Journal of Computer Science

Original Research Paper

# Performance Evaluation of Machine Learning-Based Algorithms to Predict the Early Childhood Development Among Under Five Children in Bangladesh

[1]Md. Ismail Hossain, [2]Iqramul Haq, [3,4]Ashis Talukder, [5]Sharmin Suraiya, [6]Mofasser Rahman, [1]Ahmed Abdus Saleh Saleheen, [7]Md. Injamul Haq Methun, [1]Md. Jakaria Habib, [1]Md. Sanwar Hossain, [5]Md. Iqbal Hossain Nayan and [8]Sadiq Hussain

[1]*Department of Statistics, Jagannath University, Bangladesh*
[2]*Department of Agricultural Statistics, Sher-e-Bangla Agricultural University, Bangladesh*
[3]*Department of Statistics Discipline, Khulna University, Bangladesh*
[4]*National Centre for Epidemiology and Population Health, Australian National University, Canberra, ACT 2600, Australia*
[5] *Department of Pharmacy, Northern University, Bangladesh*
[6]*Department of Agribusiness and Marketing, Sher-e-Bangla Agricultural University, Bangladesh*
[7]*Department of Statistics Discipline, Tejgaon College, Bangladesh*
[8]*Examination Branch, Dibrugarh University, India*

*Article history*
Received: 19-10-2022
Revised: 04-02-2023
Accepted: 10-02-2023

Corresponding Author:
Sadiq Hussain
Examination Branch,
Dibrugarh University, India
Email: sadiq@dibru.ac.in

**Abstract:** In this research, an effort has been made to apply a number of classifiers to predict Early Child Development (ECD) in the context of Bangladesh using the Bangladesh multiple indicator cluster survey, 2019 data set (i.e., to evaluate which sort of algorithm best identifies ECDI). To predict the ECD, nine well-known machine learning algorithms were applied, including Linear Regression (LR), Random Forest (RF), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Naïve Bayes (NB), Least Absolute Shrinkage and Selection Operation (LASSO), Classification Trees (CT), AdaBoost and Neural Network (NN). Children aged 48-59 months who were female, attending early education, reading three or more children's books, having playthings, having normal nutritional status, and were not disabled had a higher percentage of completing at least three childhood development domains, according to the bivariate analysis results. We found several performance parameters for the classification of early childhood development, including the following: Accuracy (LR) = 67.87%, AUC (LR) = 67.49%; Accuracy (RF) = 67.23%, AUC (RF) = 67.19%; Accuracy (SVM) = 67.37%, AUC (SVM) = 67.64%; Accuracy (NB) = 67.55%, AUC (NB) = 66.80%; Accuracy (LASSO) = 68.04%, AUC (LASSO) = 67.75. Based on the results of this investigation, LASSO regression predicts the ECD in Bangladeshi children moderately better than any other machine learning method utilized in this study.

**Keywords:** Early Childhood Development, ML Algorithm, LASSO Regression, Bangladesh

## Introduction

The most important stage of a child's life is the healthy era of Early Childhood Development (ECD), a process of growth in physical, cognitive, social-emotional, and language skills. ECD determines individual academic, behavioral, and economic successes (Black *et al*., 2017; Daelmans *et al*., 2017). Due to time and financial limitations, this stage should be seen as a window of opportunity to maximize the development of the child's physical, social, emotional, and language or cognitive domains; otherwise, success may be more challenging later in life (Irwin, 2007; Walker *et al*., 2011).

Previous research conducted in low- and middle-income nations found that every year, almost 249 million children fall short of their developmental potential due to a lack of resources (Lu *et al*., 2016; McCoy *et al*., 2016; McDonald and Rennie, 2011). Recent cross-national studies in LMICs have suggested that factors like child age, breastfeeding, household characteristics, malnutrition, poor health, a lack of stimulating learning environments, and mothers' marital status can all have an impact on Early

Childhood Development (ECD) (Po *et al*., 2011; Kleimola *et al*., 2015; Grantham-McGregor *et al*., 2007). Additionally, a number of biological risk factors, including shortened gestational periods (Gutbrod *et al*., 2000; Espel *et al*., 2014), low birth weights (Tong *et al*., 2006; Gill *et al*., 2013; Donald *et al*., 2019; Sania *et al*., 2019; Upadhyay *et al*., 2019), anemia (Sungthong *et al*., 2002) and stunting (Haile *et al*., 2016; Woldehanna *et al*., 2017), are known to contribute to poor cognitive performance, which is prevalent in LMIC. The development of the brain and later cognitive performance are significantly impacted by poor nutrition, which is one of the causes of stunting (de Onis and Branca, 2016) and early childhood diseases including diarrhea (Georgieff, 2007; Georgieff *et al*., 2018; Niehaus *et al*., 2002; Lorntz *et al*., 2006; Kvestad *et al*., 2015).

The sustainable development goals (4.2.1) incorporate early childhood development as an indicator of global progress (ECPC, 2021; UN, 2021). The world health organization has pledged to lead a worldwide strategy for the health of women, children, and adolescents from 2016 to 2030. As a result, UNESCO, UNICEF, the world bank, and numerous institutions are directed to promote early childhood development (Richter *et al*., 2017). Children may develop to their full potential with the support of five nurturing care components, including safety, opportunities for learning, responsive caregiving, good health, and enough nourishment (WHO, 2018). In addition to these components, a number of factors, including socioeconomic status, continued breastfeeding, and parental education, have consistently been linked to children's cognitive development (Lee *et al*., 2016; Duc, 2009; Ribe *et al*., 2018; Roberts *et al*., 1999; Christensen *et al*., 2014). According to a recent study from Bangladesh, the level of education and wealth of the mother had a substantial impact on the literacy and numeracy abilities of children under the age of five (Hossain *et al*., 2023).

As ECD is an important indicator of global growth, governments and policymakers should take appropriate policy interventions to improve the Early Childhood Development (ECD) index. Various statistical analysis has been used in the past to assess the level of development of under-five children. A study conducted in 2021 based on three low- and middle-income countries (Bangladesh, Costa Rica, and Ghana), used a binary logistic regression model to estimate the effects of socioeconomic and demographic factors on under-five childhood development (Haq *et al*., 2021). This study revealed that urban children in Costa Rica were more likely to develop early than children from rural areas, whereas Ghana had different results (Haq *et al*., 2021). However, it is important to accurately estimate how young infants under the age of five will develop. Machine learning, a scientific approach to studying vast volumes of data to uncover associations, is now widely employed in a variety of disciplines, including social sciences, healthcare, and medicine and it may be used to develop models for predictions. Only a few research have been conducted prior to this one that attempted to predict early childhood growth using machine learning techniques. In this study, we, therefore, used nine well-known machine learning algorithms: Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Nave Bayes (NB), Least Absolute Shrinkage and Selection Operation (LASSO), Classification Trees (CT), AdaBoost and Neural Network (NN). We compared these algorithms based on their performance metrics and the value of the area under the ROC curve. To the best of our knowledge, this is the first study in Bangladesh to predict the ECD status of children under the age of five using such a strategy.

## Materials and Methods

### *Data Sources*

This analysis of this study was based on the Bangladesh multiple indicator cluster survey, 2019, a nationally representative cross-sectional survey in which data on mothers and children were gathered by the Bangladesh Bureau of Statistics (BBS), and financial support was provided by UNICEF Bangladesh.

