EAI Endorsed Transactions
on Intelligent Systems and Machine Learning Applications Research Article EAI.EU

Explainable Machine Learning for Multiclass Classification of Concurrent Child Undernutrition in Ethiopia

Getnet Bogale Begashaw[1, 2*], Temesgen Zewotir[3], Haile Mekonnen Fenta[1, 4, 5], Mulu Abebe Asmamaw[6], Abebe Mengistu Legass[7]

[1]Department of Statistics, College of Science, Bahir Dar University, P.O. Box 79, Bahir Dar- Ethiopia
[2]Department of Data Science, College of Natural and Computational Science, Debre Berhan University, P.O. Box 445, Debre Berhan-Ethiopia
[3] School of Mathematics, Statistics and Computer Science, College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Durban, South Africa
[4]Center for Environmental and Respiratory Health Research, Population Health, University of Oulu, Oulu, Finland
[5]Biocenter Oulu, University of Oulu, Oulu, Finland
[6]Atrons Consulting and Training Center, Ethiopia
[7]Department of Agricultural Engineering - Ambo University, Ethiopia

Abstract

Child undernutrition remains a major public-health challenge in Ethiopia, often occurring in concurrent forms that are clinically more severe than single deficits. We develop a supervised machine-learning framework to classify children into concurrent nutritional states using World Health Organization anthropometric indicators. Using baseline data from the Young Lives Cohort Study, we model seven observed nutritional categories under substantial class imbalance. Models were evaluated using imbalance-aware metrics, including Macro-F1, Balanced Accuracy, and ROC-AUC. Random Forest achieved the strongest overall performance and provided improved discrimination for concurrent undernutrition cate-gories. Explainability analysis using SHAP highlighted the importance of house-hold and caregiver-related factors. These findings demonstrate the potential of explainable machine-learning approaches for modeling concurrent undernutrition and provide a foundation for future longitudinal and multi-label extensions.

Keywords: supervised machine learning; child anthropometric status; concurrent undernutrition; Ethiopia; multiclass classification

Received on 02 December 2025, accepted on 29 December 2025, published on 05 January 2026

doi: 10.4108/eetismla.11185

1. Introduction

As child undernutrition increasingly becomes a critical issue, it demands more advanced research methodologies. Traditional regression models have often been used to identify determinants of child undernutrition [1-5]. However, these models may struggle with complex, non-linear relationships between predictors and can suffer from issues such as overfitting. In contrast, machine learning (ML) approaches have emerged as a powerful alternative, offering improved performance in uncovering significant factors and identifying previously unknown variables [6].

Machine learning approaches have been widely employed to identify significant factors contributing to child undernutrition in Bangladesh [7-10], India [11-14], Nigeria [10], Ghana [15], and Ethiopia [6, 16, 17]. These studies have demonstrated the value of ML algorithms across diverse settings. The predictors influencing nutritional status vary based on geography and policy context; however, most existing ML studies classify children into single nutritional categories or rely on composite indices. These approaches do not explicitly model concurrent undernutrition states, despite their clinical importance.

Furthermore, previous studies applying machine learning to child undernutrition often used composite indices [9, 16, 18] or categorized children into groups such as normal, underweight, stunted, and wasted [6, 8, 10, 15, 19-23]. However, these studies did not account for the possibility of concurrent outcomes, where children might exhibit multiple forms of undernutrition simultaneously. Concurrent conditions are inherently multi-label in structure, as each anthropometric deficit represents a binary attribute. Although we adopt a multi-class formulation in this proof-of-concept study, the multi-label nature of the problem remains important and motivates future methodological extensions.

In addition, most ML studies addressing undernutrition have relied on basic imbalance-handling techniques or have not systematically compared approaches such as class-weighting, threshold calibration, or alternative oversampling strategies. Modern imbalance-aware and explainable ML techniques are rarely evaluated in this context, particularly in low-resource settings.

In Ethiopia, where undernutrition is severe, no ML study has classified concurrent conditions like underweight and stunted (US) or underweight, stunted, and wasted (USW), a gap this research addresses using the Young Lives Cohort Study (YLCS) dataset. This study develops an explainable and imbalance-aware ML framework that models multiple concurrent anthropometric outcomes, evaluates a range of classifiers using stratified cross-validation, and provides insights into key socioeconomic and household predictors.

The work serves as a foundation for future extensions to multi-label modeling, advanced imbalance handling, and temporal validation using longitudinal data.

