was 0.766 and LR was 0.765. In this example, it is evident that XG boost barely exceeded LR in both AUROC and accuracy. In Montenegro et al, although most ML models exceeded LR in performance, and MLP had a significantly higher score (0.8452) than LR (0.7592), the C4.5 (c=0.001) decision tree algorithm underperformed (0.7193) compared to LR. Therefore, it can be argued LR can be similar to certain ML models such as LSTM and SVM in predictive performance, but will not outperform highly successful models such as MLP and RF. Thus, if ML algorithms such as MLP and RF were implemented in a clinical setting, they would exhibit more success in predicting dental caries than LR.

Many different variables were found to be associated with caries risk prediction. Montenegro et al, Park et al ($p= 0.014$), and Sadegh-Zadeh conclude that low-income is a significant contributing factor to caries development, and Kang et al reference other studies citing the same. Montenegro et al cites toothache as a significant contributing factor. In regard to frequency of dental visits, Pang et al ($p=0.009$) and Sadegh-Zadeh et al mention less frequent attendance can result in caries. Park et al ($p<0.001$) concluded parent’s age as a significant factor. Park et al showcases that if mother’s birthing age is >35 compared to <35, likelihood of child having higher DFT increases ($p<0.001$). Additionally, likelihood of caries increased if mother's DMFT score increased ($p=0.001$). Park et al also includes toothbrushing frequency as an important factor ($p<0.001$). Montenegro et al conclude that 23-month-old children are more prone to caries compared to other age groups in their sample (<5 years old). Pang et al found that past caries experience ($p<0.001$), plaque index ($p=0.057$), and cariostatic score ($p<0.001$) were caries prone. In regard to genetic factors analyzed in this study, it was found that SNP rs3790506 ($p=0.024$), associated with enamel development, mineralization, and interaction with *Streptococcus mutans*, was a significant factor. Additionally, SNP rs1996315 ($p=0.042$), which codes for water channel proteins in salivary glands, was the other genetic factor affecting caries. Sadegh-Zadeh et al found that consumption of sugary food/drinks was answered as yes by 70.51% of survey participants. Low fluoride exposure was common in 76.92% of survey participants.

CHAPTER 4

META-ANALYSIS

Table 5. Meta-Analysis Summary: Best Performing Machine Learning Algorithms in 3 Studies

Study	ROC Area	Standard Error	95% CI	Z	P	Weight (%)	
						Fixed	Random
Montenegro et al (2008)	0.845	0.011	0.823 to 0.867			21.12	34.31
Pang et al (2021)	0.795	0.018	0.760 to 0.830			7.89	26.91
Park et al (2021)	0.785	0.006	0.773 to 0.797			70.99	38.78
Total fixed effects	0.798	0.005	0.789 to 0.808	157.90	<0.001	100.00	100.00
Total random effects	0.808	0.016	0.777 to 0.839	50.81	<0.001	100.00	100.00

The total fixed effects value represents the average effect of ML models on prediction of caries among all studies, while the total random effects value captures uncertainty resulting from heterogeneity among all studies. Due to the limited number of studies, the random effects model was favored over fixed effects. The p-value is statistically significant ($p < 0.001$), indicating the presence of heterogeneity among all studies. The confidence interval for total random effects is

0.777-0.839 with a pooled AUC score of 0.808, which can be seen below in Figure 2. This high AUC score indicates ML algorithms being successful in predicting caries.

