A systematic review on research trends, datasets, algorithms, and frameworks of children's nutritional status prediction

Liliana Swastina[1], Bahbibi Rahmatullah[2], Aslina Saad[2], Hussin Khan[3]
[1]Department of Information System, Faculty of Science and Technology, Sari Mulia University, Banjarmasin, Indonesia
[2]Faculty of Computing and Meta-Technology, Sultan Idris Education University, Tanjong Malim, Perak, Malaysia
[3]Faculty of Tourism Management, Universiti Teknologi MARA, Penang, Malaysia

Article Info

Article history:

Received Jul 17, 2023
Revised Oct 8, 2023
Accepted Dec 14, 2023

Keywords:

Classification
Machine learning
Malnutrition
Prediction
Public-health

ABSTRACT

The monitoring of children's nutritional status serves as a crucial tool for assessing the health of both children and society as a whole. In this regard, machine learning (ML) has been employed to predict nutritional status for monitoring purposes. This topic has been extensively discussed. However, the question remains as to which algorithm or ML framework can yield the highest accuracy in predicting the nutritional status of children within a specific region. Furthermore, determining the appropriate dataset for predictions is also crucial. Therefore, this review aims to identify and analyze the research trends, dataset characteristics, algorithms, and frameworks utilized in studies pertaining to the nutritional status of children under the age of five from 2017 to early 2022. The selected papers focus on the application of ML techniques in predicting nutritional status. The findings of this research reveal that the Bangladesh demographic and health survey 2014 dataset is among the popular choices for ML applications in this field. The most commonly employed algorithms include neural networks, random forests, logistic regression, and decision trees which demonstrated promising performance. Lastly, the data preprocessing stage within a framework plays a significant role in models aimed at predicting nutritional status.

Corresponding Author:

Bahbibi Rahmatullah
Faculty of Computing and Meta-Technology, Sultan Idris Education University
Perak, Malaysia
Email: bahbibi@meta.upsi.edu.my

1. INTRODUCTION

Monitoring children's nutrition is a fundamental instrument for measuring the health of a community [1]. Children's nutrition has a profound impact on their health, cognitive development, and academic performance. Malnutrition significantly hinders motor skills, cognitive and behavioral development, weakens immunity, and increases morbidity and mortality rates [2], [3]. Providing nutrition and healthcare services to disadvantaged children, especially those in the toddler to school-aged range, is critical. Children who have experienced childhood malnutrition, particularly those with low birth weight, face disadvantages compared to those who receive adequate nutrition and live in a healthy environment [4].

Nutritional status serves as a vital indicator for diagnosing children's health [5]. Physiological differences between nourished and undernourished children manifest as variations in body size and ratios, which are referred to as anthropometry. Anthropometric measurement tests, when validated, can be utilized to determine nutritional status [6]. In populations where both undernutrition and overnutrition coexist (known as the double burden of malnutrition (DBM)), the population faces additional challenges. Low- and middle-income countries are particularly prone to DBM [7], [8]. It is estimated that nearly one-third of children in these countries experience malnutrition, and malnutrition-related issues contribute to approximately 20% of total child deaths worldwide [9]. Anthropometric indicators such as weight for height z score (WHZ) are widely employed, but it is important to consider additional parameters such as weight for age z score (WAZ) for underweight and height for age z score (HAZ) for stunting [10]–[12]. In recent years, there has been a growing interest in utilizing machine learning (ML) techniques in children's development, including predicting nutritional status [13], [14]. ML has the potential to extract valuable patterns from large datasets, thereby enabling accurate predictions of children's nutritional status. However, identifying an effective ML framework for this purpose, as well as determining the appropriate dataset, remains a challenge.

In light of these considerations, the primary objectives of this study are threefold. Firstly, it aims to provide an overview of the various approaches employed in nutritional status prediction. Secondly, it aims to comprehensively review the relevant prior research, thereby elucidating the trends, datasets, and methodologies utilized in the application of ML techniques to predict children's nutritional status. Lastly, this study emphasizes the latest advancements in assessing children's nutritional status. By addressing these objectives, this research aims to contribute to the current understanding of the ML-based prediction of nutritional status in children, facilitating the development of more accurate and effective approaches in this field.

2. METHOD

A systematic review approach was employed for this study. Systematic reviews are characterized by their structured and rigorous methods, which involve formulating specific research questions (RQ), systematically identifying relevant papers, critically appraising the papers, analyzing the reported data, and presenting the synthesized results [15]. The search process comprised several activities, including the digital libraries selection, formulation of the search string, initial pilot search, refinement of the search string, and retrieval of an initial list of primary studies, as depicted in Figure 1. To ensure methodological rigor, a review protocol was developed to guide the review process. It includes the establishment of criteria (inclusion and exclusion) for study selection. An eligibility assessment was conducted to determine the suitability of the identified studies [16]. Finally, a data synthesis process was undertaken to systematically analyze and summarize the findings.