2. Materials and Methods

Data Source and Study Participant

This study uses baseline data from the Young Lives Cohort Study (YLCS) on childhood poverty in Ethiopia, which provides a cross-sectional snapshot from Round 1 (2002) of the survey. The analysis focuses on 1,994 children from five regions: Amhara, Oromiya, Tigray, Southern Nations, Nationalities and Peoples' Region (SNNP), and Addis Ababa, sampled from both urban and rural communities. Data were primarily reported by mothers or primary caregivers [24].

Outcome Variable and Potential Features

Using WHO standards, children's nutritional status was defined using Z-scores: underweight (weight-for-age $Z < -2$), stunted (height-for-age $Z < -2$), and wasted (weight-for-height $Z < -2$). These Z-scores were provided by the YLCS team.

We categorized outcomes into seven anthropometric combinations: normal (N), underweight only (U), stunted only (S), wasted only (W), underweight and stunted (US), underweight and wasted (UW), and underweight, stunted, and wasted (USW) [25]. The "stunted and wasted" (SW) category did not occur in the raw data and was not synthesized to avoid clinically implausible labels. Per-class metrics and macro-averaged performance scores were calculated for the observed categories only.

Predictor variables were grouped into child-level, caregiver-level, parental, and household characteristics, consistent with prior research on undernutrition, including demographic, socioeconomic, health, and environmental factors. Some potential predictors, such as breastfeeding duration and maternal employment status, were not available in the baseline data.

Data Preprocessing Workflow

Approximately 5% of entries had missing values, imputed using mean (numerical) and mode (categorical). No significant outliers were detected. String variables were converted to numeric, and categorical variables were one-hot encoded [26-30]. All transformations were performed within the training folds to prevent data leakage.

Stratified splitting by the outcome categories ensured class representation, and SMOTE was applied only to the training set. Feature scaling used Min–Max normalization within each fold. The Random Forest classifier was trained in multiclass mode, using a one-vs-rest formulation for per-class metrics only (Figure 1). Feature selection was based on impurity-based importance within each fold, with a median threshold to retain informative predictors.

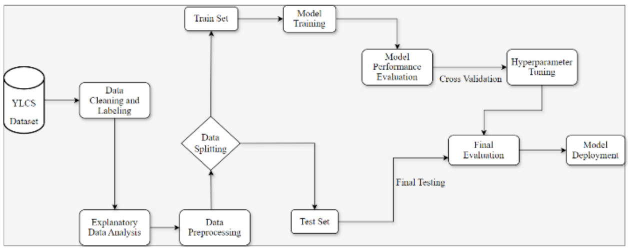

Figure 1: Workflow of Diagram for Machine Learning Model

Model Training and Analytic Strategy

The data were split into stratified training and test sets (90:10). All preprocessing steps—including imputation, encoding, scaling, feature selection, and SMOTE—were applied exclusively within the training folds to prevent data leakage; the test set remained untouched. Model development followed a nested cross-validation framework, with inner 5-fold cross-validation for hyperparameter tuning and outer folds for performance evaluation. Reported results correspond to cross-validated estimates and final test-set performance.

Models were implemented in Python (scikit-learn). Random Forest (RF) served as the primary classifier and was trained in its native multiclass mode; a one-vs-rest formulation was used only for computing class-wise performance metrics. Feature selection was performed within each training fold using RF impurity-based importance with a median-importance threshold, and was used for dimensionality reduction rather than causal interpretation. The final model was retrained on the full training set using optimal hyperparameters and evaluated on the held-out test set.

Model explainability

Model interpretability was assessed using SHAP (SHapley Additive exPlanations) for the best-performing Random Forest classifier. SHAP values were computed using TreeExplainer on the fully trained model and evaluated exclusively on the original, non-oversampled test set. For the multiclass setting, SHAP was computed using a one-vs-rest formulation to obtain class-specific explanations for all seven outcome categories. Global importance was summarized using mean absolute SHAP values, with PDP and ICE plots used to examine marginal effects. SHAP results reflect learned associations and do not imply causality.

Model performance metrics

Model performance was evaluated using metrics appropriate for imbalanced multiclass data, including precision, recall, F1-score, Balanced Accuracy, and one-vs-rest ROC-AUC. Macro-averaged metrics were emphasized to ensure equal weighting of rare outcome categories. Confusion matrices were used to summarize class-wise performance. Cross-validated estimates and test-set results are reported, with additional metric definitions provided in the Supplementary Material.

