Early Detection of Childhood Malnutrition Using Survey Data and Machine Learning Approaches

Md. Thoufiq Zumma
Department of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
thoufiq15-10968@diu.edu.bd

Md. Anikur Rahaman
Department of Computer Science
Washington University Science and Technology
NewYork, USA
anikura.student@wust.edu

Mohammad Abdul Muneem
Department of Computer Science
Washington University Science and Technology
NewYork, USA
moabdul.student@wust.edu

Obyed Ullah Khan
Department of Computer Science
Wilmington University,
NewYork, USA
okhan001@my.wilmu.edu

Nuzhat Noor Islam Prova
Department of Seidenberg School of CSIS
Pace University
NewYork, USA
np23474n@pace.edu

Abstract— **To the health and well-being of children, including topics such as the need of sleep, the forms and causes of malnutrition, puberty, food allergies, and iron deficiency. Children's physical, mental, and emotional development depends heavily on sleep, with age-specific sleep needs fluctuating. Malnutrition may result from eating too little, absorbing nutrients poorly, being unwell, living in poverty, or not knowing enough about nutrition. Food allergies and iron insufficiency are two examples of illnesses whose onsets might differ depending on environmental exposures and genetics. By providing proper diet, getting medical attention when necessary, and encouraging healthy behaviors, parents play a crucial role in supporting their children's health. For children to grow and develop properly, they need meals high in fruits, vegetables, milk, protein, calcium, iron, and fiber. Key advantages of a balanced diet include aiding bone formation, fostering heart health, and encouraging regular bowel motions and digestion. In general, children's general health and wellness depend on caregivers being aware of their dietary requirements and giving them the proper attention. The findings of the research imply that machine learning algorithms may be useful in understanding and predicting the prevalence of malnutrition in children. Furthermore, these algorithms could provide perceptive data that might help politicians create effective policies.In this research, we make use of five different machine learning algorithms. With a decision tree, the accuracy is 0.92%, followed by svm, which provides 0.91%, KNN, which provides 0.84%, GBM classifier, which provides 0.79%, and random forest, which provides 0.65%.**

Keywords— Child, Malnutrition, Health, Wellbeing, Machine learning, Decision tree, SVM

I. INTRODUCTION

Millions of children worldwide suffer from childhood malnutrition, which has long-term health effects. It is a serious issue for public health. Early detection of malnutrition is necessary for prompt care and the prevention of major health consequences. Traditional methods of diagnosing malnutrition often rely on physical traits like height and weight, but they may not always properly consider the nuances and complexity of malnutrition risk factors. It has become more and more common in recent years to use survey data and machine learning methods to enhance early detection efforts. Surveys are a common tool used to collect copious amounts of data about socioeconomic level, dietary habits, access to healthcare, and other relevant characteristics that may increase the risk of malnutrition. Machine learning algorithms provide powerful analytical tools to assist examine this vast and complex data in order to identify patterns, correlations, and prediction models that can aid in the early identification of childhood malnutrition. Survey data combined with machine learning techniques might lead to more individualized and accurate approaches for identifying children who are at-risk being developed by researchers. Healthcare professionals would be able to better manage resources and take preventive action as a result. Additionally, the use of machine learning in this setting may improve scalability and accessibility, particularly in resource-constrained areas where traditional screening methods may not be accessible or practical. However, concerns including data quality, model interpretability, and ethical difficulties must be adequately addressed in order to ensure the accuracy and moral use of these technologies. All things considered, the combination of survey data and machine learning offers a great deal of promise for enhancing early detection efforts and, ultimately, the health outcomes for children who are at risk of malnutrition. Although there are many different types of malnutrition, some prevalent ones include not getting enough vitamins and minerals, being overweight or obese, and any noncommunicable diseases linked

to food that may arise from these problems. By 2020, it is projected that 45 million children will be wasted, or too thin for their height,38.9 million children will be overweight or obese, and 149 million children under the age of five would be stunted, or too short for their age. Malnutrition is thought to be the cause of mortality in around 45% of instances affecting children under the age of five. These are especially prevalent in middle-class or lower-class countries. The prevalence of juvenile obesity and overweight is likewise rising at record rates in these same countries. The global burden of malnutrition has a detrimental effect on people's economic, social, and medical well-being as well as that of their families, communities, and nations. These effects are significant and persistent. Malnutrition may be prevalent in every country on the earth and takes many different forms. The eradication of malnutrition in all of its manifestations is one of the biggest issues affecting public health globally. Malnutrition is more likely to occur in situations when poverty is also quite significant. Individuals who are impoverished are more likely to have several types of malnourishment. Malnutrition also increases health care costs, decreases productivity, and slows economic growth all of which may prolong the cycle of poverty and ill health.