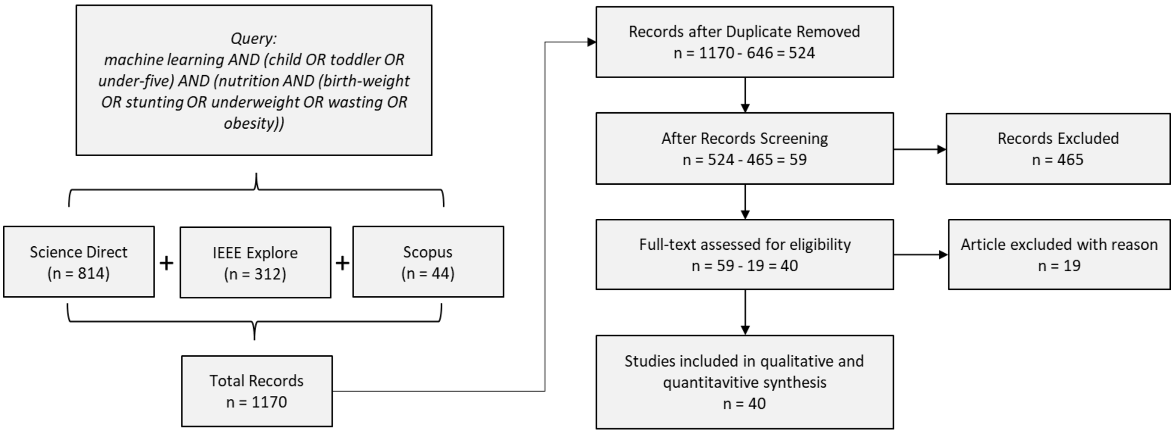

Figure 1. A comprehensive literature selection framework and its outcomes

2.1. Research questions

To maintain focus and clarity in the review, specific RQ were formulated using the population, intervention, comparison, outcomes, and context (PICOC) criteria [17]. These RQs serve as guiding principles for the review process and are presented in Table 1. The PICOC framework ensures that the review addresses the relevant aspects related to the population under study, the interventions or predictors being examined, the comparisons made, the outcomes of interest, and the specific contextual factors considered in the literature. By formulating these RQs, the review aims to systematically explore and analyze the existing literature to provide meaningful insights and contribute to the understanding of the topic at hand.

Table 1. RQ for the systematic literature review

RQ	Questions
RQ1	What are the main research topics selected by researchers in the field of nutritional prediction?
RQ2	Which journal(s) is considered the most significant in the area of nutritional prediction?
RQ3	What are the most commonly used datasets for nutritional prediction?
RQ4	What are the frequently utilized methods for nutritional prediction?
RQ5	Which method demonstrates the best performance when applied to nutritional prediction?
RQ6	What are the proposed frameworks for nutritional prediction?

2.2. Identification

To ensure comprehensive coverage of the literature, a systematic search was conducted across popular literature databases in the field. The selected databases included ScienceDirect (sciencedirect.com), Scopus (scopus.com), and IEEE Xplore (ieeexplore.ieee.org). These databases were chosen based on their reputation and extensive coverage of relevant scholarly publications. The search process involved querying the databases using the title, keyword, and abstract search criteria. The search was limited to publications within the specified timeframe of 2017 to 2022. Both journal papers and conference proceedings were considered for inclusion in the review, as they often provide valuable insights and research findings in the field. Furthermore, to streamline the review process and ensure consistency, the search was restricted to articles published in English, given its status as a widely recognized language of scholarly communication. By implementing these search criteria, the review aimed to gather a broad and diverse range of relevant studies from these databases, enabling a comprehensive analysis of the literature in the field.

2.3. Data extraction for eligibility

Data extraction from the selected primary studies was conducted to address the RQ. The data extraction process aimed to gather relevant information and properties from the studies that would contribute to answering the RQ. The specific properties and data elements to be extracted were identified based on the RQ and the analysis requirements as shown in Table 2. To ensure a comprehensive extraction of relevant information, the data extraction was performed iteratively. This iterative process allowed for the systematic collection of data elements that aligned with the RQ. Multiple properties and data points were considered to provide a comprehensive understanding of the topic and facilitate a thorough analysis [18]. By extracting and collecting the pertinent data from the selected primary studies, this review aimed to gather the necessary information to address the RQ and provide valuable insights into the nutritional status of children and the ML application.

2.4. Quality assessment and data synthesis

In order to enhance the interpretation of synthesis findings and establish the robustness of the conclusions drawn, a thorough assessment of study quality (eligibility) is conducted. Data synthesis involves the consolidation of evidence from the selected studies to address the RQ at hand. While individual pieces of evidence may hold limited strength, their collective aggregation serves to reinforce key points. This review encompasses both quantitative and qualitative data, which have been synthesized using diverse strategies tailored to address various research inquiries.

Table 2. Data extraction properties mapped to questions of research

Property	Question of research
Research publishers and trends	RQ1, RQ2
Nutritional datasets	RQ3
Nutritional prediction methods	RQ4, RQ5
Nutritional prediction frameworks	RQ6

3. RESULTS AND DISCUSSION

The initial query search yielded a total of 1,170 articles, distributed across different databases: 814 articles from ScienceDirect, 312 articles from IEEE Xplore, and 44 articles from Scopus. After the rigorous selection process, review papers and other irrelevant studies were subsequently eliminated. Ultimately, a final set of 40 articles underwent comprehensive full-text reading for analysis and synthesis.

3.1. Research topics in nutritional prediction

Several topics related to nutritional prediction have received a lot of attention, as shown in Figure 2; body mass index [13], [14], development [19], low birth weight [4], [20]–[22], malnutrition [23]–[29], mobidity

[30], nutrition [31], nutritional status [10], [11], [32]–[38], obesity [39]–[44], stunting [45]–[51], and undernutrition [52], [53]. 26 papers (65%) were sourced from 17 journals with quartile index (Q1-Q4) and scimago journal rank (SJR) values. The remaining papers originated from non-SJR journals or conference proceedings. Table 3 provides an overview of the distribution of these journals based on their SJR values and quartile index. The ranking of journal publications is determined by their SJR rating. Notably, Plos One emerged as the journal with the highest number of publications pertaining to nutritional status prediction. The papers predominantly were published in the Q1-Q3 quartiles of health journals. However, from an ML perspective, these predictions are mostly found in the Q4 quartile, indicating the potential for improvement in this area and the role played by ML techniques in advancing the field of study. Frameworks that prioritize accuracy and deliver precise results are highly valued to enhance the effectiveness of such predictions.

