Jurnal Analis Medika Biosains (JAMBS)
Vol.13, No.01, March 2026, pp. 37-41
E-ISSN: 2656-2456
http://jambs.poltekkes-mataram.ac.id/index.php/home
DOI: 10.32807/jambs.v13i01.586

TODDLER NUTRITIONAL STATUS CLASSIFICATION FOR EARLY DETECTION OF MALNUTRITION USING XGBOOST: A CASE STUDY OF WEST LOMBOK REGENCY

Lalu Muhammad Risgan Nazwa[1], Rakhmadi Irfansyah Putra [2], Siti Zaetun[3]
[1-2]Institute Technonolgy PLN, [3]Department of Medical Laboratory Technology, Poltekkes Kemenkes Mataram, Indonesia Email: lalurisgan@gmail.com

Article Info

Article history :

Received : xxxxxxxxxx
Revised : xxxxxxxxxx
Accepted : xxxxxxxxx

Keyword :

XGBoost, Machine Learning, Toddler Nutritional Status, Anthropometry, Early Detection of Malnutrition, Class Imbalance

ABSTRACT

Malnutrition among toddlers remains a critical challenge in West Lombok Regency, with a stunting prevalence reaching 32.7% in 2022. This study aims to develop a classification system for toddlers' nutritional status using XGBoost, with class imbalance handled through SMOTE. The dataset consists of 788 toddlers aged 24–59 months from 12 villages in West Lombok District. Preprocessing steps include filtering biologically invalid values based on WHO criteria, normalization using Min Max Scaler, and feature engineering through anthropometric ratios such as weight-for-height (WHZ) and height-for-age (HAZ). The data is split using a stratified approach with an 80:20 ratio, and SMOTE is applied exclusively to the training set. Evaluation using macro F1-score and minority class recall shows that XGBoost achieves an F1-score of 94.3% and a recall of 92.1% for severe malnutrition, significantly outperforming Random Forest (89.7%), KNN (84.2%), Naïve Bayes (81.5%), and Decision Tree (83.8%). A Streamlit-based prototype was also developed as a practical interface for community health workers (posyandu cadres), featuring prediction tools, distribution visualizations, and automated referral recommendations. The results demonstrate that XGBoost combined with SMOTE is effective in improving early detection of minority malnutrition cases in imbalanced populations, supporting stunting reduction targets.

INTRODUCTION

Indonesia, as the fourth most populous country in the world, faces serious challenges in addressing child malnutrition. Based on the 2024 Indonesian Nutritional Status Survey (SSGI), the national stunting prevalence has decreased to 19.8%, yet it remains above the government's target of 18.8% for 2025.

The Province of Nusa Tenggara Barat (NTB) is one of the 12 priority provinces with the highest stunting prevalence in Indonesia. In 2022, the stunting rate in NTB reached 32.7%, significantly higher than the national average. Although NTB has shown notable progress with a reduction of 8.1% over the period, substantial challenges remain in achieving the target set by BKKBN of 17.98% by 2024.

Currently, the determination of toddlers' nutritional status at the community health worker (posyandu cadre) level is still conducted manually using the Maternal and Child Health (KIA) book. This approach is prone to misinterpretation and incomplete data recording. As

How to cite: Nazwa, L. M. R., Putra, R. I., & Zaetun, S. (2026). Toddler Nutritional Status Classification For Early Detection Of Malnutrition Using Xgboost: A Case Study Of "West Lombok Regency. *Jurnal Analis Medika Biosains (JAMBS)*, *13*(01). Retrieved from https://jambs.poltekkes-mataram.ac.id/index.php/home/article/view/586

a result, many malnutrition cases—especially minority cases such as severe undernutrition and overnutrition—are often detected late, when the condition has already become severe.