### *Sample Design and Sample Size*

Two rounds of sample selection using the stratified cluster sampling technique were conducted. In the initial stage, the sampling unit-3220 Enumeration Areas (EAs) was chosen using the sampling frame from the 2011 Bangladesh population and housing census. From each sample of EAs, a systematic sample of 20 households was taken and 64,400 households were chosen in the second step. From the chosen families, 24,686 mothers and carers who had children under the age of five were eligible. 23,099 mothers or other carers were successfully questioned as a result. Our target age group for this study was children between the ages of 36 and 59 months. As a result, we did not include babies under 36 months in the analysis of this study. Therefore, we excluded children under 36 months from this study analysis. That is, our final sample size was 9442 children aged 36-59 months. Figure 1 depicts the whole procedure.

### *Dependent Variable*

The Early Childhood Development (ECD) status was the outcome variable. A child was said to be considered "developed" and given the code "1" if they were able to develop at least three of the four domains literacy and numeracy, physical development, social and emotional development, and learning the four. The youngster was coded as "0" if not developed and otherwise. Based on earlier research, this approach was developed (Hossain *et al*., 2021). The resulting binary variable is regarded as the main outcome variable.

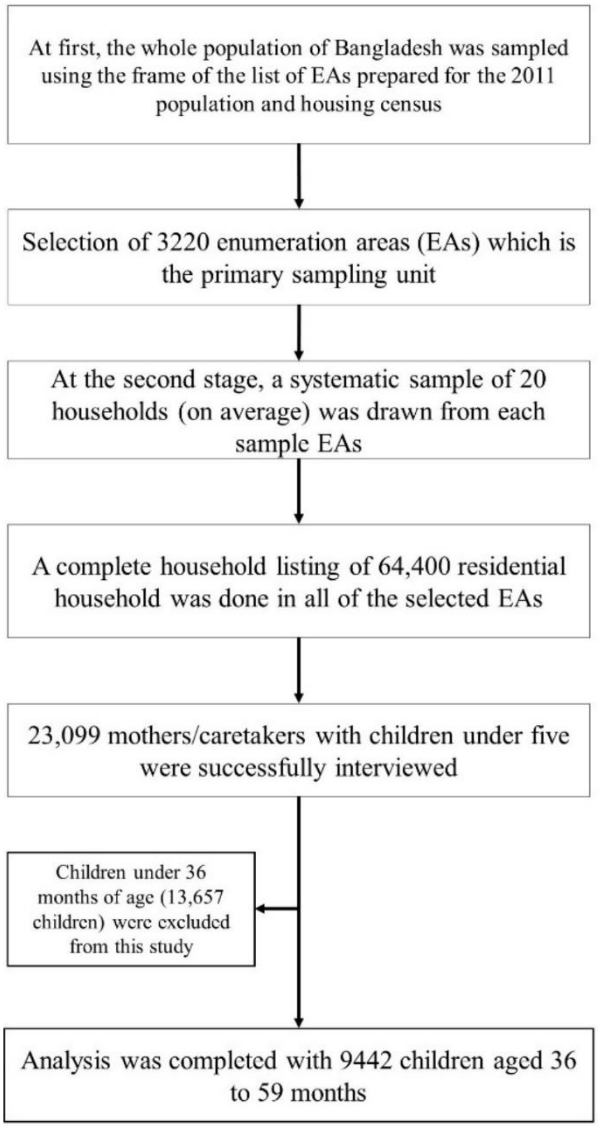


**Fig. 1:** Sample selection procedure

### *Explanatory Variables*

A set of covariates related to the early childhood development status of Bangladeshi children was considered for several machine learning algorithms. These variables included socio-demographic and economic related factors, factors related to children, quality of care, and maternal and household factors. The complete list of predictor variables and their coding are found detailed in Table 1.

### *Statistical Analysis*

To present the summary of the explanatory variables, we use a frequency distribution with percentages. A chi-square test was applied to investigate the relationship between the dependent and explanatory variables. Chi-square statistics are mathematically described as:

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(Observed\ frequencey_i - Expected\ frequencey_i\right)^2}{Expected\ frequencey_i} \quad (1)$$

To predict the outcome variable in the multivariable setting, we employed nine different supervised machine learning methods such as Logistic Regression (LR), Naïve Bayes (NB), Linear Discriminant Analysis (LDA), Random Forest (RF), Support Vector Machine (SVM), Least Absolute Shrinkage and Selection Operation (LASSO), Neural Network (NN), Classification Trees (CT) and AdaBoost as well as we assessed their efficiency using model assessment parameters.

Until now, many proposed classification techniques have been used in public health patterns with high classification accuracy. Many supervised algorithms are being used to calculate the size of the training data set. For example, linear regression, Naïve Bayes, support vector machines, etc., are used for smaller training data sets.

**Table 1:** Description of explanatory variables

| Covariates | Measures |
|---|---|
| Child age | 1 = 36-47; 2 = 48-59 |
| Child sex | 1 = Male; 2 = female |
| Early child education | 1 = Early attending; 2 = late attending |
| Reading children books | 1 = ≥3; 2 = <3 |
| Having play things | 1 = Yes; 2 = No |
| Children underweight status | 1 = Severe; 2 = moderate; 3 = normal |
| Children stunting status | 1 = Severe; 2 = moderate; 3 = normal |
| Child disability | 1 = Yes; 2 = No |
| Maternal marital age | 1 = Early marriage; 2 = late marriage |
| Maternal education level | 1 = No education; 2 = primary; 3 = secondary and above |
| Child ever born | 1 = 1; 2 = 2-3; 3 = >3 |
| Maternal disability | 1 = Yes; 2 = No |
| Household head education level | 1 = No education; 2 = primary; 3 = secondary and above |
| Mass media | 1 = Exposed; 2 = not exposed |
| Residence | 1 = Urban; 2 = rural |
| Division | 1 = Barisal; 2 = Chattogram; 3 = Dhaka; 4 = Khulna; 5 = Mymensingh; 6 = Rajshahi; 7 = Rangpur; 8 = Sylhet |
| Wealth status | 1 = Poorest; 2 = Poor; 3 = Middle; 4 = Rich = 5 = Richest |
| Water source | 1 = Improved; 2 = not improved |

However, Naïve Bayes produces better results for larger training data sets in some cases. Again, if we want to increase the accuracy, more time will be needed to train the training sample. In these cases, algorithms with tuning parameters such as random forest, neural network, boosting, etc., produce a more precise result. But if we choose speed over time, then using classification trees, logistic regression will be better. This study applied the LASSO regression to ignore multicollinearity if it was present in the data. A detailed description of the algorithms used is available in the literature (Vasquez *et al*., 2016; Wu and Zhao, 2011; Jang *et al*., 2015; Buntine, 1992; Anuse and Vyas, 2016; Talukder and Ahammed, 2020).