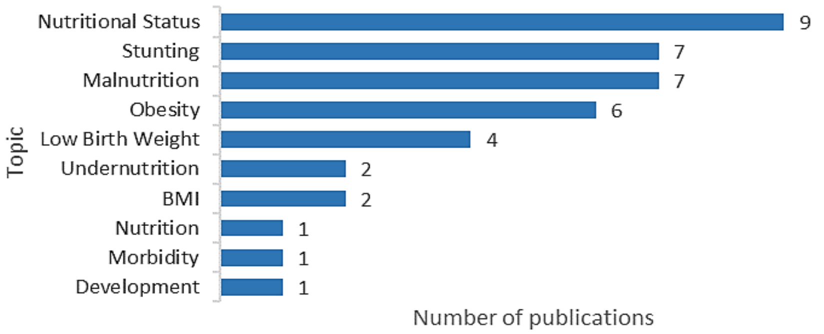

Figure 2. Research topics in nutritional prediction

Table 3. SJR of selected papers

Journal publications	Σ	SJR	Q category
American Journal of Preventive Medicine	1	2.208	Q1 in Public Health
International Journal of Medical Informatics	1	1.135	Q1 in Health Informatics
Maternal Child Nutrition.	1	1.086	Q1 in Nutrition and Dietetics
Public Health Nutrition	1	0.904	Q1 in Nutrition and Dietetics
Plos One	6	0.832	Q1 in multidisciplinary
BMC Pediatrics	1	0.676	Q1 in Pediatrics, Perinatology, and Child Health
Nutrition	1	1.002	Q2 in Nutrition
BMC Med Inform Decision Making	2	0.833	Q2 in Health Informatics
BMC Nutrition	2	0.617	Q2 in Nutrition and Dietetics
Informatics for Health and Social Care	1	0.398	Q2 in Health Informatics
Clinical Epidemiology and Global Health	1	0.371	Q3 in Public Health
International Journal of Advanced Computer Science & Applications	2	0.284	Q3 in Computer Science
Control Engineering and Applied	1	0.251	Q3 in Computer Science
Studies in Computational Intelligence	1	0.237	Q4 in Computer Science
International Journal of Applied Mathematics	1	0.272	Q4 in Computational Theory and Mathematics
Journal of Physics: Conference Series	2	0.210	Q4 in Artificial Intelligence
Journal of Theoretical and Applied Information Technology	1	0.195	Q4 in Computer Science

3.2. Datasets used for nutritional status

Since 2017, there has been a significant increase in paper publication and the utilization of open datasets in nutritional status prediction research. The demographic and health survey (DHS) project, specifically the Bangladesh DHS 2014 dataset [54], has been widely employed in this study domain. Open datasets typically encompass low- and middle-income countries, with a particular focus on Asia and Africa. Researchers investigating children's nutritional status in Bangladesh have identified a set of variables that are considered potential risk factors. These variables encompass various aspects as indicated in Table 4. While Khan *et al.* [48] incorporated nearly all of these features, it is important to note that only three papers claim to predict nutritional status with the highest level of accuracy: [11], [25], [29].

3.3. Prediction methods

Figure 3 summarizes the techniques used in various studies on predicting nutritional status. It's important to note that the number of research papers is not mutually exclusive, meaning that some papers may cover and compare multiple algorithms. Among the 40 articles reviewed, the dominant algorithm type used was classification, which was employed in 30 out of 40 articles. Other algorithm types utilized include clustering and regression, with 4, 3, and 3 articles respectively. Among the 15 algorithms analyzed, four algorithms were found to be most frequently used in predicting nutritional status: logistic regression (LR), random forest (RF), decision tree (DT), and Naïve Bayes (NB). However, it does not necessarily mean that these four algorithms yield the highest accuracy. Despite its relative rarity, the extreme gradient boosting (XGBoost) algorithm outperforms RF in some studies.

Talukder and Ahammed [29] used RF, k-nearest neighbor (KNN), LR, linear discriminant analysis (LDA), and support vector machine (SVM) to classify children's nutritional status in Bangladesh. They found that the RF algorithm gave the best results with 68.51% accuracy, 94.66% sensitivity, and 69.76% specificity. Similarly, Hemo and Rayhan [25] tested RF and DT and concluded that RF is more accurate. Fenta *et al.* [53] compared elastic net (E-Net), LR, ridge regression, lasso, and neural network (NN), and found that RF performed the best. Meanwhile, Shahriar *et al.* [12] reported that NN achieved the best results among NN, SVM, and RF. Ferdowsy *et al.* [40] stated that LR had the highest accuracy among KNN, SVM, RF, DT, mutually beneficial learning, and NB. However, several researchers found that XGBoost performed better than the other algorithms when combined with the LR, RF, DT, and NB algorithms [43], [48].

