ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 14, March 2017

# Recognising risk factors for early Childhood Caries using association rule mining

Ms.J.Athidhiya,
UG Student
Department of CSE,
Kongu Engineering College,
Erode,Tamilnadu
adhijaganathan11@gmail.com

Ms.D.Dharani,
UG student
Department of CSE,
Kongu Engineering College,
Erode,Tamilnadu
dharanitamil13@gmail.com

Mr.S.Sankar,
UG student
Department of CSE,
Kongu Engineering College,
Erode,Tamilnadu

Ms.N.Sasipriyaa,
Assistant Professor,
Department of CSE,
Kongu Engineering College,
Erode,Tamilnadu
nsasipriyaa@gmail.com

**ABSTRACT**

*Background and objective:* Early childhood caries (ECC) is a possibly severe disease affectingchildren all over the world. Association rule mining is used to extract moreinformation about ECC. Artificial Bee Colony algorithm is used to optimize the best rules.

*Methods:* ECC data was collected in a cross-sectional analytical study of the 10% sampleof preschool children in the South Backa area (Vojvodina, Serbia). Association rules wereextracted from the data by association rule mining. Risk factors were mined from thehighly ranked association rules

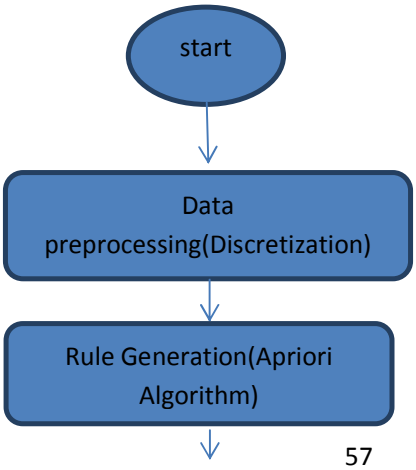
*Results:* Discovered dominant risk factors consist of male gender, frequent breastfeeding (withother risk factors), high birth order, language, and low body weight at birth. Low healthawareness of parents was significantly associated to ECC only in male children

*Conclusions:* The discovered risk factors are mostly confirmed by the literature, which upholds the value of the methods.

## 1.Introduction

Early childhood caries (ECC) is a special form of caries of theprimary dentition that affects the teeth after eruption, hasa rapid progress, and later leads to a number of symptoms and complications.The American Academy of Paediatric Dentists (AAPD) defined ECC as “the presence of one or moredecayed (noncavitated or cavitatedinjuries), missing (due tocaries), or filled tooth surfaces in any primary tooth in a childunder the age of six”. Available data, often grouped into different age ranges, indicate dramatic variation in prevalenceacross different parts of the world.In our paper we find that,ECC is widely spread in Serbia, it has not beena frequent research topic. According to the data used in thepresent study, ECC prevalence in preschool children in theSouth Backa area (SBA) of the Province of Vojvodina, Serbia,was 30.5%. In 1998, it was reported that the ECC prevalence in three-year-old children in SBA was 22.07%. It appears that ECC in Serbia may be on increase,which could be linked to the rapid decrease in living standards,therapeutic approach to ECC treatment, as well as to specificdemographic, psychosocial and behavioural characteristics ofthe environment. For these reasons, researchers need ECCmodels that correspond to the current epidemiological situation in the affected area. ARM is a group of DM techniques used to detect associations in data, may be utilized in an ECC study to detect relationships between ECC and potential risk factors.

ARM may produce an enormous number ofrules, but this may be alleviated by rule pruning and ranking. Therefore, in this exploratory study, we first performedARM to generate ECC-related association rules (ARs) and then,with the goal of uncovering ECC risk factors and their interaction, we pruned, ranked, and inspected the highly rankedassociation rules. More information about the data set, ARM methods and ABC optimization method is provided in section 2. The uncoveredrisk factors, as well as the comparison of our findings tothose from other studies and regions, are presented in section 3.



ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 14, March 2017

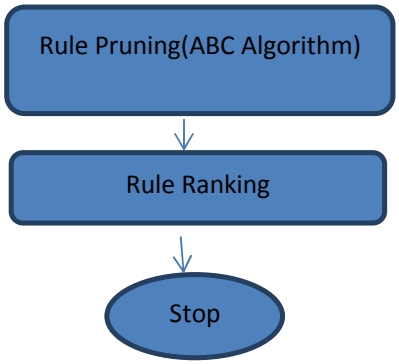



**Fig 1. Framework of proposed system**

## 2. Material and methods

### *2.1. Data Set*

*2.1.1. Data collection*

The aim of the data collection was to investigate potential risk factors and the prevalence of ECC, as well as the degree of severity among different social and ethnic groups of preschool children in SBA, Republic of Serbia. The survey was a cross-sectional analytical study of the 10% sample of preschool children in SBA, aged 13–71 months of both sexes, and different ethnicities, social status, and human environmental (urban, rural) groups. The presence or absence of ECC was recorded depending on the presence of no-cavity caries (white spot lesions), or cavity caries. Christo Ananth et al. [5] discussed about a system,the effective incentive scheme is proposed to stimulate the forwarding cooperation of nodes in VANETs. In a coalitional game model, every relevant node cooperates in forwarding messages as required by the routing protocol. This scheme is extended with constrained storage space. A lightweight approach is also proposed to stimulate the cooperation.All primary teeth were examined and caries was recorded using WHO recognized indices DMFT and DMFS, which are calculated as the total number of decayed (D), missing due to caries (M), and filled (F) teeth (T) and surfaces (S), respectively. The diagnosis and the clinical form of ECC were defined by dental check-up according to the Wyne's modified criteria.Epidemiological data of the different social and ethnic groups, as well social status, habits, attitudes, behaviour and health knowledge, were obtained by the interview of the parents of the examined children through a series of closed questions. Erupted or congenitally missing teeth were excluded from the DMFT and DMFS scores.

*2.1.2. Data overview*

In the collected data set, there are 341 individual records collected at different locations in SBA. Besides the binary variable denoting ECC presence/absence, there are 35 categorical variables carefully selected by the domain experts as potential risk factors(Table 1).

### *2.2. Association rule overview*

An association rule (AR) in this study is of the form $A \Rightarrow B$, where $A$ (the left-hand-side or the antecedent) and $B$(the right-hand-side or the consequent) are disjoint sets of attribute-value pairs (AVPs).

e.g., attribute A = value A1, attribute B = value B1 ⇒ attribute C = value C1.

The rule is valid for a case from the data set if all the AVPs from the rule hold true for the particular case. Support (SUPP)of a rule denotes the percentage of cases from the whole dataset for which the rule is valid. Confidence (CONF) of a rule indicates the same but only within the subset of cases satisfying the antecedent of the rule. Lift of a rule is the ratio of the observed support of the rule and the expected support if the antecedent and the consequent were independent. ARM is performed on the ECC data set featuring 341 cases and 36 categorical variables. A restriction that is imposed on the format of the generated rules is that the consequent of each rule

**Table 1 – Explanatory variables in the ECC data set.**

| |
|---|
| **Child-related variables** |
| Ethnicity |
| Age |
| Gender |
| Serbian language |
| Birth order |
| Birth weight |
| Breastfeeding |
| Breastfeeding frequency |
| Breastfeeding during night |
| Bottle feeding |
| Infant formulas |
| Additional food sweetening |
| Fluoride supplements |
| Fluoride toothpaste |
| Oral hygiene |
| Tooth brushing |
| Diarrhea during infancy |
| Medical syrups |
| First dentist visit |
| **Mother-related variables** |
| Age |
| Marital status |
| Ethnicity |
| Serbian language |
| Number of children |
| Education level |
| Employment status |
| Sweets during pregnancy |
| Fluoride supplements during pregnancy |
| Oral health during pregnancy |
| Health awareness |
| **Father-related variables** |
| Health awareness |
| **Family-related variables** |
| City |
| Quality of housing |
| Housing conditions |
| Household monthly income |