II. Literature Review

The goal of this study was to use the capabilities of machine learning models for the purpose of identifying and predicting important factors that contribute to malnutrition [1]. This research found that, in comparison to other models, automated machine learning algorithms and Tabular Deep Learning frameworks performed faster and more efficiently. They also outperformed other models in terms of accuracy (up to 96.46%) and AUC-ROC scores (up to 99.95%), two important factors in classification problems. This work represents a significant advancement in the use of machine learning to identify possible malnutrition identification, treatment, and prevention. Within [2] They apply a wide range of machine learning algorithms in this study to estimate prediction algorithms for the causes of stunting in children. For this research, data from the 2016 Ethiopian Demographic and Health Survey were utilized. Five machine learning (ML) algorithms eXtreme gradient boosting, k-nearest neighbors (k-NN), random forest, neural network, and generalized linear models were used to predict the sociodemographic risk factors for undernutrition in Ethiopia. A total of 9471 children under the age of five participated in the study. The best prediction algorithm (xgbTree) reveals that a number of characteristics, including time to water source, anemia history, techniques could be used to predict under-five undernutrition in Ethiopian administrative zones [3]. Among the six machine learning techniques we examined, three were selection operator, least absolute shrinkage, and logistic regression. These models' sensitivity, specificity, accuracy, and area under the curve were used to gauge how successful they were. In order to ascertain the malnutrition state of a kid older than 59 months, the goal of this study is to classify malnutrition using a deep learning approach to predictive modeling on significant malnutrition factors [4]. This research clarifies the classification of the malnourished state by using a prediction algorithm. The most accurate approach for determining stunting, underweight, and wasting is the artificial neural network (ANN). The best scientific approach to address it is to use a deep learning-based method, and this is applicable to both policy makers and medical professionals. Within [5] This study presents an overview of the importance of adequate nutrition and how it relates to immune, neurological, and cardiovascular disorders all of which are linked to some degree of nutritional deficiency and from which 60–80% of patients are chosen for a more in-depth assessment. we discuss the potential integration of technological innovations, namely those centered on machine learning, into electronic health records to provide healthcare providers with a decision-support tool for diagnosing and treating patients who are more susceptible to malnourishment. In [6], a machine learning-based prediction model for malnutrition is developed using the available characteristics of the Indian Demographic and Health Survey (IDHS) dataset. The model is then compared with features identified in the literature. According to our study, there are important characteristics that machine learning approaches uncover that have not been covered by the literature. The chance that these traits would explain malnutrition was then ascertained using logistic regression. Within [7] In order to predict the severity of edematous malnutrition in children between the ages of 1 and 59 months old in the context of Afghanistan, a machine learning-based model was proposed for the present experiment. Two hospitals in Afghanistan provided the malnutrition data, which was then examined using the Random Forest, J48, and Naive Bayes classifiers, in that order. With an accuracy of 97.14%, the Random Forest approach produced impressive results, whereas the J48 algorithm's accuracy of 94.51% was relatively mediocre. In [8], an analysis was conducted using data from the 2016 Ethiopian Demographic and Health Survey. Five machine learning (ML) techniques eXtreme gradient boosting, k-nearest neighbors (K-NN), random forest, neural network, and generalized linear models were considered to predict the sociodemographic risk factors for undernutrition in Ethiopia. According on which of the three categories a baby falls into, [9] This article's decision tree model categorizes them as either normal, acutely malnourished, or severely malnourished, depending on when they are between the ages of 0 and 59 months. The Gini index is a commonly used impureness indicator in decision tree models. Using the Gini index, the accuracy attained for stunting, wasting, and underweight conditions using the decision tree is 82.22%, 72.23%, and 78.35%, respectively. The goal of [10] is to provide a succinct synopsis of the potential uses of data mining methods in clinical diagnosis. There is reason to think that data mining, which is presently being used more often in clinical diagnostics, might be advantageous to this field of research.