Table 4. Possible risk factors based on the DHS dataset [54]

Features	Rahman *et al.* [11]	Das and Gulsan [24]	Hemo and Rayhan [25]	Talukder and Ahammed [29]	Bhowmik and Das [46]	Khan *et al.* [48]	Mansur *et al.* [49]	Sultana *et al.* [51]
Child								
Age	v	v	v	v	v	v	v	v
Sex	v	v	v		v	v	v	
Size at birth			v					v
Birth Order	v	v	v				v	
Twin	v							
Current breastfeeding status						v		
Vitamin A in the last 6 months						v		
Maternal								
Age	v		v			v		
Education	v	v	v	v	v	v	v	v
Current working status	v					v	v	v
Occupation		v						
Media exposure			v			v		
Nutritional status						v		
BMI		v	v	v	v		v	v
Age at first birth					v	v	v	v
Preceding birth interval				v	v	v		v
Current pregnancy status						v		
Birth order of index child						v		
Currently breastfeeding			v					
Duration of breastfeeding								v
Frequency of antenatal visits during pregnancy								v
Paternal								
Age						v		
Education	v	v				v	v	v
Occupation		v						v
Household								
Size						v		
Number of living children		v				v		v
Number of children alive with mother						v	v	
Decision-maker regarding child’s healthcare						v	v	
Religion		v				v		
Toilet facility	v	v				v		
Drinking water	v	v						
Wealth index	v	v	v	v	v	v	v	v
Community								
Place of residence	v	v	v	v	v	v	v	v
Division/Province/Regency	v	v	v	v	v	v	v	v

3.4. Proposed frameworks

Several papers have proposed ML frameworks for predicting nutritional status. These frameworks encompass a range of tools and techniques for data pre-processing, model training, and model evaluation. However, obtaining important and significant features can be a challenging task within this process. Among the papers reviewed, it was found that 85% of them stated performing data pre-processing, but only 35%

provided detailed explanations of how they carried out this step. Furthermore, 55% of the papers utilized a specific framework by running a particular algorithm, while the remaining papers conducted comparisons between different frameworks. Three papers stood out as they explained the frameworks they used and received high citations; [11], [12], [40]. The frameworks employed by these papers are depicted in Figures 4(a)-(c). It is interesting to note that despite using the same 2014 Bangladesh demographic and health survey (BDHS) dataset, each study identified a different algorithm as the best performer. This underscores the impact of data processing methods, including feature selection, on algorithm performance.

According to Rahman *et al.* [11], the chi-square analysis was used to assess the association between various explanatory variables and malnutrition. The study utilized 10-fold cross-validation along with SVM, RF, and LR. The results indicated that RF exhibited higher accuracy compared to the others. Shahriar *et al.* [12] highlighted that some characteristics have a limited impact on malnutrition. Data were labeled and features were extracted using anthropometric measurements followed by the primary sampling unit (PSU) to obtain certain attributes from the dataset. However, due to noisy and string data, data cleaning was performed using mean imputation and conversion to numeric format. Feature scaling was achieved through mean normalization. The results demonstrated that NN outperformed SVM, DT, and RF in terms of accuracy. According to Ferdowsy *et al.* [40], 80% of the dataset was used for training. Accuracy was calculated thrice: first on the initial processed data, then after applying principal component analysis (PCA), and finally on the unprocessed data using the algorithm. The findings indicated that the LR algorithm achieved the highest accuracy. These studies highlight the importance of considering various factors such as data preprocessing, feature selection, and choice of algorithm. The discrepancies in performance may arise due to the differences in treatment at each stage of data processing.

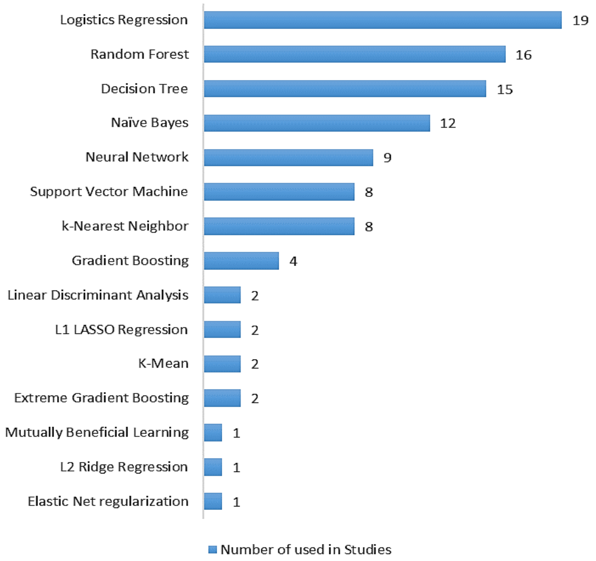

Figure 3. ML algorithms applied in nutritional status prediction studies

3.5. Challenges and opportunities

Feature selection is crucial in identifying the most significant predictors. The choice of algorithms for feature selection and division of datasets for training and testing (hold-out or cross-validation) are important considerations. The number of cross-validation folds also impacts the performance. Additionally, selecting appropriate metrics for evaluating performance, such as area under the curve (AUC), receiver operating characteristic (ROC), and root mean square error (RMSE), is essential. The frameworks proposed in [11], [12], [40], provide valuable starting points but can still be subject to improvements. Identifying the features that have a significant effect on nutritional status is also challenging. Researchers have found that demographic characteristics, including household socioeconomic status and parents' education, significantly influence the prevalence of malnutrition [25], [53]. These features were included in [11], [25], [29]. It is important to acknowledge that data behavior can differ within local communities. Therefore, it is essential to test and validate nominated frameworks using local data before full implementation. Beyond the ML prediction paradigm, these features are important as they guide policy interventions at the household and community levels, aiming to address child nutrition vulnerabilities and reduce disparities [1]. Overall, addressing these stated challenges can pave the way for effective intervention strategies to improve nutritional status.

