"A Dual Approach: Naïve Bayes and Decision Tree for Malnutrition Prediction "

[1]Manaswini P, [2]Adarsh S, [3] Parikshith Samanth, [4]Sai Srusti A S
Information Science and Engineering, Malnad College of Engineering
Hassan-573202, India
Email id:manaswiniprakash0@gmail.com, adarsh1308.s@gmail.com, parikshithsamanth@gmail.com, gowdasrusti@gmail.com

Abstract— **The system of support for medical decisions for the prediction of malnutrition in children deals with a critical problem of malnutrition in children under the age of five, especially in developing countries where malnutrition significantly affects health and future economic productivity. The proposed system uses key anthropometric and clinical parameters, including age, gender, height, weight, weight for age (WAZ), heights for age (HAZ) and weight score (WHZ) to assess the state of nutrition. Two machine learning algorithms, naive bayes and decision -making tree have been implemented since zero to support real -time prediction using dynamic data sets. The method of extraction of gross force elements was used to identify the most influential attributes affecting malnutrition. Experimental evaluation showed that the naive Bayes model achieved an accuracy of 94% with a minimum calculation time and exceeded the model of the decision tree that reached 85%. The system integrates Visual Studio as a frontend and SQL server as a backend, which demonstrates practical usability for healthcare professionals and public health organizers.**

Keywords— *Data Science, Malnutrition, Machine Learning, Naïve Bayes, Decision Tree, Health Informatics, Classification*

I. INTRODUCTION

In today's world, nutrition is more important and children should take adequate nutrition for adequate growth, development and survival [3]. The current system is a manual process of child analysis and leads to less accurate results and not a suitable method for prediction of malnutrition. The current system includes tedious tasks and means more time and is more expensive [4]. There are many factors that affect malnutrition of children such as age, sex, height, weight, WAZ, WHZ, etc. Detection of malnutrition is important in today's world. The current system is a manual process of child analysis and leads to less accurate results and not a suitable method for prediction of malnutrition. There is no automation to predict the state of malnutrition of children.
The system classifies children in stunted state, low, unnecessary and nutritional edema. We use parameters such as age, sex, height, weight, WAZ, WHZ, etc. Classifiers are used to predict the child's malnutrition, such as naive bayes and decision tree. We collect data files from online sources such as "Kaagle.com". This can be developed as a useful application for doctors. We use effective technologies such as "Visual Studio" and "SQL Server" to develop applications. Data science algorithms are generally based on mathematics and statistics [5]. The main advantage of automatic learning, traditional software is that there is no written code that shows the system to make a decision to choose the right object between two different, because it is difficult to handle every situation in the object [6]. Automatic learning works for it. Use the data to take intelligent decisions, provide future predictions or detect anomalies. Today it is used in various applications such as autonomous cars, virtual assistants, search engines, object identification, advertising, predictive analysis, etc. It can be classified in three types: supervising learning, learning and strengthening. The system uses data science for prediction.

II. RELATED WORKS

In this section, we summarize and examine current research related to infant malnutrition prediction the usage of AI/ML algorithms. These researches examine the numerous algorithms, datasets, and methodologies used to expect and classify child malnutrition.

A. *Paper Title: Efficient Machine Learning for Malnutrition Prediction among under-five children in India Authors: Saksham Jain, Tayyibah Khanam, Ali Jafar Abedi,Abid Ali Khan Year: 2022*

Description: Infant malnutrition is taken into consideration to be one of the main causes of infant mortality and malnutrition. This take a look at turned into aimed to leverage the advantages supplied by way of device gaining knowledge of models in phrases of figuring out and accurately predicting great elements of malnutrition. For this look at, the kids's recode files from the Indian Demographic and fitness Survey (IDHS) datasets from 2005-2006 and 2015-2016 were used. To have a look at the nutritional status of youngsters aged 0-59 months, this have a look at seems at stunting (top-for-age), wasting (Weight-for height), and concurrent stunted losing (peak-for-age-Weight for-top). Ordinary device mastering models, Tabular Deep getting to know frameworks, H2O base fashions, and automobile ML models are the 4 sorts of machine studying fashions hired in our research. This research located that automated system getting to know algorithms and Tabular Deep gaining knowledge of frameworks.
Methodology: Regular Machine Learning models, Tabular Deep Learning frameworks, H2O base models, and AutoML models
Limitations: Existing studies on child malnutrition prediction the usage of AI/ML algorithms have boundaries, including using smaller datasets that may not absolutely constitute the trouble. Those models are also often now not suitable for actual-time packages and keep in mind simplest a restricted set of factors, which may also affect the accuracy of predictions.

