

<sup>1</sup>Firoz Hasan<sup>2</sup>Rubina Khatun<sup>3</sup>Dewan Mamun  
Raza<sup>4</sup>Aynul Hasan Nahid<sup>5</sup>Md Rakibul Hasan  
Bhuiyan<sup>6</sup>Hasan Moon

## Predictive Modeling of Child Malnutrition and Factor Analysis using Machine Learning Approach for Early Detection and Intervention



**Abstract:** - Child malnutrition is a condition characterized by inadequate intake of nutrients necessary for proper growth and development in children under the age of five. There are many different things that may lead to malnutrition, such as inadequate food consumption, a lack of variety in one's diet, and inadequate care practices. Children who suffer from malnutrition may experience stunted growth, decreased cognitive development, weakened immunity, and increased vulnerability to disease. Malnutrition can also have long-term consequences, such as reduced work productivity and increased healthcare costs. This study's objective was to evaluate, through the use of methods from machine learning, the degree to which children in a certain location are impacted by under nutrition. A survey of children in the area who were under the age of five and were cross-sectional in nature provided the information that was used for the dataset. Age, gender, weight, height, and a variety of socioeconomic and environmental factors were among the elements that were collected as variables. Also taken into consideration were a number of other environmental factors. Every algorithm compares favorably to the results produced by the Logistic algorithm. The accuracy of this approach, which is 0.82%, is significantly higher than that of previous algorithms.

**Keywords:** Child Malnutrition, Machine Learning ,Nutritional Assessment, Socioeconomic and Environmental Factors

### I. INTRODUCTION

Child malnutrition, a condition characterized by the insufficient or excessive intake of nutrients necessary for children's growth, is a global health issue. It can manifest as under nutrition, over nutrition, or micronutrient deficiencies. These nutritional imbalances, whether due to a lack of or an excess of essential nutrients, affect children worldwide, with a higher prevalence in low- and middle-income countries where access to nutritious food and healthcare services is limited. Addressing child malnutrition is crucial due to its long-term adverse effects on children's health and development. Malnutrition leads to stunted growth, developmental delays, weakened immune systems, and increased vulnerability to infections. These health impairments not only impact the individual child but also hinder the broader economic and social development of communities and nations. As noted in various studies, the effects of malnutrition are especially pronounced in regions affected by poverty, food insecurity, and limited access to healthcare services, which makes the problem both urgent and multidimensional [12].

Previous research has explored the different forms of malnutrition and their causes. For example, under nutrition arises from a chronic deficiency in calories, protein, or essential nutrients, while over nutrition results from an excessive intake of calories, leading to obesity and related health problems. Studies such as Black et al. (2013) have emphasized the impact of poverty, inadequate sanitation, and food scarcity in exacerbating malnutrition. Micronutrient deficiencies, such as a lack of vitamins and minerals, are also prevalent in impoverished regions. Despite the increasing focus on malnutrition, current intervention strategies often fail to address its root causes effectively, particularly in low-resource settings [3], [7].

<sup>1</sup> \*Corresponding author: [firoz.cse@diu.edu.bd](mailto:firoz.cse@diu.edu.bd), Daffodil International University, Dhaka, Bangladesh

<sup>2</sup> [babyru17@gmail.com](mailto:babyru17@gmail.com), University of Rajshahi, Rajshahi, Bangladesh

<sup>3,4,5,6</sup> [dewan.cse@diu.edu.bd](mailto:dewan.cse@diu.edu.bd); [aynul.cse@diu.edu.bd](mailto:aynul.cse@diu.edu.bd); [official.bhuiyanrakib@gmail.com](mailto:official.bhuiyanrakib@gmail.com); [hasanmoon9768@gmail.com](mailto:hasanmoon9768@gmail.com) Daffodil International University, Dhaka, Bangladesh

While the literature has provided a strong foundation for understanding the factors driving child malnutrition, gaps remain in addressing the multidimensional nature of the problem. Specifically, there is a need for studies that investigate how interventions targeting nutrition, healthcare, and sanitation can be integrated for greater effectiveness. The present study seeks to address these gaps by analyzing the outcomes of malnutrition interventions in areas with limited resources. By doing so, it aims to offer new insights into scalable solutions that can be adapted to diverse contexts, improving the impact of public health efforts in combating child malnutrition [6], [8].

The wellbeing of individual children as well as the general growth and prosperity of communities and countries depends on efforts to eliminate child malnutrition. Societies may contribute to healthier populations, better educational results, and increased economic production in the long term by guaranteeing sufficient nutrition throughout the crucial early years of life.

## II. RELATED WORKS

There has been many studies and researches where people have been predicting child malnutrition using the algorithms of machine learning, like decision tree, support vector machine, random forest etc. for the collection and classification of data. In [1] Using machine learning to identify for efficient prevention, treatment, and Recognition of target populations is a promising area of study, and this research represents a significant step in that direction. Applying Tabular Deep Learning's framework, we can achieve high levels of accuracy (up to 96.46%) .In [2] In order to estimate predictive algorithms for the causes of childhood stunting, they use a wide variety of machine learning techniques. This study made use of data from the 2016 Ethiopian Demographic and Health Survey. Extreme gradient boosting, k-NN, random forest, neural network, and generalized linear models were the five ML methods used to make the predictions. In [3] there were six different ML approaches that we considered. These models' performance was assessed using their sensitivity, specificity, accuracy, and area under the curve. The RF method was chosen as the top ML model based on a variety of metrics. In [4] The ANN technique offers the best accuracy for diagnosing loss. This research improves our understanding of how malnutrition is categorized by providing a predictive model.