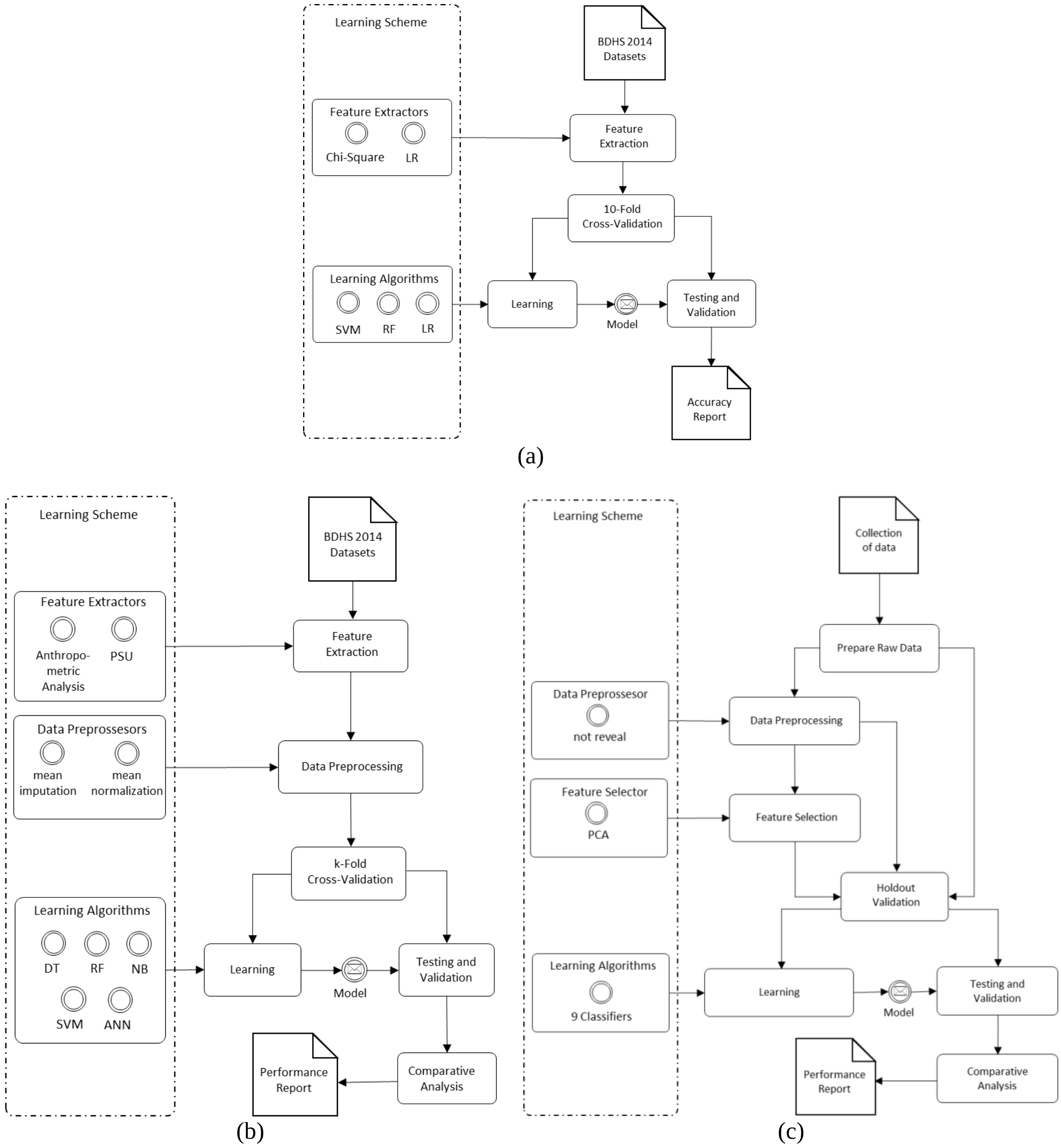

Figure 4. Frameworks proposed for the prediction nutritional status from (a) Rahman *et al.* [11], (b) Shahriar *et al.* [12], and (c) Ferdowsy *et al.* [40]

4. CONCLUSION

ML algorithms have emerged as a prominent approach for predicting nutritional status. Through the identification of relevant features, it has become possible to accurately predict nutritional status. The BDHS dataset from 2014 has been widely recognized and used as an open dataset for nutritional status prediction, with three subsets proposed for specific studies. Several ML methods, including LR, RF, XGBoost, and NN, have demonstrated their suitability and high performance in predicting nutritional status. Three frameworks for predicting nutritional status have been highlighted, but it is crucial to validate their performance with local data before full implementation. Additionally, there is an opportunity for novel modifications and improvements in these frameworks to enhance their accuracy and applicability. Overall, the utilization of ML algorithms, the identification of relevant features, and the availability of open datasets have paved the way for significant advancements in predicting nutritional status. Continued research and refinement of these approaches will contribute to better understanding and intervention strategies for improving nutritional outcomes.

ACKNOWLEDGEMENTS

The authors gratefully acknowledged the Sari Mulia University, Indonesia, and Sultan Idris Education University, Malaysia for their support.

REFERENCES

[1] B. Lareno, L. Swastina, and F. Tan, "The mapping of malnutrition and stunting through web-based support system," *Asia Pacific Institute of Advanced Research (APJCECT)*, vol. 6, no. 2, pp. 30–39, 2020.

