World Journal of Food Science and Technology
2023; 7(2): 24-29
http://www.sciencepublishinggroup.com/j/wjfst
doi: 10.11648/j.wjfst.20230702.12
ISSN: 2637-6016 (Print); ISSN: 2637-6024 (Online)

Predicting 7-24 Months Childs Infectious Disease and Responsible Supplementary Food for Infectious Disease: A Machine Learning Approach

Md Nesar Uddin Sorkar[1, *], Md. Roquib Uddin Sorkar[2]

[1]Department of Statistics, Shahjalal University of Science and Technology, Sylhet, Bangladesh

[2]Department of Civil Engineering, Southern University Bangladesh, Chittagong, Bangladesh

Email address:

nesarsust@gmail.com (Md Nesar Uddin Sorkar), roquibsorkar116@gmail.com (Md. Roquib Uddin Sorkar)

[*]Corresponding author

To cite this article:

Md Nesar Uddin Sorkar, Md. Roquib Uddin Sorkar. Predicting 7-24 Months Childs Infectious Disease and Responsible Supplementary Food for Infectious Disease: A Machine Learning Approach. *World Journal of Food Science and Technology*. Vol. 7, No. 2, 2023, pp. 24-29. doi: 10.11648/j.wjfst.20230702.12

Received: February 6, 2023; **Accepted**: February 23, 2023; **Published**: May 25, 2023

Abstract: Supplementary foods are foods that babies consume in addition to breast milk. Supplementary food is essential for a baby's health. Many malnutrition problems in babies are caused by a lack of supplementary feeding. However, not all supplementary foods are beneficial to a baby's health. The goal of this research was to identify the supplementary foods that cause infectious diseases in babies. Simultaneously, it attempted to predict infectious diseases in babies. This secondary data was collected from BDHS 2014. Here various methods such as percentage distribution, association test, logistic regression, and association rule mining have been used to find out the responsible factors for infectious diseases. At the same time, Decision Tree, Random Forest, Support Vector Machine, and Naiv Bays have been used to predict infectious diseases in babies. According to association test, association rule mining, and logistic regression, it can be said that babies who eat juices, pumpkin or carrot, liver or heart, lentils or nuts, other liquids, potatoes, bread or noodles, plain water, and other foods are more likely to be infected with infectious diseases. On the other hand, babies who eat tinned milk, mango or papaya, and baby formula are less likely to get an infectious disease. Furthermore, for this data, Random Forest is the best classifier. Therefore, it can be said that these significant variables may be responsible for the infectious disease of babies. The government and numerous NGOs should make people aware of this significant supplementary food so that future generations can be disease-free.

Keywords: Prediction, Infectious, Disease, Supplementary, Food, Machine, Learning

1. Introduction

Supplemental feeding is defined as the provision of extra food outside the usual ratio of the diet of children or families [1]. Children were fed micronutrient-rich foods locally to improve their nutritional status [2]. Supplemental feeding with ready-to-use therapeutic foods promotes children to better growth than standard fortified cereal foods [3]. Mixed feeding (breast milk and supplementary food) had the highest prevalence of diarrhea (15.2%) [4]. Early supplementary feeding might result in health issues including diarrhea, respiratory tract infections, and other complications [5]. Growth slowed starting at the age of three months when the prevalence of infectious diseases gradually increased and appropriate supplementary foods were lacking [6]. In most regions, there were high rates of diarrhea due to insufficient water supplies [7]. Supplementary feeding before 6 months showed a significant association with acute diarrhea (AOR = 6.49, 95% CI: 2.01-20.96) [8]. Infants aged 6-8 months who were given supplemental foods were more likely to have diarrhea than those who were not given supplementary foods [9].

The purpose of this study was to determine which supplementary foods cause infectious diseases in babies. It also attempted to predict infectious diseases in babies.