In [5] Care providers could benefit from a decision support system for identifying and managing patients at higher risk of malnutrition if electronic health records included algorithms based on machine learning. Malnutrition, from which 60-80% of individuals are chosen for a more comprehensive assessment. In [6] According to the findings, machine learning methods can pick out important details that are missing from the existing literature. The likelihood that these factors explain malnutrition was then calculated using logistic regression. In [7] Two hospitals in Afghanistan provided data on cases of malnutrition, In contrast to the J48 algorithm's mediocre performance (94.51% accuracy), the Random Forest approach produced amazing results (97.14%).In [8] Data from the 2016 Ethiopian Demographic and Health Survey was used in this analysis. To forecast the socio demographic risk variables for under nutrition in Ethiopia, five machine learning (ML) techniques, including eXtreme gradient boosting, knearest neighbors (k-NN), random forest, neural network, and generalized linear models, were taken into consideration. In [9] this article presents a decision tree model that classifies babies between the ages of 0 and 59 months as either normal, acutely malnourished, or severely malnourished depending on which of the three categories they fall into. Decision tree models frequently use the Gini index as their standard impureness metric.

When utilizing the Gini index, the accuracy gained using the decision tree for stunting is 82.22%, while the accuracy acquired for wasting is 72.23%, and the accuracy obtained for underweight is 78.35%.In [10] considering the algorithms' accuracy ratings, they sought for the best ones. Based on different machine learning algorithm performances, Random Forest and Logistic Regression produced the greatest results, achieving accuracy rates of 91.11% and 89.88%, as well as train accuracy rates of 1.000 and 0.847. We created this dataset to better aggregate information so that users may take primary information like what the complexity might be, what kind of health specialists they should visit, what prevention's/precautions strategies they can do to lessen risks, etc. That pertains to Child Malnutrition. This is useful, but there is a risk of receiving false information, which could lead to undue mental worry and, in the worst case, physical damage.

### III. METHODOLOGY

We've been using our typical process flow chart to give an overview, but we'll go over some of the algorithms with equations and diagrams to help you understand how they work. The process for the full research endeavor is included in this part along with a quick synopsis of the complete research effort.

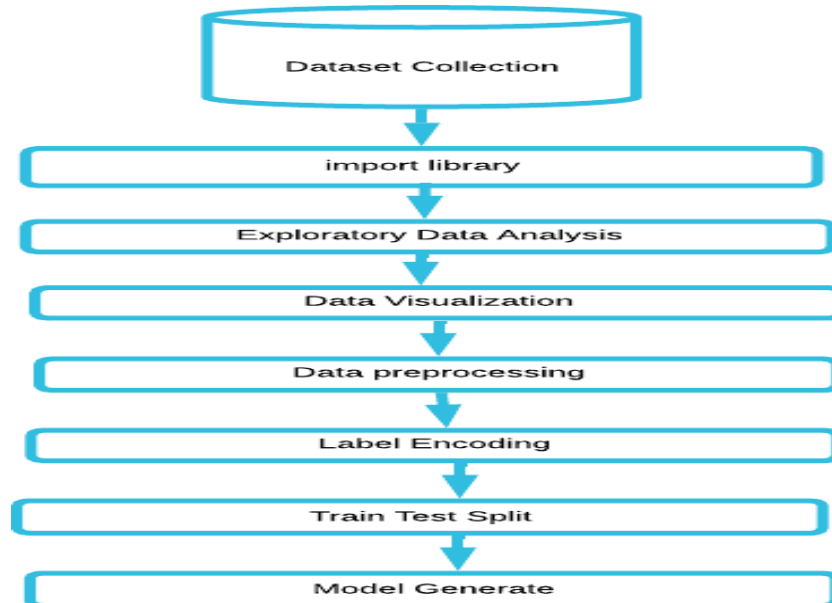


Figure 1. Working Methodology

#### 3.1 Data Collecting

The data set was collected through the use of survey questionnaires. A questionnaire is used to gather information from respondents about their attitudes, experiences, and opinions. Questionnaires are commonly used in market research, social and health sciences, and related fields. We invite the children and their parents to provide their own replies after seeing these questions online using a Google form that we use publish specific questions. And we gather, prepare, and use the information we learn about their replies for our own study. All participants' names are kept secret while they express their ideas. We are collecting information from a Google data set. We pose a few questions, and the audience members' responses are entered into our main Excel spreadsheet as rows.

I. Gender - What is your gender? (nominal: 'Male' or 'Female')

II. Age - What is age? (nominal: 18-25, 25-35, 35-40)

III. Child Sleep - A child should sleep? (nominal: Environment, Family care, Physical well-being, Quantity of intake food)

IV. Food Allergies - At what ages do food allergies, iron deficiency and occur in children? (nominal: 3 to 6 year, 1 to 3 year, 6 to 9 year)

V. Early growing stage - During the early growing stage of a child the cause of malnutrition according to you is- (nominal: Poverty, Lack of parental awareness, Physical sickness)

VI. Child nutrition effect - Child nutrition can also effect by (nominal)

VII. Early growing stage - During the early growing stage of a child the cause of malnutrition according to you is- (nominal: Poverty, Lack of parental awareness, Physical sickness)

VIII. Puberty most of the child need - During puberty most of the child need to take proper amount of nutrients because (nominal: They have started to growing, the growth rate is massive in this period, Boys and girls have some individual changes in growth)