[2] R. Kusumaningrum, T. A. Indihatmoko, S. R. Juwita, A. F. Hanifah, K. Khadijah, and B. Surarso, "Benchmarking of multi-class algorithms for classifying documents related to stunting," *Applied Sciences*, vol. 10, no. 23, pp. 1–13, 2020, doi: 10.3390/app10238621.

[3] B. Rahmatullah and J. A. Noble, "Anatomical object detection in fetal ultrasound: computer-expert agreements," in *Communications in Computer and Information Science*, Berlin, Heidelberg: Springer, 2014, pp. 207–218. doi: 10.1007/978-3-642-54121-6_18.

[4] S. M. A. I. Pollob, M. M. Abedin, M. T. Islam, M. M. Islam, and M. Maniruzzaman, "Predicting risks of low birth weight in Bangladesh with machine learning," *PLoS ONE*, vol. 17, no. 5, pp. 1-12, 2022, doi: 10.1371/journal.pone.0267190.

[5] C. Verdich *et al.*, "Changes in body composition during weight loss in obese subjects in the NUGENOB study: comparison of bioelectrical impedance vs. dual-energy X-ray absorptiometry," *Diabetes and Metabolism*, vol. 37, no. 3, pp. 222–229, 2011, doi: 10.1016/j.diabet.2010.10.007.

[6] J. Słowik *et al.*, "Nutritional Status Assessment in Children and Adolescents with Various Levels of Physical Activity in Aspect of Obesity," *Obesity Facts*, vol. 12, no. 5, pp. 554–563, 2019, doi: 10.1159/000502698.

[7] R. E. Black *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, doi: 10.1016/S0140-6736(13)60937-X.

[8] B. M. Popkin, C. Corvalan, and L. M. G. -Strawn, "Dynamics of the double burden of malnutrition and the changing nutrition reality," *The Lancet*, vol. 395, no. 10217, pp. 65–74, 2020, doi: 10.1016/S0140-6736(19)32497-3.

[9] The Development Initiatives, "Global nutrition report: shining a light to spur action on nutrition," Bristol, UK: Development Initiatives, 2018.

[10] S. N. Ariyadasa, L. K. Munasinghe, S. H. D. Senanayake, and N. A. S. Fernando, "Data mining approach to minimize child malnutrition in developing countries," in *International Conference on Advances in ICT for Emerging Regions, ICTer 2012-Conference Proceedings*, IEEE, 2012, pp. 225-225, doi: 10.1109/ICTer.2012.6423030.

[11] S. M. J. Rahman *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*, vol. 16, no. 6 June 2021, pp. 1-11, 2021, doi: 10.1371/journal.pone.0253172.

[12] M. Shahriar, M. S. Iqubal, S. Mitra, and A. K. Das, "A deep learning approach to predict malnutrition status of 0-59 month's older children in Bangladesh," in *Proceedings - 2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology, IAICT 2019*, IEEE, 2019, pp. 145–149. doi: 10.1109/ICIAICT.2019.8784823.

[13] A. Saad, S. Samuri, B. Rahmatullah, and M. C. Mustafa, "The Trend of Body Mass Index (BMI) Changes among Malaysian Children and the Prediction at 48 Months Old," *Asia-Pacific Journal of Research in Early Childhood Education*, vol. 15, no. 2, pp. 187–206, 2021, doi: 10.17206/apjrece.2021.15.2.187.

[14] S. Zare, M. R. Thomsen, R. M. Nayga, and A. Goudie, "Use of Machine Learning to Determine the Information Value of a BMI Screening Program," *American Journal of Preventive Medicine*, vol. 60, no. 3, pp. 425–433, 2021, doi: 10.1016/j.amepre.2020.10.016.

[15] S. Gopalakrishnan and P. Ganeshkumar, "Systematic reviews and meta-analysis: Understanding the best evidence in primary healthcare," *Journal of Family Medicine and Primary Care*, vol. 2, no. 1, pp. 9-14, 2013, doi: 10.4103/2249-4863.109934.

[16] S. A. Anter, B. Rahmatullah, S. H. Khaleefah, and K. F. Tamrin, "A general framework for selecting appropriate criteria of student as research assistant using fuzzy delphi method," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 4, pp. 839–846, 2021, doi: 10.11591/ijai.v10.i4.pp839-846.

[17] S. Keele, "Guidelines for performing systematic literature reviews in software engineering," United Kingdom: EBSE Technical Report, 2007.

[18] N. A. Ghafar, B. Rahmatullah, N. A. Razak, F. H. A. Muttalib, M. H. Adnan, and L. L. Sarah, "Systematic Literature Review on Digital Courseware Usage in Geography Subjects for Secondary School Students," *Journal of ICT in Education*, vol. 10, no. 1, pp. 26–39, 2023.

[19] J. Bansod, M. Amonkar, A. Naik, T. Vaz, M. Sanke, and S. Aswale, "Prediction of Child Development using Data Mining Approach," *International Journal of Computer Applications*, vol. 177, no. 44, pp. 13–17, 2020, doi: 10.5120/ijca2020919955.

[20] U. Hange, R. Selvaraj, M. Galani, and K. Letsholo, "A data-mining model for predicting low birth weight with a high AUC," in *Studies in Computational Intelligence*, Springer International Publishing, 2018, pp. 109–121. doi: 10.1007/978-3-319-60170-0_8.