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 14, March 2017

| SNo | Rules | Support | Confidence | Lift |
|---|---|---|---|---|
| 1 | {CHILD_SERBIAN_LANGUAGE,HOUSING CONDITION,BIRTH WEIGHT,CHILD ORAL HYGIENE,DIARRHEA DURING INFANCY} => {ECC} | 0.33333333 | 0.70240597 | 1.01682229 |
| 2 | {CHILD_SERBIAN_LANGUAGE,BIRTH WEIGHT,CHILD ORAL HYGIENE,DIARRHEA DURING INFANCY,MOTHER HEALTH AWARNESS} => {ECC} | 0.30409036 | 0.70370404 | 1.01911737 |
| 3 | {HOUSING CONDITION,BIRTH WEIGHT,CHILD FLUORIDE SUPPLEMENTS,CHILD ORAL HYGIENE,DIARRHEA DURING INFANCY,MOTHER HEALTH AWARNESS} => {ECC} | 0.31286055 | 0.70482517 | 1.01797557 |
| 4 | {HOUSING CONDITION,BIRTH WEIGHT,CHILD ORAL HYGIENE,DIARRHEA DURING INFANCY,SWEETS DURING PREGNANCY,MOTHER HEALTH AWARNESS} => {ECC} | 0.30116096 | 0.70357143 | 1.01616637 |
| 5 | {CHILD_SERBIAN_LANGUAGE,MARITAL STATUS,BIRTH WEIGHT,CHILD ORAL HYGIENE,DIARRHEA DURING INFANCY,MOTHER HEALTH AWARNESS} => {ECC} | 0.30116096 | 0.70463768 | 1.01770501 |
| 6 | {MARITAL STATUS,HOUSING CONDITION,BIRTH WEIGHT,CHILD ORAL HYGIENE,DIARRHEA DURING INFANCY,MOTHER HEALTH AWARNESS} => {ECC} | 0.31286055 | 0.70278912 | 1.01503746 |

contains a single AVP denoting either ECC presence or absence. The generated rules are ranked to facilitate the identification of relevant relationships between ECC and potential risk factors.

### *2.3. Rule generation*

ARM is performed using the Apriori algorithm implementation in Rstudio for arules package. We mine for two groups of rules whose consequent contains only the ECC variable:

• Set1 rules for the presence of ECC, where SUPP ≥ 0.05and CONF ≥ 0.6; and

• Set2 rules for the absence of ECC, where SUPP ≥ 0.15and CONF ≥ 0.6.

In both cases, the starting support threshold was set to 0.3(30% of all cases), but had to be gradually decreased in order to increase the number of rules for later inspection. The support threshold for R1 rules had to be further decreased because the data set is imbalanced in favour of ECC absence, i.e., ECC is present in only 104 cases (30.5%). 0.6(60%) was set as the confidence threshold in order to find AVPs that are more often associated with ECC (either presence or absence) than they are not, i.e., associated in at least 60% of the cases containing the evaluated AVPs and not associated in at most 40% of the same cases.

### *2.4. Rule Ranking*

As ARM usually results in a large number of association rules, the notion of rule interestingness is important when evaluating the generated ARs.

In order to evaluate the OSet1and OSet2 rules, we combined several objective measures of rule interestingness (OMORI) using average ranking for rules, and ranked OSet1 and OSet2 rules separately. In this manner, we obtained two ranked rule lists: ROSet1 (ranked OSet1) and ROSet2(ranked OSet2).Out of many OMORIs, we considered five: support, conviction, phi, cosine, and odds ratio. The usefulness of individual OMORIs, as indicated by their agreement with expert estimation, may depend the domain of application and the actual topic. As there is not a set of recommended measures, we had initially designed values for 16 common OMORIs, but, owing to strong correlation between various measures, had to ignore the highly correlated measures through correlogram analysis.

