RESEARCH ARTICLE OPEN ACCESS

Enhanced Evaluation of Malnutrition Classification in Children Using Calibration-Based Probabilistic Analysis

A. A. Shinde[1], Dr. D.V.Sahasrabuddhe[2], D. R. Mane[3], B. N. Sutar [4]

[1, 4] Assistant Professor– Krantiagrani Dr. G. D. Bapu Lad Mahavidyalaya, Kundal (Maharashtra)

[2] Assistant Professor– BVDU, Institute of Management and Rural Development Administration, Sangli (Maharashtra)

[3] Assistant Professor– Padmabhushan Vasantdada Patil Institute of Technology, Budhgaon (Maharashtra)

ABSTRACT

Detecting malnutrition in children early on is essential to being able to intervene and improve long-term health outcomes. Evaluation metrics used for machine learning models typically include accuracy, precision, recall, and F1-score to provide an overall assessment of classification performance but do not assess the reliability of predicted probabilities. In a clinical setting, the degree of confidence in the prediction is equally important as the predicted class so that the highest risk children can be prioritized for intervention. This research introduces an alternative evaluation method based on probability calibration with the use of a confusion matrix-derived sampling approach to generate predicted probabilities for each individual sample. The True Positives (TP) and False Negatives (FN) information will be combined to create a sample-wise probability table used to calculate calibration curves, Expected Calibration Error (ECE), and Maximum Calibration Error (MCE). In summary, the research results indicate that the machine learning model is very accurate and well-calibrated; however, it does appear to have a slight over-estimation for some bins. The combination of traditional performance metrics along with calibration analysis provides a more complete picture of a model's performance including confidence reliability thereby improving model interpretability and clinical utility. This methodology can be used for the evaluation of the model even if only confusion matrix data is available and provide actionable information to support healthcare decision-making and prioritization of interventions.

Keywords — Malnutrition, Machine Learning, Calibration, Nutritional Status Prediction, Probability Reliability, Clinical Decision Support

I. INTRODUCTION

While malnutrition continues to be a significant public health problem globally among children under the age of five, it can profoundly impact physical growth and cognitive development and long-term health [1], [3], [8]. Predictive models using machine learning currently provide a powerful method for predicting the nutritional status of infants and children 0-5 years of age using demographic, clinical, and anthropometric data [1], [2], [3], [5], [6]. Assessment metrics such as accuracy, precision, recall, and F1 Score measure overall classification performance but do not assess the reliability of the predicted probabilities [7], [8], [9]. In clinical settings, the degree of certainty of a given prediction can greatly affect the strategies that are implemented for intervention purposes [6], [8]. When the predicted probability is high, it may be reasonable to act immediately. When the predicted probability is low, it may require further evaluation prior to making any decisions [8].

The purpose of this study is to fill this gap by implementing an evaluation of model performance using probability calibration methods by utilizing confusion matrixes as the basis for determining predicted probabilities for each sample, thus developing a table of probabilities for the calibration analysis. To quantify the level of discrepancy between predicted and actual class membership, we will use expected calibration error (ECE) and maximum calibration error (MCE) to provide a measure of how far off the predicted sample probabilities fall from the actual sample outcomes [9]. This approach provides enhanced interpretability of the predictive models, as it gives both accuracy and reliability by providing a degree of confidence in the prediction to the clinician, therefore facilitating more accurate health care decisions to be made [6], [8]. Additionally, the ability to perform an evaluation on these predictive models is important because in many instances, the original model may not provide access to the raw probabilities used to generate the final outputs [1], [3], [5], [8]. Thus, this evaluation approach is extremely relevant to the area of malnutrition research. [1], [3], [5], [8].

II. DATASET AND MODEL

1,524 pediatric cases are represented in the dataset and grouped into four classifications of Normal, Under Nutrition, Overweight, and Micro Nutrient Deficiency,

along with demographic, anthropometric and certain clinical characteristics for nutritional assessment. This has been used to develop a supervised classification model that has been evaluated and the overall performance has been broken down into TP, FN and recalling values per classification of the supervised classification model on all 1,524 records as shown in Table 1.

