Malaysian Journal of Public Health Medicine 2023, Vol. 23 (2):187-197

ORIGINAL ARTICLE

# PREDICTION CLASSIFICATION AND MODELLING USING DECISION TREE WITH ORDERED REGRESSION AND ITS APPLICATION TO SOCIO-BEHAVIORAL FACTORS ASSOCIATED WITH TOOTHBRUSHING FREQUENCY IN CHILDREN

**Hazik Bin Shahzad[1,4], Wan Muhamad Amir W Ahmad*[2], Mohamad Nasarudin Adnan[2] and Anas Imran Arshad[3,5]**

*[1]Dental Public Health Unit, School of Dental Sciences, Universiti Sains Malaysia, Health Campus, Kota Bharu, Kelantan, Malaysia*
*[2]Department of Biostatistics, School of Dental Sciences, Universiti Sains Malaysia, Health Campus, Kota Bharu, Kelantan, Malaysia*
*[3]Paediatric Dentistry Unit, School of Dental Sciences, Universiti Sains Malaysia, Health Campus, Kota Bharu, Kelantan, Malaysia*
*[4]Department of Community and Preventive Dentistry, Rashid Latif Dental College, Lahore, Pakistan*
*[5]Department of Paediatric Dentistry, Rashid Latif Dental College, Lahore, Pakistan*

***Corresponding author: Muhamad Amir W Ahmad**
**Emai: wmamir@usm.my**

## ABSTRACT

*Toothbrushing is considered the best self-care behavior for the prevention of oral diseases. Brushing teeth twice a day is considered the social norm, but the development of such habits is dependent on psychosocial, economic, and environmental factors. Recognizing the significance of statistical modeling in medical sciences, this study will use decision trees and ordinal regression to predict frequency of toothbrushing in children. The methodology will be harmonized in the R syntax. The study illustrated the development of the method using 527 observations from WHO oral health questionnaire for children. Before regression analysis, the clinical relevance and significance of each of the 28 variables will be assessed using decision tree analysis and tested for accuracy. The classification obtained will be used as an input for the ordinal regression modeling. According to decision tree analysis, smoking, maternal education, dietary habits, history of toothache and self-rated tooth health contributed significantly to the children's overall toothbrushing frequency. These six variables were used as input for ordinal regression analysis and the developed syntax was used to assess the goodness of fit for the model. Our proposed method achieves the highest level of forecasting precision possible. The process is an alternate to ordinal regression modelling as the selection of appropriate variables is based on computational analysis, forecasting the importance of the independent variables chosen for the final model. This process demonstrates the possibility of developing prediction models which can then be used to formulate clinical hypothesis and inform future researchers.*

***Keywords***: *Toothbrushing, decision tree analysis, bootstrap, ordinal logistic regression, R*

## INTRODUCTION

Globally, dental diseases have the highest prevalence among all other diseases. According to the Global Burden of Disease study in 2019, an estimated 40% of the world population has some form of oral disease with caries in the permanent dentition being the most prevalent[1]. Developed countries have shown improvements in oral health, but many populations in low and middle income countries, are still at high risk for oral diseases including caries and periodontal diseases[2]. Dental diseases cause a lot of discomfort and distress.[3] Poor oral health impacts the ability to eat, sleep, speak and interact with others[4]. Treatment is usually costly and lengthy, requiring leave from work or school, creating an imbalance in work life or educational attainment[5]. In Pakistan, around 90% of oral diseases remain untreated[6]. According to the 2004 WHO report, the mean DMFT (Decayed Missing and Filled Teeth) score for children 12-15 years old was 1.38[7].

Maintaining good oral hygiene is of the utmost importance for prevention of oral diseases such as caries and periodontitis[8]. Oral diseases are multifactorial depending on socioeconomic factors and preventive practices[9, 10]. An individual's daily life activities play an important part in prevention of oral diseases, emphasizing the relevance of dental health education in daily life[11]. Oral diseases are preventable, and the easiest way to maintain good oral hygiene is regular toothbrushing, if possible with fluoridated toothpaste[12] and lifestyle habits of low sugar diet, and routine dental visits[13]. The maintenance of good oral hygiene can be considered regulated and predictable[14]. Toothbrushing is the most important selfcare activity for maintenance of oral health[15, 16]. Toothbrushing behavior is similar to other health promoting behaviors such as eating healthy or being physically active[17]. The motor skills required are developed at an early age[18]. Frequency of brushing and its quality are the most important elements of toothbrushing[14]. Evidence shows that twice a day brushing prevents oral diseases[15]. According to oral health guidance for parents in the United Kingdom, supervised brushing until at least seven years of age, twice a day with fluoridated toothpaste is recommended and should begin as soon as the first tooth erupts[19]. Low frequency of brushing, later age of commencement of brushing and lack of parental involvement are shown to be strongly associated with oral diseases[20]. Oral health behaviors are influenced by psychosocial, economic and environmental factors[21, 22]. These include a broad context of life style habits such as diet and activity[23]. Also included are parental factors such as the educational role and extent of parents[24]. Socialization in the form of physical and social contact teaches norms of behavior to a child. This in turn creates an image of self during pre-adolescence and is an ideal time for creating interest in healthy habits[25]. Ensuring healthy habits is both beneficial to the individual and the standardized functionality of the society[26].

