Machine Learning for Early Detection of Growth Faltering in Children Under Five

Mahdi Atallah, Yannik Tim Krone, Alexandre Jean Marcel Mourot

***Abstract*—Abstract—Malnutrition remains a critical global health issue, particularly in underserved regions with limited healthcare infrastructure. This study leverages machine learning to predict malnutrition status at future medical visits using a dataset of 16 million consultations from Burkina Faso, provided by Terre des Hommes. Key features include anthropometric, demographic, and clinical data.**

Among evaluated models, LightGBM achieved the best performance, particularly when using data from three visits and focusing on non-improving sequences. The final model, tested on the entire dataset, demonstrated robust accuracy and predictive capability, by achieving an accuracy of 96.14% and 96.68%, and an F1 score of 52.15% and 76.54% on multi-class and binary targets respectively. This highlights the potential of machine learning to enhance early detection and intervention for malnutrition.

I. Introduction

Malnutrition is a major global health issue, contributing to over 45% of deaths in children under five [1]. The burden is most severe in resource-limited regions, with Asia accounting for over half of stunted children and Africa over one-third [2].

In low-resource settings where malnutrition is most prevalent, healthcare systems often face substantial challenges. Consequently, there is a critical need for tools that can enable early detection and facilitate targeted, timely interventions.

Machine learning presents a promising solution by leveraging historical medical data to predict malnutrition outcomes at future visits. Such predictive models can help healthcare systems prioritize resources, identify children at the highest risk, and intervene proactively.

A. Problem Statement

Our objective is to explore the potential of machine learning models to predict the malnutrition status at a child's next medical visit using data from prior consultations.

To address this challenge, we utilize a dataset provided by *Terre des Hommes*, which contains approximately 16 million records of medical consultations for children under the age of five in Burkina Faso. The dataset includes the following features:

- **Child demographics**: e.g. age, gender, and region.
- **Anthropometric data**: e.g. weight, height, and mid-upper arm circumference (MUAC).
- **Derived indicators**: e.g. Z-scores for weight-for-height and height-for-age.
- **Malnutrition states**: severe malnutrition (with or without complications), moderate malnutrition, or no malnutrition.
- **Clinical conditions and feeding practices**: health-related observations and feeding behaviors.

B. Relevant Work

Machine learning has proven effective in predicting malnutrition outcomes in children under five in resource-limited settings, though existing studies rely on static datasets with single time-point predictions.

- Talukder and Ahammed [3] applied various machine learning algorithms to the Bangladesh Demographic and Health Survey (BDHS) data. Their study found Random Forest to be the most effective, achieving an accuracy of 68.51% and a recall of 94.66%.
- Similarly, Bitew *et al.* [4] used machine learning to predict undernutrition in Ethiopian children, emphasizing the importance of anthropometric and demographic features.

C. Contributions

Our main contributions include:

- Extending the previous approach to longitudinal data by evaluating machine learning models to predict malnutrition status at subsequent visits.
- Providing software to enable reproducibility and future analysis.

II. Methodology

A. Exploratory Data Analysis

1) Sample Distributions:

- **Malnutrition Status Distribution**: The distribution of malnutrition states in the dataset is highly imbalanced, with more severe conditions being rarer, as expected in medical data. The percentages in Table I correspond to all recorded consultations.

Table I
PERCENTAGE DISTRIBUTION OF MALNUTRITION STATES IN THE DATASET

Malnutrition State	Percentage (%)
Severe malnutrition (with complications)	0.41%
Severe malnutrition (without complications)	1.12%
Moderate malnutrition	3.28%
No malnutrition	94.92%

- **Frequency of Visits**: Half of the children attended only one consultation, while others had between 2 and 6 visits, with 6 being the maximum. The distribution is illustrated in Figure 1.

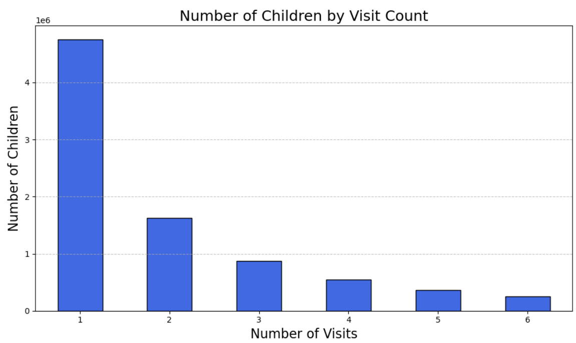

Figure 1. Distribution of the time gaps separating successive medical visits

- **Intervals Between Consecutive Visits**: The visit time distribution shows that children typically return for follow-up visits within a short timeframe after their initial consultation.
- **Transitions in Malnutrition Status**: We analyzed transitions between malnutrition states across consecutive visits to understand how children's nutritional statuses evolve over time. The transitions are visualized in Figure 3.
  To improve clarity, transitions where children remained in the *no malnutrition* state across visits were excluded, as these dominated the data and obscured other meaningful patterns.
  Key findings include:
  - The most frequent transitions show improvements, such as children moving from *moderate malnutrition* to *no malnutrition*.