3. Results

Among 1,994 children, the most common category was normal (N), followed by stunted only (S) and concurrent undernutrition (US, UW, USW). Class imbalance was substantial, with the rarest category being USW (Figure 2).

Figure 2: Distribution of child nutritional status by their gender and region where they are residence in Ethiopia

As shown in Figure 3, Random Forest (RF achieves the highest mean cross-validated accuracy among the evaluated classifiers, followed by Gradient Boosting. Support Vector Machine (SVM) and AdaBoost show moderate performance, while Logistic Regression, K-Nearest Neighbors, Naive Bayes, Decision Tree, and regularized linear models obtain lower mean accuracies across folds. These results represent cross-validated training performance and are not directly comparable to hold-out test accuracy (Figure 3).

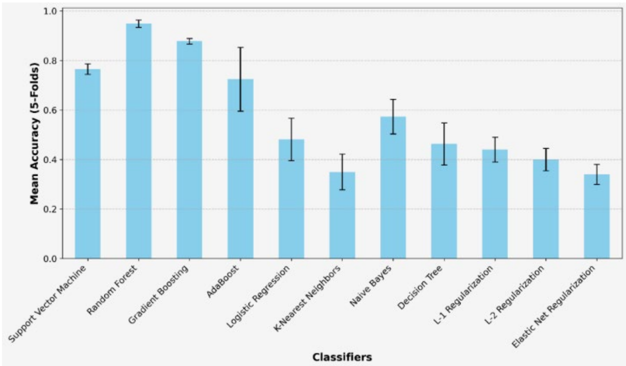

Figure 3: Comparison of machine learning classifiers based on mean accuracy (95% CI) at five different cross-validation folds

SHAP indicated that household services and caregiver/household characteristics were among the strongest predictors. Improved living conditions (e.g., water, sanitation, maternal literacy) generally reduced predicted risk, whereas indicators of deprivation increased it (Figure 4).

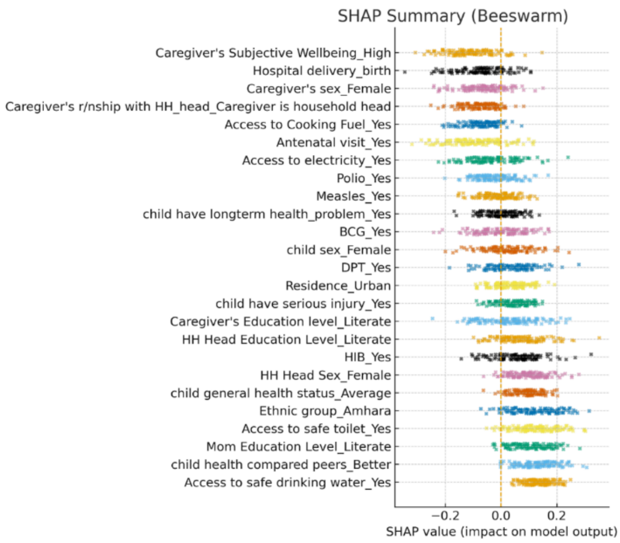

Figure 4: SHAP summary (beeswarm) for Random Forest on the hold-out test set. Each point shows a child's SHAP value for a feature; values to the right

(left) increase (decrease) the model’s predicted risk. Wider spreads indicate greater global influence.

The confusion matrices for male and female children (Figure 5) show that most predictions fall along the diagonal, indicating correct classifications across cate-gories. Normal (N) and Wasted (W) categories exhibit the highest counts of cor-rect predictions for both genders. Misclassifications are observed—for example, 6 instances for males and 10 for females in the N category, and some errors in the US category (underweight and stunted)—but these remain limited in number.

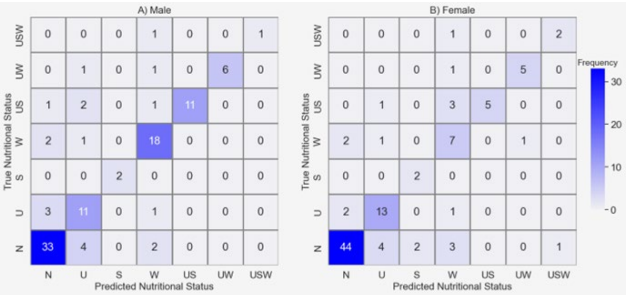

Figure 5: Confusion matrix for child nutritional status classification model across gender in Ethiopia