B. *Paper Title: Ensemble Approach for Early Prediction of Malnutrition Level of Children: A Case study on Children under Five Years Old Authors: H.M.C. Nirmani, U.P. Kudagamage Year: 2024*

Description: Malnutrition, an international health difficulty affecting youngsters, leads to extreme brief- and lengthy-time period results on their boom and improvement, in particular within the essential 1–60 months age range. This observe aims to increase a predictive version for early detection of malnutrition the usage of ensemble mastering strategies, based on data from 574 kids underneath five inside the Lunugala location. After information preprocessing, the dataset was cut up into 70% education and 30% checking out sets, and nine gadget gaining knowledge of algorithms, consisting of SVM, selection Tree, Random woodland, and XGBoost, had been evaluated.

Methodology: Support Vector Machine (SVM), Logistic Regression (LR), AdaBoost and Extreme Gradient Boosting (XGBoost) were used.

Limitations: Many studies focus entirely on constructing ML models for toddler malnutrition prediction, however these fashions are regularly not well matched with actual-time use. In addition they tend to rely upon smaller datasets, which can limit the robustness of the predictions. Moreover, in some instances, the evaluation of the ML fashions is not performed very well, main to an incomplete know-how in their overall performance and capability obstacles.

C. *Paper title: Implementation of Hybrid Bat Algorithm-Ensemble on Side Effect Prediction: Case Study Metabolism and Nutrition Disorders Authors: Dzaky Raihan Ahmad, Jondri, Isman Kurniawan Year: 2024*

Description: Drug aspect effects, ranging from moderate issues like nausea to intense outcomes, pose vast health dangers. Contemporary detection methods, reliant on scientific trials, are luxurious and often omit capability results. This examine addresses those challenges through the usage of a Hybrid Bat algorithm (HBA) for feature choice and an ensemble technique combining Random forest, AdaBoost, and XGBoost to are expecting drug side effects, that specialize in metabolism and nutrients issues.

Methodology: Hybrid Bat Algorithm (HBA), Random Forest, AdaBoost, and XGBoost

Limitations: A few studies goal to expect drug side effects, with a focus on metabolism and vitamins issues, but they're unable to as it should be expect child malnutrition. These fashions regularly generate much less correct consequences, as they're now not specifically designed for malnutrition prediction, proscribing their applicability and effectiveness in addressing infant malnutrition issues. These models lack key indicators specific to children such as MUAC and age height, which reduces their accuracy in the prediction of malnutrition. In addition, their limited assessment of socio -economic and developmental factors further limits their importance in assessing childhood malnutrition.

D. *Paper Title: Machine Learning Approaches for Prediction of Nutrition Deficiency among Women of Different Age Groups Authors: Javeria Ali , Waseemullah, Masood Ahmed Khan ,Najeed Ahmed Khan Year: 2022*

Description: Malnutrition is a sizable fitness problem among girls in underdeveloped areas, in particular in Pakistan, due to constrained healthcare access. This take a look at aims to predict nutritional deficiencies the usage of system getting to know models like SVM, Random forest, Logistic Regression, KNN, and Naïve Bayes. Outcomes display Random woodland accomplished the best accuracy, making it the maximum dependable for future predictions. Testing on similar datasets showed the model's reliability for assessing women's health conditions.

Methodology: Support vector machine, random forest, Logistic Regression, KNN algorithm used

Limitations: Some studies focus on predicting nutrition deficiencies for women, with age-based constraints, making them unsuitable for child malnutrition prediction. Additionally, the datasets used in these studies are often incompatible with the needs of child malnutrition models, limiting their relevance and accuracy for this specific issue.