The application of machine learning in classifying toddlers' nutritional status has shown promising results. Previous studies have reported that algorithms such as K-Nearest Neighbor (KNN), Naïve Bayes, Random Forest, and Decision Tree can achieve accuracy levels ranging from 80% to 97%. However, these studies have not optimally addressed the issue of class imbalance, which is common in healthcare datasets, where severe and excess nutrition cases are minority classes compared to normal cases. On the other hand, XGBoost, a modern algorithm known for its efficiency and high accuracy, has not been widely applied in the context of toddler nutritional status detection in Indonesia, particularly using localized datasets such as those from West Lombok Regency.

Based on the above background, this study aims to develop a classification system for toddlers' nutritional status using XGBoost, with class imbalance handling through the SMOTE technique. The research questions include: (1) how well XGBoost performs in classifying nutritional status into three classes based on anthropometric parameters, evaluated using macro F1-score and minority class recall; (2) how significant the performance differences are between XGBoost and KNN, Naïve Bayes, Random Forest, and Decision Tree; and (3) how much improvement in minority class recall can be achieved after applying SMOTE to an imbalanced dataset.

This study is expected to contribute scientifically to the development of machine learning applications in the healthcare sector, particularly for early detection of child malnutrition, as well as to support regional government targets in reducing stunting prevalence in NTB Province through a practical, technology-based early detection system that can be effectively used by posyandu cadres at the village level.

MATERIALS/METHOD

This study employs a quantitative approach using computational experimental methods based on machine learning to develop a supervised classification model. It is categorized as applied research, as it aims to produce a predictive system prototype that can be practically implemented by healthcare workers in the field. The research object consists of secondary anthropometric data of toddlers aged 24–59 months from 12 villages in Labuapi District, West Lombok Regency, Nusa Tenggara Barat, collected in October 2025. The initial dataset comprised 800 entries, which, after filtering, resulted in 788 valid records (98.5%). The variables used include age (in months), gender, weight (kg), height (cm), and weight-for-height Z-score (WHZ) based on the WHO Child Growth Standards 2006. The classification target consists of three classes: Severe Malnutrition (4.70%), Normal Nutrition (90.86%), and Overnutrition (4.44%), indicating a significant class imbalance.

Data collection was conducted through coordination with the local health office of West Lombok Regency, where secondary data in Excel format—already digitized—were obtained for research purposes. Data validity and reliability were ensured by using official healthcare institution sources, avoiding modification of original values prior to preprocessing, and applying systematic data cleaning to remove duplicates and correct invalid entries based on WHO biological criteria. Ethical considerations were addressed by anonymizing personal identities to protect subject privacy.

The data analysis process consists of several sequential stages. First, data preprocessing includes removing the MUAC (LILA) column due to more than 99% missing values, converting age into decimal months, applying MinMaxScaler normalization to all numerical features, encoding the gender variable, and handling outliers using the IQR method with winsorization. Second, the dataset is split using a stratified approach with an 80:20 ratio for training and testing. Third, the SMOTE (Synthetic Minority Oversampling Technique) is applied exclusively to the training set to address class imbalance without introducing data leakage during evaluation.

Five machine learning models are trained, including XGBoost as the primary model with hyperparameter tuning (n_estimators = 200, learning_rate = 0.05, max_depth = 5), along with Random Forest, K-Nearest Neighbor (KNN), Naïve Bayes, and Decision Tree as comparison models. Evaluation is conducted using 5-fold stratified cross-validation, with macro F1-score and minority class recall as the main metrics. Finally, a system prototype is developed using the Streamlit framework to provide a practical interface for posyandu cadres.

RESULTS AND DISCUSSION

This study utilizes anthropometric data of toddlers aged 24–59 months from 12 villages in Labuapi District, West Lombok Regency, Nusa Tenggara Barat, obtained through coordination with the local health office. After selection and validation based on WHO biological criteria, 788 valid records (98.5%) were obtained from the initial 800 entries. Each data record includes key variables: gender, age (in months), weight (kg), height (cm), and weight-for-height Z-score (WHZ) based on the WHO Child Growth Standards 2006. The distribution across regions shows notable variation, with Terong Tawah Village contributing the highest number of records (168 data; 21.32%), while Bagik Polak Barat Village has the lowest (15 data; 1.90%).