Table 2 –Sample association rules generated

### 2.5 Rule Pruning

While generating frequent item sets from a large dataset using association rule mining, computer takes too much time and more number of rules are generated. This can be improved by using artificial bee colony algorithm (ABC). The Artificial bee colony algorithm is an optimization algorithm based on the foraging behavior of artificial honey bees. In this paper, artificial bee colony algorithm is used to generate high qua.lity association rules for finding frequent item sets from large data sets

In the ABC algorithm, each food source is a possible solution for the problem under consideration and the nectar amount of a food source represents the quality of the solution represented by the fitness value.

The number of food sources is same as the number of employed bees and there is exactly one employed bee for every food source. This algorithm starts by associating all employed bees with randomly generated food sources (solution). In each iteration, every employed bee determines a food source in the neighborhood of its current food source and evaluates its nectar amount (fitness).

The $i^{th}$ food source position is represented as $X_i = (x_{i1}, x_{i2}. . . x_{id})$. $F$ (Xi) refers to the nectar amount of the food source located at Xi. If an employed bee's new fitness value becomes better than the best fitness value achieved so far, then the employed bee moves to this new food source abandoning the old one, otherwise it remains in its old food source. When all employed bees have finished this process, they share the

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 14, March 2017

fitness information with the onlookers inside the hive, each of which selects a food source according to the probability. The probability depends on the quality of the food source. With this scheme, good food sources will get more onlookers than the bad ones. Each bee will search for better food source around neighbourhood patch for a certain number of cycles (limit), and if the fitness value will not improve within limit number of cycles, then that bee becomes scout. The procedure is continued until the termination criterion is attained.

**PSEUDO CODE FOR ABC ALGORITHM WITH EXAMPLE**

Main steps of the algorithm are given below:

1. Initialize the food source positions.
2. Each employed bee produces a new food source in their food source site and exploits the better source.
3. Each onlooker bee selects a source depending on the quality of her solution, produces a new food source in selected food source site and exploits the better source.
4. Determine the source to be abandoned and allocate its employed bee as scout for searching new food sources.
5. Memorize the best food source found so far.
6. Repeat steps 2-5 until the stopping criterion is met.

In order to use the ABC algorithm, the following points must be addressed: initial population, fitness value, employed, onlooker, and scout bees. Here Initial population is generated using randomly generated transactions.

The objective of this fitness function is maximization. The larger the particle support and confidence, the greater the strength of the association, meaning that it is an important association rule. The fitness function is as follows:

Fitness(k) = confidence(k)* log(support(k)*length(k) + 1

Fitness(k) is the fitness value of association rule type k. Confidence(k) is the confidence of association rule type k. Support(k) is the actual support of association rule type k. Length(k) is the length of association rule type k. Table 3 represents the top three best rules generated using ABC optimization algorithm.

**Table 3 - Best rules**

| No | Best Rules |
|---|---|
| 1 | {CHILD_SERBIAN_LANGUAGE,HOUSING CONDITION,BIRTH WEIGHT,CHILD ORAL HYGIENE,DIARRHEA DURING INFANCY} => {ECC} |
| 2 | {CHILD_SERBIAN_LANGUAGE,BIRTH WEIGHT,CHILD ORAL HYGIENE,DIARRHEA DURING INFANCY,MOTHER HEALTH AWARNESS} => {ECC} |
| 3 | {HOUSING CONDITION,BIRTH WEIGHT,CHILD FLUORIDE SUPPLEMENTS,CHILD ORAL HYGIENE,DIARRHEA DURING INFANCY,MOTHER HEALTH AWARNESS} => {ECC} |

## 3. Results and discussion

Rule generation resulted in 18 Set1 rules and 88142 Set2 rules. After pruning, there remained 15 OSet1 rules and 6306 OSet2 rules. We present all AVPs featured in the antecedents of the OSet1 rules. One AVP may appear in multiple rules and one rule may contain multiple AVPs in its antecedent. After ruleranking, we obtain the ROSet1 and ROSet2 lists.We discuss the top three ROSet1 rules in the following subsections. The top three ROSet2 rules share threechild-related AVPs in their antecedents: understands andspeaks Serbian, born as the first one, and birth weight greaterthan 2500 g. Other AVPs include comfortable housing, motheremployed, and mother rarely consuming sweets duringpregnancy. As opposed to the ROSet1 rules, the ROSet2 rules havehigher support (from 0.27 to 0.30) and confidence (greater than80%), but lower lift (about 1.2), i.e., they cover more cases withmore confidence but they are not far from the expected.