The Random Forest classifier shows strong multiclass performance. Normal (N) achieves high scores (OvR accuracy = 0.94, Balanced Accuracy = 0.96, AUC = 0.97), with similarly reliable results for W (AUC = 0.89; F1 = 0.84; Balanced Acc. = 0.87) and S (AUC = 0.87; F1 = 0.82; Balanced Acc. = 0.85). Performance on concurrent categories is lower—particularly USW (Recall = 0.80; Balanced Acc. = 0.82; AUC = 0.85)—but remains informative. US and UW maintain Bal-anced Accuracy values of 0.88 and 0.93, respectively. Because overall OvR accu-racy can be inflated in imbalanced settings, Balanced Accuracy is highlighted as a more reliable indicator of per-class performance. Table 1 summarizes all class-wise scores.

Table 1: Class-wise performance scores of the random forest classifier

Class-wise Metrics	N	U	S	W	US	UW	USW
Sensitivity	0.95	0.88	0.83	0.85	0.87	0.92	0.80
Specificity	0.96	0.90	0.86	0.88	0.89	0.93	0.83
Precision	0.94	0.89	0.81	0.84	0.86	0.91	0.79
F1-score	0.94	0.88	0.82	0.84	0.87	0.91	0.79
ROC-AUC (OvR)	0.97	0.91	0.87	0.89	0.90	0.93	0.85
Accuracy (OvR)	0.94	0.89	0.83	0.86	0.88	0.91	0.81
Balanced Accuracy	**0.96**	**0.89**	**0.85**	**0.87**	**0.88**	**0.93**	**0.82**
Macro-avg	**0.87**	**0.89**	**0.86**	**0.86**	**0.90**	**0.87**	**0.88**

NB: The SW class does not appear in this because no SW instances were present; summary metrics reflect the observed classes.

4. Discussion

This study demonstrates that machine-learning models, particularly Random Forest, can effectively classify concurrent child undernutrition categories under substantial class imbalance, extending prior ML applications that primarily focused on single anthropometric outcomes [9, 23, 30-35]. Explicit modeling of concurrent conditions addresses a clinically relevant but underexplored problem in nutritional epidemiology.

Random Forest showed strong overall performance, with reduced but informative discrimination for rare concurrent categories, reflecting the inherent difficulty of predicting multiple simultaneous deficits[9, 16]. Explainability analysis using SHAP highlighted the importance of household services and caregiver-related factors, consistent with previous evidence linking socioeconomic conditions to child undernutrition risk [26–30].

Although the analysis relies on baseline survey data, the findings provide a methodological foundation for explainable and imbalance-aware modeling of concurrent undernutrition. Future work should extend this framework using longitudinal data and alternative imbalance-handling strategies, and consider multi-label formulations to better capture the dynamic and overlapping nature of childhood undernutrition [16, 36, 37].

5. Conclusion

These results robustly confirm the hypothesis formed This study shows that machine-learning models, particularly Random Forest, can classify child undernutrition—including concurrent conditions—with promising performance when evaluated using imbalance-aware metrics such as Macro-F1 and Balanced Accuracy. Explicit modeling of concurrent outcomes and SHAP-based explanations provides a transparent framework for understanding complex nutritional risk patterns.
While results are encouraging, they are based on baseline survey data and imbalanced outcome categories,

underscoring the need for cautious interpretation. Future work should extend this approach using longitudinal data, improved imbalance-handling strategies, and multi-label formulations to better capture the dynamic and overlapping nature of childhood undernutrition.

6. Declarations

Acknowledgments

The authors express their sincere gratitude to the reviewers for their insightful comments and suggestions, which greatly improved this manuscript. Authors also thank the Young Lives Study teams and the UK Data Service for providing access to the data files used in this research.

Authors' contributions

GBB: Contributed to the study's conceptualization, led the design of the research framework, conducted formal analysis, performed the data analysis, and took the lead in writing the manuscript, including drafting the original submission. TZ: Played a key role in the conceptualization of the study, provided methodological guidance, supervised the research process, reviewed and edited the manuscript, and approved the final version. HMF: Provided expert insights into the methodology, assisted in refining the data collection and analysis process, contributed to the software application, and actively participated in writing the report, ensuring high-quality analysis and reporting throughout the research. MAA: Led the end-to-end data management process, implemented data preprocessing workflows, developed and optimized the machine-learning models, and contributed extensively to Python coding, model training, and hyperparameter tuning. AML: Contributed to the revision of the manuscript, supported the documentation of the analysis and results, and assisted with verification and validation of the analytical procedures to ensure accuracy and clarity.