It is essential to focus on the factors associated with the frequency of toothbrushing in children. The influencing factors will provide alternate approaches targeted towards adolescents to promote oral health maintenance. It can only be done by finding the exact health related behaviors that influence frequency of toothbrushing. No study exploring all these factors simultaneously has been conducted to identify tooth brushing frequency and its association with a wide range of socioeconomic and behavioral characteristics. The present study should therefore be helpful for the implementation of a more targeted, appropriate, and efficient prevention program for toothbrushing. The current study will create the ordinal regression model using the result obtained from decision tree analysis using a machine learning algorithm to explore frequency of toothbrushing with behavioral and environmental factors. The decision tree isolates specific levels of individuals' perceptions and attitudes and reveals how they convert to different frequencies of toothbrushing. The other objective of this research is to develop the R syntax for ordinal logistic regression modeling, considering decision tree analysis. In the future, it is anticipated that this programming will facilitate optimal decision-making outcomes for the researcher and contribute to dental health education towards school children and adolescents.

## METHOD

### Study design

This study was a secondary data analysis from a school-based survey in Lahore, Pakistan, administering the standardized WHO oral health questionnaire for children[27]. It collected 527 observations on oral health, social factors, environmental factors and behaviors of children aged 10 to 18. Data was collected via online distribution of the questionnaire in a non-probability convenience sampling method[28].

### Tooth Brushing

This dependent variable was recorded as an ordinal variable with three modalities: “Brushing less than once a day”, “Brushing once a day” or “Brushing more than once a day.” This cutoff was chosen to match the international recommendations for tooth brushing[29].

### Explanatory variables

The following socio-demographic and behavioral data were available along with toothbrushing; gender, age, self-rated tooth and gum health, toothache or discomfort in last 12 months, visits to the dentist in past 12 months, reason for last visit to dentist, object used to clean teeth, use of fluoridated toothpaste, daily life problems associated with teeth (satisfaction with appearance of teeth, other children making fun of teeth, missed school due to toothache, difficulty in chewing), eating habits (fruits, biscuits, fizzy drinks), use of tobacco products and parental education.

### Statistical analyses

Figure 1 shows a step-by-step process of the statistical analysis.

### Decision Tree

Decision trees are useful method for prediction classification, and facilitate sequential decision problems[30]. In the medical field, the decision makers encounter chronological choices, depending on chance and varying in outcomes. Decision trees are the best way to graphically show this kind of information. Decision trees are used for predictive analysis in which the model is trained based on some dataset and then used for predictive purposes on other data of the same kind[31]. This means that the decision tree model can now predict new or unseen data that would estimate which class the unseen data might belong to[32]. This has helped medical professionals in making better decisions while also assisting in how decisions can be made and what could happen. Like an ordinary tree, decision trees consist of nodes and branches. There are three types of nodes: (a) decision node, (b) chance nodes, and (c) terminal (leaf) nodes.

### Conditional Inference Tree

Conditional inference trees, or recursive binary splitting decision tree was first proposed by Hothorn et al.[33]. Conditional inference decision tree modelling is an effortlessly executed and deductive statistical method that integrates complex data well. It examines the relationship of a single variable with potentially multiple explanatory variables[34]. It is superior to conventional linear regression methods because it does not rely on underlying assumptions of linearity and address the concerns of overfitting the sample. It employs formal statistical inferences at every stage of the splitting process and separates variable selection from the process.

There are three steps involved; the first is variable selection, the second chooses the splitting methodology and the third is the continuous repetition of the previous two steps. The variable to split is determined by measuring the association for every variable against the outcome. After selecting the splitting variable, the best point of split is calculated. If the global null hypothesis is rejected, then the node is split for the variable with the strongest relationship to the outcome. Accuracy is the most suitable measure for the performance evaluation of decision trees. Estimating accuracy is important for several reasons, such as verifying if a model is reliable for future predictions[35]. The accuracy of a classifier or predictor is normally estimated with the help of a confusion matrix, which is a useful tool for analyzing how well the classifier can recognize the different classes[36, 37]. All 28 parameters of the WHO oral health questionnaire for children were tested for the decision tree classification.