### *3.1. Child birth order and age*

It has been found that the third and every next born child in a family has a higher risk for ECC than the previous children, which is similar to the findings presented in [14].Our study also showed a direct correlation between child's age and prevalence of ECC, which was detected in other studies as well . As the number of teeth was increasing due to age, it was logical to expect an increase of caries, as well as the severity of disease in environmentally unchangeable conditions such as eating habits, oral hygiene, and fluoride use

### *3.2. Language*

In the top three ROSet2 rules (ECC absence), there is an AVP about child's lack of fluency in Serbian, the official language of the country. In the last official census from 2011, 25 categories fornationality were defined and the multiethnic population of SBA (Province ofVojvodina)

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 14, March 2017

was pointed out .The highest prevalence was noted in Roma children, followed by Ruthenian, Slovakian, and other nationalities who, according to our data, did not speak Serbian fluently. The spokenlanguage is the basis of interpersonal communication, education, socialization, and equal participation in all provinces of social life. Considering these facts, there is a need for understanding the problem of multilingualism of the minority population groups living in Vojvodina. In our study, we noted that children who could not understand Serbian, apart from ethnicity, were more significantly exposed to ECC as opposed to children who could understand and speak Serbian. This translates to a more difficult approach to health information and mass media (television, radio, newspapers)and consequently to a lower level of health education, and increased possibility for ECC occurrence. As a result of the language barriers, the low level of understanding between parent–child–dentist may lead to distrust and feeling of discrimination among parents themselves and their children as well.

### *3.3. Child birth weight*

The prevalence of the moderate, middle and severe form ofECC with complications was higher in children who were bornwith body mass less than 2500 g as opposed to children ofnormal weight at birth. The relationship between ECC andlow birth weight was also identifiedin . Prematurely bornchildren mostly had lower birth weight and accordingly a significantly higherprevalence of linear enamel defects (LEDs)and other developmental anomalies. They were also morepredisposed to frequent diseases and rapid progression of ECC. It was stated that teeth with LEDs were moreaffected with ECC when compared to healthy teeth. Davies explained that the teeth with LEDs are more predisposedto caries because they had rough surface with the presenceof the expressed respite on the enamel which could increaseplaque accumulation and mutans streptococci colonization.The studies of Seow et al. and Douglass et al. alsoindicated the significance of the LEDs as a predisposed factorfor ECC and the high prevalence of this defect of up to 62% inprematurely born children with very low birth weight.

### *3.4. Housing Condition*

Housing Condition is an important factor as the children's habit will also be a reason for the disease. The above pruned rules shows that in most of the cases housing condition and the quality of housing also leads to ECC.

### *3.5.Diarrhea during infancy*

Most of the rules consists of diarrhea during infancy as one of the risk factors which causes ECC.It may results in loss of health and the children may not get some of the nutrients.As a result it leads to loss of body weight, which is also another risk factors for ECC.

## 4. Conclusion

The described methodology could be useful in ECC research because the resulting patterns include associations of potential risk factors. As ECC is a multifactorial disease, researchers need to combine factors to explain the risk of ECC appearance. Researchers have attempted to expand basic microbiological models for ECC development, and to include various social, demographic and behavioural factors such as ethnicity, family income, maternal education level, family status and parental knowledge, as well as inappropriate nursing habit, since the reason for disease appearance is still unknown. The main strengths of our study when compared to other DM-based studies on ECC are: (i) examination of a different set of potential risk factors; (ii) usage of different techniques, namely ARM; and (iii) better readability of the results. Moreover, as opposed to using DM for ECC prediction, which is the main goal of other studies, we focus on ECC description. The used approach may lead to a list of potential risk factors for the analyzed environment. We identified the following ECC-related patterns: (i) male children; (ii) male children whose parents are not well informed about dental health;(iii) children who were frequently breastfed; (iv) lack of fluency in Serbian(the official language); and (vi) low child's body weight at birth. Contrary to the expectations, the relationship between parent health awareness and ECC was significant only in male children. The sometimes-questioned relationship between breastfeeding and ECC seems to hold true when breastfeeding is frequent and coupled with other factors. As discussed in Section 3, the remaining risk factors are confirmed or recognized in other studies.In our future work, we plan to formally model the domain knowledge about ECC, and use it during rule selection and interpretation. We also intend to experiment with summarization of ARs into a decision tree and check if that wouldlead to understandable predictive models.