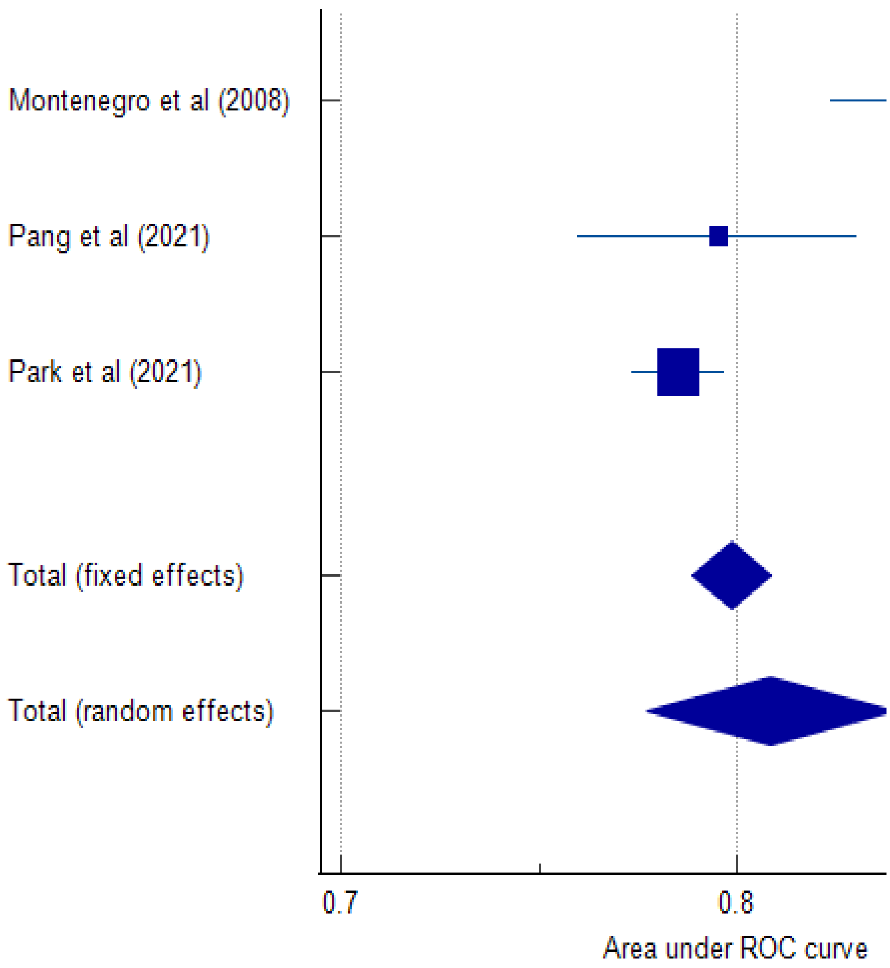


Figure 2. Forest Plot: AUC values of the Best Performing ML Algorithms in 3 Studies

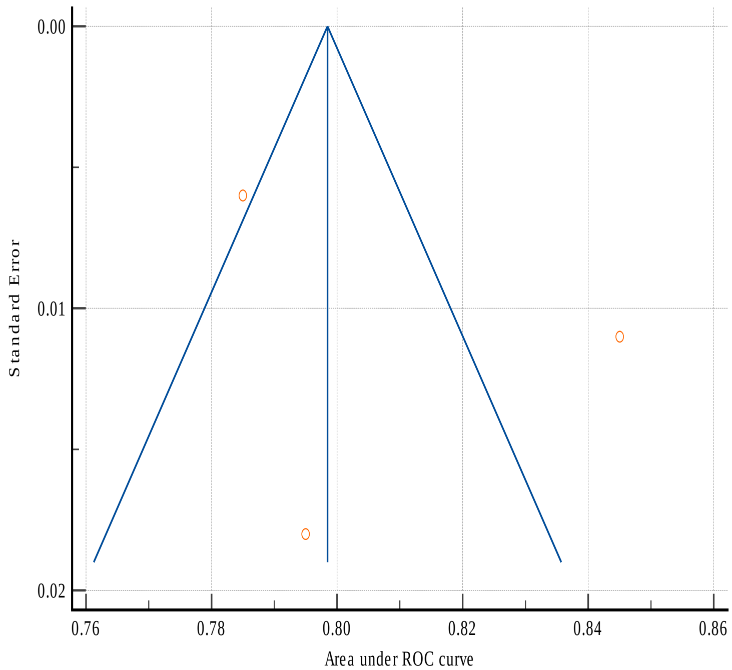


Figure 3. Funnel Plot showing publication bias of ML Algorithms

Table 6. Publication Bias Test: ML Algorithms i3 Studies

Egger's test	
Intercept	3.6324
95% CI	-69.6810 to 76.9457
Significance level	P = 0.6423
Begg's Test	
Kendall's Tau	0.3333
Significance level	P = 0.6015