2.1 Comparison and Motivation for Our Approach

Although previous studies have used naive bayes and algorithms of decision -making trees for medical diagnostic tasks, many depend on static data sets and previously constructed automatic learning libraries, which limit adaptability in real -time environments. On the contrary, the present work implements both the models of Bayes and decision -making trees to operate with dynamic data sets, offering greater flexibility and transparency. This approach allows more fine control over data preprocessing, model logic and adaptability in real world environments, such as malnutrition prediction in various field conditions.

In addition, existing literature often lacks a mechanism for selecting robust characteristics adapted to malnutrition factors. To overcome this, a strategy for the extraction of gross force characteristics was integrated to identify the most influential attributes, thus improving both the precision and interpretability of the system. These improvements address the key limitations observed in previous methodologies and demonstrate better suitability for real -time medical decisions support systems.

III. PROPOSED WORK

Proper nutrition is an essential component for the survival, growth and development of children. Malnutrition is a global problem in current life. The main objective of the proposed system is to predict the state of malnutrition of children under five years. The system classifies children in stunted state, low, wasted and nutritional edema. We use parameters such as age, gender, height, weight, waz, whz, etc. Classifiers used for the prediction of child malnutrition. We use efficient classifiers such as the Naïve Bayes and decision tree. We collect data sets from online sources such as "kaagle.com". We can develop this as a useful application for doctors. We use efficient technologies such as "Visual Studio" and "SQL Server" for the development of applications.

IV. METHODOLOGY

4.1 Data Science: Data Science is a process of reading data from extraordinary prospective s and extraction of useful facts from the processed facts. Information science applied on n quantity of fields and used to resolve real international troubles. Records technology helps many strategies.

Classification Rules (Classifiers): Basically class is used to categorise each object in a fixed of information into one of the predefined set of training or businesses. Class methods make use of mathematical strategies for problem fixing

Ex: Worker statuses in an employer (leaves or stay)

To predict which cutting-edge employees are possibly to leave inside the destiny.

Naive Bayes and decision -making algorithms have been selected for this study due to their effectiveness in solving numerical data, simplicity of interpretation and suitability for small and large medical data sets. These models are known for their rapid calculation time and the ability to handle more independent functions that match well with the multi -factor nature of malnutrition prediction. Their proven performance in problems with the classification of medical classification also supports their use in clinical decision -making systems in real time.

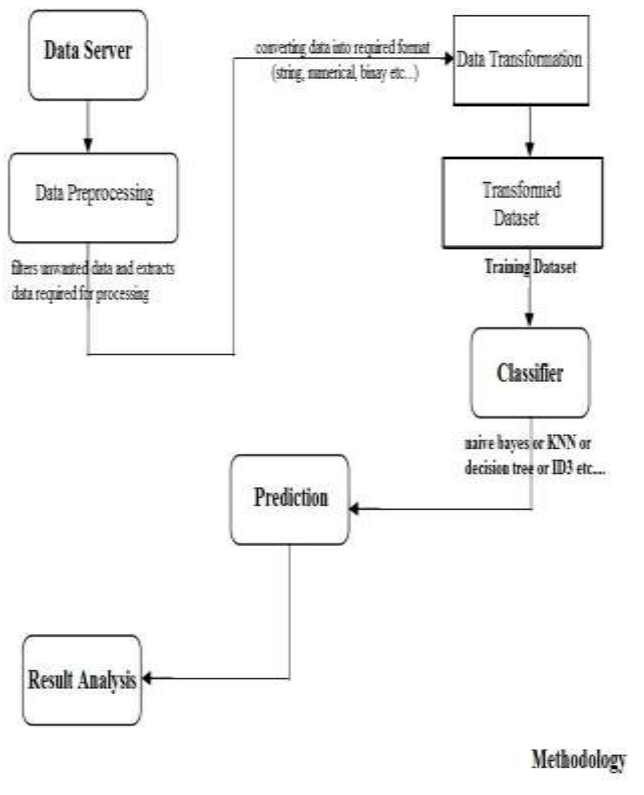


Fig.1. Methodology

4.2 Naïve Bayes algorithm

The classification process using the Bayesian algorithm was carried out through the following steps:

Step 1: Acquisition and preparation of data: The data set was collected from storage systems that include structured databases, cloud storage and Excel sheets. The data were prepared to address missing values, normalize scales and format categorical variables.