### *Data Preparation and Model Evaluation*

We first verify the original data set to see if there are any missing values and if so, we remove them before fitting these nine machine learning classifiers. The complete data set is only split into training and testing since there are so few research data points. Here, the ML algorithm is applied to 70% of the total sample, chosen at random, known as the training data set and the remaining 30% of the sample, known as the test data set, is used for verification. The classifiers are trained using a training data set and the model parameters are estimated and assessed using a test data set. 10-fold cross-validation was performed on the training set to see how well it performed on the test set. The accuracy outcomes of the 10-fold cross-validation were displayed using the violin plot, a mix of a box plot, and a kernel density plot. There is no need to normalize categorical data because all explanatory factors were categorical. Since each value has distinct meaning when a feature is categorical, normalizing these features will change their meaning. We encoded the categorical variables as 0 and 1 before fitting the models. Categorical variables are either 0 or 1 when we encode them. As a result, there is little scale variation and no need to use normalization or standardization processes.

The following metrics were used to assess how well the Machine Learning (ML) algorithm performed: The confusion matrix, the Receiver Operating Characteristic (ROC), and the Area Under the Curve (AUC). According to the convention, a confusion matrix can predict four different outcomes: TR = True Positives whether the model correctly identified the ECD class, TN = True Negatives whether the model correctly identified the non-ECD class, FP = False Positives whether the model incorrectly identified non-ECD classes as ECD and FN = False Negatives whether the model incorrectly identified ECD classes as non-ECD. These four possible outcomes are typically used to produce a number of performance indicators, such as accuracy, precision, recall, and the F1-score, to rate the classifier. These measurements can be computed mathematically as follows:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

Both the expected and actual results were used to create the ROC curves. The AUC of the ROC for the test data sets has been averaged in order to assess the discriminating capacity of the ML algorithms (Liu *et al*., 2016). Theoretically, the AUC lies between 0 and 1, with a perfect classifier being able to use the highest value of 1. Given that 0.5 is typically the lowest restriction for random classification, an AUC larger than 0.5 can at least partially distinguish between cases and non-cases (Liaw and Wiener, 2002). We utilized Cohen's kappa statistic in addition to these metrics because it is a more accurate way to assess how well two rating agencies agree. It is determined by using the data set's actual and expected classifications. Cohen's kappa statistic has a value of ≤1. The R-language (version 4.0.0) was utilized for data analysis.

## Results

Among the 9442 mothers or caregivers with children, 51.6% have a male child, and 48.4% have a female child. This study included children between 36 and 59 months of age. So, we categorize children's ages into 36-47 and 48-59 months. Table 2 confirms that the proportion of children between 36 and 47 months is 50.7% and the remaining children (49.3%) are between 48 and 59 months. However, approximately 81.1% of children are late for school. It is observed that 88% of children have read fewer children's books, whereas only 12% have read three or more children's books. The highest number of children with toys (96.9%). We also found that the stunted and underweight status of most children is normal. Table 2 also shows that almost all (97.5%) of the children have no disability. More than half (62.9%) of their mothers got married at an early age and 61.9% of them are secondary or above educated. The mothers who had given birth to 2-3 children were 42 and 98.3% of the mothers had no disability. Slightly, more than half (53.4%) of mothers do not have mass media exposure and the overwhelming majority of mothers live in rural areas (81.4%). The largest group of respondents (21%) comes from the Chattogram division and their wealth status is the poorest (25.5%). Slightly, more than two-fifths (41%) of household heads have secondary or higher education and 96.9% of households have improved water sources.

**Table 2:** Distribution of the percent of socio-demographic and economic characteristics

| Variables | Frequency | Percentage |
|---|---|---|
| Child age (month) | | |
| 36-47 | 4790 | 50.7 |
| 48-59 | 4652 | 49.3 |
| Child sex | | |
| Male | 4873 | 51.6 |
| Female | 4569 | 48.4 |
| Early child education | | |
| Early attending | 1782 | 18.9 |
| Late attending | 7660 | 81.1 |
| Reading children books | | |
| ≥3 | 1134 | 12.0 |
| <3 | 8308 | 88.0 |
| Having playthings | | |
| Yes | 9148 | 96.9 |
| No | 294 | 3.1 |
| Children stunting status | | |
| Severe | 755 | 8.0 |
| Moderate | 1925 | 20.4 |
| Normal | 6762 | 71.6 |
| Children underweight status | | |
| Severe | 441 | 4.7 |
| Moderate | 1888 | 20.0 |
| Normal | 7113 | 75.3 |
| Child disability | | |
| Yes | 238 | 2.5 |
| No | 9204 | 97.5 |
| Maternal marital age | | |
| Early marriage | 5940 | 62.9 |
| Late marriage | 3502 | 37.1 |
| Maternal education level | | |
| No education | 1255 | 13.3 |
| Primary education | 2340 | 24.8 |
| Secondary and above | 5847 | 61.9 |
| Child ever born | | |
| 1 | 2300 | 24.4 |
| 2-3 | 3979 | 42.1 |
| >3 | 3163 | 33.5 |
| Maternal disability | | |
| Yes | 162 | 1.7 |
| No | 9280 | 98.3 |
| Mass media | | |
| Exposed | 4396 | 46.6 |
| Not exposed | 5046 | 53.4 |
| Residence | | |
| Urban | 1754 | 18.6 |
| Rural | 7688 | 81.4 |
| Division | | |
| Barisal | 826 | 8.7 |
| Chattogram | 1981 | 21.0 |
| Dhaka | 1807 | 19.1 |
| Khulna | 1314 | 13.9 |
| Mymensingh | 573 | 6.1 |
| Rajshahi | 1031 | 10.9 |
| Rangpur | 1116 | 11.8 |
| Sylhet | 794 | 8.4 |
| Wealth status | | |
| Poorest | 2405 | 25.5 |

**Table 2:** Continue

| Variables | Frequency | Percentage |
|---|---|---|
| Poor | 2001 | 21.2 |
| Middle | 1769 | 18.7 |
| Rich | 1744 | 18.5 |
| Richest | 1523 | 16.1 |
| HH education level | | |
| No education | 2773 | 29.4 |
| Primary education | 2816 | 29.8 |
| Secondary and above | 3853 | 40.8 |
| Water source | | |
| Improved | 9149 | 96.9 |
| Not improved | 293 | 3.1 |

HH = Household Head

Table 3 reveals the association between the percentage distribution of the ECD and socio-demographic characteristics. It also represents factors significantly associated (p-value <0.001 or p-value <0.01) with the ECD, which includes age, child sex, early childhood education, reading children's books, playing things, children's stunting status, underweight status, disability, educational level, exposure to the media, division, wealth status and level of household head education.

The age of the child is significantly and positively associated with early childhood development in Bangladesh, where 80% of children aged 48-59 months were shown to have developed in at least three domains. Child sex also has a significant and positive association with early childhood development, where 70.1% of males and 76.5% of females were shown to be developed.

This study found a significant association between early childhood education and childhood development in Bangladesh. The highest percentage (84.8%) of the children who attended school earlier in their lives is shown to be more developed in at least three domains than others. Again, there is a positive relationship between children reading books and playing with things and early child development and it was discovered that 83.8% of children who have read more than three books are followed by 73.5% of children who have toys to play with.

Children’s nutrition status is also significantly and positively associated with early childhood development. Children who had a normal stunting status (75.1%) and those who had a normal underweight status (74.3%) were shown to develop, followed by 74% of children who did not have a disability. Children whose mothers had secondary and above secondary education were shown to develop, followed by mothers who had exposure to mass media (approximately 76% each). Children in the Rangpur division show the highest percentage of child development in at least three domains (81.1%), as well as children from households with the richest wealth status (82.5%) and secondary and above-secondary educated household heads (77.8%).

