International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 November 2023*

Nutrition Deficiency Prediction using Machine Learning Techniques

U. Hemavathi
Assistant professor, Department of Computer Science and Engineering,
Vel Tech RangarajanDr.Sagunthala R&D Institute of Science and Technology,
Chennai, India
e-mail: uhemav@gmail.com

T.Veeramakali
Associate Professor , Department of Data Science and Business Systems,
School of Computing, SRM Institute of Science and Technology,
Kattankulathur, Chennai, India
e-mail: drveeramakali@gmail.com

S. Prabu
Assistant Professor , Department of Computing Technologies, School of Computing,
SRM Institute of Science and Technology, Kattankulathur,
Chennai, India
e-mail: drprabucse@gmail.com

B. S. Deepa Priya
Associate Professor, Department of Computer Science and Engineering,
Bannari Amman Institute of Technology,
Sathyamangalam, India.
e-mail: deepapriya@bitsathy.ac.in

Abstract— Despite the fact that many developing nations have experienced economic progress, Nutrition- deficiency remains a pervasive problem in the society, with millions of impoverished people's diets lacking in essential macro and micronutrients essential for optimal human health. Lack of awareness of food consumed daily causes Nutrition deficiency among general population, data from multiple health records are used for research and prediction. It investigates the importance of a well-balanced diet for our daily life. The Healthy Food Diversity Index (HFDI) is a supplement to the popular Household Dietary Diversity Score (HDDS). It's a tool for determining the diversity of household food. The HDDS has been established as a reliable source of information, but it has several limitations as a measure of dietary diversity that is linked to nutritional quality. In this paper, various machine learning techniques such as Random Forest classifier (RF), Support-Vector Machine (SVM), Linear Discriminant Analysis (LDA) and Logistic Regression (LR) are used to predict Nutrition-Deficiency using house hold risk factors and they compared their Accuracy, Sensitivity and Specificity. The predictions were also compared to the anthropometric classifications used by the National school feeding program to prove the efficiency of the proposed approach.

Keywords- Nutrition Deficiency, Machine Learning, Logistic Regression, Random Forest, Prediction, HFDI, HDDS, BMI

I. INTRODUCTION

Nutrition deficiency is defined as a person's intake of energy and nutrients are insufficient to meet their demands for optimal health. People with Nutrition-Deficiency can become too short for their age, dangerously thin, and lacking in vitamins, Macro nutrients (Carbohydrates, Proteins and Fats) and minerals (micronutrient deficient), someone can have multiple types of nutrition deficiency at the same time [1,2].

Despite being easily preventable, Nutrition-Deficiency kills 3.1 million children each year, making it the world's leading cause of death among children. Only 24 countries account for 80% of all occurrences of child malnutrition, making this an extremely concentrated problem. Even among adults not being aware of the food that are consumed are the main reason for exceeding in multiple nutrition deficiency and other health problems [3-5].

Nutrition is a critical component in reaching a variety of associated health and development objectives. Globally, better nutrition will reduce child and maternal mortality, enhance educational outcomes, and boost productivity and economic growth, whereas bad nutrition will perpetuate a cycle of poor health and poverty [6-8].

II. RESEARCH BACKGROUND

Following are some of the research papers which came up with various methods for Nutrition-Deficiency prediction and classification. Talukder A and Ahammed B [10] conducted a survey using machine learning algorithms for predicting malnutrition status among the children under the age of five in Bangladesh. The following parameters are noted during the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 November 2023*

evaluation as, nutrition deficiency had a lot of influence factors like mother's BMI index, Parent's education, disease history in the family and Educational status of the parents. The results demonstrated that RF algorithm was moderately superior to any other Machine Learning algorithms used for experimental analysis. Khare S., et.al [11]., investigated the nutritional deficiency of children using supervised classification Techniques. They used the dataset of Indian Demographic data for their investigation. The logistic regression was used to identify the relevant feature which is suitable to identify the malnutrition. The results indicated that mother education has strong relation with child nutritional state. This method mostly depends on parent's socio-economic status. Hence, proper prediction wasn't able to be calculated for other features [16-18]. Md. Merajul Islam et.al., proposed a machine learning based algorithm for prediction of Malnutrition among women in Bangladesh. They used Naïve Bayes, SVM, Decision Tree and Artificial Neural Networks for experimental analysis. The following features such as age, religion, wealth index, education were used and accuracy is calculated and compared against various machine learning models [22].