Funding

This research has no specific grant from any funding agency, commercial, or not-for-profit sectors to report.

Availability of Data and Materials

The dataset used in this study was obtained from the Young Lives Study. Access to the data can be obtained either by completing the form available at Young Lives Data Access, selecting the dataset "Young Lives: Rounds 1-5 constructed files, 2002-2016" (and then extract only 2002 or baseline survey only where authors used in this study), or by creating a user account through the UK Data Service, subject to their terms and conditions. Additionally, the survey questionnaires for round 1 is available through the following link: Young Lives Round 1 Questionnaires.

Consent for Publication

As this study utilizes publicly available datasets, individual consent for publication was not required. The data used in this research were anonymized, ensuring that no personal identifiers were included, in accordance with ethical research practices.

Competing Interests

The authors declare that they have no competing interests regarding the publication of this manuscript.

Abbreviation

AB: AdaBoost, AUC: Area Under the Curve, GB: Gradient Boosting, KNN: K-Nearest Neighbors, LR: Logistic Regression, LSTM: Long Short-Term Memory, ML: Machine Learning, N: Normal, NB: Naive Bayes, RF: Random Forest, S: Stunted only, SMOTE: Synthetic Minority Oversampling Technique, SVM: Support Vector Machine, SW: Stunted and Wasted, WHO: World Health Organization, U: Underweight only, US: Underweight and Stunted, USW: Underweight, Stunted, and Wasted, UW: Underweight and Wasted, W: Wasted Only, YLCS: Young Lives Cohort Study.

References

[1] Fenta, H.M., et al., Determinants of stunting among under-five years children in Ethiopia from the 2016 Ethiopia demographic and Health Survey: Application of ordinal logistic regression model using complex sampling designs. Clinical Epidemiology and Global Health, 2020. 8(2): p. 404–413.

[2] Kasaye, H.K., et al., Poor nutrition for under-five children from poor households in Ethiopia: Evidence from 2016 Demographic and Health Survey. PloS one, 2019. 14(12): p. e0225996.

[3] Kassie, G.W. and D.L. Workie, Determinants of under-nutrition among children under five years of age in Ethiopia. BMC Public Health, 2020. 20: p. 1–11.

[4] Toma, T.M., et al., Factors associated with wasting and stunting among children aged 06–59 months in South Ari District, Southern Ethiopia: a community-based cross-sectional study. BMC nutrition, 2023. 9(1): p. 34.

[5] Woldeamanuel, B.T. and T.T. Tesfaye, Risk factors associated with Under-Five stunting, wasting, and underweight based on Ethiopian demographic health survey datasets in Tigray Region, Ethiopia. Journal of nutrition and metabolism, 2019. 2019(1): p. 6967170.

[6] Bitew, F.H., C.S. Sparks, and S.H. Nyarko, Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia. Public health nutrition, 2022. 25(2): p. 269–280.

[7] Shahriar, M.M., et al. A Deep Learning Approach to Predict Malnutrition Status of 0-59 Month's Older Children in Bangladesh. in 2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT). 2019. IEEE.

[8] Rahman, S.J., et al., Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach. Plos one, 2021. 16(6): p. e0253172.

[9] Talukder, A. and B. Ahammed, Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh. Nutrition, 2020. 78: p. 110861.

[10] Kawo, M.A., et al., PREDICTING UNDERNUTRITION RISK FACTORS USING MACHINE LEARNING TECHNIQUES IN NIGERIAN UNDER FIVE CHILDREN. 2024.