**DOI: 10.3844/jcssp.2023.641.653**

**Table 3:** Analysis of the early childhood development track in Bangladesh among under-five children according to various sociodemographic factors

| | Developmentally to track in at least three domains | | | |
|---|---|---|---|---|
| Variables | Yes (%) | No (%) | $\chi^2$ value | p-value |
| Child age | | | | |
| 36-47 | 66.6 | 33.4 | 217.647 | <0.001 |
| 48-59 | 80.0 | 20.0 | | |
| Child sex | | | | |
| Male | 70.1 | 29.9 | 48.775 | <0.001 |
| Female | 76.5 | 23.5 | | |
| Early child education | | | | |
| Early attending | 84.8 | 15.2 | 152.295 | <0.001 |
| Late attending | 70.5 | 29.5 | | |
| Reading children books | | | | |
| ≥3 | 83.8 | 16.2 | 73.656 | <0.001 |
| <3 | 71.7 | 28.3 | | |
| Having playthings | | | | |
| Yes | 73.5 | 26.5 | 13.196 | <0.001 |
| No | 63.9 | 36.1 | | |
| Children stunting status | | | | |
| Severe | 67.5 | 32.5 | 45.408 | <0.001 |
| Moderate | 68.6 | 31.4 | | |
| Normal | 75.1 | 24.9 | | |
| Children underweight status | | | | |
| Severe | 64.2 | 35.8 | 27.694 | <0.001 |
| Moderate | 71.0 | 29.0 | | |
| Normal | 74.3 | 25.7 | | |
| Child disability | | | | |
| Yes | 37.8 | 62.2 | 155.627 | <0.001 |
| No | 74.1 | 25.9 | | |
| Maternal marital age | | | | |
| Early marriage | 72.8 | 27.2 | 6.220 | 0.01 |
| Late marriage | 73.8 | 26.2 | | |
| Maternal education level | | | | |
| No education | 68.0 | 32.0 | 66.275 | <0.001 |
| Primary education | 68.7 | 31.3 | | |
| Secondary and above | 76.1 | 23.9 | | |
| Child ever born | | | | |
| 1 | 76.1 | 23.9 | 15.090 | <0.001 |
| 2-3 | 72.8 | 27.2 | | |
| >3 | 71.5 | 28.5 | | |
| Maternal disability | | | | |
| Yes | 64.8 | 35.2 | 5.882 | 0.01 |
| No | 73.3 | 35.2 | | |
| Mass media | | | | |
| Exposed | 75.8 | 24.2 | 29.108 | <0.001 |
| Not exposed | 70.9 | 29.1 | | |
| Residence | | | | |
| Urban | 75.1 | 24.9 | 3.97 | 0.02 |
| Rural | 72.7 | 27.3 | | |
| Division | | | | |
| Barisal | 67.6 | 32.4 | 195.627 | <0.001 |
| Chattogram | 75.5 | 24.5 | | |
| Dhaka | 80.9 | 19.1 | | |
| Khulna | 68.5 | 31.5 | | |
| Mymensingh | 62.5 | 37.5 | | |
| Rajshahi | 70.2 | 29.8 | | |
| Rangpur | 81.1 | 18.9 | | |
| Sylhet | 64.0 | 36.0 | | |

**Table 3:** Continue

| | | | | |
|---|---|---|---|---|
| Wealth status | | | | |
| Poorest | 68.4 | 31.6 | 103.971 | <0.001 |
| Poor | 70.6 | 29.4 | | |
| Middle | 73.6 | 26.4 | | |
| Rich | 74.2 | 25.8 | | |
| Richest | 82.5 | 17.5 | | |
| HH education level | | | | |
| No education | 70.6 | 29.4 | 70.330 | <0.001 |
| Primary education | 69.5 | 30.5 | | |
| Secondary and above | 77.8 | 22.2 | | |
| Water source | | | | |
| Improved | 73.4 | 26.6 | 6.774 | 0.006 |
| Not improved | 66.6 | 33.4 | | |

HH = Household Head

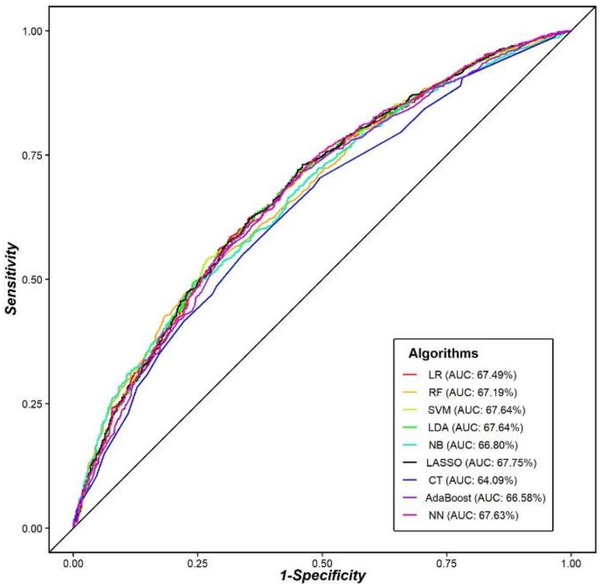


**Fig. 2:** Area under the curve of all algorithms

The prediction results along with the performance parameters for each algorithm are shown in Table 4 and Fig. 2. In Table 4, we observed that the accuracy of the LR classifier is 67.87%. That is, the LR algorithm is approximately 68% correct for prediction. The precision and recall of the fitted model are 80.30 and 74.34%, respectively. Whereas the F1-score is 77.20%. Using the random forest algorithm, we got an accuracy of 67.23% in the test data set with a precession of 75.09% and a recall of 75.16%. Here, the F1-score is 77.05%. In the support vector machine classifier, the cost/capacity parameter is the model tuning parameter generally chosen by cross-validation. In this study, the value of the cost/capacity parameter was 1 and the final accuracy was 67.37% with 80.32% precession and 73.42% recall. The F1-score in this case was 76.71%. Using linear discriminant analysis, the accuracy in the test data set was seen as 67.66% with precession and recall of 80.72 and 73.32%, respectively and the F1-score was 76.85%.

The Naïve Bayes technique demonstrated 67.55% accuracy in predicting the early developmental state of children, with a precision of 79.35% and a recall of 77.25%, according to the test observation data. Finally, we consider the L1 regularization-based lasso regression. Lambda and alpha are the two model parameters in this case. In this investigation, we obtained accuracy, precision, and recall of 68.04, 80.64 and 74.14% respectively. A classification tree helped us achieve a test data accuracy of 64.16 percent. Precision, recall, and F1 in this instance are valued at 79.55, 68.69 and 73.73%, respectively. These percentages for AdaBoost are 66.63% for accuracy, 80.99% for precession, 71.10% for recall, and 75.73% for F1-score. We predicted early childhood development status with 67.62% accuracy using an artificial neural network. Precession, recall, and F1-score are also present and their respective values are 80.88, 73.03, and 76.76%.