- IX. Breast feeding- The most important nutrition source of an infant or baby is breast feeding. Without breast feeding they can be grown up but do you think it can be (nominal: Malnutrition, no if proper nutrients are getting by other foods so it's ok, without breast breastfeeding growth will not be affected but it could enhance the development)
- X. Girl Child - A girl child needs to have best care during puberty (nominal: Attractions towards uncertainty, Due to her mental health, Physiological change)
- XI. Common malnutrition- The most common malnutrition of children in our country (nominal: Rickets, Marasmus, Kwashiorkor )
- XII. Parental awareness- Do you prefer parental awareness in case of a sound health of a child (nominal: Strongly agree, Agree, Disagree)
- XIII. Diarrhea disease - Diarrhea is a very common disease of infant. According to you it happens due to (nominal: Lack of cleanliness to the instruments that are used to feed them, Unhealthy environment, Overfeeding)
- XIV. Body growth - Some people think that feeding a child more and more can increase their body growth. But according to you it is a (nominal: Cause of malnutrition, Overweight, Dysfunctioning of a child growth)
- XV. Nutrition child - Nutrition of a child related to (nominal: Quantity of nutrients, Including vegetables, fruits, Milk as a supplement )
- XVI. Grain foods- Whole grain foods such as breads, cereals that are high in fibre (nominal: Helps bone growth, making stool, helps in digestion, Helps in bowel function)
- XVII. Diet - Toddlers/preschoolers should have following nutrients in their diet (nominal: Calcium for bone growth, Milk for the growth but sometimes it causes milk-allergy, playing can accelerate their growth)

### 3.2 Exploratory Data Analysis (EDA)

This chapter analyzes our data in a variety of ways. We looked at the dataset's total number of columns, unique values, and null values. It is necessary to import a dataset before starting an exploratory data analysis since this phase occurs first. The dataset has 414 rows and 18 columns altogether. There are no null values in our data collection, which is being checked for the existence of null values.

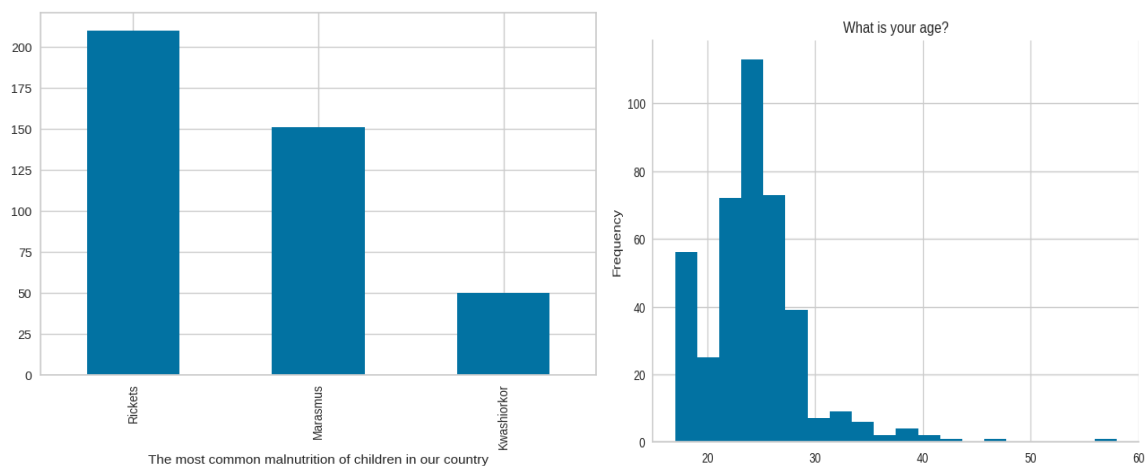


Figure 2: Categories of Malnutrition and Age Distribution

Here are some general facts on the three categories of malnutrition mentioned on above figure 2, when the body doesn't have enough calcium, phosphorus, or vitamin D, the condition known as rickets can arise. It may cause muscular weakness, growth problems, and abnormalities of the bones. Marasmus is a severe form of malnutrition that is typically brought on by a diet that is low in calories and protein and is characterized by weight loss, muscle atrophy, and a weakened immune system. Kwashiorkor, a severe form of malnutrition, can arise from a lack of protein in the diet. It may cause an enlarged belly, a skin rash, and a weakened immune system, among other symptoms. The percentage of people within a population or a particular group that are divided into several age

groups is known as age distribution. It offers a glimpse of the population’s age distribution and is frequently shown as a frequency or percentage distribution. Age pyramids in figure 3, a graphical depiction of the distribution of different age groups in a population, are a common way to demonstrate age distribution.

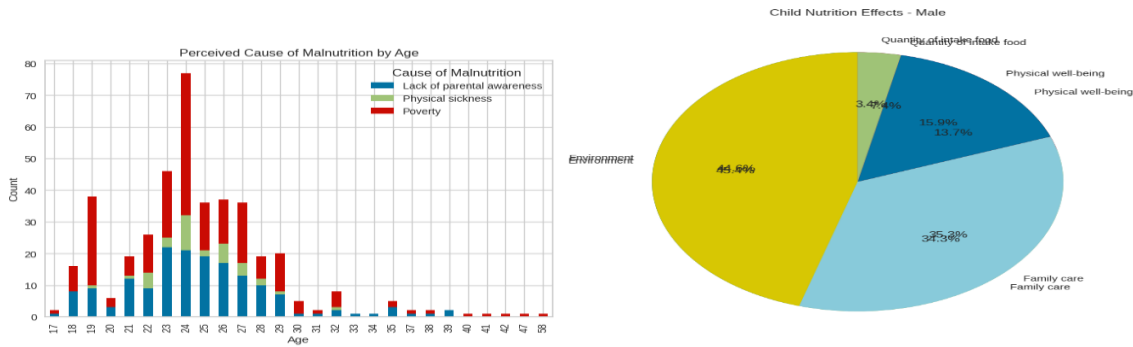


Figure 3: Perceived Cause of Malnutrition and Nutrition Effects

The birth and mortality rates in different age groups vary, which causes the pyramid form. The ratio of men and women in a population’s various age groups is referred to as the age-wise gender distribution. Insights regarding gender distribution at various phases of life are provided, and it also aids in the analysis of gender discrepancies and trends across distinct age cohorts. The proportion of population who fall into certain age groups is known as the age-wise gender distribution. In addition to analyzing gender discrepancies and trends across distinct age cohorts, it offers insights into the gender composition at various phases of life.