III. Proposed Method

The authors of the foundation paper proposed four machine learning algorithms such as Random Forest (RF), Support-Vector Machine (SVM), Linear Discriminant Analysis (LDA) and Logistic Regression (LR) for predicting Nutrition-Deficiency from the sample [9,12-15]. This analysis uses individual and household risk factors for experimental investigation. The performances of the algorithms are compared with various metrics such as Accuracy, Sensitivity and Specificity. The accuracy also compared to the anthropometric classifications used by the National school feeding program.

Anthropometry indicators such as the Body Mass Index (BMI) is commonly used to classify nutritional status since they are relatively noninvasive and quick to calculate. So, we have considered Body Mass Index (BMI) as our anthropometric indicator and House hold metrics as another factor to find the nutritional status.

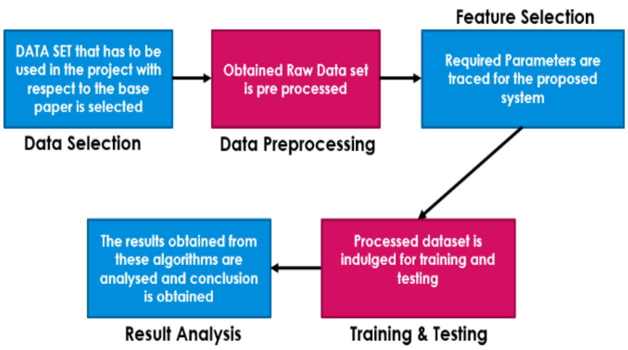

Figure 1. Block Diagram of Proposed Method.

Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 1", even at the beginning of a sentence.

TABLE I. Table Type Styles

Category	Feature Name	Description	Variable Type
Target Variable	NObesity	Based on EMI	Categorical
Eating habits	FAVC	Frequent consumption of high caloric food	Categorical
Eating habits	FCVC	Frequency of consumption of vegetables	Ordinal
Eating habits	NCP	Number of main meals	Ordinal
Eating habits	CAEC	Consumption of food between meals	Ordinal
Eating habits	CH20	Consumption of water daily	Ordinal
Eating habits	CALC	Consumption of alcohol	Ordinal
Physical Conditioning	SCC	Calories consumption monitoring	Categorical
Physical Conditioning	FAF	Physical activity frequency	Ordinal
Physical Conditioning	TUE	Time using technology devices	Ordinal
Physical Conditioning	MTRANS	Transportation	Categorical
Physical Conditioning	SMOKE	Smoking Yes or No	Categorical
Responder Characteristics	Family History with Overweight	Yes or No	Categorical
Responder Characteristics	Gender	Gender is Male or Female	Categorical
Responder Characteristics	Age	Age in years	Integer
Responder Characteristics	Height	Height in meters	Float
Responder Characteristics	Weight	Weight in kg	Float

IV. Experimental Results

A. *Calculation of Confusion matrix*

The performance of the supervised classifiers is analyzed using confusion matrix. It uses four terms as True Positive, True Negative, False Positive and False Negative. TP and TN values are high for better performing classifiers [19-21]. Table 2 shows the illustration of confusion matrix.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 November 2023*

TABLE II. ILLUSTRATION OF CONFUSION MATRIX

N= total predictions	Actual :NO	Actual : YES
Predicted : NO	True Negative (TP)	False Positive (FP)
Predicted : YES	False Negative (FN)	True Positive (TP)

TABLE III. PERFORMANCE EVALUATION OF VARIOUS CLASSIFIERS

ML Techniques	**Confusion matrix**			
	True Positive (TP)	*False Positive (FP)*	*False Negative (FN)*	*True Negative (TN)*
LR	40	4	0	44
DT Classifier	39	4	1	44
RF Classifier	38	5	2	43
SVM	39	5	0	44
LDA Analysis	34	9	2	43

Table 3 represents the confusion matrix of various classifiers. From Table 3, it is observed that, Logistic regression yields better performance compared to other models considered for analysis. Table 4 shows the Accuracy, Precision, Recall, F1-score and AUC-ROC scores of various machine learning algorithms on BMI dataset.

TABLE IV. PERFORMANCE METRIC OF VARIOUS CLASSIFIERS ON BMI DATASET

ML Techniques	**Performance metrics (%)**				
	Accuracy	*Precision*	*Recall*	*F1-Score*	*AUC-ROC score*
LR	94.9	95	94.9	94.9	98.9
DT Classifier	91	91	91	91	94.7
RF Classifier	92.3	92.2	92.3	92.2	99.2
SVM	95.3	95.4	95.3	95.3	99.5
LDA	79.9	80.8	79.9	80.1	96.6

From the above results, it is observed that Support Vector Machine yields the accuracy 95.3%, precision 95.4, recall 95.3, F1-score 95.3 and AUC-ROC score of 99.5 which is greater than all the other algorithms.