With an accuracy of 68.04%, LASSO regression had the best performance among the nine classifiers. Numerous researchers utilize the value of the area under the curve to assess the model's performance even though accuracy is a metric for model evaluation. The curve produced by the area under the curve incorporates all of the cut points because it is a graph that tracks sensitivity and 1-specificity. So, if the model performance is compared to the area under the curve value, it will undoubtedly be better and more accurate than the accuracy. In terms of AUC (Fig. 2), the best model is the LASSO regression (AUC = 67.75%). That is, this model gives more accurate predictions than the remaining eight algorithms. However, in this study, we have found approximately the same results based on accuracy or AUC, however, in terms of the cut point, we give priority to AUC.

**Table 4:** Performance evaluation for nine ML algorithms

| Model name | Accuracy (95% CI) | Cohen's *k* | Precession | Recall | F1 |
|---|---|---|---|---|---|
| Train data set | | | | | |
| LR | 67.31 (66.16, 68.44) | 0.2354 | 80.84 | 72.50 | 76.45 |
| RF | 67.29 (66.15, 68.42) | 0.3325 | 88.80 | 63.28 | 73.90 |
| SVM | 67.67 (66.53, 68.80) | 0.2453 | 81.21 | 72.63 | 76.68 |
| LDA | 67.69 (66.54, 68.81) | 0.2497 | 81.46 | 72.30 | 76.60 |
| NB | 66.52 (65.37, 67.66) | 0.2039 | 79.56 | 73.00 | 76.14 |
| LASSO | 67.76 (66.62, 68.89) | 0.2482 | 81.32 | 72.63 | 76.73 |
| CT | 65.51 (64.35, 66.65) | 0.2293 | 81.54 | 68.33 | 74.35 |
| AdaBoost | 66.28 (65.12, 67.42) | 0.2443 | 82.02 | 69.05 | 74.98 |
| NN | 67.34 (66.19, 68.47) | 0.2462 | 81.46 | 71.68 | 76.26 |
| Test data set | | | | | |
| LR | 67.87 (66.11, 69.59) | 0.2305 | 80.30 | 74.34 | 77.20 |
| RF | 67.23 (65.47, 68.96) | 0.1990 | 75.09 | 75.16 | 77.05 |
| SVM | 67.37 (65.61, 69.10) | 0.2260 | 80.32 | 73.42 | 76.71 |
| LDA | 67.66 (65.90, 69.38) | 0.2362 | 80.72 | 73.32 | 76.85 |
| NB | 67.55 (65.79, 69.27) | 0.2083 | 79.35 | 75.25 | 77.25 |
| LASSO | 68.04 (66.29, 69.76) | 0.2386 | 80.64 | 74.14 | 77.26 |
| CT | 64.16 (62.36, 65.93) | 0.1831 | 79.55 | 68.69 | 73.73 |
| AdaBoost | 66.63 (64.86, 68.37) | 0.2308 | 80.99 | 71.10 | 75.73 |
| NN | 67.62 (65.86, 69.34) | 0.2386 | 80.88 | 73.03 | 76.76 |

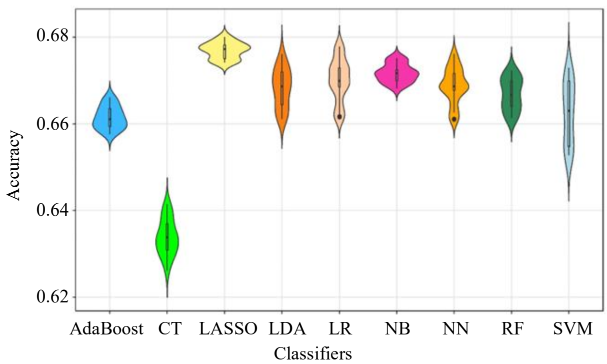


**Fig. 3:** Violin plot shows cross-validation results for each machine learning classifier

The relationship between accuracy and the nine classifiers is depicted in Fig. 3. The distribution of the data in each classifier is depicted by the violin plot's shaded regions. Figure 3 shows that the highest mean accuracy was provided by LASSO, which was then followed by LR, LDA, NN, NB, SVM, RF, AdaBoost, and CT. The whole distribution of 10-fold accuracy may be shown in this violin plot, unlike the box plot.

## Discussion

The study highlighted the early childhood development of under five-year-old children according to their sociodemographic status and these factors had a significant effect on development status. In a related study, (Hossain *et al*., 2021) found that early child development was significantly influenced by the child's age, sex, mother's education, early childhood education, wealth status, nutrition status of the child, reading children's books, having toys and the mother's functional difficulties (Hossain *et al*., 2021). This study can provide an opportunity for researchers and policymakers to experience pertinent highlights from a larger perspective and can help to highlight the importance of key measures. In addition, compared to male children (43.5%), female children who were exposed to violent acts but had enough nourishment and belonged to a poor family were more likely to have their development on track (58.1%) (Athena Infonomics, 2022).

In this study, we employed machine learning methods to predict Bangladesh's early child development status using LR, RF, SVM, LDA, NB, LASSO, CT, AdaBoost, and NN. Based on the accuracy of the early child development prediction rate for the Bangladesh MICS data set for 2019, the current analysis sought to determine which performed better. Furthermore, we were unable to locate a study that used a variety of traditional machine learning techniques to forecast Bangladeshi children's early life development status. The application of data mining methods (Miller *et al*., 2020; Gera and Goel, 2015; Ansari and Gupta, 2012; Bansod *et al*., 2020) and the machine learning strategy in the study of early childhood development data sets have only been the subject of a modest amount of research (Bizzego *et al*., 2023). Studies have been done on the application of ML in the field of neurodevelopmental problems to predict infants who would be born prematurely (Baker and Kandasamy, 2022) and predict early intervention to promote cognitive development (Bowe *et al*., 2022).

With a prediction accuracy of 68.04% among the nine methods, LASSO regression outperformed the other eight algorithms. The analysis's findings also demonstrated that LASSO regression and neural networks produced the same outcomes in terms of Cohen's (Cohen's = 0.2386). When machine learning must deal with the multiclassification problem, Cohen's statistics are a very helpful metric. Researchers occasionally discover that most models have accuracy values that are roughly the same, which does not give a complete view of the model's performance. They try to use Cohen's agreement to interpret this case. The Cohen's $k$ value in this study indicates that the LASSO regression and neural network have achieved better performance than other LR, RF, SVM, LDA, NB, CT, and AdaBoost. Again, based on the accuracy value criterion, we found the highest accuracy value for the LASSO regression model. One thing is that there is very little difference in LASSO and NN accuracy values, but we still highlight that LASSO is the best model for prediction. Although the difference is very small, there is a big difference in terms of interpretation. That is, because NN works in a non-linear fashion, we cannot easily interpret the NN model. On the other hand, LASSO is a type of regression analysis that facilitates variable selection and regularization. That’s why, though there was very little difference in terms of accuracy performance, the LASSO is best for classifying and predicting early childhood development in Bangladesh. A study on the CHICA dataset by Dugan *et al*. (2015) showed that the ID3 model was the best at predicting early childhood obesity with an accurate prediction of 85% (Dugan *et al*., 2015). Another study applied a Classification and Regression Trees (CART) machine learning approach to predict ECD in Rwanda (Athena Infonomics, 2022). In terms of Early Childhood Caries (ECC), four different machine-learning models (XGBoost, random forest, light GBM, and logistic regression) were used and it was observed that there was no significant difference between the AUROC values of the four models (Park *et al*., 2021).