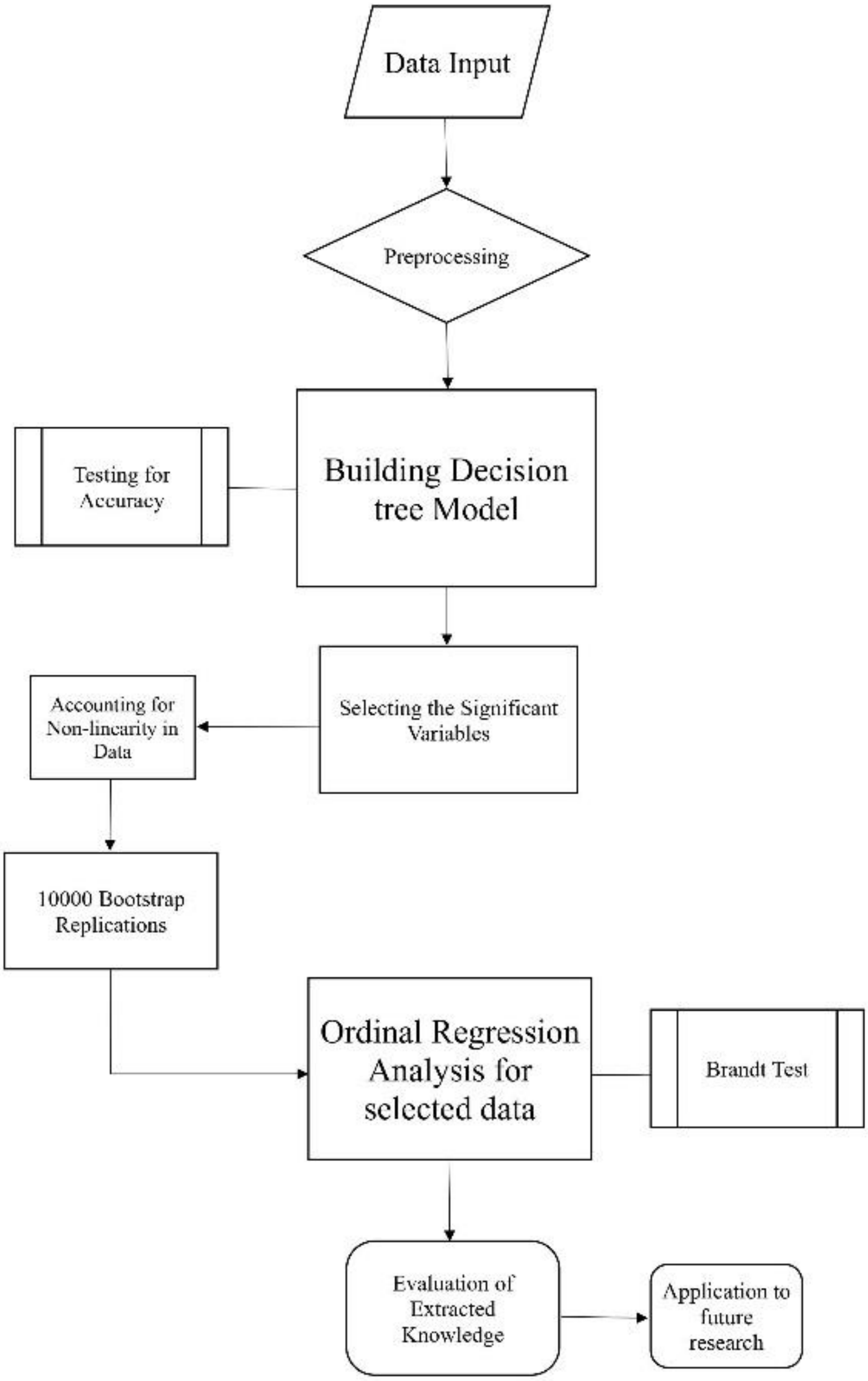


**Figure 1: Diagram showing the different steps of the utilized approach.**

**Bootstrap**

The bootstrap method was developed by Efron in 1979 using a resampling approach[38]. The bootstrap procedure begins with the original sample drawn from the population of interest. The next step is replicating the original sample multiple times to create a new population while keeping the old one in mind. The bootstrap selects many samples that are then replaced by a random sampling technique, resulting in a fresh sample from the beginning. Random sampling with substitution produces a pseudo population that is not identical to the original sample. The bootstrap computes statistics for each sample as the sample is drawn using replacements.

**Ordinal Regression**

Ordinal regression, also called ordinal classification, is a type of regression analysis used for predicting a variable whose value exists on an arbitrary scale where only the relative ordering between different values shows a natural order between labels[39]. In this study, toothbrushing frequency is on an ordinal scale, and we will use the decision tree classification to fit the ordinal regression. The decision tree classification gave six significant variables: smoking, toothache, mothers education, self-rated tooth health (SRTH) fresh fruit consumption and coffee with sugar consumption. As toothache had two distinct levels of split in the decision tree, it had a non-linear effect on the final odds ratios. Therefore, toothache was split for regression analysis as the likelihood ratio was p=2.2e-16. The maximum likelihood method will be used to estimate the regression parameter's value. Brant test showed that the assumption for parallel regression is correct. The model for ordinal for each cutoff is given by "$\text{Logit}(P(Y\leq 1) = \varepsilon + x_i \beta_i$"[40].

Where: ε = Threshold or intercept

$\beta_i$ = Parameter in the model

$x_i$ = Set of factors or independent variables.

**R Syntax**

The statistical analysis was done with R studio software package (4.2.1, R Core Development Team), using the package 'party'. This study combines multiple statistical methodologies into single syntax to get the best results.

**Decision tree with ordinal regression and bootstrapping analysis using R syntax**

```
# Installing required packages
library(caret)
library(class)
library(tree)
library(caTools)
library(party)
library(partykit)
library(ConfusionTableR)
# Data Input
library(readxl)
TB <- read_excel("TB.xlsx")
# Converts Column from Numeric to Ordered.
str(TB)
TB$Brushing <- ordered(TB$Brushing)

# Step 1: Decision Tree Model
# Decision tree using all the data
tree <- ctree(TB$Brushing~., data = TB)
plot(tree)
# Calculating the prediction and accuracy for the data
pred = predict(tree, TB)
confusionMatrix(pred, TB$Brushing)

# Step 2: Ordinal Regression Model for Selected Variables
library(Mass)
library(ordinal)
library(erer)
library(brant)
# Converting Smoking and Toothache to factors for effect modification
TB$Toothache <- ordered(TB$Toothache)
TB$Smoking <- ordered(TB$Smoking)
# Bootstrapping data for 10000
mydata <- rbind.data.frame(TB, stringsAsFactors = FALSE)
iboot <- sample(1:nrow(mydata),size=10000, replace = TRUE)
Bootdata <- mydata[iboot,]
# Fitting ordered logit model
m <- polr(Brushing ~ Smoking + Toothache + MothersEducation + FreshFruits + SRTH + SugarCoffee , data = Bootdata, Hess = TRUE)
# Viewing a summary of the model and testing for proportional odds
summary(m)
brant(m)
# Calculating P Values
(ctable <- coef(summary(m)))
p <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
(ct <- cbind(ctable, `p value` = p))
# Calculating the Odds Ratio and 95% Confidence interval
(ci <- confint(m))
exp(coef(m))
exp(cbind(OR = coef(m), ci))
```