III. Proposed Methodology

The complete study method is covered here. Different approaches may be used to solve each analysis. The survey data was collected first. After generating and analyzing the dataset, we must eliminate null values and columns not related to our

study. Select a machine learning algorithm. Since we're using five machine learning approaches, we need a data set before we can build the model and fit the algorithm. Using this data, the model is trained. This underpins feature selection. Information was used to construct training and testing sets. Also known as the test and train data sets. A large amount of the data needed to assess our model is left after fitting the data into different machine learning algorithm models and training using a training data set. Next, model accuracy is assessed. We've been using our normal process flow chart to present an overview, but we'll use equations and diagrams to explain specific methods. The research project procedure and a short overview are in this part.

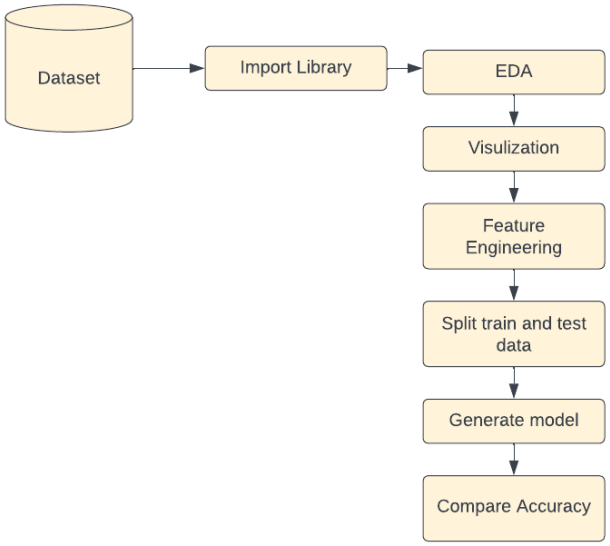

Fig.1. The proposed methodology of the work

A. Dataset

The survey questionnaire, which was sent using the Google Form platform, asked respondents—children in particular—to express their views, experiences, and opinions on pertinent themes. This provided the dataset used in the research. This systematic technique of gathering data is consistent with standard procedures in a number of disciplines, including the social and health sciences and market research. There were 20 columns in the questionnaire that each represented a distinct component of the respondents' answers. The dataset offered a sizable sample size for study, including 1500 rows. The dataset was divided into training and testing sets, with 70% of the data used for training the model and the remaining 30% set aside for evaluating its prediction abilities, in order to assure validity and robustness. Notably, in order to preserve confidentiality and ethical norms, all participant names were made anonymous. The CSV-formatted information is the basis for the use of machine learning methods to forecast malnutrition in children, hence enabling early identification and intervention approaches.

B. Dataset Discriptions

Children's sleep: Children’s sleep to needs vary by age. Infants need 14–17 hours of sleep per day, toddlers 11–14, and school-aged youngsters 9–11 hours. Sleep helps a kid grow and improve their immune system. Sleep improves mood, attention, and cognition in children, making it vital to their mental and emotional health. Childhood malnutrition may have several causes, particularly in early development. Lack of food or a balanced diet may cause malnutrition in children. The child's family may not have enough food or the youngster may be a finicky eater and not receive enough nutrients. Poor nutrition absorption may occur in children with diseases that hinder nutritional absorption. Children with cystic fibrosis or celiac disease may have problems absorbing vitamins and minerals. Children with infections or diseases may not eat as much, making it challenging for them to get appropriate nourishment.

Poverty: Families may lack access to nutritious food or cannot afford medical care for their children. A major nutritional risk factor.

Insufficient information: Parents and caregivers may lack nutrition and child development knowledge to provide appropriate nourishment at appropriate times. Ignorance may cause malnutrition. Malnutrition typically has several causes, so remember that. Preventing and curing child malnutrition requires identifying and addressing its causes. Female kids need the best care throughout puberty since it's vital for their physical and mental development. Puberty brings menstruation and breast development. Parents, caregivers, and healthcare professionals must be aware of the emotional and social concerns that may accompany these changes to support and care for girls. Food allergies and iron deficiency may occur in children at different ages depending on genetics, environment, and diet.

Food allergies: It may occur at any age, although they are more common in youngsters. Food allergies afflict 2-4% of adults and 5-8% of children in the US, according to the American College of Allergy, Asthma, and Immunology. Kids' top food allergies include cow's milk, eggs, peanuts, tree nuts, soy, wheat, fish, and shellfish. Some youngsters grow out of food sensitivities, while others have them permanently. Iron deficiency is more common in young children. The CDC reports that 9% of American children aged 1-2 are iron deficient. Due to their rapid growth and development, newborns require more iron-rich diets since they have less iron at birth. Iron deficiency may induce anemia, which can cause fatigue, weakness, and developmental delays in children.