Previous studies in Bangladesh observed that RF algorithms were the best predictors in terms of accuracy when predicting childhood anemia, while SVM identified some important characteristics when used to predict the status of child diarrhea in Bangladesh (Khan *et al*., 2019; Maniruzzaman and Abedin, 2020). In a different context, neural networks were poorer at predicting student academic performance than linear regression and support vector regression (Obsie and Adem, 2018). In another study, evidence from early childhood asthma persistence (Bose *et al*., 2021) and early childhood obesity (Pang *et al*., 2021) and researchers observed that XGBoost was found to be one of the best performing models in their study.

It is obvious that the area under the curve value is better and more accurate than accuracy when comparing the model's performance with that number. Interestingly, the top model in our investigation in terms of AUC was LASSO regression (AUC = 67.75%) and this model makes predictions that are more accurate than those made by the other eight machine learning algorithms. We emphasize that for predicting early infant development status in Bangladesh, LASSO regression outperforms all other supervised classifiers. On the other hand, hybrid Nave Bayes and decision tree algorithms offer the highest level of accuracy for predicting infant development (Ling *et al*., 2003; Ambili and Afsar, 2016).

There are several limitations to this study. There are many other features connected to the child development index in addition to the ones we utilized, but we were unable to use them because the data we used did not have that information. Combining those elements could increase the prediction power. The study's low AUC score, which shows that we need to further train our algorithm, is another drawback. Due to cross-sectional data, we were unable to demonstrate a true cause-and-effect link. More research is necessary because the performance of the prediction model using contemporary methods like deep learning must be enhanced.

## Conclusion

This study demonstrated that ML algorithms may be used to predict children's developmental status based on common criteria, which can help in the creation of interventions to improve Bangladeshi children's developmental status. For the purpose of this study, we compared nine supervised machine learning algorithms to forecast the development of Bangladeshi kids between the ages of 36 and 59 months. The LASSO regression outperforms the other nine algorithms in terms of accuracy and Area Under the Curve (AUC). This study suggests the use of LASSO regression for the Early Child Development (ECD) index and academics or researchers should pay attention to continuing this study in the future. Once more, this outcome will be useful for policymakers to adopt various initiatives to assure early child development in impoverished nations.

## Acknowledgment

We would especially like to thank UNICEF for allowing us to use Bangladesh multiple indicator cluster survey from http://mics.unicef.org/ for our study.

## Funding Information

This research received no specific grant from any funding agency, commercial entity, or any profit and non-profit organization.

## Data Availability Statement

Data associated with this study is available at the United Nations Children's Fund (UNICEF) at https://mics.unicef.org/surveys.

## Compliance with Ethical Standards

In this study, we used MICS-2019 data downloaded from the MICS. A special thanks go to UNICEF for allowing us to conduct our research using Bangladesh MICS from http://mics.unicef.org/.During the data collection, MICS maintained the privacy and confidentiality of respondents and at the beginning of individual interviews, respondents were given an informed consent form. Therefore, as this study was focused on public use, no further ethical approval was required for our data analysis.

## Author's Contributions

**Md. Ismail Hossain:** Conceptualization; methodology, formal analysis; data curation, written reviewed, and edited.

**Iqramul Haq:** Conceptualization; methodology, written reviewed, and edited.

**Ashis Talukder:** Methodology, written reviewed, and edited.

**Sharmin Suraiya, Mofasser Rahman, Md. Iqbal Hossain Nayan and Sadiq Hussain:** Written reviewed and edited.

**Ahmed Abdus Saleh Saleheen:** Data curation, reviewed and edited.

**Md. Injamul Haq Methun:** Data curation, written reviewed, and edited.

**Md. Jakaria Habib and Md. Sanwar Hossain:** Methodology, written.

## Ethics

In Bangladesh, secondary cross-sectional and nationally representative data from the MICS-2019 survey were used. MICS respected the respondents' privacy and confidentiality while gathering the data, and each got an informed permission form prior to their individual interview. No additional ethical approval was needed for our data analysis because this study was intended for public use.

## References

Ambili, K., & Afsar, P. (2016). A Prediction Model for Child Development Analysis using Naïve Bayes and Decision Tree Fusion Technique-NB Tree. *International Research Journal of Engineering and Technology (IRJET)*, *3*(07). https://www.irjet.net/archives/V3/i7/IRJET-V3I779.pdf

Ansari, A. Q., & Gupta, N. K. (2012, November). Automatic diagnosis of asthma using neurofuzzy system. In *2012 4th International Conference on Computational Intelligence and Communication Networks* (pp. 819-823). IEEE. https://ieeexplore.ieee.org/abstract/document/6375228

Anuse, A., & Vyas, V. (2016). A novel training algorithm for convolutional neural network. *Complex & Intelligent Systems*, *2*(3), 221-234. https://doi.org/10.1007/s40747-016-0024-6

Athena Infonomics. (2022). Athena Infonomics Utilizing CART Machine-Learning Models in Early Childhood Development (ECD). https://www.athenainfonomics.com/?portfolio=utilizing-cart-machine-learning-models-in-early-childhood-development-ecd

Baker, S., & Kandasamy, Y. (2022). Machine learning for understanding and predicting neurodevelopmental outcomes in premature infants: A systematic review. *Pediatric Research*, 1-7. https://doi.org/10.1038/s41390-022-02120-w

Bansod, J., Amonkar, M., Naik, A., Vaz, T., Sanke, M., & Aswale, S. (2020). Prediction of Child Development using Data Mining Approach. *International Journal of Computer Applications*, *177*(44), 13-17.

Bizzego, A., Gabrieli, G., Lim, M., Rothenberg, W. A., Bornstein, M. H., & Esposito, G. (2023). Predictors of Early Childhood Development: A Machine Learning Approach. In *Parenting and Child Development in Low-and Middle-Income Countries* (pp. 210-239). Routledge.

Black, M. M., Walker, S. P., Fernald, L. C. andersen, C. T., DiGirolamo, A. M., Lu, C., ... & Grantham-McGregor, S. (2017). Early childhood development coming of age: Science through the life course. *The Lancet*, *389*(10064), 77-90. https://doi.org/10.1016/S0140-6736(16)31389-7

Bose, S., Kenyon, C. C., & Masino, A. J. (2021). Personalized prediction of early childhood asthma persistence: A machine learning approach. *PloS One*, *16*(3), e0247784. https://doi.org/10.1371/journal.pone.0247784

Bowe, A. K., Lightbody, G., Staines, A., & Murray, D. M. (2022). Big data, machine learning and population health: predicting cognitive outcomes in childhood. *Pediatric Research*, 1-8. https://doi.org/10.1038/s41390-022-02137-1

Buntine, W. (1992). Learning classification trees. *Statistics and Computing*, *2*, 63-73. https://doi.org/10.1007/BF01889584

Christensen, D. L., Schieve, L. A., Devine, O., & Drews-Botsch, C. (2014). Socioeconomic status, child enrichment factors and cognitive performance among preschool-age children: Results from the Follow-Up of Growth and Development Experiences study. *Research in Developmental Disabilities*, *35*(7), 1789-1801. https://doi.org/10.1016/j.ridd.2014.02.003