Step 2: Calculation of attributes probability: For each class label in the data set (for example, types of malnutrition), the probability of each attribute value was calculated using the probability formula for m-problem to avoid zero frequency problems in dispersed data scenarios.

Step 3: Application of the M-Estimate Formula

P(Ai|Vj) = n + mP / n+m

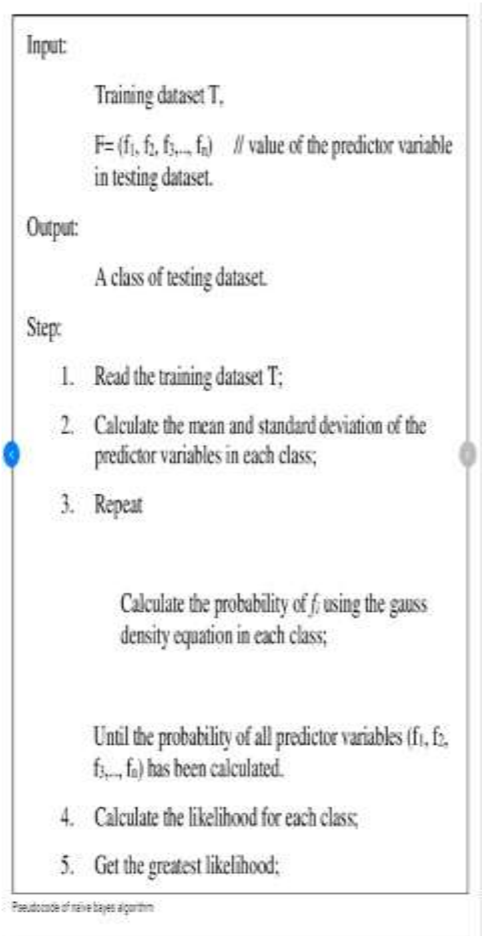
Where:

- n is the total number of training samples where the class is Vj
- nc is the number of samples where the attribute Ai appears within class Vj.
- P is the prior estimate of P(Ai|Vj)
- m is the equivalent sample size determining the strength of prior belief

Step 4: Probabilistic classification. The subsequent probability for each class was calculated using:

$$P(\text{Class} \mid \text{Attributes}) \propto P(\text{Class}) \times \prod_{i=1}^{n} P(A_i \mid \text{Class})$$

Step 5: Final class prediction. The class with the most calculated subsequent probability was selected as a predicted result for the entry data instance. This method allows efficient classification, particularly in real -time environments where speed and interpretability are crucial.



Input:

Training dataset T,

F= ($f_1, f_2, f_3,..., f_n$) // value of the predictor variable in testing dataset.

Output:

A class of testing dataset.

Step:

1. Read the training dataset T;
2. Calculate the mean and standard deviation of the predictor variables in each class;
3. Repeat

   Calculate the probability of f_i using the gauss density equation in each class;

   Until the probability of all predictor variables ($f_1, f_2, f_3,..., f_n$) has been calculated.
4. Calculate the likelihood for each class;
5. Get the greatest likelihood;

Pseudocode of naive bayes algorithm

Fig.2. Naïve Bayes Algorithm Flow

4.3 Decision Tree algorithm

Step 1: Scan the data set (storage servers)

Step 2: For each attribute, calculate the gain [number of occurrences]

Step 3: Let A_Best be the attribute of the greatest gain [the highest count]
Step 4: Create a decision -based decision node - recovery of nodes [records] where attribute values coincide with A_Best.
Step 5: Resort to the sub-lists [patient list] and calculate the results count [results], called sub-nodes. Based on the highest count, we classify the new node.