Disorders may develop at different ages, and some kids never do. Parents and other caregivers may prevent and cure these diseases by feeding their children a balanced diet, monitoring their growth, and visiting a doctor if they have concerns.

Some basic facts about three malnutrition types: Rickets is a condition caused by insufficient vitamin D, calcium, or phosphorus in the body. It may cause development problems, bone deformities, and muscular weakness. Marasmus is a severe malnutrition characterized by weight loss, muscle wasting, and weakened immune system, often caused by a low-calorie, low-protein diet. Lack of protein in the diet may lead to kwashiorkor,

a severe type of malnutrition. An expanded belly, skin rash, and an impaired immune system may ensue. Parents are frequently the first to defend their child's health. Observing changes in their child's behavior or physical symptoms might help them get the right medical treatment. Parental knowledge helps avoid health difficulties by encouraging good diets, exercise, and cleanliness. For parents to stay up to date on their child's health, they must schedule routine well-child exams with a physician. To find out more about their child's growth and health, they may also consult reputable publications, websites, and educational resources.

Quantity of nutrients: For healthy growth, children need protein, carbs, good fats, vitamins, and minerals. The amount required depends on the child's age, weight, and activity level. A child's daily nutrition intake is critical to their health. Vitamins, minerals, and fiber—all necessary for optimal health—are found in fruits and vegetables. To ensure they receive the nutrients they need, encourage kids to consume a variety of colorful fruits and vegetables. Iron and calcium are abundant in spinach and kale, whereas vitamin C is abundant in oranges and strawberries.

Food supplement: Vitamin D and calcium are abundant in milk and are vital for the development of strong bones. Children who don't get enough milk or dairy products may become vitamin deficient. Some children cannot consume milk because of lactose intolerance or allergies; in these cases, they need to find other foods to eat or take supplements of calcium and vitamin D. Their feeding gear is filthy: If don't properly sterilize bottles or wash your hands before handling feeding equipment, might introduce hazardous bacteria or viruses that cause diarrhea in babies.

Unsanitary conditions: Diarrheal illnesses may be more common in babies who live in unclean or crowded environments. Eating or drinking tainted food or water might result in diarrhea.

Overfeeding: A baby's digestive tract may get overworked by overeating or frequent feeding, which may result in diarrhea. Diarrhea may also occur in infants who are fed incorrect diets or are introduced to solid food at an early age. A child does not necessarily develop faster just because they eat more food. Obesity, malnutrition, and aberrant growth may result from overfeeding.

When a child is overfed without receiving adequate nutrients for healthy development, malnutrition may result. Malnutrition and stunted growth may result from a diet rich in calories but low in protein, vitamins, and minerals.

Promoting regularity: Fiber aids in controlling bowel movements. It adds volume to stools, which facilitates their passage through the digestive system and helps treat constipation and encourage regularity. Slowing the breakdown of meals with fiber might potentially lower blood sugar levels and decrease appetite between meals.

Satiety: Fiber-rich foods are fuller, which may help regulate weight and calorie consumption. Whole grains and fiber may lower the risk of heart disease, stroke, and other chronic diseases, according to research. Fiber lowers "bad" LDL cholesterol, a heart disease risk.

Calcium: Early bone formation requires calcium. Milk, yoghurt, cheese, fortified plant-based milks, tofu, and leafy greens are calcium-rich.

Protein: Protein is needed for tissue growth, repair, and immune system function. Protein-rich foods include lean meats, poultry, fish, beans, lentils, eggs, dairy, and nuts.The body uses iron to make red blood cells and deliver oxygen. Red meat, poultry, fish, beans, lentils, fortified cereals, and dark leafy greens are iron-rich.

Fruits and vegetables: These provide fiber, vitamins, and minerals. Choose a range of colors and kinds to provide youngster a diversity of nutrients.

C. Feature Engineering

Label-encoded categorical data is a popular practice. This approach assigns a number to alphabetical labels. Learn about label encoding challenges and how to solve them in Python using the scikit-learn package. When a categorical feature's order is irrelevant, there are many types of one-hot encoding that may take up a lot of memory. Data labeling is a crucial component of data preparation for machine learning, especially for supervised learning, where input and output data are both categorized and labeled to serve as a learning foundation for further data processing. Data labeling is a crucial component of data preparation for machine learning, especially for supervised learning, where input and output data are both categorized and labeled to act as a learning foundation for further data processing.