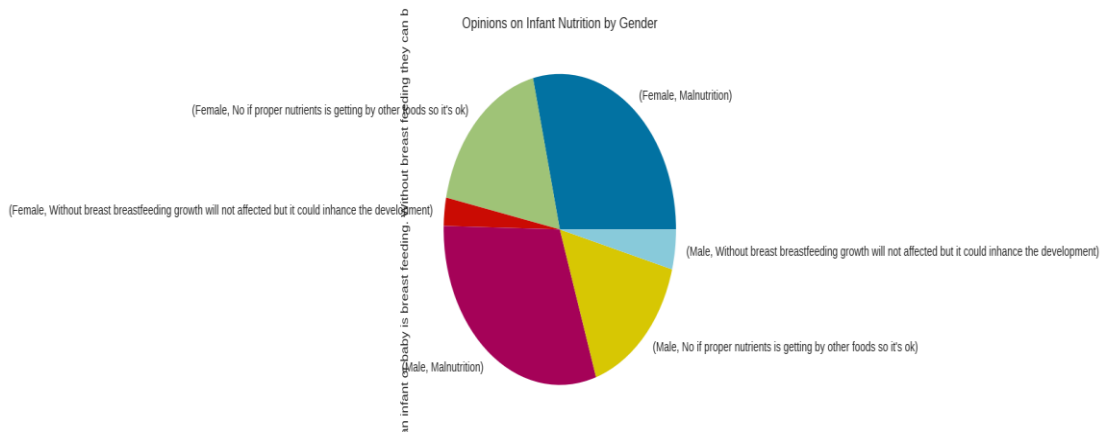


Figure 4: Gender Variation on others Effect

Data with category labels are often encoded. The alphabetical labels are given a number using this technique. Learn about the challenges of label encoding and how to utilize Python’s scikit-learn module to solve them. There are several types when the order of a categorical feature is irrelevant since one-hot encoding might use a lot of memory. Data labeling is a crucial component of data pretreatment for ML, especially for supervised learning, where input and output data are both classified and labeled to serve as a learning foundation for further data processing.

### 3.3 Model Implementation

To ensure the best possible working of machine learning algorithms it needs to work with some key parameters. Each and every task requires a different model based on the type of data and work is being dealt with. Hence, it is crucial to adjust the model’s parameters to increase its utility and accuracy. In our work we have tried to ensure to tune all the models with adequate parameter values and plump for the foremost value for our models. We outline how machine learning techniques can be used to identify student Mental Health. The five techniques that we will look at are decision trees, logistic regression, K-Nearest Neighbors (KNN), linear regression, and naive Bayes. Logistic regression models the probability that a given input  $x$  (which could include features like age, weight, height, diet, family care, etc.) belongs to a particular class (e.g., 0 for well-nourished, 1 for moderately malnourished, or 2 for severely malnourished). The model assumes a linear relationship between the input features  $x = [x_1, x_2, \dots, x_n]$  and the log-odds of the probability  $P(y = i | x)$ , for each

class  $i$ . For a binary classification, this is written as:  $P(y = 1 | x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}$  Where:  $\beta_0$  is the bias (intercept) term.  $\beta_1, \dots, \beta_n$  are the coefficients corresponding to the features. At each node of the decision tree, the characteristic with the maximum information gain is selected as the best split since it offers the most useful data for categorization. Decision trees can efficiently split the data and increase classification accuracy by choosing attributes with a higher information gain. Information gain essentially aids decision trees in locating the most informative aspects for the best possible decision-making.

$$\text{childmalnutrition} = 0 + \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n \quad (1)$$

The SVM equation seeks to maximize the margin between classes in order to locate an ideal hyper plane that divides data points. Where  $y(x)$  is the predicted class label for a given data point  $x$ ,  $w$  denotes the weight vector orthogonal to the hyper plane,  $b$  denotes the bias term, and  $\text{sign}()$  denotes the sign function that determines the proper class label based on the predicted value. Where  $d$  is the weight vector's is the feature vector (e.g., age, weight, height, dietary intake).  $b$  is the bias term.

$$f(x) = wx + b \quad (2)$$

A data point's class label is predicted by the k-nearest neighbors (KNN) algorithm using the labels of its closes neighbors.

$$A = (a_1, \dots, a_n) \text{ and } B = (b_1, \dots, b_n) \quad (3)$$

$$d(x, x_i) = \sqrt{\sum_{j=1}^n (x_j - x_{ij})^2} \quad (4)$$

The equation  $\text{Gain}(T, X) = \text{Entropy}(T) - \text{Entropy}(T, X)$  is not directly applied to splitting characteristics in random forest. An altered version known as "Gini impurity" or "Gini index" is usually used. Random forest uses the Gini impurity to evaluate attribute priority and choose the appropriate splits for each decision tree in the ensemble.  $\text{Gain}(T, X) = \text{Entropy}(T) - \text{Entropy}(T, X)$  indicates the information gain utilized in decision tree algorithms.

$$\text{Gini}(P) = \sum_{i=1}^n p_i (1 - p_i) = 1 - \sum_{i=1}^n (p_i)^2 \quad (5)$$

$$\hat{y} = \text{mode}(T_1(x), T_2(x), \dots, T_N(x)) \quad (6)$$

GaussianNB, which estimates the likelihood  $p(x|C)$  based on the mean and variance of the training data for each class, assumes that  $p(x|C)$  follows a Gaussian (normal) distribution. During the classification process, the marginal probability  $p(x)$  is roughly estimated as a constant for all classes, and the prior probability  $p(C)$  is determined as the percentage of training samples that belong to class  $C$ .

$$P(\text{malnutrition} | X_1, \dots, X_n) = \frac{P(X_1 | \text{malnutrition}) \dots P(X_n | \text{malnutrition})}{P(X_1) P(X_2) \dots P(X_n)} \quad (7)$$

#### IV. EXPERIMENTAL RESULT ANALYSIS

The performance metrics of a classification algorithm, especially a Decision Tree, are shown in the figure below for various value names or categories. Precision, recall, and F1- score, which are frequently used metrics to gauge the effectiveness of classification models, have been used to evaluate the method.