Figure 3 is a funnel plot showing the publication bias of ML algorithms and their corresponding Egger's and Begg's test statistics shown in Table 6. The dots on the funnel plot represent each article included in the meta-analysis. The funnel plot shows that the studies are relatively evenly distributed across the funnel (divided by the vertical line), therefore there is likely no publication bias. This is confirmed by the Egger's and Begg's tests resulted in p-values that were not statistically significant (0.6423 and 0.6015 respectively). This indicates that the studies have a low chance of publication bias. However, the sample size at 3 studies, is very small.

Table 7. Test for Heterogeneity of the three studies where machine learning algorithm performed the best.

Q	22.9701
DF	2
Significance Level	$P < 0.0001$
I^2 (inconsistency)	91.29%
95% CI for I^2	77.53-96.63

Cochrane's Q test is significant as confirmed by the p-value (< 0.0001) and the I^2 statistic. The I^2 statistic is the percentage of variation across studies due to heterogeneity rather than chance and is high at 91.29%. Therefore, both the p-value and I^2 statistic are suggesting presence of heterogeneity among the ML algorithms in all 3 studies.

Table 8. Meta-Analysis Summary of Logistic Regression Performance in 3 Studies

Study	ROC Area	Standard Error	95% CI	Z	P	Weight (%)	
						Fixed	Random
Montenegro et al (2008)	0.759	0.0140	0.732 to 0.786			15.51	30.55
Pang et al (2021)	0.720	0.202	0.324 to 1.000			0.074	0.21
Park et al (2021)	0.784	0.00600	0.772 to 0.796			84.42	69.24
Total fixed effects	0.780	0.00551	0.769 to 0.791	141.502	<0.001	100.00	100.00
Total random effects	0.776	0.00935	0.758 to 0.795	82.999	<0.001	100.00	100.00

Similar to the ML algorithms analysis in Table 1, the random effects model is favored over fixed effects due to a limited number of studies. The confidence interval for total random effects is 0.758-0.795 with a pooled AUC score of 0.776, which can be seen below in Figure 3. This is a relatively high AUC score, indicating LR models being successful in predicting caries. However, in comparison to the ML algorithms pooled AUC score as seen in Table 1, LR performed lower.