## RESULTS

**Descriptive Statistics**

In terms of sampling structure, there were more female participants comprising 62.3% of all participants. Participants mean age was 14.95 (95% CI 14.75-15.15). The sample included 160 (30.3%) who brushed less than once a day, 103 (19.5%) brushed once a day and 264 (50.1%) brushed more than once a day. 58.0% of the mothers had a university degree, while 46.4% of fathers had a university degree. 310 (58.8%) of the participants had experienced frequent toothache and only 55.7% of the participants considered their tooth health to be very good and excellent.

**Decision Tree Results**

The outcome of the decision tree analysis is plotted in figure 2. Out of the twenty-eight variables, six variables played a significant role ($p<0.05$) in the categorization of toothbrushing frequency. The conditional inference tree analysis identified smoking (split: never/seldom-regular), toothache (split: rarely / occasionally / often), mothers' education (split: not educated-college / university), self-rated tooth health (split: poor-good / very good-excellent), fresh fruit consumption (split: never-monthly / weekly-everyday) and coffee with sugar (split: never-weekly / everyday) as significant factors that could discriminate between frequency of brushing. Frequency of toothbrushing was associated with smoking habits in children ($p<0.001$). Children who have never smoked are further divided by the presence of toothache. Rare toothache, with good self-reported tooth health, led to higher frequency of brushing twice a day, while occasional toothache was further divided by mothers' education level. Highly educated mothers and good self-rated tooth health also brought about higher twice a day brushing frequency, while less educated mothers and low fruit consumption had brushing less than once daily. On the right side of figure 2, children who smoke seldom or regularly, are again divided by mothers' education. The low education level of mothers and occasional or more toothache, has high frequency of toothbrushing less than once a day. Similarly, smokers, with highly educated mothers, but who drink coffee with sugar regularly, have a high frequency of brushing less than once a day. The ordinal regression modelling will use these six variables as feeder for further modelling.

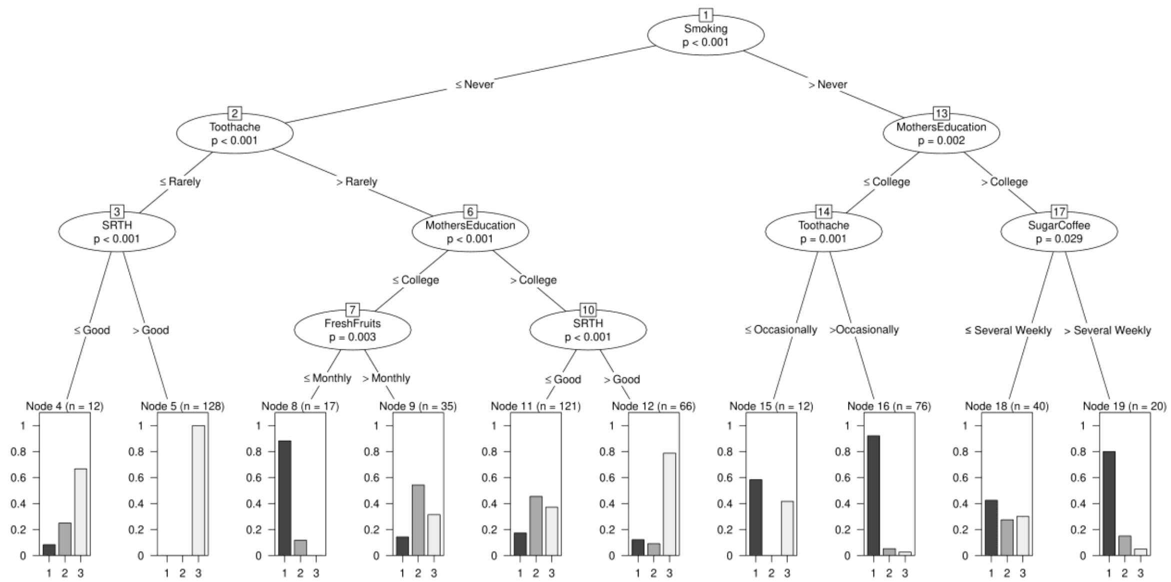


**Figure 2: Decision Tree Analysis**

**Prediction using the results of Decision tree analysis.**

Table 1 shows the results of the prediction done by the decision tree analysis. The confusion matrix and statistics assist in calculating the accuracy value obtained. Table 2 shows the summary of statistics for the decision tree analysis. The accuracy rate obtained for the prediction is 73.4%.

According to table 3, the positive prediction values for Class:1 is 75.7%, Class:2 is 47.4% and Class:3 is 91.2%. The table contains various other statistical summaries.