[11] Jain, S., et al. Efficient Machine Learning for Malnutrition Prediction among under-five children in India. in 2022 IEEE Delhi Section Conference (DELCON). 2022. IEEE.
[12] Khare, S., et al., Investigation of nutritional status of children based on machine learning techniques using Indian demographic and health survey data. Procedia computer science, 2017. 115: p. 338–349.
[13] Vasu, S.R., et al. Features Explaining Malnutrition in India: A Machine Learning Approach to Demographic and Health Survey Data. in Advanced Computing: 10th International Conference, IACC 2020, Panaji, Goa, India, December 5–6, 2020, Revised Selected Papers, Part I 10. 2021. Springer.
[14] Von Grafenstein, L., Mapping Real-Time Child Malnutrition in India Using Machine Learning. 2023.
[15] Anku, E.K. and H.O. Duah, Predicting and identifying factors associated with undernutrition among children under five years in Ghana using machine learning algorithms. Plos one, 2024. 19(2): p. e0296625.
[16] Fenta, H.M., T. Zewotir, and E.K. Muluneh, A machine learning classifier approach for identifying the determinants of under-five child undernutrition in Ethiopian administrative zones. BMC Medical Informatics and Decision Making, 2021. 21: p. 1–12.
[17] Kebede Kassaw, A., et al., The application of machine learning approaches to determine the predictors of anemia among under five children in Ethiopia. Scientific Reports, 2023. 13(1): p. 22919.
[18] Van, V.T.S., et al., Predicting undernutrition among elementary schoolchildren in the Philippines using machine learning algorithms. Nutrition, 2022. 96: p. 111571.
[19] Shen, H., H. Zhao, and Y. Jiang, Machine learning algorithms for predicting stunting among under-five children in Papua New Guinea. Children, 2023. 10(10): p. 1638.
[20] Chilyabanyama, O.N., et al., Performance of machine learning classifiers in classifying stunting among under-five children in Zambia. Children, 2022. 9(7): p. 1082.
[21] Khan, J.R., J.H. Tomal, and E. Raheem, Model and variable selection using machine learning methods with applications to childhood stunting in Bangladesh. Informatics for Health and Social Care, 2021. 46(4): p. 425–442.
[22] Ndagijimana, S., et al., Prediction of stunting among under-5 children in Rwanda using machine learning techniques. Journal of Preventive Medicine and Public Health, 2023. 56(1): p. 41.
[23] Begum, N., M.M. Rahman, and M. Omar Faruk, Machine learning prediction of nutritional status among pregnant women in Bangladesh: Evidence from Bangladesh demographic and health survey 2017–18. Plos one, 2024. 19(5): p. e0304389.
[24] Oxford, A Guide to Young Lives Research. , Y. Lives, Editor. 2017.
[25] Begashaw, G.B., T. Zewotir, and H.M. Fenta, Multistate Markov chain modeling for child undernutrition transitions in Ethiopia: a longitudinal data analysis, 2002–2016. BMC Medical Research Methodology, 2024. 24(1): p. 283.
[26] Genuer, R., J.-M. Poggi, and C. Tuleau-Malot, Variable selection using random forests. Pattern recognition letters, 2010. 31(14): p. 2225–2236.
[27] Janitza, S., G. Tutz, and A.-L. Boulesteix, Random forest for ordinal responses: prediction and variable selection. Computational Statistics & Data Analysis, 2016. 96: p. 57–73.
[28] Rodriguez-Galiano, V.F., et al., An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS journal of photogrammetry and remote sensing, 2012. 67: p. 93–104.
[29] Breiman, L., Random forests. Machine learning, 2001. 45: p. 5–32.
[30] Fenta, H.M., et al., Factors of acute respiratory infection among under-five children across sub-Saharan African countries using machine learning approaches. Scientific Reports, 2024. 14(1): p. 15801.
[31] Yuliansyah, H., et al., Comparison and analysis of classification algorithm performance for nutritional status data. International Journal of Computer Applications, 2020. 176(20): p. 14–20.
[32] Khudri, M.M., et al., Predicting nutritional status for women of childbearing age from their economic, health, and demographic features: A supervised machine learning approach. PloS one, 2023. 18(5): p. e0277738.
[33] Fatmawati, M., B.A. Herlambang, and N.Q. Nada, Random Forest Algorithm for Toddler Nutritional Status Classification Website. Journal of Applied Informatics and Computing, 2024. 8(2): p. 428–433.
[34] Selemani, B., D. Machuve, and N. Mduma, Machine learning model for predicting fetal nutritional status. Computational Ecology and Software, 2024. 14(1): p. 68.
[35] Turjo, E.A. and M.H. Rahman, Assessing risk factors for malnutrition among women in Bangladesh and forecasting malnutrition using machine learning approaches. BMC nutrition, 2024. 10(1): p. 22.
[36] Begashaw, G.B., T. Zewotir, and H.M. Fenta, A deep learning approach for classifying and predicting children's nutritional status in Ethiopia using LSTM-FC neural networks. BioData Mining, 2025. 18(1): p. 11.
[37] Begashaw, G.B., T. Zewotir, and H.M. Fenta, Dynamic Bayesian network modeling for longitudinal data on child undernutrition in Ethiopia (2002–2016). Frontiers in Public Health, 2024. 12: p. 1399094.