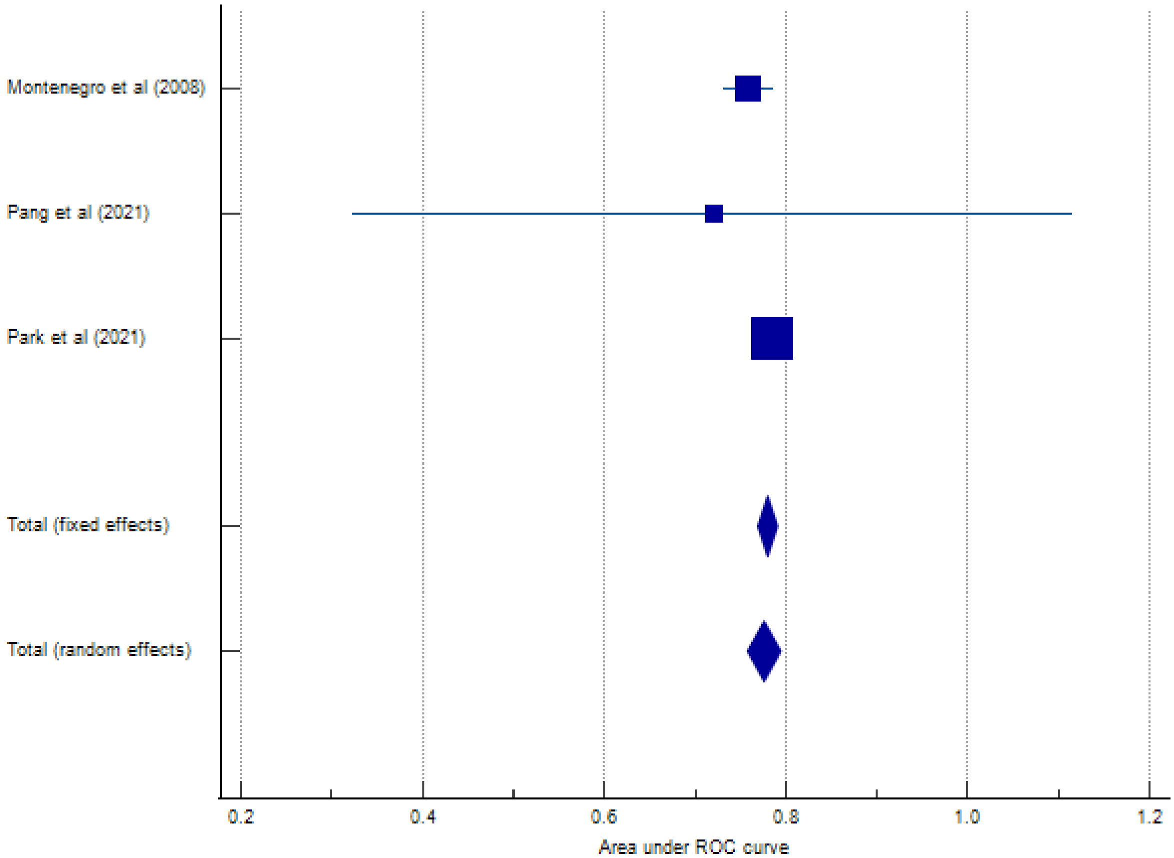


Figure 4. Forest Plot of AUC values of Logistic Regression in 3 Studies

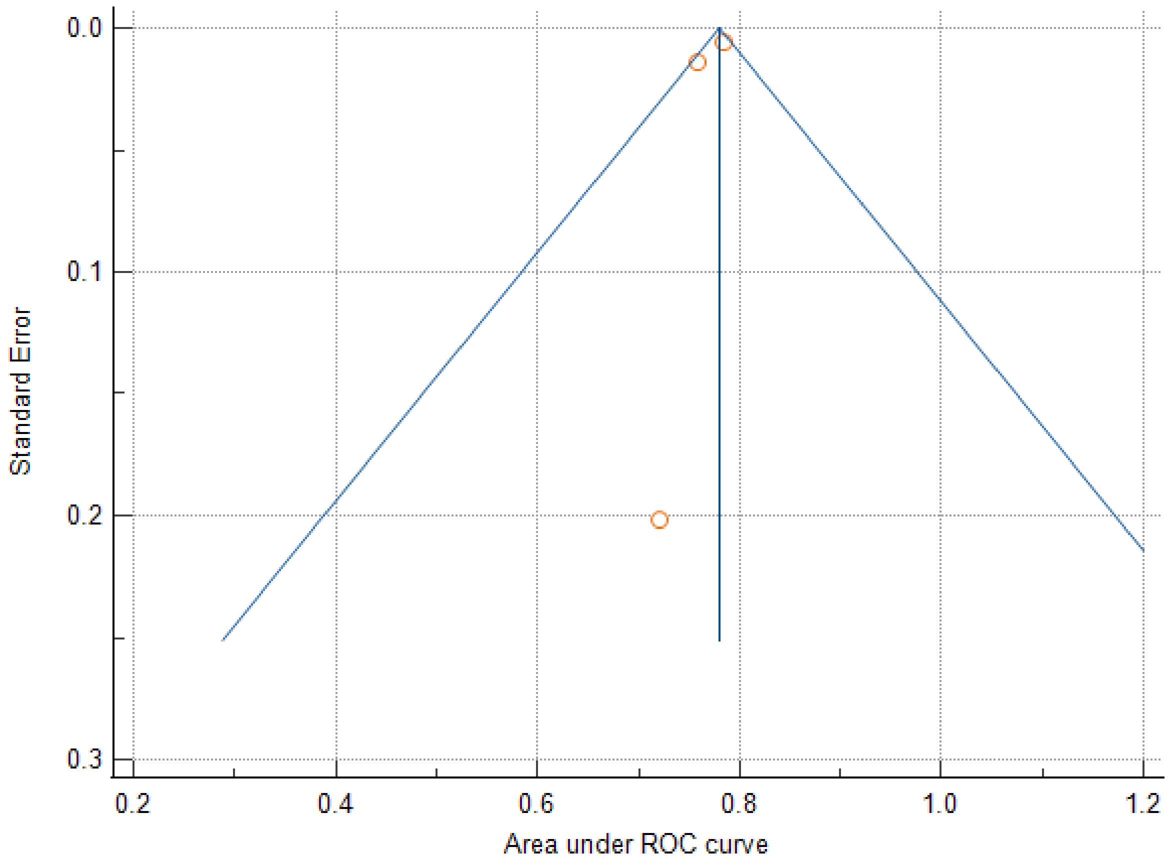


Figure 5. Funnel Plot LR Showing Publication Bias of LR Models

Table 9. Publication Bias Test

Egger's Test	
Intercept	-0.9536
95% CI	-15.9893 to 14.0822
Significance level	P = 0.5682
Begg's Test	
Kendall's Tau	-0.3333
Significance level	P = 0.6015

Egger's test and Begg's test have p-values of 0.5682 and 0.6015 respectively, which are statistically not significant. This indicates the studies have low chance of publication bias. The funnel plot is relatively symmetrical, therefore there is likely no publication bias. However, the sample size is very small.

Table 10. Test for Heterogeneity

Q	2.7825
DF	2
Significance level	P = 0.2488
I^2 (inconsistency)	28.12%
95% CI for I^2	0.00 to 97.59

Cochrane's Q test is not significant as confirmed by a p-value of 0.2488, indicating homogeneity among all studies. The I^2 value at 28.12% is within the 0.00 to 97.59 CI range, but is low, so it is another parameter confirming homogeneity among the LR models in all 3 studies.

Table 11. Summary of Meta-Analysis

Analysis	**Machine Learning Algorithms**	**Statistical Models (LR)**
Pooled AUC score (based on random effects)	0.808 ($p < 0.001$)	0.776 ($p < 0.001$)
Confidence Interval	0.777 to 0.839	0.758 to 9.795
Publication Bias (Egger's test and Begg's test, respectively)	No (0.6423 and 0.6015)	No (0.5682 and 0.6015)

Table 11. (continued)

Cochrane's Q Test for Heterogeneity	Significant ($p < 0.0001$)	Not significant ($p = 0.2488$)
Higgin's I^2 test of Heterogeneity	91.29%	28.12%