Since label encoding is a part of Python's data preparation, we will import the LabelEncoder class and use the preprocessing module from the sklearn package, Make a LabelEncoder () instance and store it in the labelencoder variable or object.Fit and convert may be used to change categorical data into a numerical value.

IV. Results and Discussion

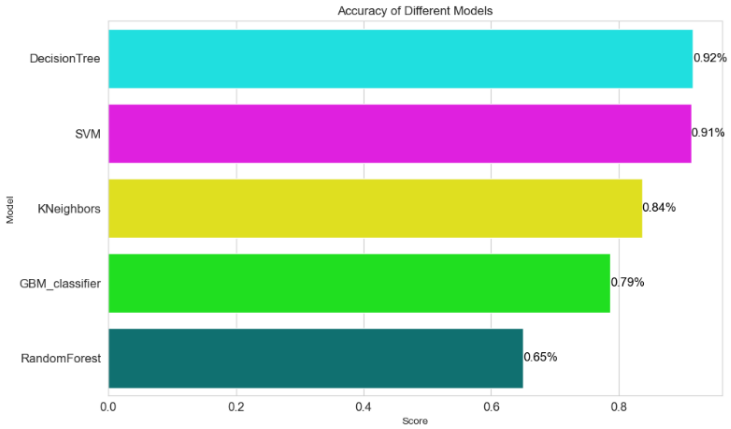

Fig.2. Accuracy of all algorithm

TABLE I. PERFORMANCES OF DIFFERENT CLASSIFIERS

Algorithm	Value name	Precision	Recall	F1-Score	Accuracy
Decision Tree	Ecological conditions	0.92	0.93	0.92	0.92
	Family care	0.92	0.93	0.92	
	Physical health	0.95	0.89	0.92	
	Serving size	0.82	0.86	0.84	
SVM	Ecological conditions	0.92	0.93	0.92	0.91
	Family care	0.95	0.93	0.94	
	Physical health	0.87	0.86	0.87	
	Serving size	0.78	0.86	0.82	
KNN	Ecological conditions	0.82	0.89	0.85	0.84
	Family care	0.84	0.84	0.84	
	Physical health	0.88	0.71	0.78	
	Serving size	0.90	0.72	0.80	
GBM Classifier	Ecological conditions	0.78	0.85	0.80	0.79
	Family care	0.80	0.80	0.80	
	Physical health	0.96	0.57	0.71	
	Serving size	0.77	0.67	0.72	
RandomForest	Ecological conditions	0.65	0.79	0.71	0.65
	Family care	0.62	0.67	0.64	
	Physical health	0.82	0.34	0.48	
	Serving size	1.00	0.19	0.33	

Five distinct machine learning methods were tested for performance in this project: Random Forest, Gradient Boosting Machine (GBM) classifier, K-Nearest Neighbors (KNN), Decision Tree, and Support Vector Machine (SVM). On a dataset, each algorithm was tested and trained, and the results showed what each method's accuracy was. Based on the findings, the Decision Tree algorithm had the greatest accuracy (92%). This implies that the Decision Tree model was very effective in locating the underlying patterns and relationships in the data in order to provide accurate predictions. Not far behind, with 91% accuracy, was the SVM algorithm. Even though SVM performed somewhat worse than Decision Tree, it was still a solid option for the task at hand because of its strong predictive abilities. The KNN algorithm performed mediocrely, falling short of the top two performers with an accuracy rate of 84%. The performance of the nearest neighbor method, which KNN employs to make predictions, could have been influenced by the special characteristics of the dataset. 79% of the time, the GBM classifier was accurate. Lastly, with an accuracy of just 65%, the Random Forest algorithm was the least accurate. Even though Random Forest is an ensemble method that aggregates predictions from several decision trees, it did not perform as well as the other algorithms on this particular dataset.