Table 4 summarizes the outcome of the ordinal regression model, integrating the inputs from the decision tree analysis. It was discovered that children who never smoked were 0.29 times less likely to be in higher category of toothbrushing when compared to children who seldom or regularly smoked. Children who rarely have toothache were 0.10 times less likely to brush twice a day, compared to children who had occasional toothache. Children with occasional toothache were 2.2 times more likely to be in higher category of toothbrushing compared to children with occasional toothache. Mothers with lower education were 0.43 times less likely to have children with better brushing habits compared to mothers with higher degrees. Dietary habits of children showed that low consumption of fresh fruits had 1.12 times higher likelihood of being in higher category of toothbrushing compared to high fruit consumption. Children drinking coffee with sugar less times a week were

0.93 times less likely to be in higher category of brushing compared to frequent drinkers of coffee with sugar. Children considering their tooth health as low were 1.43 times more likely to have better brushing habits than children considering their tooth health as good or excellent.

There are two distinct thresholds designated as cut 1 and 2. These thresholds can be used to predict probabilities of a patient brushing category, given certain characteristics. The proposed ordered logistic regression models for the various cut-off points can be represented by separate equations as,

**Ordinal logistic regression for Cut 1** (Less than once a day VS Once a day, More than once a day)

$$\text{Logit } (P(Y \leq 1)) = 1.3374 - 1.2042 \times \text{Smoking} - 2.2874 \times \text{Toothache} \leq \text{rarely} + 0.7914 \times \text{Toothache} \leq \text{occasionally} + 0.4386 \times \text{MothersEducation} + 0.1163 \times \text{FreshFruits} - 0.0620 \times \text{SuagrCoffee} + 0.3587 \times \text{SRTH} \quad (1)$$

**Ordinal logistic regression for Cut 2** (Less than once a day, Once a day VS More than once a day)

$$\text{Logit } (P(Y \leq 2)) = 2.8369 - 1.2042 \times \text{Smoking} - 2.2874 \times \text{Toothache} \leq \text{rarely} + 0.7914 \times \text{Toothache} \leq \text{occasionally} + 0.4386 \times \text{MothersEducation} + 0.1163 \times \text{FreshFruits} - 0.0620 \times \text{SuagrCoffee} + 0.3587 \times \text{SRTH} \quad (2)$$

**Table 1:Confusion matrix for prediction analysis**

| | | Reference | | |
|---|---|---|---|---|
| | | Less than once a day | Once a day | More than once a day |
| Predicted | **Less than once a day** | 125 | 20 | 20 |
| | **Once a day** | 26 | 74 | 56 |
| | **More than once a day** | 9 | 9 | 188 |

**Table 2: Statistics for decision tree analysis**

| **Statistics** | **Value** |
|---|---|
| **Accuracy** | 0.7343 |
| **95% CI** | 0.6944-0.7716 |
| **P-Value** | <2.2e-16 |
| **Kappa** | 0.5921 |

**Table 3: Summary of statistics by class**

| | **Class:1 Less than once a day** | **Class:2 Once a day** | **Class:3 More than once a day** |
|---|---|---|---|
| **Sensitivity** | 0.7812 | 0.7184 | 0.7121 |
| **Specificity** | 0.8910 | 0.8066 | 0.9316 |
| **Positive Prediction Value** | 0.7576 | 0.4744 | 0.9126 |
| **Negative Prediction Value** | 0.9033 | 0.9218 | 0.7632 |
| **Prevalence** | 0.3036 | 0.1954 | 0.5009 |
| **Detection Rate** | 0.2372 | 0.1404 | 0.3567 |
| **Detection Prevalence** | 0.3131 | 0.2960 | 0.3909 |
| **Balanced Accuracy** | 0.8361 | 0.7625 | 0.8218 |

Table 4: Parameter estimate for ordinal regression model

| Response | OR | Coeff | St Error | t-Value | P-value | 95% CI |
|---|---|---|---|---|---|---|
| Smoking | 0.2999 | -1.2042 | 0.0337 | -35.656 | 1.8e-278 | 0.2805-0.3203 |
| Toothache ≤Rarely | 0.1015 | -2.2874 | 0.0667 | -34.285 | 1.3e-257 | 0.0888-0.1154 |
| Toothache ≤Occasional | 2.2066 | 0.7914 | 0.0604 | 13.086 | 3.9e-39 | 1.9605-2.4853 |
| Mothers Education | 1.5506 | 0.4386 | 0.0204 | 21.485 | 2.1e-102 | 1.4902-1.6144 |
| Fresh Fruits | 1.1233 | 0.1163 | 0.0182 | 6.371 | 1.8e-10 | 1.0839-1.6144 |
| Coffee with Sugar | 0.9398 | -0.0620 | 0.0157 | -3.945 | 7.9e-05 | 0.9112-0.9692 |
| SRTH | 1.4314 | 0.3587 | 0.0203 | 17.620 | 1.7e-69 | 1.3755-1.4898 |
| Cut1 | - | 1.3374 | 0.1459 | 9.167 | 4.8e-20 | - |
| Cut2 | - | 2.8369 | 0.1491 | 19.031 | 9.3e-81 | - |