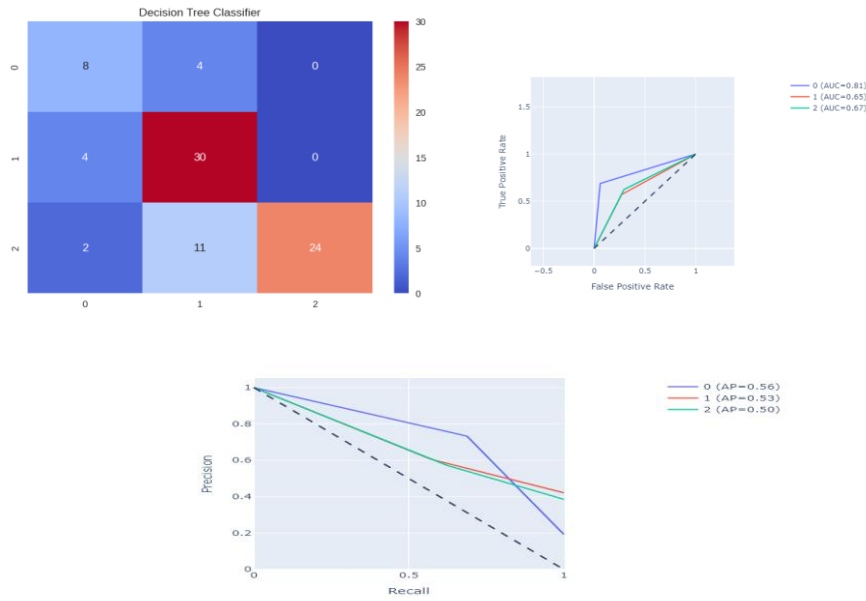


Figure 5: Performance Analysis for Decision Tree

The confusion matrix illustrates the performance of a decision tree in figure 5 classifier applied to assess child malnutrition, with three categories: well-nourished (0), moderately malnourished (1), and severely malnourished (2). The rows represent the actual labels, and the columns show the predicted labels. The model correctly classified 12 well-nourished, 26 moderately malnourished, and 18 severely malnourished children. However, there were some errors, such as 14 severely malnourished children being misclassified as moderately malnourished. The color gradient visually emphasizes correct predictions in red and misclassifications in shades of blue, providing insights into the model's ability to differentiate among malnutrition levels. Additionally, the evaluation of this decision tree model using Precision-Recall (PR) and ROC curves sheds light on its performance across the three classes (0, 1, and 2). The PR curve indicates that the model performs best for Class 0, with an Average Precision (AP) of 0.56, while it performs slightly worse for Classes 1 and 2 (AP = 0.53 and 0.50, respectively). The ROC curve reveals how well the model distinguishes between true and false positives, with Class 0 achieving the highest Area Under the Curve (AUC) of 0.81, suggesting stronger classification for well-nourished children. Classes 1 and 2, with AUCs of 0.65 and 0.67, indicate poorer performance. In summary, while the model performs better at classifying well-nourished children, it faces challenges in accurately distinguishing between moderate and severe malnutrition cases.

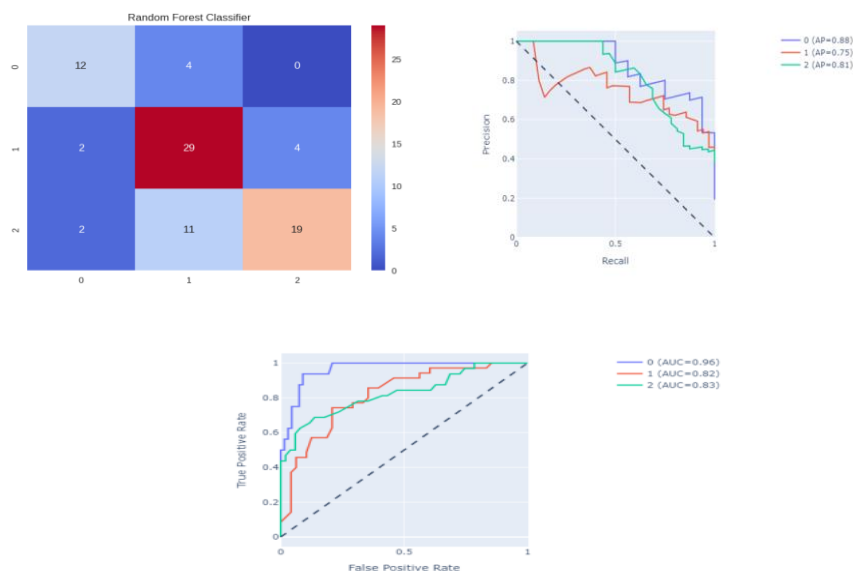


Figure 6: Performance Analysis for Random Forest Classifier

The Random Forest Classifier in figure 6 correctly identified 12 well-nourished, 29 moderately malnourished, and 19 severely malnourished cases, though some misclassifications occurred, such as two well-nourished and two severely malnourished children being classified as moderately malnourished, and 11 severely malnourished children misclassified as moderately malnourished. The Precision-Recall (PR) curve further evaluates the model's performance, with Class 0 (well-nourished) achieving the highest Average Precision (AP = 0.88), followed by Class 2 (severely malnourished) with AP = 0.81, and Class 1 (moderately malnourished) with AP = 0.75. This demonstrates that the model performs best in identifying well-nourished children, with slightly lower but still reasonable performance in detecting moderate and severe malnutrition.