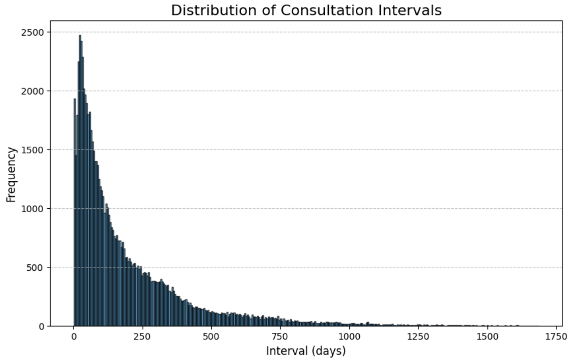

Figure 2. Distribution of the time gaps separating successive medical visits

- Recurrences of malnutrition, such as *moderate malnutrition* to *moderate malnutrition* and *severe malnutrition* to *severe malnutrition*, indicate cases where children fail to recover fully.
- Some transitions reflect worsening conditions, such as children moving from *no malnutrition* to *severe malnutrition*.

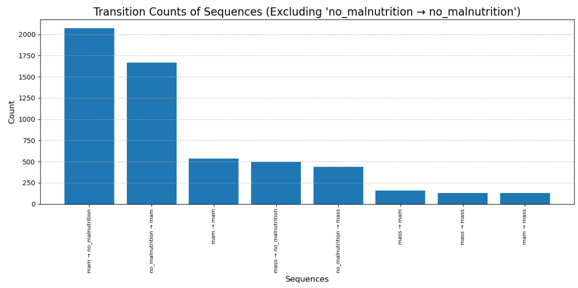

Figure 3. Transitions Between Malnutrition States Excluding Cases Remaining in the No Malnutrition State.

2) Missing values: Most features fall within the 0–10% range of missing (NaN) values, indicating a relatively low proportion of missing data across the dataset. The figure below provides a detailed breakdown of features grouped by their percentage of missing values.

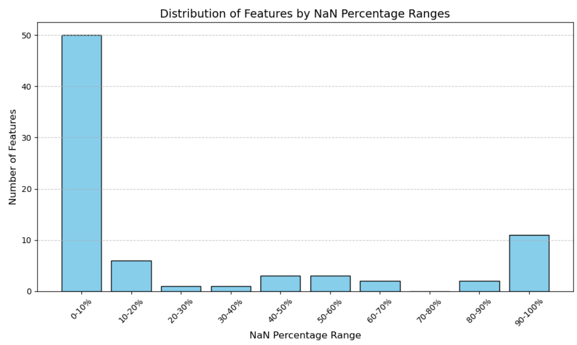

Figure 4. Proportion of Non-NaN Values per Feature.

3) Correlation Analysis:

- **Correlation Between Manually Entered and Calculated Features:** We analyzed the most correlated pairs of manually entered and calculated features, as visualized in Figure 5. The aim was to determine whether features not directly involved in malnutrition calculations, such as feeding practices, clinical observations, or demographic data, exhibit meaningful relationships with derived metrics like Z-scores or MUAC change.
  - Feeding-related features, such as nighttime breastfeeding, appetite during tests, and breastfeeding frequency (e.g., more than 12 times daily), show correlations with calculated Z-scores and MUAC changes.
  - Demographic and clinical indicators, such as the child's sex and the presence of edema, exhibit weaker correlations but still suggest possible links with malnutrition-related metrics.

The analysis revealed generally weak correlations, which may seem non-intuitive given the relevance of some features to malnutrition outcomes. However, these correlations are not negligible and may still highlight meaningful associations. Additionally, it is possible that these relationships exhibit non-linear dependencies, which are not captured by the correlation metric.