## DISCUSSION

The discussion of this paper is divided into two phases; first, the discussion focuses on the development of the technique, and second, the discussion focuses on the findings. The primary goal was to create, test, and validate a combination of the decision tree, bootstrap, and ordinal regression for developing and implementing medical statistical strategies. The suggested hybrid model is based on the assumption of proportional odds for the outcome category along with the proposition that the bootstrap procedure increases accuracy. With increased availability of monitoring data, numerous characteristics can be accounted for in statistical analysis. Here, a decision tree has been used for the selection of the most relevant input variables to further build the model. This process demonstrates the possibility of developing prediction models which can then be used to formulate clinical hypothesis and inform future researchers. The most perplexing task of any research is selecting the appropriate input parameters, made easy by this methodology. The process is an alternate to ordinal regression modelling as the selection of appropriate variables is based on computational analysis, forecasting the importance of the independent variables chosen for the final model. Researchers might utilize decision trees to determine the most significant independent variable when developing an ordinal regression model and also to test a clinical hypothesis. To evaluate the accuracy of the established approach, the predictive model is applied to the original data, and its output is compared. The results help in calculating the accuracy, sensitivity and specificity of the prediction made. In our analysis, an accuracy of 73.4% was achieved, with a positive prediction value for Class 1 at 75%, Class 2 at 47% and Class 3 at 91%. The most significant variables obtained from the decision tree classification were smoking habits, toothache, maternal education, SRTH, fresh fruit and coffee with sugar consumption, while the other variables are being excluded from the classification.

Toothbrushing is the most fundamental self-care behavior for prevention of oral diseases. The habit of brushing at least twice a day is a social norm, and it is common practice for dentists and professional organizations to offer such advice. According to Kamppi et al, regular adult smokers had higher odds of oral diseases. Smoking was statistically associated with other harmful habits such as low tooth brushing frequency, frequent snacking, eating sweets and consuming carbonated drinks[41]. Ellison et al found that children who smoked were more concerned about their bad breath and stained teeth[42]. Looking at it in terms of oral hygiene practice, it can be inferred that children who smoked would have higher frequency of brushing to prevent teeth staining and bad breath. In our results, children who never smoked were at lower odds for brushing twice daily. Looking at children who smoked seldomly or regularly, they being children would certainly want to hide the smell or their bad breath from grownups and inadvertently have more frequent brushing habits than non-smoking children.

Multiple studies have corroborated the role of mothers in a child's oral health. Whether it be that the mothers are less educated, or are less concerned about their own oral health, leading to less concern of the child's oral health, or mothers underestimating the importance of oral health, eventually all would lead to negative oral health outcomes for the child[43-46]. This may further manifest in terms of a child's dietary habits. Studies have shown a variety of junk foods and frequent snacking lead to higher oral diseases. Similarly, children with such eating habits also have reduced odds of frequent toothbrushing[44-46]. In our study regular coffee with sugar added lowered odds (0.93) for having better toothbrushing habits while consumption of fresh fruits increased odds (1.12) for better oral hygiene habits. Fresh fruits having both fructose instead of sucrose and fiber, acting itself as a cleaning agent, leads to a healthier oral cavity. It can be assumed that dietary consumption can be attributed to parental income and education, where having good income leads to better availability of healthier foods. Inversely, working mothers have a detrimental impact on oral health of child as the parent is not available to supervise the child[47]. González Martínez *et al*. reported that when parents do not monitor their children during toothbrushing, there is a greater risk of the development of oral diseases. This relationship of maternal education and toothbrushing frequency of their children showed higher odds (1.55) for low educated mothers and the child having good brushing habits. This may be due to higher educated mothers being employed and having less time to oversee their child daily.

In our study, the children who have never, or rarely experienced toothache had lower odds of higher frequency of toothbrushing, while children who experienced occasional toothache were more regular and frequent in their toothbrushing habits. Toothbrushing is considered the first line of defense for any oral problem, so it would make sense that children would turn to toothbrushing when they feel discomfort or pain in the oral cavity. A lack of perceived needs is an important factor towards self-care behavior[7]. Similarly, Michiko Furuta found tooth brushing frequency to be associated to self-rated oral health[48]. Self-rated oral health covers domains of functionality and aesthetics both. Appearance of teeth is found to be a source of concern in western as well as Asian samples[49]. Literature suggests that there is a strong link between self-perceived oral needs and psychosocial wellbeing[50]. In our study, those with lower SRTH were more likely to be in better toothbrushing category compared to those with higher SRTH. Children who are more concerned about the status of their teeth, including colour, staining, functionality, would have better brushing habits to counter the effects.

This study suggests that interventions to enhance tooth brushing behavior are still needed. On the individual level, research and interventions should focus on parental involvement and dietary habits of children. Additionally, there should be attention to oral health education concerning smoking habits and dental visitation related to toothache. Decision tree inference of which variables to use in making a forecast proves the usefulness of this approach to health care planning. By combining a decision tree with ordered regression analysis, we may improve the clinical application of risk variables. The technique simplifies the most difficult aspect of any research project, which is choosing the right input parameters. In this instance of research, the decision tree method accurately evaluates variables to be selected with care for the final model. Predictive modelling for forecasting dental health behaviors is crucial for public health professionals as healthcare costs continue to rise due to the prevalence of non-communicable chronic illnesses. This research will aid medical professionals in spotting patients who are at risk of oral diseases. This research also has direct applications to the field of dental public health. For dentists, there is potential clinical benefit for using decision splitting to better understand the patient and cater their preventive education accordingly.

## CONCLUSION

The R syntax integrates application with the concept of a developed methodology-based approach. This research output demonstrated the findings and their application to managing oral disease and their risk factors. The model's sensitivity was greater than 70% and specificity was greater than 80%. We can conclude that our proposed method produces excellent results with the highest possible forecasting precision. The parameters of smoking, maternal education, dietary intake, and self-perceived oral health contributed significantly to a patient's overall self-care habit of toothbrushing, demonstrating that the methodology and findings were developed successfully. Interventions to enhance tooth brushing behavior are still needed with focus on parental involvement and better dietary habits of children. Additionally, attention should be given to oral health education concerning smoking habits and dental visitation related to toothache.