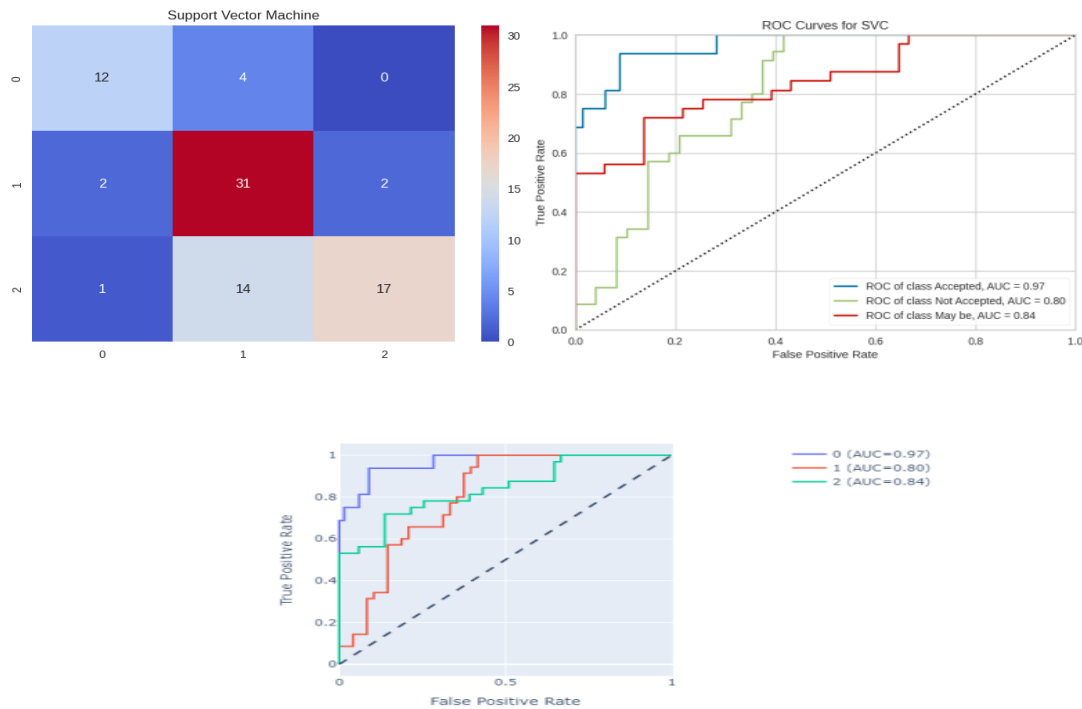


Figure 7: Performance Analysis for Support Vector Machine

The ROC curves for SVM in figure 7 demonstrate the classifier's ability to distinguish between these classes, with the highest performance for the well-nourished class (AUC = 0.97), followed by the moderately malnourished (AUC = 0.80), and severely malnourished (AUC = 0.84). The confusion matrix highlights the classifier's performance, showing that it correctly classifies most well-nourished children (12 correct), moderately malnourished (31 correct), but has more misclassifications in the severely malnourished category, where 14 children were wrongly predicted as moderately malnourished. This suggests that while the SVC performs well overall, distinguishing between moderate and severe malnutrition could be improved.