Babies often suffer from diarrhea and blood in the stools. Many babies die each year as a result of these diseases. These diseases are now a threat to babies. So it has become very essential to find out which foods are causing these diseases in babies.

This study uses data from Bangladesh Demographic and Health Survey (BDHS) 2014. Since the data from BDHS 2014 has been used here, this study will be able to represent all the babies of Bangladesh. So it can be said that this study will cover the whole of Bangladesh.

No one has worked out which supplementary food is responsible for the infectious disease of the baby. So here’s an attempt to work with it.

Since BDHS 2014 data has been used here, it represents all the babies in Bangladesh. Once the study is completed, all the babies in Bangladesh will be beneficiaries.

2. Methods

2.1. Data Source

This secondary data was gathered from the BDHS 2014. Around the age of six months, an infant's energy and nutrient requirements begin to exceed those provided by breast milk, necessitating the use of supplemental foods to meet those requirements [10]. Here is an attempt to find out which supplementary foods are responsible for the infectious diseases of babies. So in this study, data has been taken from only 7-24-month-old babies.

2.2. Study Area

Bangladesh has widespread poverty and is one of the world's most densely inhabited countries. Alluvial soil covers the majority of the terrain, which is low and flat. With moderate winters and hot, humid summers, Bangladesh's tropical climate is dominated by seasonal monsoons. Bangladesh is 85 meters above sea level, and the fertile delta is frequently hit by natural calamities including floods, cyclones, tidal bores, and drought [11].

2.3. Study Design and Sample Size

A nationally representative data set from the 2014 BDHS was used. The BDHS adopts stratified random sampling with two stages of selection. To begin, 600 enumeration areas (EAs) were chosen with a probability proportional to their size, with 207 urban and 393 rural EAs. Second, 17,989 households were randomly selected following systemic sampling, and 17,300 of them were successfully questioned. In all, 17,863 married women aged 15 to 49 years old and mothers of 7886 children younger than 5 years old were questioned regarding pregnancy, postnatal care, immunization, and health issues including diseases and foods were questioned [12].

2.4. Outcome Variable

The infectious disease for this study is the outcome variable. It has been developed by the union of diarrhea and blood in the stools variable. If a child's mother reported that the baby had diarrhea or blood in the stools in the two weeks before the survey, the baby was considered to be suffering from diarrhea or blood in the stools.

2.5. Predictor Variables

For this study, juice, tinned milk, other foods, eggs, pumpkin or carrot, mango or papaya, liver or heart, fish or shellfish, lentils or nuts, other liquid, bread or noodles, potato, meat, leafy vegetable, any other Fruits, plain water, baby formula, and yogurt are predictor variables.

2.6. Statistical Analysis

To determine the responsible factors for infectious diseases, several methods such as percentage distribution, association test, logistic regression, and association rule mining were used. Decision Tree, Random Forest, Support Vector Machine, and Naiv Bays have all been used to predict infectious diseases in babies. SPSS 23 and R 3.6.3 have been used for data analysis.

2.7. Variable Selection

Important variables have been extracted through the backward elimination technique.

3. Results

3.1. Percentage Distribution

Table 1 demonstrates that juice is consumed by 27% of babies, tinned milk by 21%, and plain water by 98 percent. Furthermore, 81% of babies consume bread or noodles, 26% eat eggs, 42% eat potatoes, and 33% eat vegetables.

Table 1. *Percentage of Used Supplementary Food.*

Variants (Yes)	Frequency (%)
Juice	54 (16.27%)
Tinned milk	70 (21.08%)
Other food	146 (43.98%)
Eggs	88 (26.51%)
Pumpkin or carrot	22 (6.63%)
Mango or papaya	66 (19.88%)
Liver or heart	18 (5.42%)
Fish or shellfish	98 (29.52%)
Lentils or nuts	22 (6.63%)
Other liquid	48 (14.46%)
Bread or noodles	272 (81.93%)
Potato	142 (42.77%)
Meat	58 (17.47%)
Leafy vegetable	112 (33.73%)
Any other fruits	76 (22.89%)
Plain water	324 (97.59%)
Baby formula	28 (8.43%)
Yogurt	12 (3.61%)

3.2. Association Test

According to the association test, Table 2 shows that the p-value of juices, tinned milk, liver or heart, lentils or nuts, other liquids, and yogurt variables is less than 0.1. So it can be said that these variables are statistically significant. Therefore, it appears that juices, tinned milk, liver or heart, lentils or nuts, other liquids, and yogurt variables are associated with suffering from infectious diseases.