Daelmans, B., Darmstadt, G. L., Lombardi, J., Black, M. M., Britto, P. R., Lye, S., ... & Richter, L. M. (2017). Early childhood development: The foundation of sustainable development. *The Lancet*, *389*(10064), 9-11. https://doi.org/10.1016/S0140-6736(16)31659-2 de Onis, M., & Branca, F. (2016). Childhood stunting: a Global Perspective. Maternal and Child Nutrition. *12*(1),12-26. https://doi.org/10.1111/mcn.12231

Donald, K. A., Wedderburn, C. J., Barnett, W., Nhapi, R. T., Rehman, A. M., Stadler, J. A., ... & Stein, D. J. (2019). Risk and protective factors for child development: An observational South African birth cohort. *PLoS Medicine*, *16*(9), e1002920. https://doi.org/10.1371/journal.pmed.1002920

Dugan, T. M., Mukhopadhyay, S., Carroll, A., & Downs, S. (2015). Machine learning techniques for prediction of early childhood obesity. *Applied Clinical Informatics*, *6*(03), 506-520. http://dx.doi.org/10.4338/ACI-2015-03-RA-0036

ECPC. (2021). 2030 Sustainable Development Goals. *Early Childhood Peace Consortium*. https://ecdpeace.org/work-content/2030-sustainable-development-goals

Espel, E. V., Glynn, L. M., Sandman, C. A., & Davis, E. P. (2014). Longer gestation among children born full term influences cognitive and motor development. *PLoS One*, *9*(11), e113758. https://doi.org/10.1371/journal.pone.0113758

Georgieff, M. K. (2007). Nutrition and the developing brain: nutrient priorities and measurement. *The American Journal of Clinical Nutrition*, *85*(2), 614S-620S. https://doi.org/10.1093/ajcn/85.2.614S

Georgieff, M. K., Ramel, S. E., & Cusick, S. E. (2018). Nutritional influences on brain development. *Acta Paediatrica*, *107*(8), 1310-1321. https://doi.org/10.1111/apa.14287

Gera, M., & Goel, S. (2015, May). A model for predicting the eligibility for placement of students using data mining technique. In *International Conference on Computing, Communication & Automation* (pp. 114-117). IEEE. https://ieeexplore.ieee.org/abstract/document/7148355

Gill, S. V., May-Benson, T. A., Teasdale, A., & Munsell, E. G. (2013). Birth and developmental correlates of birth weight in a sample of children with potential sensory processing disorder. *BMC Pediatrics*, *13*, 1-8. https://doi.org/10.1186/1471-2431-13-29

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. (2007). Developmental potential in the first 5 years for children in developing countries. *The Lancet*, *369*(9555), 60-70. https://doi.org/10.1016/S0140-6736(07)60032-4

Gutbrod, T., Wolke, D., Soehne, B., Ohrt, B., & Riegel, K. (2000). Effects of gestation and birth weight on the growth and development of very low birthweight small for gestational age infants: A matched group comparison. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *82*(3), F208-F214. http://dx.doi.org/10.1136/fn.82.3.F208

Haile, D., Nigatu, D., Gashaw, K., & Demelash, H. (2016). Height for age z score and cognitive function are associated with Academic performance among school children aged 8-11 years old. *Archives of Public Health*, *74*(1), 1-7. https://doi.org/10.1186/s13690-016-0129-9

Haq, I., Hossain, M., Zinnia, M. A., Hasan, M. R., & Chowdhury, I. A. Q. (2021). Determinants of the Early Childhood Development Index among children aged< 5 years in Bangladesh, Costa Rica and Ghana: A comparative study. *Eastern Mediterranean Health Journal*, *27*(11).

Hossain, M. I., Haq, I., Hossain, M. S., Habib, M. J., Islam, F. B., Roy, S., & Rahman, M. (2023). Factors associated with early literacy and numeracy development among children under five years in Bangladesh: Multivariate two-level mixed effect approach. *International Journal of Social Economics*, *50*(3), 345-358. https://doi.org/10.1108/IJSE-10-2021-0595

Hossain, M. I., Haq, I., Zinnia, M. A., Mila, M. S., & Nayan, M. I. H. (2021). Regional variations of child development index in Bangladesh. *Heliyon*, *7*(5), e07140. https://doi.org/10.1016/j.heliyon.2021.e07140

Irwin, L. (2007). Early child development: A powerful equalizer. Final report for the World Health Organization's Commission on the social determinants of health. http://www.who.int/social_determinants/resources/ecd_kn_report_07_2007. pdf. https://ci.nii.ac.jp/naid/10026507958/

Jang, W., Lee, J. K., Lee, J., & Han, S. H. (2015). Naive Bayesian classifier for selecting good/bad projects during the early stage of international construction bidding decisions. *Mathematical Problems in Engineering*, *2015*. https://doi.org/10.1155/2015/830781

Khan, J. R., Chowdhury, S., Islam, H., & Raheem, E. (2019). Machine learning algorithms to predict the childhood anemia in Bangladesh. *Journal of Data Science*, *17*(1), 195-218. https://doi.org/10.6339/JDS.201901_17(1).0009

Kleimola, L. B., Patel, A. B., Borkar, J. A., & Hibberd, P. L. (2015). Consequences of household air pollution on child survival: Evidence from demographic and health surveys in 47 countries. *International Journal of Occupational and Environmental Health*, *21*(4), 294-302. https://doi.org/10.1179/2049396715Y.0000000007

Kvestad, I., Taneja, S., Hysing, M., Kumar, T., Bhandari, N., & Strand, T. A. (2015). Diarrhea, stimulation and growth predict neurodevelopment in young North Indian children. *PloS One*, *10*(3), e0121743. https://doi.org/10.1371/journal.pone.0121743

Duc, L. T. (2009). The Effect of Early Age Stunting on Cognitive Achievement among Children in Vietnam. [online] Oxford: Young Lives. https://www.younglives.org.uk/sites/default/files/migrated/YL-WP45-Le-EarlyAgeStunting.pdf

Lee, H., Park, H., Ha, E., Hong, Y. C., Ha, M., Park, H., ... & Kim, Y. (2016). Effect of breastfeeding duration on cognitive development in infants: 3-year follow-up study. *Journal of Korean medical science*, *31*(4), 579-584. https://doi.org/10.3346/jkms.2016.31.4.579

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18-22. https://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf

Ling, C. X., Huang, J., & Zhang, H. (2003). AUC: A better measure than accuracy in comparing learning algorithms. In *Advances in Artificial Intelligence: 16th Conference of the Canadian Society for Computational Studies of Intelligence, AI 2003, Halifax, Canada, June 11-13, 2003, Proceedings 16* (pp. 329-341). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-44886-1_25

Liu, B., Fang, L., Liu, F., Wang, X., & Chou, K. C. (2016). iMiRNA-PseDPC: microRNA precursor identification with a pseudo distance-pair composition approach. *Journal of Biomolecular Structure and Dynamics*, *34*(1), 223-235. https://doi.org/10.1080/07391102.2015.1014422