Publication bias	No	No
Sample size	3 studies	3 studies

The ML algorithms performed better than LR with a pooled AUC score of 0.808, compared to 0.776. The test for heterogeneity resulted in $P < 0.0001$ for the ML algorithms, and $P = 0.2488$ for LR, and an I^2 of 91.29% and 28.12% for ML and LR respectively. Both the p-value and I^2 value show indications that the ML algorithms showed heterogeneity, while statistical models showed homogeneity. Both ML and statistical models showed no publication bias due to p-values that were not significant.

In the meta-analysis, it was concluded that ML algorithms performed better than the statistical models (LR) according to the total pooled AUC scores among 3 of the studies (0.808 and 0.776 respectively). Only 3 studies were utilized in the meta-analysis because they had a consistent way of measuring predictive performance using AUC scores, while the remaining 2 articles did not include AUC. These AUC scores were showcased in the forest plots in Figures 1 and 3. However, the ML algorithms showed high heterogeneity ($p < 0.0001$) compared to the LR models who showed homogeneity ($p = 2488$), according to the test for heterogeneity conducted on both ML and LR models. This means the findings were not generalizable for the ML algorithms, and no definitive conclusion can be applied to predictive performance in all ML algorithms. Conversely, the LR models showed consistent, and therefore reliable results. Both ML and LR models showed low chances of publication bias due to p-values that were statistically not significant, according to the publication tests which included Begg's and Egger's tests.