Table 2. *Association between Infectious Disease and Supplementary Food.*

Variants (Yes)	Infectious disease (Yes)	χ^2	p-value
Juice	34 (20.73%)	4.7476	0.0293
Tinned milk	28 (17.07%)	3.1339	0.0767
Other food	74 (45.12%)	0.1728	0.6777
Eggs	50 (30.49%)	2.6377	0.1044
Pumpkin or carrot	14 (8.54%)	1.911	0.1668
Mango or papaya	28 (17.07%)	1.6025	0.2055
Liver or heart	14 (8.54%)	6.1325	0.0133
Fish or shellfish	48 (29.27%)	0.0097	0.9215
Lentils or nuts	16 (9.76%)	5.1303	0.0235
Other liquid	32 (19.51%)	6.6946	0.0097
Bread or noodles	138 (84.15%)	1.0775	0.2993
Potato	76 (46.34%)	1.6879	0.1939
Meat	34 (20.73%)	2.3916	0.122
Leafy vegetable	56 (34.15%)	0.0245	0.8755
Any other fruits	38 (23.17%)	0.0143	0.9048
Plain water	162 (98.78%)	1.9521	0.1624
Baby formula	12 (7.32%)	0.5233	0.4694
Yogurt	10 (6.1%)	5.736	0.0166

3.3. Influence of Supplementary Food on Infectious Disease

According to the logistic regression, Table 3 states that the p-value of juice, tinned milk, pumpkin or carrot, mango or papaya, liver or heart, lentils or nuts, and other liquid variables is less than 0.1. So it can be said that these variables are statistically significant. The odds ratio of juice here is 2.53, which means that babies who drink juice are 2.53 times more likely to get the infectious disease than babies who do not drink juice. The odds ratio of mango or papaya is 0.354, which means that babies who eat mango or papaya are 0.354 times less likely to be infected with infectious diseases than all babies who do not eat mango or papaya. Similarly, the odds ratio of other variables can also be explained.

Therefore, it is evident that babies who eat juices, pumpkin or carrot, liver or heart, lentils or nuts, and other liquids are more likely to be infected with infectious diseases. On the other hand, babies who eat tinned milk, mango, or papaya are less likely to get an infectious disease.

Table 3. *Influence of Supplementary Food on Infectious Disease.*

Covariates (Yes)	OR	95% C.I. for OR		p-value
		Lower	Upper	
juice	2.539	1.222	5.274	.013
tinned milk	.428	.230	.796	.007
other food	.859	.519	1.423	.555
eggs	1.594	.870	2.922	.131
pumpkin or carrot	2.575	.862	7.692	.090
mango or papaya	.354	.171	.731	.005
liver or heart	6.800	1.514	30.530	.012
fish or shellfish	1.360	.753	2.457	.308
lentils or nuts	4.273	1.462	12.488	.008
other liquid	1.871	.897	3.904	.095
bread or noodles	1.139	.579	2.241	.707
potato	1.340	.765	2.347	.307
meat	.829	.369	1.860	.649
leafy vegetable	.760	.434	1.328	.335
any other fruits	.608	.328	1.125	.113
plain water	2.437	.455	13.051	.298
baby formula	.559	.224	1.391	.211
yogurt	3.831	.741	19.808	.109

Note: Variable Characteristics Level 'No' is considered as Reference Group

3.4. Association Rule Mining

Figure 1 shows that the bubble of rule-1 is very darkly colored. Therefore, it can be said that all babies who eat potatoes suffer from infectious diseases. Similarly, it seems that the bubbles of rule-2 to rule-7 are also darkly colored. Therefore, it can be said that all the babies who do not eat tinned milk, mango, or baby formula, suffer from more infectious diseases. On the other hand, babies who eat bread or noodles, plain water, and other food are more prone to infectious diseases.