Lorntz, B., Soares, A. M., Moore, S. R., Pinkerton, R., Gansneder, B., Bovbjerg, V. E., ... & Guerrant, R. L. (2006). Early childhood diarrhea predicts impaired school performance. *The Pediatric Infectious Disease Journal*, *25*(6), 513-520. https://doi.org/10.1097/01.inf.0000219524.64448.90

Lu, C., Black, M. M., & Richter, L. M. (2016). Risk of poor development in young children in low-income and middle-income countries: An estimation and analysis at the global, regional and country level. *The Lancet Global Health*, *4*(12), e916-e922. https://doi.org/10.1016/S2214-109X(16)30266-2

Maniruzzaman, I. S., & Abedin, M. (2020). Prediction of Childhood Diarrhea in Bangladesh using Machine Learning Approach. *Insights Biomed Res*, *4*(1), 111-116. https://doi.org/10.36959/584/456

McCoy, D. C., Peet, E. D., Ezzati, M., Danaei, G., Black, M. M., Sudfeld, C. R., ... & Fink, G. (2016). Early childhood developmental status in low-and middle-income countries: National, regional and global prevalence estimates using predictive modeling. *PLoS Medicine*, *13*(6), e1002034. https://doi.org/10.1371/journal.pmed.1002034

McDonald, L. A., & Rennie, A. C. (2011). Investigating developmental delay/impairment. *Paediatrics and Child Health*, *21*(10), 443-447. https://doi.org/10.1016/j.paed.2011.02.008

Miller, A. C., Garchitorena, A., Rabemananjara, F., Cordier, L., Randriamanambintsoa, M., Rabeza, V., ... & Ratsifandrihamanana, L. (2020). Factors associated with risk of developmental delay in preschool children in a setting with high rates of malnutrition: A cross-sectional analysis of data from the IHOPE study, Madagascar. *BMC Pediatrics*, *20*(1), 1-11. https://doi.org/10.1186/s12887-020-1985-6

Niehaus, M. D., Moore, S. R., Patrick, P. D., Derr, L. L., Lorntz, B., Lima, A. A., & Guerrant, R. L. (2002). Early childhood diarrhea is associated with diminished cognitive function 4 to 7 years later in children in a northeast Brazilian shantytown. *The American Journal of Tropical Medicine and Hygiene*, *66*(5), 590-593. https://es.ircwash.org/sites/default/files/Niehaus-2002-Early.pdf

Obsie, E. Y., & Adem, S. A. (2018). Prediction of student academic performance using neural network, linear regression and support vector regression: A case study. *International Journal of Computer Applications*, *180*(40), 39-47.

Pang, X., Forrest, C. B., Lê-Scherban, F., & Masino, A. J. (2021). Prediction of early childhood obesity with machine learning and electronic health record data. *International Journal of Medical Informatics*, *150*, 104454. https://doi.org/10.1016/j.ijmedinf.2021.104454

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*, *18*(16), 8613. https://doi.org/10.3390/ijerph18168613

Po, J. Y., FitzGerald, J. M., & Carlsten, C. (2011). Respiratory disease associated with solid biomass fuel exposure in rural women and children: Systematic review and meta-analysis. *Thorax*, *66*(3), 232-239. http://dx.doi.org/10.1136/thx.2010.147884

Ribe, I. G., Svensen, E., Lyngmo, B. A., Mduma, E., & Hinderaker, S. G. (2018). Determinants of early child development in rural Tanzania. *Child and Adolescent Psychiatry and Mental Health*, *12*(1), 1-8. https://doi.org/10.1186/s13034-018-0224-5

Richter, L. M., Daelmans, B., Lombardi, J., Heymann, J., Boo, F. L., Behrman, J. R., ... & Lancet Early Childhood Development Series Steering Committee. (2017). Investing in the foundation of sustainable development: Pathways to scale up for early childhood development. *The Lancet*, *389*(10064), 103-118. https://doi.org/10.1016/S0140-6736(16)31698-1

Roberts, E., Bornstein, M. H., Slater, A. M., & Barrett, J. (1999). Early cognitive development and parental education. *Infant and Child Development: An International Journal of Research and Practice*, *8*(1), 49-62. https://doi.org/10.1002/(SICI)1522-7219(199903)8:1<49::AID-ICD188>3.0.CO;2-1

Sania, A., Sudfeld, C. R., Danaei, G., Fink, G., McCoy, D. C., Zhu, Z., ... & Fawzi, W. (2019). Early life risk factors of motor, cognitive and language development: A pooled analysis of studies from low/middle-income countries. *BMJ Open*, *9*(10), e026449. http://dx.doi.org/10.1136/bmjopen-2018-026449

Sungthong, R., Mo-suwan, L., & Chongsuvivatwong, V. (2002). Effects of haemoglobin and serum ferritin on cognitive function in school children. *Asia Pacific Journal of Clinical Nutrition*, *11*(2), 117-122. https://doi.org/10.1046/j.1440-6047.2002.00272.x

Talukder, A., & Ahammed, B. (2020). Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh. *Nutrition*, *78*, 110861. https://doi.org/10.1016/j.nut.2020.110861

Tong, S., Baghurst, P., & McMichael, A. (2006). Birthweight and cognitive development during childhood. *Journal of Paediatrics and Child Health*, *42*(3), 98-103. https://doi.org/10.1111/j.1440-1754.2006.00805.x

UN. (2021). Goal 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all. https://unstats.un.org/sdgs/report/2017/goal-04/

Upadhyay, R. P., Naik, G., Choudhary, T. S., Chowdhury, R., Taneja, S., Bhandari, N., ... & Bhan, M. K. (2019). Cognitive and motor outcomes in children born low birth weight: A systematic review and meta-analysis of studies from South Asia. *BMC Pediatrics*, *19*, 1-15. https://doi.org/10.1186/s12887-019-1408-8

Vasquez, M. M., Hu, C., Roe, D. J., Chen, Z., Halonen, M., & Guerra, S. (2016). Least absolute shrinkage and selection operator type methods for the identification of serum biomarkers of overweight and obesity: Simulation and application. *BMC Medical Research Methodology*, *16*(1), 1-19. https://doi.org/10.1186/s12874-016-0254-8

Walker, S. P., Wachs, T. D., Grantham-McGregor, S., Black, M. M., Nelson, C. A., Huffman, S. L., ... & Richter, L. (2011). Inequality in early childhood: Risk and protective factors for early child development. *The Lancet*, *378*(9799), 1325-1338. https://doi.org/10.1016/S0140-6736(11)60555-2

Woldehanna, T., Behrman, J. R., & Araya, M. W. (2017). The effect of early childhood stunting on children's cognitive achievements: Evidence from young lives Ethiopia. *Ethiopian Journal of Health Development*, *31*(2), 75-84. https://www.ajol.info/index.php/ejhd/article/view/167766

WHO. (2018). Nurturing care for early childhood development: A framework for helping children survive and thrive to transform health and human potential. https://apps.who.int/iris/bitstream/handle/10665/272603/9789241514064-eng.pdf?sequence=1&isAllowed=y

Wu, P., & Zhao, H. (2011). Some analysis and research of the AdaBoost algorithm. In Intelligent Computing and Information Science: International Conference, ICICIS 2011, Chongqing, China, January 8-9, 2011. Proceedings, Part I (pp. 1-5). *Springer Berlin Heidelberg*. https://doi.org/10.1007/978-3-642-18129-0_1