V. CONCLUSION

Research on children's health and evaluation of machine learning algorithms show that ensuring optimal health throughout childhood requires addressing several factors such as nutrition, sleep patterns, and sickness avoidance. The findings point to the intricate relationships between a variety of factors that influence children's health outcomes, emphasizing the need for all-encompassing approaches that consider both cognitive and physical development. Comparing machine learning algorithms produced encouraging results: SVM and Decision Tree algorithms fared well in predictive modeling for healthcare models. These findings may increase the effectiveness of medical initiatives aimed at enhancing children's health by having significant ramifications for early diagnosis and intervention strategies. Subsequent research endeavors may focus on many significant subjects in order to enhance predictive modeling capabilities and expand our understanding of pediatric health. To begin with, more investigation into the specific traits and factors influencing the differences in performance across different machine learning algorithms might provide valuable data for enhancing and refining the algorithms. It could also be simpler to develop trustworthy prediction models that consider a variety of socioeconomic and demographic factors that have an impact on children's health with larger and more diverse datasets from longitudinal research. Moreover, merging state-of-the-art technologies such as deep learning and natural language processing might provide novel approaches for interpreting complex medical data.

REFERENCES

[1] M. 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,” 2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), Jul. 2019. doi:10.1109/iciaict.2019.8784823

[2] S. Kar et al., “Prediction of child malnutrition using machine learning,” 2021 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON), Dec. 2021. doi:10.1109/iemecon53809.2021.9689083

[3] R. Rahutomo, I. Nurlaila, A. S. Perbangsa, and B. Pardamean, “Database Management System Design with time series modification for child growth and malnutrition monitoring in the regency of Serdang Bedagai,” 2020 International Conference on Information Management and

Technology(ICIMTech),Aug.2020.doi:10.1109/icimtech50083.2020.9211170

[4] S. Jain, T. Khanam, A. J. Abedi, and A. A. Khan, "Efficient machine learning for malnutrition prediction among under-five children in India," 2022 IEEE Delhi Section Conference (DELCON), Feb. 2022. doi:10.1109/delcon54057.2022.9753080

[5] A. Trivedi et al., "Height estimation of children under five years using depth images," 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Nov. 2021. doi:10.1109/embc46164.2021.9630461

[6] A. R. Lakshminarayanan et al., "Malnutrition detection using Convolutional Neural Network," 2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII), Mar. 2021. doi:10.1109/icbsii51839.2021.9445188

[7] C. Fereno, S. Shankar, S. Corr, J. D. Pinter, and L. Herz, "A novel optimization approach for developing foods that address childhood malnutrition," 2019 IEEE Global Humanitarian Technology Conference (GHTC), Oct. 2019. doi:10.1109/ghtc46095.2019.9033043

[8] S. Bonaldo et al., "Portable digital stadiometer for assessing the degree of childhood malnutrition in low-income countries," 2021 IEEE International Humanitarian Technology Conference (IHTC), Dec. 2021. doi:10.1109/ihtc53077.2021.9698950

[9] R. Sasmita, M. Sam'An, L. Assafaat, A. Faturrohman, and Safuan, "Improved malnutrition classification: A grey wolf optimization-based feature selection approach with Random Forest Model," 2023 Eighth International Conference on Informatics and Computing (ICIC), Dec. 2023. doi:10.1109/icic60109.2023.10381980

[10] D. C. Garnica et al., "Machine learning-based classification of children affected by malnutrition using multimodal SMRI and DTI Brain Images," 2023 19th International Symposium on Medical Information Processing and Analysis (SIPAIM), Nov. 2023. doi:10.1109/sipaim56729.2023.10373493

[11] C. Senthilsingh, L. Sujihelen, R. G. Franklin, L. K. Joshila Grace, and A. Sivasangari, "Growth monitoring of children and pregnant women using IOT devices," 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), Dec. 2023. doi:10.1109/icacrs58579.2023.10405379

[12] R. Rahutomo, G. N. Elwirehardja, N. Dominic, B. Caesario, and B. Pardamean, "Database management system design improvement for child stunting data collection in multiple observation areas," 2022 International Conference on Information Management and Technology (ICIMTech), Aug. 2022. doi:10.1109/icimtech55957.2022.9915209

[13] T. Mehmood, M. Sadiq, and M. Aslam, "Filter-based factor selection methods in partial least squares regression," IEEE Access, vol. 7, pp. 153499–153508, 2019. doi:10.1109/access.2019.2948782

[14] A. N. Mallick et al., "Automatic pasteurized formula milk preparation machine with automatic sterilized containers," 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Jul. 2022. doi:10.1109/embc48229.2022.9871811

[15] A. Chakrapani et al., "A pandemic-specific 'emergency essentials kit' for children in the migrant bop communities," 2020 IEEE International Smart Cities Conference (ISC2), Sep. 2020. doi:10.1109/isc251055.2020.9239078