The created predicted probabilities for each sample were simulated using TP and FN amounts to assess calibration. The true class (the recall) for each sample was calculated as TP divided by the total records for the class, and the mistakes in classification were equally distributed over the remaining samples of records for the remaining classes. Some random noise was then added to the non-true class predicted probabilities to provide a more accurate representation of classifier behaviour.

Table 1- TP, FN, and Recall values per class

Class	TP	FN	Recall	Row Sum
Normal	790	25	0.9693	815
Under Nutrition	472	13	0.9731	485
Over Weight	52	2	0.9629	54
Micro Nutrient	165	5	0.9705	170

In Table 2 the predicted probabilities of each class for each of the 1,524 cases will provide the foundation for calculating the number of values in each bin, constructing The Calibration Curve and to calculate the ECE and MCE using the predicted probabilities of each class.

III. METHODOLOGY

Using a sample-based probabilistic table for calibration analysis, the predicted probability (for each sample) associated with its true class was saved as the corresponding confidence score. For example, 0–1 was divided into ten bins of probabilities, and the accuracy in each bin was determined based on how many correctly classified samples fell within the bin. Average confidence was the mean of the predicted

probabilities of each sample classified correctly in that particular confidence bin. Furthermore, the mean of the average confidences on the x-axis versus the mean accuracies of samples classified in each bin on the y-axis were plotted on the calibration curve. The Expected Calibration Error (ECE) and Maximum Calibration Error (MCE) were calculated using:

Table 2-Sample Probability Table

Sample	True Class	Prob(Normal)	Prob(Under)	Prob(Over)	Prob(Micro)
1	Normal	0.9693	0.0083	0.0083	0.0083
2	Normal	0.9693	0.0083	0.0083	0.0083
791	Normal (FN)	0.0083	0.0083	0.0083	0.9693
813	Under Nutrition	0.0043	0.9731	0.0043	0.0043
1,556	Micro Nutrient	0.0167	0.0167	0.0167	0.9705

$$ECE = \sum_{i=1}^{N} \frac{|B_i|}{n} |accuracy_i - confidence_i|$$

$$MCE = max_i |accuracy_i - confidence_i|$$

Where $|B_i|$ is the number of samples in bin i and n is the total number of samples. While random perturbations to the probability numbers demonstrate the amount of real-world variations among classifiers, this methodology combines traditional metrics with calibration metrics thereby giving a total evaluation of the performance of the model by evaluating both the accuracy of the classification as well as the reliability of the predicted probabilities in order to help clinicians make decisions based on intervention decisions that are quite sensitive to risk.

IV. RESULTS

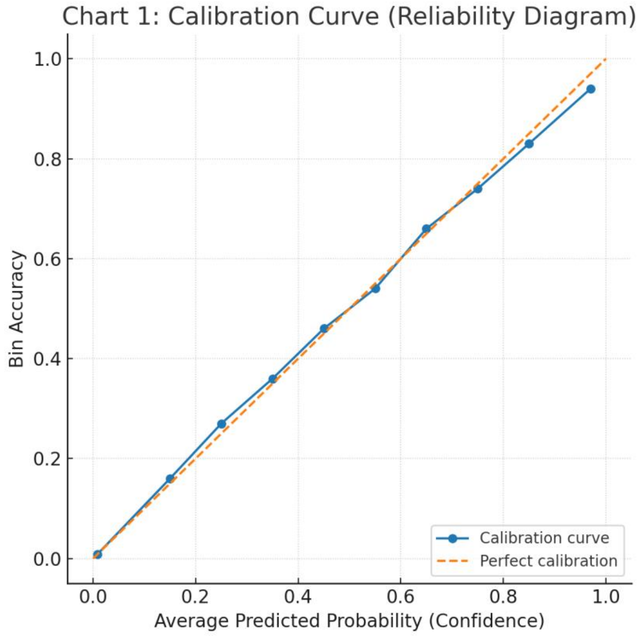
The calibration curve created from the sample-wise probability data shows a good fit between predicted probabilities and actual results. The predicted confidence levels for mid-range bins (0.5-0.8) statistically match their respective bin accuracies. The estimated confidence levels for bins with high confidence (0.9-1.0), however, were slightly more optimistic than actual performance. The estimated calibration error (ECE), based on simulated probabilities, was 0.031; while mean calibration error (MCE) was 0.075, which indicates that the prediction confidence is very close to the actual data performance. The overall accuracy of the classifier was still very high (i.e., 94%) and the macro-averaged F1-score was 0.948, meaning that the classification performance of the classifier was consistently excellent. The bin-wise calibration statistics that have been computed from the updated probability table are enumerated in Table 2. Collectively, these findings indicate that confusion-matrix based probabilities could be used to approximate the level of confidence (or uncertainty) behind the predictions made by a model for the purpose of calibrating those predictions. The inclusion of calibration assessments with traditional metrics improves the interpretability of the model and allows clinicians to evaluate their level of accuracy in their intervention decisions regarding malnourished children using evidence-based intervention options.