Input: Dataset

Output: Post-pruned decision tree with Bayes minimum risk function

```
Generate initial decision tree based on C4.5
For each node,
    If node is a parent node then
        Compute total leaf risk-rate (R_l) and parent risk-rate (R_p)
        If (R_p > R_l) then
            Convert parent node to a leaf
        End if
    End if
End for
Return the final tree
```

Fig.3. Decision Tree Algorithm Flow

4.4 Data division and validation
To guarantee a robust model evaluation, we apply K-Fold cross validation with different trains/test division relationships, including 50:50, 60:40, 70:30 and 90:10. The purpose of testing multiple divisions was to observe the stability and performance of classifiers under different training data sizes. We discovered that the 90:10 relationship produced the best performance results for both Bayes and Decision trees algorithms, particularly in terms of precision and classification consistency.

4.5 Dataset description
The dataset used in this study was obtained from Kaggle[6].
It contains approximately 3,000 records related to children under five years. Each record includes attributes such as age, gender, height, weight, whz (z score for height), waz (z score for age), make (height for age score) and other clinical parameters. The data set is labeled with four different kinds of results: 0 - Normal 1 - Stuned 2 - Low weight 3 - obese Before applying automatic learning models, we perform data preprocessing using the binning method. This technique was used to handle noise data, soft inconsistencies and address missing values. Binning helped improve the reliability of the data set by grouping continuous values in discrete intervals, allowing a better classification performance.

4.6 Description of Clinical Parameters

- Whz (weight height z score): Whz evaluates body weight in relation to height, identifying waste (acute malnutrition). A whz below -2 indicates moderate wear, and below -3 indicates severe wear. It is calculated by comparing a child's weight with the average weight of a reference population of the same height and turning this with a Z score.
- Waz (weight age z score): Waz measures body weight in relation to age. Identify the low weight state but does not differentiate between acute and chronic malnutrition. A WAZ below -2 is considered low weight, and below -3 indicates low severe weight.
- Haz (height age z score): HAZ is used to evaluate stunning (chronic malnutrition). It compares the height of the child with a standard reference population of the same age. HAZ pod -2 means stunning and below -3 means serious staffing.
- MUAC (Circumference of the upper middle arm): Muac is a simple detection tool to assess acute malnutrition. It measures the circumference of the upper arm at the midpoint between the shoulder and the elbow. A MUAC value less than 115 mm indicates severe acute malnutrition.
- Age: Age (in months or years) is a crucial factor to evaluate malnutrition, since growth standards differ with age. It is used to calculate Z scores and the understanding of development standards.
- Gender: Gender plays a role in nutritional evaluations, since growth standards are specific to sex. The calculations of the Z score differ slightly between boys and girls.
- Weight: The child's weight (in kilograms) is a fundamental entrance used in the calculations of Whz and Waz. Variations in weight are directly related to nutritional status.
- Height/length: The height (or the length for babies) is used in hac and whz calculations. Provides information on acute and chronic malnutrition.

4.7 Evaluation metrics
To guarantee a significant performance evaluation, especially in the context of medical diagnosis where erroneous classifications can have significant consequences, we use the confusion matrix to evaluate automatic learning models: the Bayesian and decision -making decision tree.
The confusion matrix provides a breakdown of:
True positive (TP): Malnourished cases predicted correctly
False positive (FP): Malnourished cases predicted incorrectly
True negative (TN): Healthy cases predicted correctly
False negatives (FN): Lost malnourished cases
This allows a more detailed understanding than raw precision alone.
Although commonly used metrics, such as F1 score, precision recovery and ROC-AUC are standard in medical data sets due to class imbalance, we prioritize applicability and real-time interpretability. Our project is intended for health support, where speed and clarity are critical for decision making. Therefore, we focus on the metrics that are aligned with the usability and simplicity in real time.
In future work, we plan to incorporate reference comparisons with models such as logistics regression and majority class predictors to compare the effectiveness of our proposed models and support a more complete evaluation through metrics such as the F1 and Roc.

4.8 System Architecture and Real-time Integration
To demonstrate the integration of Visual Studio and SQL Server in our malnutrition prediction system, we design a modular architecture that admits the processing, storage and visualization of data in real time. The system consists of three main actors: administrator, members (health workers) and the server. The architecture allows efficient interaction between the user interface and the backend prediction engine. This setting

ensures smooth data flow, allowing healthcare workers to enter the patient's details, start predictions and load the results through a user-friendly interface.