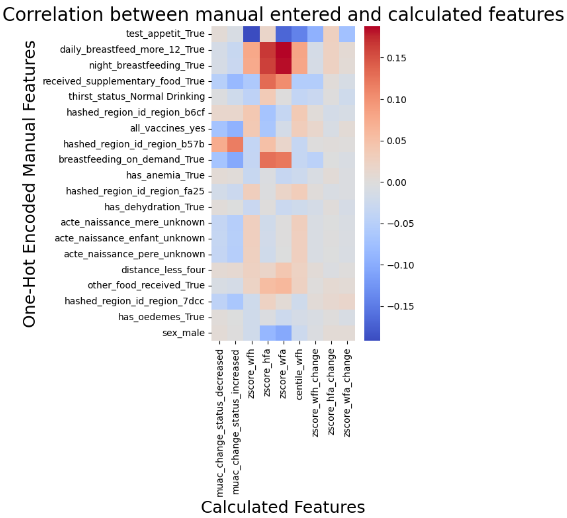

Figure 5. Top 20 Correlations Between Manually Entered and Calculated Features.

- **Correlation Between Season, Region, and Malnutrition Status** We analyzed the relationship between malnutrition and both seasonal and regional features, summarized in Figure 6:
  - The **hot season** showed a very weak positive correlation with malnutrition (coefficient ≈ 0.03), contradicting the initial intuition that reduced rainfall might increase malnutrition risk.
  - **Regional correlations** were similarly weak, ranging from -0.02 to 0.05. This may be due to limited coverage, as the dataset focuses on regions where *Terre des Hommes* operates, which could have comparable malnutrition levels.

Overall, the low correlation values suggest that seasonality and regional features are not strong predictors of malnutrition when considered individually.

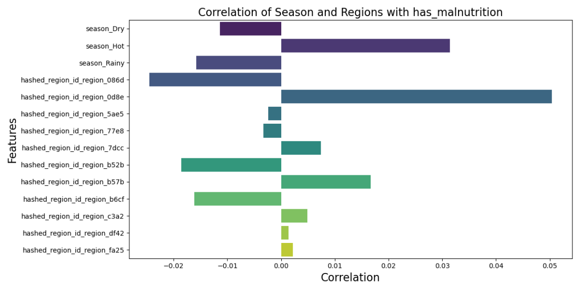

Figure 6. Correlation Between Season, Region, and Malnutrition.

B. Feature processing

1) Feature Selection: We removed redundant features and those with excessive missing values to improve model efficiency and performance.

2) Handling Missing Values: To address missing (NaN) values, we applied the following strategies:

- **Categorical Values:** Replaced missing values with the mode, i.e., the most frequently occurring value in the non-missing samples.
- **Numerical Values:** For the few instances where numerical features contained missing values, we replaced them with the mean of the non-missing samples.

3) Class balancing: To address class imbalance in the dataset, we employed Synthetic Minority Over-sampling Technique (SMOTE) which was used to generate synthetic samples for underrepresented classes to achieve a more balanced distribution.

4) Onehot encoding: Categorical variables in the dataset were encoded using one-hot encoding. This approach converts each category into a binary vector, ensuring the model can interpret these features without assuming any ordinal relationship between categories.

5) Normalization: In this case, normalization is unnecessary as the Z-scores in the dataset are already calculated based on the mean and variance of the measured metrics for the population.

6) Incorporating Temporal Dimensions: The dataset consists of individual medical visits, each with a date, but does not explicitly define inputs and targets. To model trends over time, we constructed a temporal framework by taking data from previous visits as inputs and the results of the latest visit as the target.

For the model to accurately predict malnutrition outcomes, it must understand how far into the future it is forecasting. To address this, we explored two strategies:

- **Fixed Interval Analysis:** This approach uses only visits separated by a fixed time interval, treating the latest visit as the target. However, the variability in the dataset, as shown in Figure 2, makes it difficult to choose a fixed interval that adequately captures the range of visit patterns.
- **Time-Aware Modeling:** This method includes the *time to the next visit* as an additional feature, allowing the model to account for temporal variability. While more flexible, this approach assumes knowledge of future visit timing, which may not always be feasible in practice.

To balance these challenges, we chose to predict malnutrition status within the *next four months or less*. This approach excludes extreme outliers, such as visits occurring years apart, while utilizing the majority of the dataset.

C. Machine learning methods

In this study, we employ three machine learning models to predict malnutrition outcomes. **Logistic Regression** serves as a baseline model due to its simplicity and interpretability. Additionally, we utilize **Random Forest** and **LGBMClassifier**, which were selected based on their demonstrated performance in related studies [3][4].