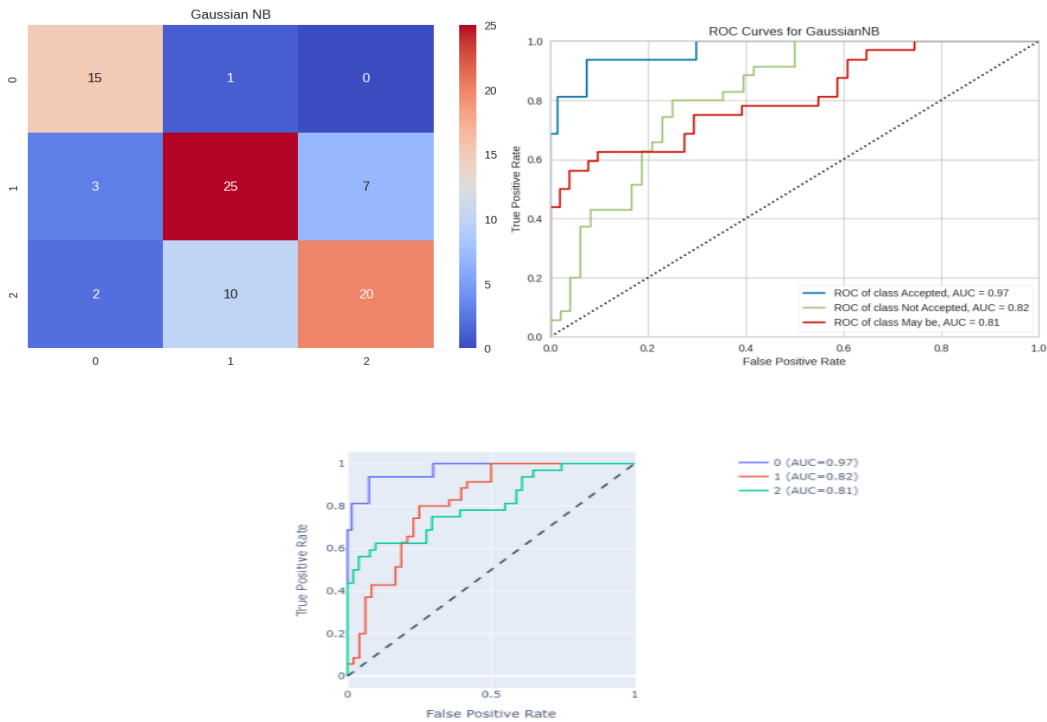


Figure 8: Performance Analysis for GaussianNB

The performance of a Gaussian Naive Bayes (GaussianNB) figure 8 classifier used to assess child malnutrition, classifying children into three categories: well-nourished (0), moderately malnourished (1), and severely malnourished (2). The first image displays ROC curves for these categories, with the model performing best for the well-nourished class (AUC = 0.97), while the performance for moderately malnourished (AUC = 0.82) and severely malnourished (AUC = 0.81) is slightly lower. The second image shows a confusion matrix that indicates the model accurately classifies most moderately malnourished cases (25 correct predictions), but there are misclassifications, especially with the severely malnourished class, where 10 children are wrongly classified as moderately malnourished. This highlights the classifier's strong performance for well-nourished children but potential areas of improvement for differentiating between moderate and severe malnutrition.

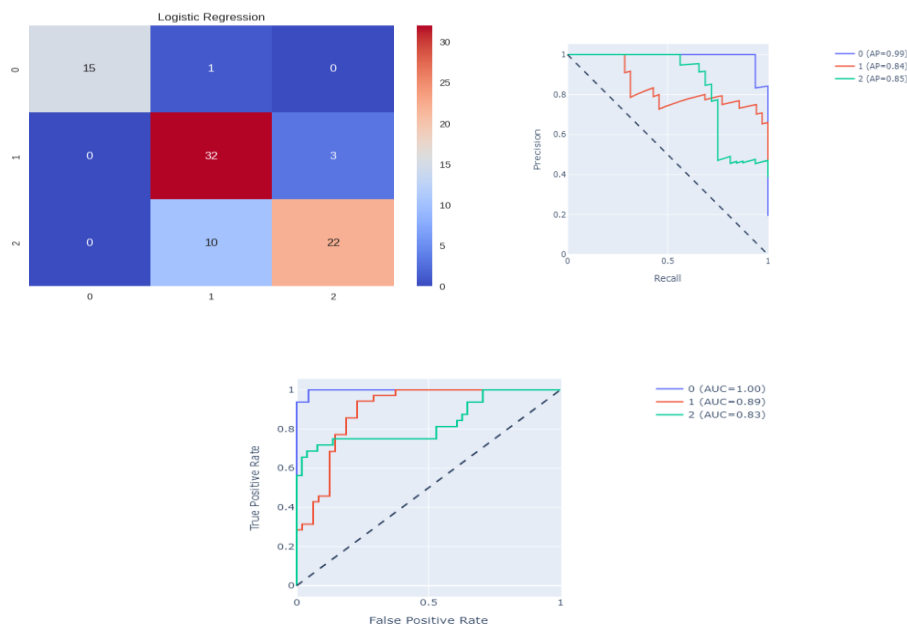


Figure 9: Performance Analysis for Logistic Regression

The Logistic Regression performs well for the moderately malnourished class (1), correctly predicting 32 cases, but struggles with distinguishing between moderately and severely malnourished children. Specifically, 10 severely malnourished children are misclassified as moderately malnourished, which reflects a challenge in separating these two classes. The model also performs well for the well-nourished class (0), with 15 correct predictions and only one misclassification. Overall, while the model is strong for moderate malnutrition, it could be improved in differentiating between severe and moderate cases.

Table 1: Result Analysis of Model Accuracy

| Model Name          | Model Accuracy |
|---------------------|----------------|
| Logistic Regression | 83.1325        |
| SVM                 | 72.2891        |
| Random Forest       | 74.9294        |
| Gaussian NB         | 72.2891        |
| Decision Tree       | 73.95432       |

Five distinct algorithms were employed in our work. The aforementioned table displays every algorithm accuracy. The Logistic Regression algorithm yields the most effective outcomes. Compared to other algorithms, it has a greater accuracy rate of 83.2%. Several algorithm analyses are included in this section. In the context of child malnutrition, factors like environment, family care, physical wellbeing, and food intake are crucial determinants that can be incorporated into a Logistic Regression model. These factors, alongside nutritional data such as the consumption of whole grains (which aid in bone growth, digestion, and bowel function), can serve as predictors of malnutrition categories. Logistic regression can help quantify how each of these factors influences the likelihood of a child being well-nourished, moderately malnourished, or severely malnourished. For example, environmental conditions, inadequate family care, and poor physical wellbeing might be significant predictors of higher odds for moderate or severe malnutrition. In contrast, a balanced diet, including fiber-rich foods like whole grains, might decrease the likelihood of falling into malnutrition categories. Critical life stages, such as those between 1 to 9 years, are also tied to specific nutritional risks like food allergies or iron deficiency, which should be accounted for in the model as part of the nutritional and health data. For instance, Logistic Regression can identify the likelihood that children with inadequate iron intake between ages 1 and 3 are more prone to severe malnutrition. Additionally, when focusing on gender-specific factors, such as puberty in girls, a Logistic Regression model can integrate physiological and mental health changes during this period, which may heighten the risk of malnutrition if proper care is not provided. Logistic regression helps in predicting the probability of malnutrition while accounting for these dynamic, interacting variables, thus offering a robust understanding of the most influential risk factors for targeted interventions.

## V. CONCLUSION

Machine learning can play a critical role in addressing child malnutrition by helping to identify and address the root causes of the problem. By analyzing large amounts of data, machine learning algorithms can help identify patterns and trends that can inform targeted interventions to prevent and treat child malnutrition. Child malnutrition is a serious issue that affects millions of children worldwide. It can have long-term negative effects on the physical and cognitive development of children, leading to lifelong health problems. Machine learning has the potential to improve our understanding of child malnutrition and develop more effective strategies for prevention and treatment, leading to better health outcomes for children around the world. In the future, Deep Convolutions Neural Network will be something that we will be able to implement.