3.5. Prediction Algorithms Characteristics

In this case, the infectious disease is predicted based on several supplementary foods. Then the accuracy, sensitivity, specificity, and value of the area under the curve are determined.

3.5.1. Accuracy

As can be seen from Figure 2 (a), the accuracy of the Decision Tree, Random Forest, Support Vector Machine, and Naive Base Classifier is about 72%, 85%, 65%, and 72%, respectively. Here random forest accuracy is highest, so in this case, random forest is the best classifier.

3.5.2. Sensitivity

Sensitivity refers to a test's ability to detect true events as true. Here Figure 2 (b) shows the highest (about 82%) sensitivity of random forest. In other words, the Random Forest can detect infectious diseases in 82 babies out of babies'. So, in this case, also random forest is the best classifier.

3.5.3. Specificity

Specificity refers to a test's ability to detect false events as false. Figure 2 (c) also shows that random forest specificity is the highest (about 88%). So here also, the Random Forest is the best classifier.

3.5.4. Area Under Curve (AUC)

Figure 2 (d) shows that the area under the curve of Random Forest is the highest. The amount is about 85%. Therefore, here also the Random Forest is the best classifier.

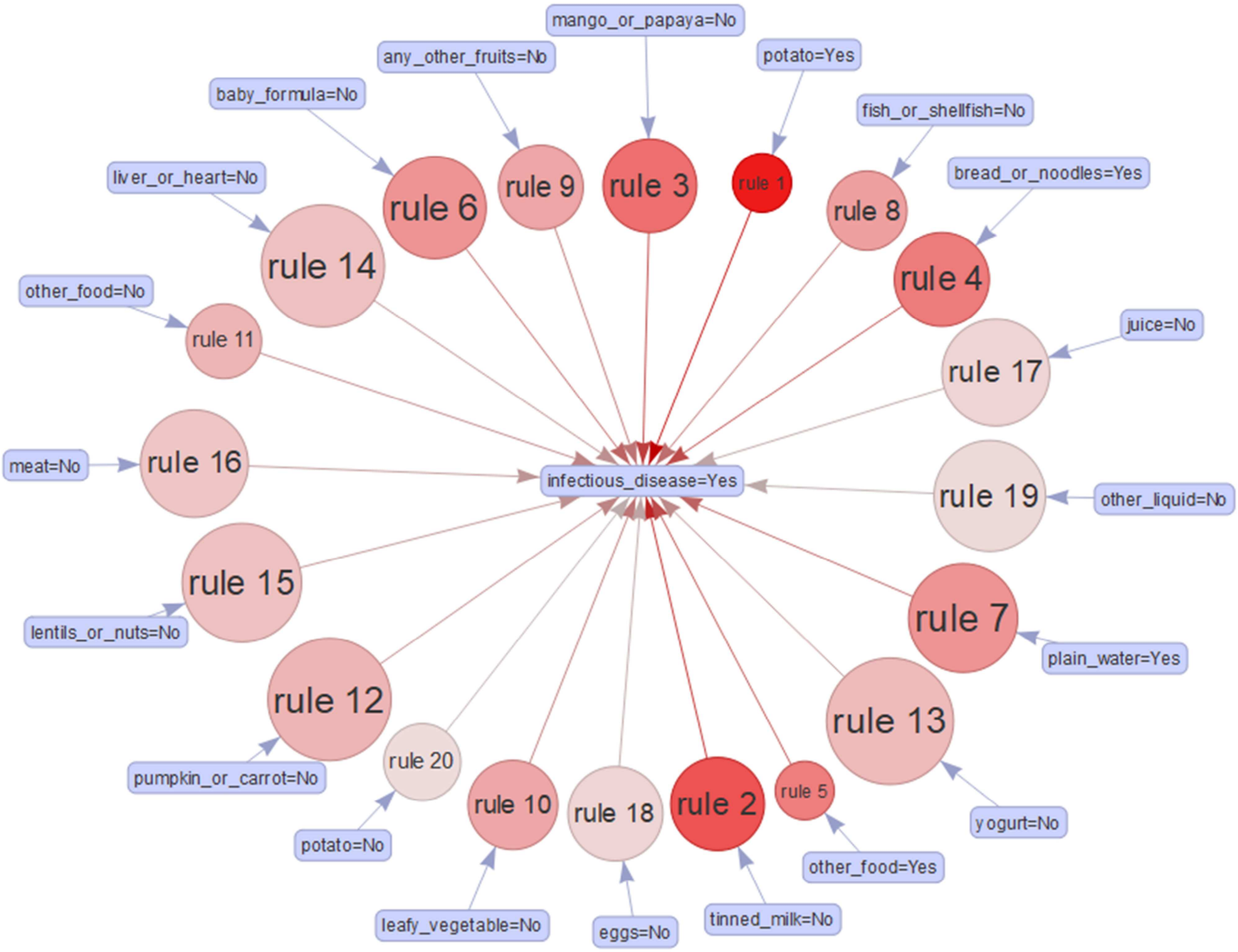
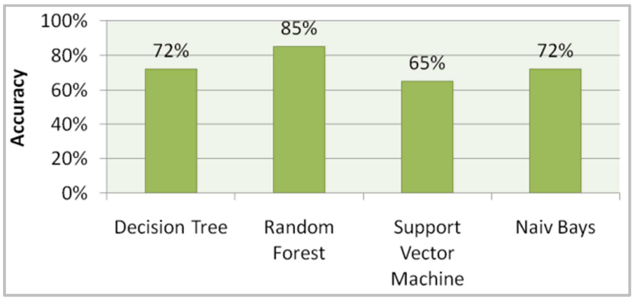
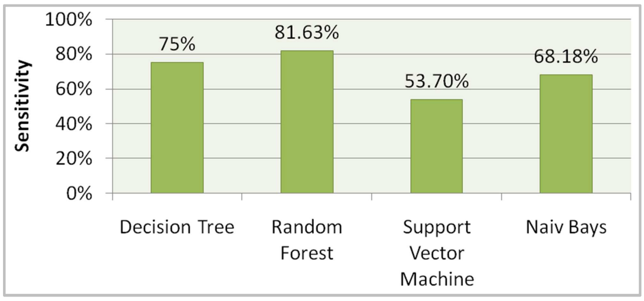