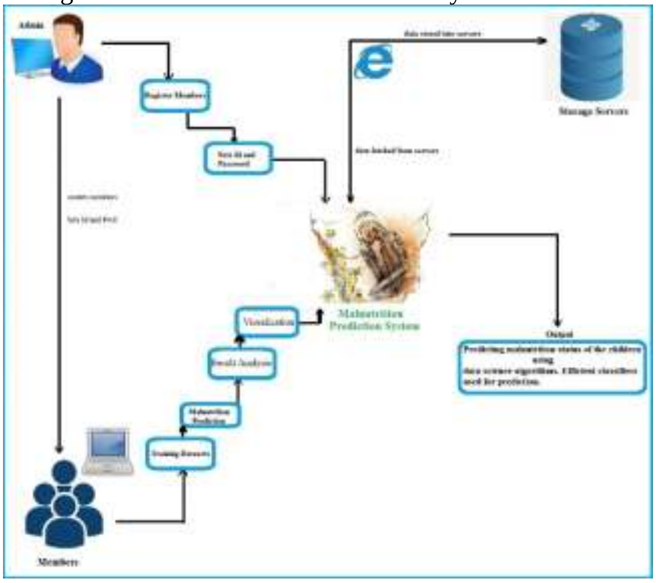


Fig.4. System Architecture of Malnutrition Prediction

Description: The administrator records users and establishes credentials for safe access. Members use the interface (developed in Visual Studio) to enter children's health data, which is stored and recovered through SQL Server. The data undergo an automatic learning classification using trained models. The results are visualized and analyzed in real time, with the exit predicted by the child's malnutrition status using efficient classifiers such as Bayesian and Decision Tree.

V. EXPERIMENT RESULTS

Malnutrition Prediction - Bayes Naive Algorithm Results
Results analysis

5.1 Naive Bayes Algorithm

The evaluation was carried out using multiple K-Fold cross validation divisions such as 50:50, 60:40, 70:30 and 90:10. The 90:10 division gave the most optimal precision results. Here we build an application in real time useful for society. This project builds with Microsoft Technologies. Trained medical data sets with the Naive Bayes algorithm and we obtained very good results. The Naive Bayes algorithm is programmed in such a way that it works for dynamic data sets. The logic of the Naive Bayes algorithm is written and is our own library. We are obtaining about 94% of the precise results and takes about 1500 mili seconds for the prediction.

Constraint	Naive Bayes Algorithm
Accuracy	94%
Time (milli secs)	1606
Correctly Classified (precision)	94%
Incorrectly Classified (Recall)	6%

TABLE I. NB RESULTS

5.2 Decision Tree Algorithm

Constraint	DT Algorithm
Accuracy	85%
Time (milli secs)	2806
Correctly Classified (precision)	85%
Incorrectly Classified (Recall)	15%

TABLE II. DECISION TREE RESULTS

5.3 COMPARATIVE ANALYSIS OF ALGORITHMS

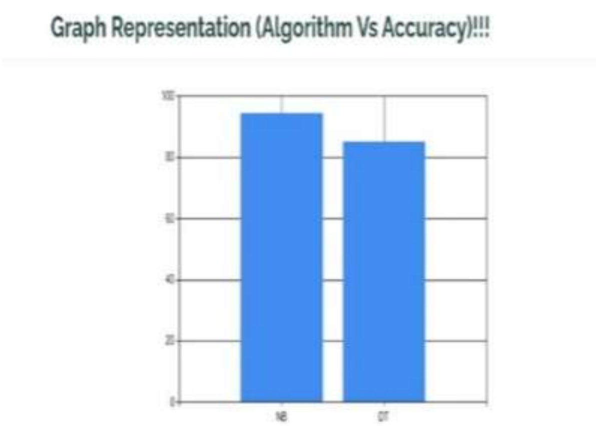


Fig.5. Comparative Analysis of ML Models

VI. CONCLUSION

Proper nutrition is an essential component for the survival, growth and development of children. Malnutrition is a global problem in current life. The main objective of the proposed system is to predict the state of malnutrition of children under five years. The system classifies children in stunted state, low, wasted and nutritional edema. We use parameters such as age, gender, height, weight, waz, beam, whz, etc. Classifiers used for the prediction of child malnutrition. The system also predicts anemia and suggests adequate diet recommendations for users. The system is an application in real time useful for the medical sector.