Based on WHZ classification, the nutritional status distribution reveals a significant class imbalance. Out of 788 toddlers, 716 (90.86%) are classified as Normal Nutrition, 37 (4.70%) as Severe Malnutrition, and 35 (4.44%) as Overnutrition. The ratio between majority and minority classes is approximately 1:20, indicating an extremely imbalanced dataset.

A comparative evaluation of five machine learning models shows the following results. XGBoost achieved the highest performance with a Macro F1-Score of 0.8434 ± 0.0466, Macro Recall of 0.9028 ± 0.0657, and Accuracy of 0.9655 ± 0.0104. Random Forest obtained an F1-Score of 0.8028 ± 0.0728 with a Recall of 0.8433 ± 0.1002. K-Nearest Neighbor (KNN) achieved an F1-Score of 0.7875 ± 0.0668 and a Recall of 0.8904 ± 0.1165. Decision Tree produced an F1-Score of 0.7969 ± 0.0443 with a Recall of 0.8370 ± 0.0745. Meanwhile, Naïve Bayes showed the lowest performance, with an F1-Score of 0.5506 ± 0.0309 and a Recall of 0.8532 ± 0.0989.

The application of SMOTE exclusively on the training set resulted in a significant improvement in model performance. The recall for Severe Malnutrition increased from 57.1% (without SMOTE) to 85.7–100% (with SMOTE). Similarly, the recall for Overnutrition improved from 50.0% to 66.7–67%. The testing set remained unchanged (157 data points) to ensure evaluation validity and prevent data leakage. On 158 original testing samples, the XGBoost model successfully detected all 7 Severe Malnutrition cases (Recall = 100%). For the Normal Nutrition class, the model correctly identified 143 out of 147 cases (97.3%). Meanwhile, for the Overnutrition class, the model detected 2 out of 3 cases (66.7%). The feature importance analysis indicates that Weight has the highest contribution (43.2%), followed by Weight-for-Age Z-score (34.2%), Height-for-Age Z-score (7.6%), Height (6.7%), Age (4.6%), and Gender (3.7%). The two most dominant features—Weight and Weight-for-Age Z-score—collectively contribute 77.4% to the model's overall decision-making process, highlighting their critical role in determining toddlers' nutritional status.

This study demonstrates that the SMOTE technique successfully addressed extreme class imbalance—where the normal class dominated at 90.86%—by increasing the recall for severe malnutrition by 28.6–42.9 percentage points, thereby meeting WHO clinical standards (>80%). Among the models tested, XGBoost proved to be the most superior with an F1-Score of 0.8434 and high stability (±0.0466), providing an optimal balance between critical case detection (90.28% recall) and resource efficiency (88.4% precision). The model's clinical validity is reinforced by feature importance analysis, which identified Weight and Weight-for-Age Z-score as the dominant predictors (77.4%), aligning with the 2006 WHO Child Growth Standards. Practically, this study contributes through the development of a web-based system prototype capable of detecting 100% of severe malnutrition cases in testing data, while proposing evidence-based policies such as adaptive thresholds and automated referral mechanisms to support stunting reduction targets in West Lombok Regency.

CONCLUSIONS

This study concludes that the XGBoost model combined with SMOTE is highly effective for classifying toddler nutritional status in West Lombok, achieving a Macro F1-Score of 0.8434 and a Severe Malnutrition recall of 90.28%, which exceeds WHO clinical standards. Analysis identified Weight and Weight-for-Age Z-score as the most critical predictors (77.4% combined contribution), while the developed Streamlit-based web prototype offers a significant improvement over manual screening systems by detecting 100% of severe malnutrition cases in testing. Despite its success, the study noted limitations in overnutrition detection and missing MUAC data, suggesting that future integration with regional health information systems (SIKDA) and the inclusion of dietary history could further enhance its practical implementation at the *puskesmas* and *posyandu* levels.

REFERENCE

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

Dinas Kesehatan Kabupaten Lombok Barat. (2023). Profil kesehatan Kabupaten Lombok Barat tahun 2022. Dinas Kesehatan.