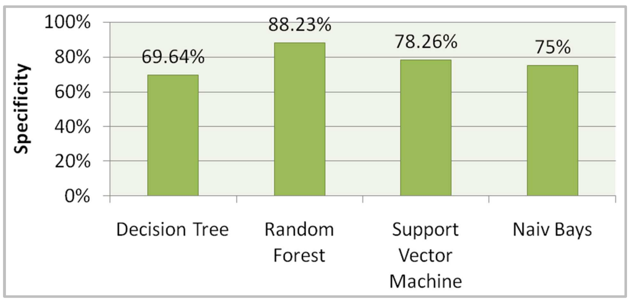
Figure 1. *Association Plot.*



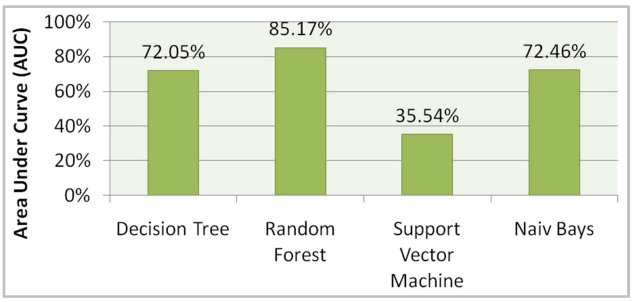
a) Accuracy



b) Sensitivity



c) Specificity



d) AUC

Figure 2. *Prediction Algorithms Characteristics.*

3.5.5. Brier Score

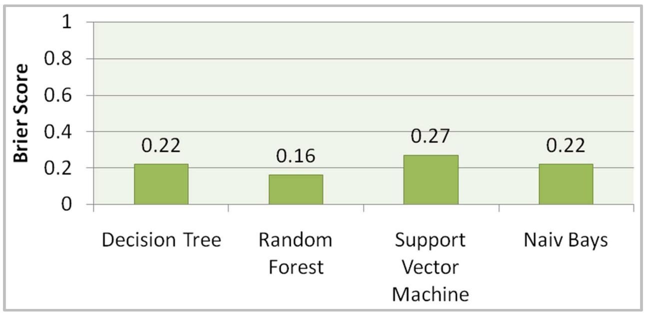


Figure 3. *Brier Score.*

A brier score is a method to validate the accuracy of a probability forecast. A probability forecast refers to a specific event. The lower the Brier Score, the better the method. In Figure 3, the Brier Score of the Random Forest is the lowest (0.16). So, in this case, the random forest is also the best classifier.

3.5.6. Comparison

From Table 4, it can be said that the value of accuracy, sensitivity, specificity, and the area under the curve of random forest is highest than the remaining three methods. The brier score of random forest is the lowest of the remaining three methods. So overall random forest is the best classifier.

Table 4. *Comparison of Different Algorithms.*

Method	Accuracy	Sensitivity	Specificity	Area Under the Curve	Brier Score
Decision tree	72%	75%	69.64%	72.05%	0.22
Random Forest	85%	81.63%	88.23%	85.17%	0.16
Support vector machine	65%	53.70%	78.26%	35.54%	0.27
Naiv Bays	72%	68.18%	75%	72.46%	0.22

4. Discussion

A higher risk of diarrhea was associated with bottle feeding and the introduction of supplemental foods [13]. The prevalence of diarrhea was higher in infants who were introduced to supplemental foods [14]. Which is similar to this study. Because, this study also found that babies who eat juices, pumpkin or carrot, liver or heart, lentils or nuts, other liquids, potatoes, bread or noodles, plain water, and other foods supplements are more likely to be infected with infectious diseases.

Supplemental feeding has a positive impact on growth [15]. It has also supported this study. Because, this study also found that babies who eat tinned milk, mango or papaya, and baby formula supplements are less likely to get an infectious disease.

5. Limitation

The data used in this study has a sample size of 332, which is very small. The result would be better if the sample size was larger.

6. Conclusion

According to association test, association rule mining, and logistic regression, it can be said that babies who eat juices, pumpkin or carrot, liver or heart, lentils or nuts, other liquids, potatoes, bread or noodles, plain water, and other foods are more likely to be infected with infectious diseases. On the other hand, babies who eat tinned milk, mango or papaya, and baby formula are less likely to get an infectious disease.

7. Recommendation

People should be made aware of this significant supplementary food by the government and various NGOs. Explain to them which supplementary foods are good for the baby's health and which are bad for the baby's health.

Conflict of Interests

The authors declare that they have no competing interests.

Acknowledgements

A lot of thanks to the Bangladesh Demographic and Health Survey for using their collected data.

References

[1] Y. Sguassero, M. de Onis, A. M. Bonotti, and G. Carroli, "Community-based supplementary feeding for promoting the growth of children under five years of age in low and middle income countries," *Cochrane Database Syst. Rev.*, 2012, doi: 10.1002/14651858.cd005039.pub3.