D. Evaluation Metrics

To comprehensively evaluate model performance, we utilized both *accuracy* and *F1 score* as metrics. The F1 score was chosen for its ability to balance precision and recall, which is critical in this context. High precision ensures that identified malnutrition cases are genuinely at risk, optimizing the allocation of limited intervention resources. Meanwhile, high recall reduces the likelihood of overlooking vulnerable children who require urgent attention.

For multi-class classification, we calculated the macro-average F1 score to equally account for performance across all classes.

E. Comparing Simplified and Multi-Class Classification Approaches

We explore the impact of simplifying the classification task by compare the results obtained using *fused malnutrition classes* against the performance of the original multi-class setup. The intuition behind this approach is that simplifying the classes may improve performance, but at the cost of reduced descriptive power.

F. Evaluation of the Visit Frequency Impact

To evaluate the influence of visit frequency on model performance, we begin by establishing a baseline predictive accuracy. We then analyze how the number of visits affects the model's ability to accurately predict malnutrition status at the next visit, providing insights into the importance of visit frequency in early detection.

G. Analysis of Non-Improving Sequences

Finally, we investigate the performance of models trained on two distinct subsets of the data:

- **All sequences:** Including both improving and worsening malnutrition states.
- **Non-improving sequences:** Focusing exclusively on cases where the child's malnutrition status either worsens or remains unchanged.

This analysis is particularly significant, as an important objective of our study is to identify at-risk individuals who may otherwise go untreated. By focusing on non-improving sequences, we aim to refine the models to prioritize children who are most vulnerable and in need of intervention.

III. Results and discussion

Most experiments were conducted on sampled subsets of the dataset due to limited computational resources, which restricted our ability to run all configurations on the full dataset. However, our best-performing model was ultimately evaluated on the entire dataset.

A. Comparison of Machine Learning Methods

We evaluated the models discussed in II-C using data corresponding to 3 visits on sample of the dataset. The results, summarized in Table II shows that the LGBM classifier achieved the highest accuracy and F1 score, slightly outperforming the Random Forest model. Both LGBM and Random Forest significantly outperformed Logistic Regression.

B. Comparison of Simplified vs. Multi-Class Classification Approaches

We compared the accuracy and F1 scores of all models for both simplified and multi-class classification tasks, as detailed in Table II. The evaluation was performed on data corresponding to 3 visits. The simplified classification performed better due to the reduced complexity, resulting in higher predictive performance. However, this improvement comes at the cost of descriptive power, as the simplified approach provides less granular information. This trade-off between performance and descriptive power raises an important avenue for future exploration.

C. Comparing the Visit Frequency Impact

We evaluated the performance of the most promising model, the LGBM classifier, using varying numbers of past visits as input. The results indicate that the model's predictive accuracy improves with more historical data about the child. However, for practicality, we limited our analysis to children with up to four visits. This decision was based on the limited number of children with five or six visits in the dataset sample, which would affect

Table II
F1 SCORE AND ACCURACY FOR BINARY AND MULTI-CLASS CLASSIFICATION

Model	Task	F1 Score (%)	Accuracy (%)
Logistic Regression	Binary	36.05%	46.62%
	Multi-Class	16.6%	44.1%
Random Forest	Binary	73.34%	96.34%
	Multi-Class	50.22%	95.91%
LGBM Classifier	Binary	**73.44%**	**96.34%**
	Multi-Class	**50.73%**	**95.96%**

the reliability of the evaluation metrics. The results are illustrated in Figure 1

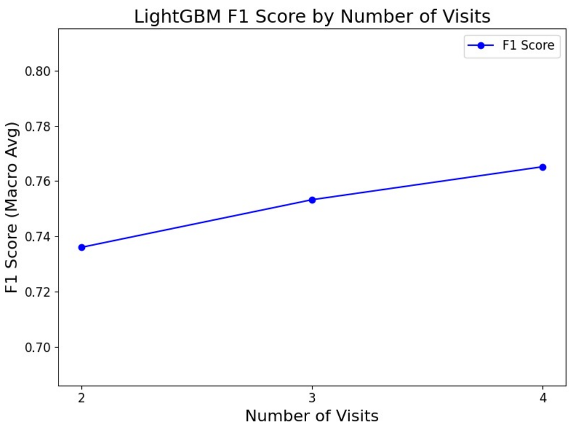

Figure 7.

D. Comparing Performance Across All Sequences and Non-Improving Sequences

The results indicate that the F1 score improves significantly when focusing on non-improving sequences. This outcome aligns with expectations, as training on non-improving sequences helps the model better capture cases of malnutrition persistence or deterioration.