### Limitations

The cross-sectional method used to collect data is limited in establishing causality between the social determinants and the tooth-brushing frequency. In addition, all the sociodemographic and behavioral information presented and used in this article were self-reported. Self-reported behavioral data are sometimes subjected to social reporting bias.

## ACKNOWLEDGMENTS

The authors would like to thank Rashid Latif Dental College for providing the research data. The authors would like to thank Ministry of Higher Education Malaysia for Fundamental Research Grant Scheme with Project Code: FRGS/1/2022/STG06/USM/02/10 and Universiti Sains Malaysia (USM).

## REFERENCES

1. Roth G. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017 (GBD 2017) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2018. *The Lancet*. 2018;392:1736-88.

2. Petersen PE. bourgeois D, Ogawa H, Estupinan-Day S, Ndiaye C. *The global burden of oral diseases and risks to oral health bulletin of the World Health Organization*. 2005;83(9):661-69.

3. Shahzad HB, Awais F, Shirazi U-e-R, Majeed HA, Rafique A, Shahbaz M. The impact of dental caries on oral health related quality of life amongst adult population in Lahore, Pakistan. *Makara Journal of Health Research*. 2020;24(1):1.

4. Pakkhesal M, Riyahi E, Naghavi Alhosseini A, Amdjadi P, Behnampour N. Impact of dental caries on oral health related quality of life among preschool children: perceptions of parents. *BMC Oral Health*. 2021;21(1):1-8.

5. Krisdapong S, Somkotra T, Kueakulpipat W. Disparities in Early Childhood Caries and its impact on oral health-related quality of life of preschool children. *Asia Pacific Journal of Public Health*. 2014;26(3):285-94.

6. Niaz MO, Naseem M, Siddiqui SN, Khurshid Z. An outline of the oral health challenges in "Pakistani" population and a discussion of approaches to these challenges. *JPDA*. 2013;21(3).

7. Khan AA, Sharea I, Ayma S, Ambreena Q, Inayatullah P, Sofia S. Oral health in Pakistan: a situation analysis. *Dev Dent*. 2004;5(2):35-44.

8. Chu R, Craig B. Understanding the determinants of preventive oral health behaviours. *Probe (Ottawa, Ont)*. 1996;30(1):12-8.

9. Collaborators GOD, Bernabe E, Marcenes W, Hernandez C, Bailey J, Abreu L, et al. Global, regional, and national levels and trends in burden of oral conditions from 1990 to 2017: a systematic analysis for the global burden of disease 2017 study. *Journal of dental research*. 2020;99(4):362-73.

10. Fisher-Owens SA, Gansky SA, Platt LJ, Weintraub JA, Soobader M-J, Bramlett MD, et al. Influences on children's oral health: a conceptual model. *Pediatrics*. 2007;120(3):e510-e20.

11. Arora A, Nargundkar S, Fahey P, Joshua H, John JR. Social determinants and behavioural factors influencing toothbrushing frequency among primary school children in rural Australian community of Lithgow, New South Wales. *BMC Research Notes*. 2020;13(1):1-8.

12. Marinho VC, Higgins J, Logan S, Sheiham A. Fluoride toothpastes for preventing dental caries in children and adolescents. *Cochrane database of systematic reviews*. 2003(1).

13. Sicca C, Bobbio E, Quartuccio N, Nicolò G, Cistaro A. Prevention of dental caries: A review of effective treatments. *Journal of clinical and experimental dentistry*. 2016;8(5):e604.

14. Savolainen JJ, Suominen-Taipale AL, Uutela AK, Martelin TP, Niskanen MC, Knuuttila ML. Sense of coherence as a determinant of toothbrushing frequency and level of oral hygiene. *Journal of periodontology*. 2005;76(6):1006-12.

15. Kumar S, Tadakamadla J, Johnson N. Effect of toothbrushing frequency on incidence and increment of dental caries: a systematic review and meta-analysis. *Journal of dental research*. 2016;95(11):1230-6.

16. Walsh T, Worthington HV, Glenny AM, Marinho VC, Jeroncic A. Cochrane Oral Health Group. Fluoride toothpastes of different concentrations for preventing dental caries. *Cochrane Database Syst Rev*. 2019.

17. Ylöstalo P, Ek E, Laitinen J, Knuuttila M. Optimism and life satisfaction as determinants for dental and general health behavior—oral health habits linked to cardiovascular risk factors. *Journal of dental research*. 2003;82(3):194-9.

18. Kerr R, Claman D, Amini H, Alexy E, Kumar A, Casamassimo PS. Evaluation of the ability of five-to 11-year-olds to brush their teeth effectively with manual and electric toothbrushing. *Pediatric Dentistry*. 2019;41(1):20-4.

19. Davies G, Davies R. Delivering better oral health-an evidence-based toolkit for prevention: a review. *Dental update*. 2008;35(7):460-4.

20. Duijster D, Verrips G, van Loveren C. The role of family functioning in childhood dental caries. *Community Dentistry and Oral Epidemiology*. 2014;42(3):193-205.

21. Duijster D, de Jong-Lenters M, Verrips E, van Loveren C. Establishing oral health promoting behaviours in children-parents' views on barriers, facilitators and professional support: a qualitative study. *BMC oral health*. 2015;15(1):1-13.