[2] P. Nestel *et al.*, "Complementary food supplements to achieve micronutrient adequacy for infants and young children," *Journal of Pediatric Gastroenterology and Nutrition*. 2003. doi: 10.1097/00005176-200303000-00005.

[3] M. P. Patel, H. L. Sandige, M. J. Ndekha, A. Briend, P. Ashorn, and M. J. Manary, "Supplemental feeding with ready-to-use therapeutic food in Malawian children at risk of malnutrition," *J. Heal. Popul. Nutr.*, 2005, doi: 10.3329/jhpn.v23i4.352.

[4] R. Chander, P. Vimesh, and S. Singh, "Prevalence of Diarrhoea in Infants and its Relation to Feeding and Weaning Practices," *JMS Ski.*, 2011, doi: 10.33883/jms.v14i1.65.

[5] F. O. Fauthrisna, M. Masrul, and E. Chundrayetti, "Hubungan Pemberian Makanan Tambahan Dini terhadap Status Gizi Bayi Usia 4-6 Bulan di Daerah Pantai Kota Padang Tahun 2013," *J. Kesehat. Andalas*, 2015, doi: 10.25077/jka.v4i3.376.

[6] U. Lindskog, B. Bjorksten, and M. Gebre-Medhin, "Infant care in rural Malawi. A prospective study of morbidity and growth in relation to environmental factors," *Ann. Trop. Paediatr.*, 1994, doi: 10.1080/02724936.1994.11747690.

[7] E. C. Okele and U. A. Onyechi, "Factors affecting weaning in rural and urban areas of South Eastern Nigeria," *Cajanus*, 1994.

[8] A. Fenta, K. Alemu, and D. A. Angaw, "Prevalence and associated factors of acute diarrhea among under-five children in Kamashi district, western Ethiopia: Community-based study," *BMC Pediatr.*, 2020, doi: 10.1186/s12887-020-02138-1.

[9] F. A. Ogbo, H. Nguyen, S. Naz, K. E. Agho, and A. Page, "The association between infant and young child feeding practices and diarrhoea in Tanzanian children," *Trop. Med. Health*, 2018, doi: 10.1186/s41182-018-0084-y.

[10] "Infant and young child feeding." https://www.who.int/news-room/fact-sheets/detail/infant-and-young-child-feeding (accessed Aug. 15, 2021).

[11] M. I. K. Imran, M. U. A. Inshafi, R. Sheikh, M. A. B. Chowdhury, and M. J. Uddin, "Risk factors for acute respiratory infection in children younger than five years in Bangladesh," *Public Health*, 2019, doi: 10.1016/j.puhe.2019.05.011.

[12] N. I. of P. R. and T.- NIPORT/Bangladesh, M. and Associates, and I. International, "Bangladesh Demographic and Health Survey 2014." Mar. 01, 2016. Accessed: Aug. 15, 2021. [Online]. Available: https://dhsprogram.com/publications/publication-fr311-dhs-final-reports.cfm

[13] F. A. Ogbo, A. Page, J. Idoko, F. Claudio, and K. E. Agho, "Diarrhoea and Suboptimal Feeding Practices in Nigeria: Evidence from the National Household Surveys," *Paediatr. Perinat. Epidemiol.*, 2016, doi: 10.1111/ppe.12293.

[14] M. V. Dhami, F. A. Ogbo, T. M. O. Diallo, and K. E. Agho, "Regional analysis of associations between infant and young child feeding practices and diarrhoea in indian children," *Int. J. Environ. Res. Public Health*, 2020, doi: 10.3390/ijerph17134740.

[15] E. Kristjansson *et al.*, "Food supplementation for improving the physical and psychosocial health of socio-economically disadvantaged children aged three months to five years," *Cochrane Database of Systematic Reviews*. 2015. doi: 10.1002/14651858.CD009924.pub2.