## VI. LIMITATION AND FUTURE WORK

There are some significant limitations to be aware of, even if this study offers relevant information on forecasting students' mental health. One of these is the extremely tiny dataset consisting of only 414 survey responses, which may limit the model's applicability in other contexts. Furthermore, there's a possibility that response bias will be

created because the data is dependent on self-reporting, which would diminish the validity of mental health assessments. To further enhance the model's reliability, further studies ought to endeavor to augment the dataset's magnitude and incorporate diverse data sources. Future studies in this field should primarily concentrate on the limits that have been discovered. This necessitates the exploration of strategies for minimizing response bias and larger-scale data collection projects in order to strengthen the model's robustness. Further investigation may focus on developing increasingly extensive machine learning models and integrating relevant variables. Studies with long-term planning have the advantage of being able to track changes in mental health over time and offer an in-depth understanding of the changing mental health of Bangladeshi students.

#### ACKNOWLEDGMENT

We wish to extend our sincere thanks to the Binary Intelligence Lab at Daffodil International University for their invaluable support in the preparation of datasets and conducting laboratory work. This research did not receive any specific funding from public, private, or nonprofit organizations.

#### REFERENCES

- [1] Chung, J., Teo, J. (2023). Single classifier vs. ensemble machine learning approaches for mental health prediction. *Brain informatics*, 10(1), 1-10.
- [2] Bieliński, A., Rojek, I., Mikołajewski, D. (2023). Comparison of Selected Machine Learning Algorithms in the Analysis of Mental Health Indicators. *Electronics*, 12(21), 4407.
- [3] Yeung, H. W., Stolicyn, A., Buchanan, C. R., Tucker-Drob, E. M., Bastin, M. E., Luz, S., ... Smith, K. (2023). Predicting sex, age, general cognition and mental health with machine learning on brain structural connectomes. *Human Brain Mapping*, 44(5), 1913-1933.
- [4] Van Mens, K., Lokkerbol, J., Wijnen, B., Janssen, R., de Lange, R., Tiemens, B. (2023). Predicting Undesired Treatment Outcomes With Machine Learning in Mental Health Care: Multisite Study. *JMIR Medical Informatics*, 11, e44322.
- [5] Jage, S., Chaudhari, S., Jatte, M., Mhatre, A., Mane, V. (2023, August). Predicting Mental Health Illness using Machine Learning. In *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)* (pp. 1-5). IEEE.
- [6] Nova, K. (2023). Machine Learning Approaches for Automated Mental Disorder Classification based on Social Media Textual Data. *Contemporary Issues in Behavioral and Social Sciences*, 7(1), 70-83.
- [7] Vaishnavi, K., Kamath, U. N., Rao, B. A., Reddy, N. S. (2022). Predicting mental health illness using machine learning algorithms. In *Journal of Physics: Conference Series* (Vol. 2161, No. 1, p. 012021). IOP Publishing.
- [8] Srividya, M., Mohanavalli, S., Bhalaji, N. (2018). Behavioral modeling for mental health using machine learning algorithms. *Journal of medical systems*, 42, 1-12.
- [9] Jain, T., Jain, A., Hada, P. S., Kumar, H., Verma, V. K., Patni, A. (2021, September). Machine Learning Techniques for Prediction of Mental Health. In *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 1606-1613). IEEE.
- [10] Shafiee, N. S. M., Mutalib, S. (2020). Prediction of mental health problems among higher education student using machine learning. *International Journal of Education and Management Engineering (IJEME)*, 10(6), 1-9. *ing Applications (ICIRCA)* (pp.1606-1613). IEEE.
- [11] Khondoker, M., Dobson, R., Skirrow, C., Simmons, A., Stahl, D. (2016). A comparison of machine learning methods for classification using simulation with multiple real data examples from mental health studies. *Statistical methods in medical research*, 25(5), 1804-1823.
- [12] Sahlan, F., Hamidi, F., Misrat, M. Z., Adli, M. H., Wani, S., Gulzar, Y. (2021). Prediction of mental health among university students. *International Journal on Perceptive and Cognitive Computing*, 7(1), 85-91.

- [13] Laijawala, V., Aachaliya, A., Jatta, H., Pinjarkar, V. (2020, June). Classification algorithms based mental health prediction using data mining. In 2020 5th international conference on communication and electronics systems (ICCES) (pp. 1174-1178). IEEE.
- [14] Khan, S. I., Islam, A., Hossen, A., Zahangir, T. I., Hoque, A. S. M. L. (2018, October). Supporting the treatment of mental diseases using data mining. In 2018 International Conference on Innovations in Science, Engineering and Technology (ICISSET) (pp. 339-344). IEEE.
- [15] M. L. Bermingham, R. Pong-Wong, A. Spiliopoulou, C. Hayward, I. Rudan, H. Campbell, and C. S. Haley, "Application of high-dimensional feature selection: Evaluation for genomic prediction in man," *Scientific Reports*, vol. 5, p. 10312, 2015.
- [16] J. Ehiabhi and H. Wang, "A systematic review of machine learning models in mental health analysis based on multi-channel multi-modal biometric signals," *BioMedInformatics*, vol. 3, no. 1, pp. 193-219, 2023.
- [17] A. Cipriani, T. A. Furukawa, G. Salanti, A. Chaimani, L. Z. Atkinson, Y. Ogawa, and J. R. Geddes, "Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: A systematic review and network meta-analysis," *Lancet*, vol. 391, no. 10128, pp. 1357–1366, 2018.