22. Aarø LE, Laberg JC, Wold B. Health behaviours among adolescents: towards a hypothesis of two dimensions. *Health Education Research*. 1995;10(1):83-93.

23. Zaborskis A, Milciuviene S, Narbutaite J, Bendoraitiene E, Kavaliauskiene A. Caries experience and oral health behaviour among 11-13-year-olds: an ecological study of data from 27 European countries, Israel, Canada and USA. *Community Dent Health*. 2010;27(2):102-8.

24. Maes L, Vereecken C, Vanobbergen J, Honkala S. Tooth brushing and social characteristics of families in 32 countries. *International dental journal*. 2006;56(3):159-67.

25. Scheerman JF, van Loveren C, van Meijel B, Dusseldorp E, Wartewig E, Verrips GH, et al. Psychosocial correlates of oral hygiene behaviour in people aged 9 to 19-a systematic review with meta-analysis. *Community dentistry and oral epidemiology*. 2016;44(4):331-41.

26. Elkin F, Handel G. The child and society: The process of socialization: McGraw-Hill Humanities, Social Sciences & World Languages; 1988.

27. Organization WH. Oral health surveys: basic methods: World Health Organization; 2013.

28. Shahzad HB, Kazmi F, Manzar S, Rashid S, Awais F, Arshad AI, et al. Dental Care use and Self-Reported Dental Problems in Children in Lahore, Pakistan. *Pakistan Journal of Medical & Health Sciences*. 2023;17(02):218-.

29. NHS. Healthy Teeth and Gums UK: NHS; 2022 [Available from: https://www.nhs.uk/live-well/healthy-teeth-and-gums/how-to-keep-your-teeth-clean/.

30. Mesarić J, Šebalj D. Decision trees for predicting the academic success of students. *Croatian Operational Research Review*. 2016;7(2):367-88.

31. Kantardzic M. Data mining: concepts, models, methods, and algorithms: John Wiley & Sons; 2011.

32. Alpaydin E. Decision Trees. 2014.

33. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical statistics*. 2006;15(3):651-74.

34. Zuur AF, Ieno EN, Smith GM. Analysing ecological data: Springer; 2007.

35. Lavanya D, Rani KU. Performance evaluation of decision tree classifiers on medical datasets. *International Journal of Computer Applications*. 2011;26(4):1-4.

36. Patro VM, Patra MR. Augmenting weighted average with confusion matrix to enhance classification accuracy. *Transactions on Machine Learning and Artificial Intelligence*. 2014;2(4):77-91.

37. Kohavi R, Provost F. Confusion matrix. *Machine learning*. 1998;30(2-3):271-4.

38. Efron B. Computers and the theory of statistics: thinking the unthinkable. *SIAM review*. 1979;21(4):460-80.

39. Liu X. Ordinal regression analysis: Fitting the proportional odds model using Stata, SAS and SPSS. *Journal of Modern Applied Statistical Methods*. 2009;8(2):30.

40. Adeleke K, Adepoju A. Ordinal logistic regression model: An application to pregnancy outcomes. 2010.

41. Tanner T, Kämppi A, Päkkilä J, Järvelin M-R, Patinen P, Tjäderhane L, et al. Association of smoking and snuffing with dental caries occurrence in a young male population in Finland: a cross-sectional study. *Acta Odontologica Scandinavica*. 2014;72(8):1017-24.

42. Ellison J, Mansell C, Hoika L, MacDougall W, Gansky S, Walsh M. Characteristics of adolescent smoking in high school students in California. *American Dental Hygienists' Association*. 2006;80(2):8-.

43. Huebner CE, Riedy CA. Behavioral determinants of brushing young children's teeth: implications for anticipatory guidance. *Pediatric dentistry*. 2010;32(1):48-55.

44. Shearer D, Thomson W, Broadbent J, Poulton R. Maternal oral health predicts their children's caries experience in adulthood. *Journal of dental research*. 2011;90(5):672-7.

45. Okada M, Kawamura M, Kaihara Y, Matsuzaki Y, Kuwahara S, Ishidori H, et al. Influence of parents' oral health behaviour on oral health status of their school children: an exploratory study employing a causal modelling technique. *International journal of paediatric dentistry*. 2002;12(2):101-8.

46. Dye BA, Vargas CM, Lee JJ, Magder L, Tinanoff N. Assessing the relationship between children's oral health status and that of their mothers. *The Journal of the American Dental Association*. 2011;142(2):173-83.

47. Baiju R, Peter E, Narayan V, Varughese JM, Varghese N. Do children of working mothers experience more dental caries? *Contemporary Clinical Dentistry*. 2018;9(4):541.

48. Furuta M, Ekuni D, Takao S, Suzuki E, Morita M, Kawachi I. Social capital and self-rated oral health among young people. *Community dentistry and oral epidemiology*. 2012;40(2):97-104.

49. Ao H, Deng X, She Y, Wen X, Wu Q, Chen F, et al. A biopsychosocial-cultural model for understanding oral-health-related quality of life among adolescent orthodontic patients. *Health and Quality of Life Outcomes*. 2020;18(1):1-8.

50. AlSagob EI, Alkeait F, Alhaimy L, Alqahtani M, Hebbal M, Gassem AAB. Impact of Self-Perceived Dental Esthetic on Psycho-Social Well-Being and Dental Self Confidence: A Cross-Sectional Study

Among Female Students in Riyadh City. *Patient preference and adherence*. 2021;15:919.