VII. FUTURE ENHANCEMENTS

In addition, a greater number of algorithms can be used and algorithms can compare to identify the efficient algorithm. More training data sets can be used for prediction.

VIII. REFERENCES

[1] "Malnutrition in Children," UNICEF, Accessed: Jun. 13, 2019.[Online].Available: https://data.unicef.org/topic/nutrition/malnutrition/

[2] "Joint child malnutrition estimates - Levels and trends (2019 edition)," World Health Organization, Jun. 6, 2019. Accessed: Jun. 13, 2019. [Online]. Available: https://www.who.int/nutgrowthdb/estimates2018/en/
[3] C. G. Victora et al., "Maternal and child undernutrition: consequences for adult health and human capital," The Lancet, vol. 371, no. 9609, pp. 340–357, 2008.
[4] Ministry of Public Health, Afghanistan, Integrated Guidelines for the Management of Acute Malnutrition, 2014.
[5] R. E. Black et al., "Maternal and child undernutrition: global and regional exposures and health consequences," The Lancet, vol. 371, no. 9608, pp. 243–260, 2008.
[6] Child Malnutrition - UNICEF Dataset, Kaggle, https://www.kaggle.com/datasets/usharengaraju/child-malnutrition-unicef-dataset.
[7] S. Nandy, M. Irving, D. Gordon, S. V. Subramanian, and G. D. Smith, "Poverty, child undernutrition and morbidity: new evidence from India," Bull. World Health Organ., vol. 83, pp. 210–216, 2005.
[8] K. Maleta, "Undernutrition," Malawi Med. J., vol. 18, no. 4, pp. 189–205, 2006.
[9] "Malnutrition: Afghanistan's silent emergency," UNICEF. Accessed: Jun. 13, 2019. [Online]. Available: https://www.unicef.org/afghanistan/nutrition
[10] Ministry of Public Health, Afghanistan, Afghanistan Nutrition Cluster: 2018 Annual Report, 2019.
[11] Ministry of Public Health (MoPH), Afghanistan, Afghanistan National Nutrition Survey, 2004.
[12] Ministry of Public Health (MoPH), Afghanistan, Afghanistan National Nutrition Survey, 2013.
[13] Ministry of Public Health (MoPH), ICF, and Central Statistics Organization, Afghanistan Demographic and Health Survey 2015, 2017.
[14] A. Higgins-Steele et al., "Stop stunting: situation and way forward to improve maternal, child and adolescent nutrition in Afghanistan," Matern. Child Nutr., vol. 12, pp. 237–241, 2016.
[15] N. Akseer et al., "Geospatial inequalities and determinants of nutritional status among women and children in Afghanistan: an observational study," Lancet Glob. Health, vol. 6, no. 4, pp. e447–e459, 2018.
[16] T. Mashal et al., "Factors associated with the health and nutritional status of children under 5 years of age in Afghanistan: family behaviour related to women and past experience of war-related hardships," BMC Public Health, vol. 8, no. 1, 2008.
[17] Ministry of Public Health, Afghanistan, Integrated Management of Acute Malnutrition: National Guidelines, Kabul: Public Nutrition Directorate, 2018.
[18] S. Khare, S. Kavyashree, D. Gupta, and A. Jyotishi, "Investigation of nutritional status of children based on machine learning techniques using Indian Demographic and Health Survey data," Procedia Comput. Sci., vol. 115, pp. 338–349, 2017.
[19] S. Das, M. Z. Hossain, and M. A. Islam, "Predictors of child chronic malnutrition in Bangladesh," Proc. Pakistan Acad. Sci., vol. 45, no. 3, pp. 137–155, 2008.