[21] W. T. Bekele, "Machine learning algorithms for predicting low birth weight in Ethiopia," *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, pp. 1-16, Sep. 2022, doi: 10.1186/s12911-022-01981-9.

[22] A. Zahirzada and K. Lavangnananda, "Implementing Predictive Model for Low Birth Weight in Afghanistan," in *KST 2021-2021 13th International Conference Knowledge and Smart Technology*, IEEE, 2021, pp. 67–72. doi: 10.1109/KST51265.2021.9415792.

[23] C. Browne *et al.*, "Multivariate random forest prediction of poverty and malnutrition prevalence," *PLoS ONE*, vol. 16, no. 9 September, pp. 1-23, Sep. 2021, doi: 10.1371/journal.pone.0255519.

[24] S. Das and J. Gulshan, "Different forms of malnutrition among under five children in Bangladesh: A cross sectional study on prevalence and determinants," *BMC Nutrition*, vol. 3, no. 1, pp. 1-12, 2017, doi: 10.1186/s40795-016-0122-2.

[25] S. A. Hemo and M. I. Rayhan, "Classification tree and random forest model to predict under-five malnutrition in Bangladesh," *Biometrics & Biostatistics International Journal*, vol. 10, no. 3, pp. 116–123, 2021, doi: 10.15406/bbij.2021.10.00337.

[26] Z. Momand, P. Mongkolnam, P. Kositpanthavong, and J. H. Chan, "Data mining based prediction of malnutrition in Afghan children," in *KST 2020 - 2020 12th International Conference on Knowledge and Smart Technology*, IEEE, 2020, pp. 12–17. doi: 10.1109/KST48564.2020.9059388.

[27] G. Moonga *et al.*, "Modelling chronic malnutrition in Zambia: A Bayesian distributional regression approach," *PLoS ONE*, vol. 16, no. 8 August, pp. 1-20, 2021, doi: 10.1371/journal.pone.0255073.

[28] R. L. Pula and R. A. Esquivel, "Malnutrition among Filipino children: a predictive analysis using data mining approach,"