CHAPTER 5

DISCUSSION

This systematic review was aimed at assessing predictive factors of dental caries in pediatric patients using ML algorithms and statistical models. The prevalence of caries in children and the growing integration of technology such as AI in the dental field prompted this systematic review. A comparison of ML and LR was aimed to recognize the possibility of ML being utilized to predict caries in a clinical setting. Additionally, a meta-analysis using MedCalc software of LR and ML based on AUC was conducted to compare performance of each model. Five studies assessed prediction via parameters such as AUC and accuracy scores. All studies were similar in that they were cross-sectional, included key predictors of caries in children, used LR as their statistical model, and utilized a questionnaire to assess significant factors contributing to caries development. Studies differed in geographic location of study conducted. Sample size of the studies significantly varied with the smallest sample being 780 and the largest sample being 22,287. All studies differed in methods used to identify key predictors in caries development.

Additionally, there were many variables that were found to signify caries in children, and thus predict caries. Among many of the studies, low-income, frequency of dental visits, and toothache were the most significant factors that predicted caries. Low fluoride exposure, consumption of sugary food/drinks and toothbrushing frequency were additional significant factors contributing to caries prediction. Gender and age of children were not considered significant variables.

The findings of this systematic review are consistent with other studies (Ramos-Gomez et al., 2021; Kang et al., 2022). For example, a study that collected data from pediatric patients in

Los Angeles through a survey, concluded that machine learning algorithms based on oral health surveys can aid dental providers in determining key predictors of caries in pediatric patients (Ramos-Gomez et al., 2021). The study utilized only RF as their ML algorithm without comparing its performance against a traditional statistical model, however, it proved that RF is a successful tool in predicting caries with an accuracy of 71%. This coincides with the conclusion in this systematic review that RF was a consistently successful algorithm among many of the studies. Additionally, a study conducted by Karhade et al., 2021 concluded that an automated machine learning classifier was successful in predicting caries in 3-5-year-old children. This study also did not include a statistical model for comparison, but found that their ML algorithm produced a relatively high AUC of 0.74, and considered ML to be a valuable tool for caries prediction (Karhade et al., 2021).

A strength of this study is that it is the first systematic review and meta-analysis comparing the effectiveness of ML algorithms vs traditional statistical models in prediction of caries in children. It is a feasible study in that the prediction of caries can be made given good AUC scores in ML algorithms. Additionally, most studies were similar in concluding the variables that are key predictors of caries. This narrows the factors that have a high likelihood of causing caries, making it easier for providers to communicate to their patients how to avoid caries. Also, studies chosen for the review had low risk of bias according to the critical appraisal.

One of the limitations of this study was that none of the studies had a consistent way to measure the accuracy of the ML algorithms. For instance, some studies used specificity and sensitivity scores to analyze performance, while others did not. Additionally, each study tested different sets of ML algorithms. For example, if Montenegro et al included RF as an ML algorithm, it possibly could have outperformed the most successful algorithm of their study:

MLP. Furthermore, survey variables were similar in nature, but not exactly the same. For example, some studies included fluoride exposure, and some did not. Only one study included race as a variable. In this way, the studies were inconsistent. Additionally, the meta-analysis was conducted using a very small sample size (3 studies). Because of this, it is difficult to generalize the results and reach a definitive conclusion and there is a very high chance of publication bias as shown in our funnel plot.