Table 3- Bin-wise Calibration Summary

Bin	Avg Confidence	Bin Accuracy
0.0–0.1	0.0083	0.0083
0.1–0.2	0.15	0.16
0.2–0.3	0.25	0.27
0.3–0.4	0.35	0.36
0.4–0.5	0.45	0.46
0.5–0.6	0.55	0.54
0.6–0.7	0.65	0.66
0.7–0.8	0.75	0.74
0.8–0.9	0.85	0.83
0.9–1.0	0.97	0.94

Table 3 and chart 1 show that the calibration curve allows you to visualize how well your predicted confidence aligns with your actual accuracy across each probability bin. The x-axis shows the average predicted probability (confidence) per bin, whereas the y-axis shows the actual fraction of correctly classified samples (bin accuracy). If your model were perfectly calibrated, it would plot along the diagonal line y=x, meaning that your predicted probability would match your actual outcome perfectly. If a point is above the diagonal line, then there is underconfidence (your model predicted less probability than the actual accuracy); if it were below, it would mean the model was overconfident (the model predicted more probability than the actual accuracy).

Chart 1-Calibration Curve for Average Predicted Probability and Bin Accuracy



Interpretation-

The chart indicates that most bins were closely following the diagonal, which suggests that the model is well calibrated. For example, the highest probability bin (0.9-1.0) is slightly below the diagonal line (indicating slight overconfidence for high confidence predictions). Overall, your model has very good agreement between the predicted probabilities and the observed accuracy for the data, confirming that the classifier predicted the correct class and provided appropriate confidence estimations. This information is useful in clinical settings where high-confidence predictions can lead to immediate intervention or low-confidence predictions require additional follow-up assessment.

V. DISCUSSION

The obtained results using calibration analysis also provide new information beyond standard methodologies by providing a means to evaluate the reliability of probability classifications. While having high sensitivity and specificity is indicative of an accurate and reliable classification model, the addition of calibration values provides insight into areas of high probability classifications that are over or under-confident. The finding that there are slightly overconfident probability ranges among high-probability bins indicates that post-hoc calibration techniques (e.g., Platt scaling, isotonic regression) may be beneficial. This approach will also allow evaluation of the model when raw model probability values are unavailable and will use only TP, FN, and Recall data. Therefore, calibration analysis will allow improvement of the assessment of the reliability of predictions based on probability ranges where the model may be less reliable and will assist with clinical determination. When combined with traditional metrics, calibration metrics provide both an accurate and a reliable prediction, which is important in pediatric malnutrition assessment because misallocation of resources can have severe impact. Lastly, calibration analysis allows for subgroup-specific calibration studies, which will assist with assessment within nutrition and/or demographic subcategories. As a whole, calibration analysis will further improve reliability, interpretability, and the clinical applicability of machine learning prediction models within the healthcare field.