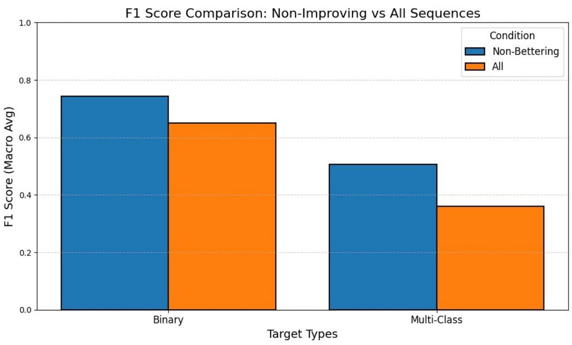

Figure 8.

E. Final Model Selection

The results include performance metrics for all combinations of the following parameters:

- **Number of visits**: 2 or 3 (i.e., 1 or 2 prior visits with the latest visit as the target).
- **Worsening**: True or False (non-improving sequences only or all sequences).
- **Target type**: Binary or Multi-Class.

This yields a total of 8 configurations. The complete results for these configurations are summarized in Table III.

To validate these findings, we reran the best-performing LightGBM model (3 visit) on the entire dataset. The final results are summarized in Table IV.

Table III
ACCURACY AND MACRO-AVERAGE SCORES FOR LIGHTGBM ACROSS TASKS AND SETTINGS

Visits	Target Type	Worsening	Accuracy (%)	Macro F1 (%)
2	Binary	True	96.27	73.60
2	Binary	False	94.23	68.04
2	Multi-Class	True	95.95	48.90
2	Multi-Class	False	93.61	39.31
3	Binary	True	96.57	75.32
3	Binary	False	94.17	65.24
3	Multi-Class	True	96.17	49.63
3	Multi-Class	False	93.72	36.41

Table IV
PERFORMANCE OF THE BEST MODEL ON THE ENTIRE DATASET

Target Type	Accuracy (%)	Recall (%)	Precision (%)	F1 (%)
Binary	96.68	68.88	98.22	76.54
Multi-Class	96.14	43.96	74.95	52.15

F. Limitations and Future Work

Despite the promising results, the recall of even the best-performing model remains suboptimal, highlighting a need for further improvement. Due to limited computational resources, we were only able to evaluate the best model on the full dataset, while other configurations were tested on subsets for experimentation. However, the provided codebase is designed to support comprehensive evaluations and can be utilized effectively with sufficient computational power.

Future work could explore the use of more advanced models, such as LSTMs or Transformers, to better capture temporal dependencies and potentially improve predictive performance in this context.

G. Ethics and Risks

No significant ethical risks were identified in this study, as the methods and outcomes are designed to enhance care delivery without causing harm or bias.

IV. CONCLUSION

While the results are promising, the recall of even the best-performing models remains an area for improvement. Additionally, due to computational constraints, some analyses were limited to subsets of the dataset. However, the provided codebase offers scalability for future work with greater computational resources.

Future research could investigate the application of more advanced models, such as LSTMs or Transformers, to better account for temporal dependencies. Furthermore, refining class balancing techniques and exploring additional feature engineering strategies could enhance the performance and utility of the predictive models.

Ultimately, this work lays a foundation for leveraging machine learning in addressing malnutrition and emphasizes the importance of data-driven approaches in improving early detection and targeted interventions in resource-limited settings.

V. ACKNOWLEDGEMENTS

We express our gratitude to *Terre des Hommes* for providing the dataset and for their valuable support throughout this project.

REFERENCES

[1] R. E. Black, C. G. Victora, S. P. Walker *et al.*, "Maternal and child undernutrition and overweight in low-income and middle-income countries," *The Lancet*, vol. 382, no. 9890, pp. 427–451, 2013.

[2] UNICEF, "Malnutrition: Regional statistics and trends," 2023. [Online]. Available: https://data.unicef.org/topic/nutrition/malnutrition/

[3] A. Talukder and B. Ahammed, “Machine learning algorithms for predicting malnutrition among under-five children in bangladesh,” *Nutrition*, vol. 78, p. 110861, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0899900720301441

[4] F. H. Bitew, C. S. Sparks, and S. H. Nyarko, “Machine learning algorithms for predicting undernutrition among under-five children in ethiopia,” *Public Health Nutrition*, vol. 25, no. 2, p. 269–280, 2022.