## References

[1] American Academy of Pediatric Dentists, Policy on earlychildhood caries (ECC): classifications, consequences, andpreventive strategies, Ref. Man. 36 (2014) 50–52.
[2] M.G. Gussy, E.G. Waters, O. Walsh, N.M. Kilpatrick, Earlychildhood caries: current evidence for aetiology andprevention, J. Paediatr. Child Health 42 (2006) 37–43.

[3] A.R. Milnes, Description and epidemiology of nursing caries,J. Public Health Dent. 56 (1996) 38–50.
[4] National Center for Health Statistics, Health, United States,2013: With Special Feature on Prescription Drugs, NationalCenter for Health Statistics, Hyattsville, MD, 2014.
[5] Christo Ananth, M.Muthamil Jothi, A.Nancy, V.Manjula, R.Muthu Veni, S.Kavya, "Efficient message forwarding in MANETs", International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE), Volume 1,Issue 1, August 2015,pp:6-9
[6] R. Naidu, J. Nunn, A. Kelly, Socio-behavioural factors andearly childhood caries: a cross-sectional study of preschoolchildren in central Trinidad, BMC Oral Health 13 (2013) 30.
[7] B. Jose, N.M. King, Early childhood caries lesions in preschoolchildren in Kerala, India, Pediatr. Dent. 25 (2003) 594–600
.[8] F. Szatko, M. Wierzbicka, E. Dybizbanska, I. Struzycka, E.Iwanicka-Frankowska, Oral health of Polish three-year-oldsand mothers' oral health-related knowledge, CommunityDent. Health 21 (2004) 175–180.
[9] K.M.G. Cari˜no, K. Shinada, Y. Kawaguchi, Early childhoodcaries in northern Philippines, Community Dent.OralEpidemiol. 31 (2003) 81–89
.[10] R.J. Schroth, P.J. Smith, J.C. Whalen, Ch. Lekic, M.E.K. Moffatt,Prevalence of caries among preschool-aged children in anorthern Manitoba community, J. Can. Dent. Assoc. 71(2005), 27-27f
.[11] L.D. Rajab, M. Hamdan, Early childhood caries and riskfactors in Jordan, Community Dent. Health 19 (2002) 224–229
[12] S.N. Kiwanuka, A.N. Åstrøm, T.A. Trovik, Dental cariesexperience and its relationship to social and behaviouralfactors among 3–5-year-old children in Uganda, Int. J.Paediatr. Dent. 14 (2004) 336–346
[13] L.M. Kaste, R.H. Selwitz, R.J. Oldakowski, J. Brunelle, D.M.Winn, L.J. Brown, Coronal caries in the primary andpermanent dentition of children and adolescents 1–17 yearsof age: United States, 1988–1991, J. Dent. Res. 75 (1996)631–641.
[14] K. Hallett, P. O'Rourke, Social and behavioural determinantsof early childhood caries, Aust. Dent. J. 48 (2003) 27–33.
[15] M. Vulovi´c, M. Carevi´c, An infective nature of dental caries,Stomatolo˘skiglasnikSrbije 45 (1998) 5–9 (Serbian).
[16] L. Powell, Caries prediction: a review of the literature,Community Dent. Oral Epidemiol. 26 (1998) 361–371.[17] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone,Classification and Regression Trees, Wadsworth, BelmontCA, 1984.