International Journal of Computer Trends and Technology, vol. 70, no. 1, pp. 20–24, 2022, doi: 10.14445/22312803/ijctt-v70i1p104.
[29] A. Talukder and B. Ahammed, "Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh," *Nutrition*, vol. 78, 2020, doi: 10.1016/j.nut.2020.110861.
[30] M. I. H. Methun, A. Kabir, S. Islam, M. I. Hossain, and M. A. Darda, "A machine learning logistic classifier approach for identifying the determinants of under-5 child morbidity in Bangladesh," *Clinical Epidemiology and Global Health*, vol. 12, pp. 1-7, 2021, doi: 10.1016/j.cegh.2021.100812.
[31] S. S. Hardiani, M. Safii, and D. Suhendro, "Application of Naive Bayes algorithm in classification of child nutrition at the simalungun health office," *International Journal of Basic and Applied Science*, vol. 10, no. 3, pp. 79–87, 2021, doi: 10.35335/ijobas.v10i3.57.
[32] 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 Computer Science*, vol. 115, pp. 338–349, 2017, doi: 10.1016/j.procs.2017.09.087.
[33] R. A. Putri, S. Sendari, and T. Widiyaningtyas, "Classification of toddler nutrition status with anthropometry calculation using naïve bayes algorithm," in *3rd International Conference on Sustainable Information Engineering and Technology, SIET 2018-Proceedings*, IEEE, Nov. 2018, pp. 66–70. doi: 10.1109/SIET.2018.8693169.
[34] T. E. Putri, R. T. Subagio, Kusnadi, and P. Sobiki, "Classification system of toddler nutrition status using Naïve Bayes classifier based on z- score value and anthropometry index," *Journal of Physics: Conference Series*, vol. 1641, no. 1, pp. 1-6, Nov. 2020, doi: 10.1088/1742-6596/1641/1/012005.
[35] S. Sendari, T. Widyaningtyas, and N. A. Maulidia, "Classification of toddler nutrition status with anthropometry using the k-nearest neighbor method," in *ICEEIE 2019 - International Conference on Electrical, Electronics and Information Engineering: Emerging Innovative Technology for Sustainable Future*, IEEE, 2019, pp. 154–158. doi: 10.1109/ICEEIE47180.2019.8981408.
[36] S. Winiarti, H. Yuliansyah, and A. A. Purnama, "Identification of Toddlers' nutritional status using data mining approach," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 1, pp. 164–169, 2018, doi: 10.14569/IJACSA.2018.090122.
[37] H. Yuliansyah, S. Winiarti, I. Arfiani, and N. Sari, "Comparison and analysis of classification algorithm performance for nutritional status data," *International Journal of Computer Applications*, vol. 176, no. 20, pp. 14–20, 2020, doi: 10.5120/ijca2020920157.
[38] A. Ridwan and T. N. Sari, "The comparison of accuracy between Naïve Bayes classifier and c4.5 algorithm in classifying toddler nutrition status based on anthropometry index," *Journal of Physics: Conference Series*, vol. 1764, no. 1, pp. 1-6, 2021, doi: 10.1088/1742-6596/1764/1/012047.
[39] A. Alqahtani, F. Albuainin, R. Alrayes, N. Al Muhanna, E. Alyahyan, and E. Aldahasi, "Obesity level prediction based on data mining techniques," *IJCSNS International Journal of Computer Science and Network Security*, vol. 21, no. 3, pp. 103-111, 2021, doi: 10.22937/IJCSNS.2021.21.3.14.
[40] F. Ferdowsy, K. S. A. Rahi, M. I. Jabiullah, and M. T. Habib, "A machine learning approach for obesity risk prediction," *Current Research in Behavioral Sciences*, vol. 2, pp. 1-9, Nov. 2021, doi: 10.1016/j.crbeha.2021.100053.
[41] R. Hammond *et al.*, "Predicting childhood obesity using electronic health records and publicly available data," *PLOS ONE*, vol. 14, no. 4, pp. 1-18, Apr. 2019, doi: 10.1371/journal.pone.0215571.
[42] D. Molina, A. De-La-Hoz, and F. Mendoza, "Classification and features selection method for obesity level prediction," *Journal of Theoretical and Applied Information Technology*, vol. 99, no. 11, pp. 2525–2536, 2021.
[43] X. Pang, C. B. Forrest, F. Lê-Scherban, and A. J. Masino, "Prediction of early childhood obesity with machine learning and electronic health record data," *International Journal of Medical Informatics*, vol. 150, pp. 1-8, 2021, doi: 10.1016/j.ijmedinf.2021.104454.
[44] A. Taghiyev, A. A. Altun, and S. Caglar, "A hybrid approach based on machine learning to identify the causes of obesity," *Control Engineering and Applied Informatics*, vol. 22, no. 2, pp. 56–66, 2020.
[45] F. M. Amin and D. C. R. Novitasari, "Identification of stunting disease using anthropometry data and long short-term memory (LSTM) model," *Computer Engineering and Applications Journal*, vol. 11, no. 1, pp. 25–36, 2022, doi: 10.18495/comengapp.v11i1.395.
[46] K. R. Bhowmik and S. Das, "On selection of an appropriate logistic model to determine the risk factors of childhood stunting in Bangladesh," *Maternal and Child Nutrition*, vol. 15, no. 1, pp. 1-10, 2019, doi: 10.1111/mcn.12636.
[47] E. Harrison *et al.*, "Machine learning model demonstrates stunting at birth and systemic inflammatory biomarkers as predictors of subsequent infant growth – a four-year prospective study," *BMC Pediatrics*, vol. 20, no. 1, pp. 1-10, 2020, doi: 10.1186/s12887-020-02392-3.
[48] J. R. Khan, M. B. Hossain, and N. Awan, "Community-level environmental characteristics predictive of childhood stunting in Bangladesh - a study based on the repeated cross-sectional surveys," *International Journal of Environmental Health Research*, vol. 32, no. 3, pp. 473–486, 2022, doi: 10.1080/09603123.2020.1777947.
[49] M. Mansur, A. Afiaz, and M. S. Hossain, "Sociodemographic risk factors of under-five stunting in Bangladesh: Assessing the role of interactions using a machine learning method," *PLoS ONE*, vol. 16, no. 8 August, pp. 1-17, 2021, doi: 10.1371/journal.pone.0256729.
[50] A. Nugroho, H. L. H. S. Warnars, F. L. Gaol, and T. Matsuo, "Trend of stunting weight for infants and toddlers using decision tree," *IAENG International Journal of Applied Mathematics*, vol. 52, no. 1, pp. 1-5, 2022.
[51] P. Sultana, M. M. Rahman, and J. Akter, "Correlates of stunting among under-five children in Bangladesh: a multilevel approach," *BMC Nutrition*, vol. 5, no. 1, pp. 1-12, 2019, doi: 10.1186/s40795-019-0304-9.
[52] 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, pp. 269–280, 2022, doi: 10.1017/S1368980021004262.
[53] H. M. Fenta, 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*, vol. 21, no. 1, pp. 1-12, 2021, doi: 10.1186/s12911-021-01652-1.
[54] National Institute of Population Research and Training (NIPORT), Mitra and Associates, and ICF International, "Bangladesh demographic and health survey 2014," Dhaka, Bangladesh, and Maryland, USA: NIPORT, Mitra and Associates, and ICF International, 2016.

BIOGRAPHIES OF AUTHORS

Liliana Swastina is a lecturer at Sari Mulia University Banjarmasin at Information System study program. Currently, she is a doctoral degree student at the Faculty of Computing and Meta-Technology at Sultan Idris Education University, Malaysia. Her research interests is big data, deep learning, machine learning, data mining, databases, and computer science. She can be contacted at email: liliana.swastina@unism.ac.id.

Bahbibi Rahmatullah is currently an associate professor in the Faculty of Computing and Meta-Technology, Sultan Idris Education University, Malaysia. She has authored a wide range of publications and has been invited to review articles for high-impact journals and conferences. Her current research interest includes image and signal processing, machine learning, analytics, child development, ICT, and education. She can be contacted at email: bahbibi@meta.upsi.edu.my.

Aslina Saad is currently an associate professor in the Faculty of Computing and Meta-Technology at Sultan Idris Education University, Malaysia. She holds a Ph.D. in Computer Science from Loughborough Uni, UK in 2011. With a strong background in information systems and IT, she brings a wealth of knowledge and expertise to her academic pursuits and is actively engaged in research projects that contribute to advancements in this field. She can be contacted at email: aslina@meta.upsi.edu.my.

Hussin Khan is currently serving as a faculty member in the Faculty of Tourism Management at UITM Penang, Malaysia. He holds a B.Sc. in Food Service Management and a Master's degree in Food Sciences. He has developed strong expertise in various areas of the food industry. His research focuses on nutritional aspects, consumer preferences, and emerging trends in the food sector. He can be contacted at email: hussi103@uitm.edu.my.