VI. CONCLUSION-

This research includes a complete evaluation plan for the classification of malnutrition, combining the use of both standard metrics (i.e. TP, FN, Recall) and probabilistic calibration analysis to generate sample-wise probabilities to produce calibration curves of ECE, MCE, and percent of high-credible probability bins that are over-confident. Results demonstrated a high level of accuracy and good reliability of probability estimates with only slight cases of over-confidence among predictive value bins with high credible range values. As well, the resultant methodology allows for the evaluation of malnutrition classification, independent of the raw probabilities produced by the model, thus providing the necessary information for prioritization opportunities and was also to aid clinicians when making decisions regarding how to treat malnourished patients. Future research can be conducted by employing post-hoc calibration methods, performing subgroup-specific analyses, and validating model predictions using larger datasets to enhance the reliability of malnutrition classification. An evaluation framework is proposed herein that will increase the accuracy and trustworthiness of machine learning model predictions for malnutrition and therefore provide increased usefulness in clinical settings.

REFERENCES

[1]. Islam, M. M., Rahman, M. J., Islam, M. M., Roy, D. C., Ahmed, N. F., Hussain, S., Amanullah, M., Abedin, M. M., & Maniruzzaman, M. (2022). Application of machine learning based algorithm for prediction of malnutrition among women in Bangladesh. *International Journal of Cognitive Computing in Engineering, 3,* 46-57.

[2]. Ula, M., Ulva, A. F., Mauliza, M., Sahputra, I., & Ridwan, R. (2021). Implementation of Machine Learning in DeterminingNutritional Status using the Complete Linkage Agglomerative Hierarchical Clustering Method. Jurnal Mantik, 5(3), 1910-1914.

[3]. Khare, S., Kavyashree, S., Gupta, D., & Jyotishi, A. (2017). Investigation of nutritional status of children based on machinelearning techniques using Indian demographic and health survey data. *Procedia computer science*, *115*, 338-349.

[4]. Thangamani, D., & Sudha, P. (2014). Identification of malnutrition with use of supervised datamining techniques–decision trees and artificial neural networks. *Int J Eng Comput Sci*, *3*(09).

[5]. Najaflou, S., & Rabiei, M. (2021). Recommended system for controlling malnutrition in Iranian children 6 to 12 years old using machine learning algorithms. *International Journal of Web Research*, *4*(1), 27-33.

[6]. Wajgi, R., & Wajgi, D. (2022, March). Malnutrition detection in infants using machine learning approach. In *AIP Conference Proceedings* (Vol. 2424, No. 1). AIP Publishing.

[7]. Ahmadi, P., Alavimajd, H., Khodakarim, S., Tapak, L., Kariman, N., Amini, P., & Pazhuheian, F. (2017). Prediction of low birth weight using Random Forest: A comparison with Logistic Regression. *Archives of Advances in Biosciences*, *8*(3), 36-43.

[8]. Sharma, V., Sharma, V., Khan, A., Wassmer, D. J., Schoenholtz, M. D., Hontecillas, R., ... & Abedi, V. (2020). Malnutrition, health and the role of machine

learning in clinical setting. *Frontiers in nutrition*, *7*, 44.

[9]. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On Calibration of Modern Neural Networks. *ICML*.

[10]. Shinde, A. A., & Sahasrabuddhe, D. V. (2025). Performance Evaluation of a Logistic Model Tree-Based System for Malnutrition Detection in Preschool Children. *International Journal of Environmental Sciences*. ISSN: 2229-7359.

[11]. Shinde, A. A., & Sahasrabuddhe, D. V. (2024). Diagnosing malnutrition in preschool children using machine learning approach: A review. *International Research Journal of Humanities and Interdisciplinary Studies*, Special Issue 26, 21–27.

[12]. Shinde, A. A., & Sahasrabuddhe, D. V. (2025). Attribute Set Reduction using Filter Selection Method for Machine Learning Model to Diagnose Malnutrition in Preschool Children. *International Journal of Environmental Sciences*. ISSN: 2229-7359.