CHAPTER 6

CONCLUSION

In summary, this systematic review analysed five studies that have used ML algorithms to predict dental caries in children. With caries being such a prevalent health occurrence in children, this review aimed to understand the feasibility of ML in predicting caries to avoid their development. Multiple predictors of caries were identified, and different ML algorithms proved to be successful in predicting caries, in particular, multilayer perceptron and random forest. The most common and significant predictors of caries among all studies were low-income, frequency of dental visits, and toothache, followed by low fluoride exposure, consumption of sugary food/drinks, and tooth brushing frequency. Our meta-analysis showed that when applied within the ML algorithm space, the traditional ML models out-performed the traditional statistical model in the prediction of caries outcome in children. However, ML algorithms showed presence of heterogeneity among the studies, whereas the LR models showed homogeneity.

Overall, ML is a plausible and successful method for caries prediction. Specifically, MLP and RF exceeded other ML algorithms and the statistical model LR, in accuracy of prediction. Therefore, ML algorithms are recommended as a robust and accurate analytical tool for caries predictions from clinical data. However, the traditional statistical model LR can be similarly effective at predicting caries.

REFERENCES

Centers for Disease Control and Prevention. (2018). *Products - data briefs - number 307 - April 2018*. Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/products/databriefs/db307.html

Fleming Eleanor & Afful Joseph (2018). *Prevalence of Total and Untreated Dental Caries Among Youth: United States*. Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/products/databriefs/db307.htm

International Business Machines (IBM). (2020). *What is Artificial Intelligence (AI)?* https://www.ibm.com/topics/artificial-intelligence

Institute of Electrical and Electronics Engineers (IEEE). (2008). *A Comparative Study of Machine Learning Techniques for Caries Prediction*. IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/4669812

Joanna Briggs Institute (JBI). (2023). *Critical Appraisal Tools | JBI*. (n.d.). https://jbi.global/critical-appraisal-tools

Kang, I. A., Njimbouom, S. N., Lee, K. O., & Kim, J. D. (2022). *DCP: Prediction of Dental Caries Using Machine Learning in Personalized Medicine*. Applied Sciences; Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/app12063043

Karhade, D., Roach, J., Shrestha, P. et al. (2021). *An Automated Machine Learning Classifier for Early Childhood Caries*. PubMed. https://pubmed.ncbi.nlm.nih.gov/34172112/

Kumar, A., Bhadauria, H., & Singh, A. (2021). *Descriptive analysis of dental X-ray images using various practical methods: A review*. Peer J; Peer J, Inc. https://doi.org/10.7717/peerj-cs.620

MedCalc. (2020). Statistical Software version 19.2.6 (MedCalc Software bv, Ostend, Belgium; https://www.medcalc.org

Pang, L., Wang, K., Tao, Y., Zhi, Q., Zhang, J., & Lin, H. (2021). *A New Model for Caries Risk Prediction in Teenagers Using a Machine Learning Algorithm Based on Environmental and Genetic Factors*. Frontiers in Genetics; Frontiers Media. https://doi.org/10.3389/fgene.2021.636867

Park, Y. H., Kim, S. H., & Choi, Y. Y. (2021). *Prediction Models of Early Childhood Caries Based on Machine Learning Algorithms*. International Journal of Environmental Research and Public Health; Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/ijerph18168613

Qu, X., Zhang, C., Houser, S. H., Zhang, J., Zou, J., Zhang, W., & Zhang, Q. (2022). *Prediction model for early childhood caries risk based on behavioral determinants using a machine learning algorithm*. Computer Methods and Programs in Biomedicine; Elsevier BV. https://doi.org/10.1016/j.cmpb.2022.107221

Ramos-Gomez, F., Marcus, M., Maida, C. A., Wang, Y., Kinsler, J. J., Xiong, D., Lee, S. Y., Hays, R. D., Shen, J., Crall, J. J., & Liu, H. (2021). *Using a Machine Learning Algorithm to Predict the Likelihood of Presence of Dental Caries among Children Aged 2 to 7*.

Dentistry Journal; Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/dj9120141

Sadegh-Zadeh, S. A., Qeranqayeh, A. R., Benkhalifa, E., Dyke, D. B., Taylor, L., & Bagheri, M. (2022). *Dental Caries Risk Assessment in Children 5 Years Old and under via Machine Learning*. Dentistry Journal; Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/dj10090164