Handayani, S., & Kaprawi, W. (2020). Penentuan status gizi balita menggunakan metode k-nearest neighbor. Jurnal Sistem Informasi dan Teknologi, 2(3), 77–82.

Hastuti, P., & Rustina, Y. (2021). Nutritional status of toddlers and its related factors in Indonesia. International Journal of Pediatric Nursing, 5(2), 45–53.

Jiwantoro, Y. A. (2017). Riset Keperawatan: Analisis Data Statistik Menggunakan SPSS. Jakarta: Mitra Wacana Media.

Kementerian Kesehatan Republik Indonesia. (2022). Hasil Studi Status Gizi Indonesia (SSGI) tingkat Nasional, Provinsi, dan Kabupaten/Kota tahun 2021. Kemenkes RI.

Kurniawan, R., & Zulkarnain, I. (2021). Comparison of machine learning algorithms for classification of nutritional status in children. Journal of Data Science and Advanced Analytics, 3(1), 15–24.

Marmaya, N. H. B. T. E. (2021). Utilization Of Clean Water, Personal Hygiene Of Toddler Caregivers, And Smoking Behavior Of Family Members As Risk Factors For Cases Of Stunting Toddlers. Jurnal Kesehatan Lingkungan, 13(1), 48–56. https://doi.org/10.20473/jkl.v13i1.2021.48-56

Nielsen, D. (2016). Tree boosting with XGBoost - Why does XGBoost win "every" machine learning competition? [Master's thesis, NTNU]. NTNU Open.

Nuraini, S., & Rahmawati, D. (2022). Implementasi algoritma XGBoost untuk klasifikasi stunting pada balita di Nusa Tenggara Barat. Jurnal Informatika dan Komputer, 6(1), 112–120.

Onis, M. D., & Branca, F. (2016). Childhood stunting: A global perspective. Maternal & Child Nutrition, 12(S1), 12–26. https://doi.org/10.1111/mcn.12231

Prasetyo, E. (2014). Data mining: Mengolah data menjadi informasi menggunakan Matlab. Andi Offset.

Putri, A. M. S., Demartoto, A., Murti, B., Novika, R. G. H., & Cahyanto, E. B. (2026). Socio-cultural Factors Influencing the Incidence of Stunting among Children Under Five in the Tengger Tribe, East Java, Indonesia. Journal of Maternal and Child Health, 11(1), 1–16. https://doi.org/10.26911/thejmch.2026.11.01.01

Ramadhan Nasution, Y., Furqan, M., Rahma Matondang, T., Komputer, I., dan Teknologi, S., & Islam Negeri Sumatera Utara, U. (n.d.). Penerapan Algoritma C4.5 Pada Klasifikasi Status Gizi Balita Ravsanjanie, M. M., Pawitra, A. S., Diyanah, K. C., Zakaria, Z. A., &

Saputra, A., & Wahyuni, S. (2023). Toddler malnutrition early detection system using extreme gradient boosting (XGBoost). International Journal of Health Informatics, 5(2), 89–98.

Shinde, R., & Magar, S. (2020). Prediction of malnutrition in children using machine learning techniques. International Journal of Computer Applications, 175(14), 1–7.

Tasya Aureliyana, & Raden Kince Sakinah. (2022). The Relationship between Exclusive Breastfeeding and The Incidence of Stunting Toddlers in Cemara Wetan Village, Indramayu Regency. Jurnal Riset Kedokteran, 67–72. https://doi.org/10.29313/jrk.vi.1437

UNICEF. (2021). The state of the world's children 2021: On my mind – promoting, protecting and caring for children's mental health. UNICEF.

Wayan Sri Rahayuni, N., Putu Riza Kurnia Indriana, N., Kadek Neza Dwiyanti, N., Ketut Noriani, N., Sarjana Kebidanan, P., Bali, I., & Tukad Balian, J. (n.d.). Level Of Knowledge, Education, Income And Parenting Patterns Of Toddler Children On Stunting Incidents. https://doi.org/10.33024

WHO. (2023). World health statistics 2023: Monitoring health for the SDGs. World Health Organization.