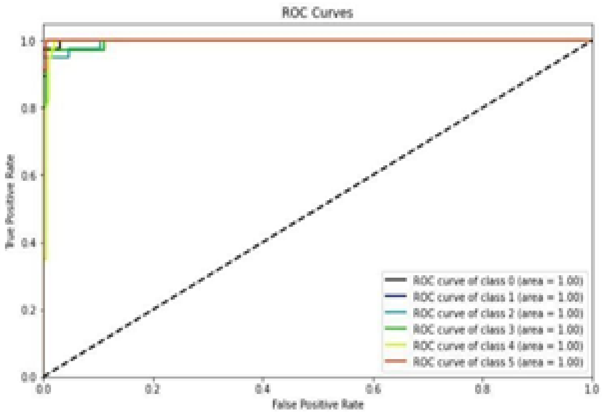

Figure 2. RoC curve of LR.

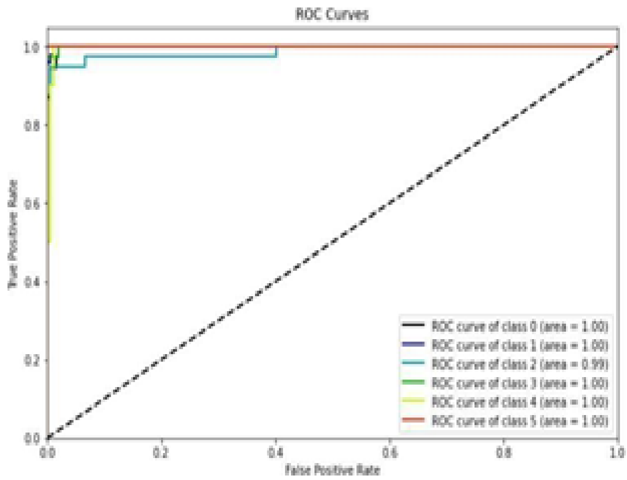

Figure 3. RoC Curve of SVM.

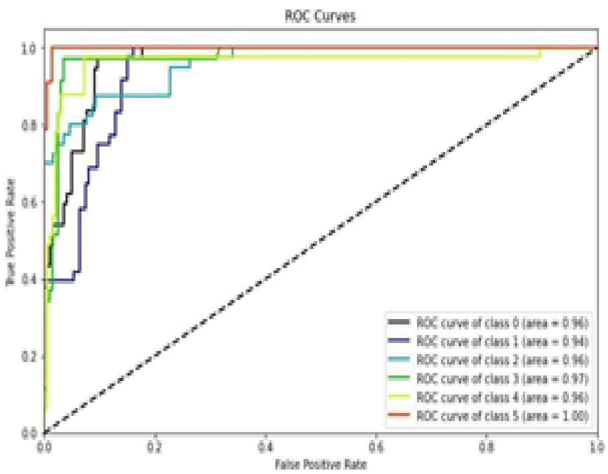

Figure 4. RoC Curve of LDA.

V. CONCLUSION

This research work concludes that machine learning can accurately predict Nutrition-Deficiency based on widely available risk variables, which can help identify children who are at risk for health and nutrition interventions. When applied to an RF model, AMDR classification indicates potential for targeting feeding beneficiaries when compared to traditional BMI-based metrics. However, when developing nutrition intervention standards, local food culture should be considered. Our work uses datasets comprising of about 500 Anthropometric data and 2100 Household data. Different feature sets were tested for classification. All extracted features were taken in consideration for final accuracy score and our method has achieved accuracy of about 94.9%, 91.0%, 92.3%, 95.3% and 79.9% (BMI analysis) for Logistic Regression, Decision Tree Classification, Random Forest Classifier , Support Vector Machine and Linear Discriminant Analysis algorithms and 91%, 95%, 95% and 85% respectively for Logistic Regression, Random Forest Classifier , Support Vector Machine and Linear Discriminant Analysis algorithms which proves the efficiency of our methods for datasets with extended features and size.

REFERENCES

[1] V. Azevedo and M. Robles, "Multidimensional targeting: Identifying beneficiaries of conditional cash transfer programs," *Social Indicators Research*, vol.. 112, pp. 447-475, 2013.

[2] L. Kuhn, S. Brosig and L. Zhang, "The brink of poverty: Implementation of a social assistance programme in rural China. *Journal of Current Chinese Affairs*, vol. 45, no. 1, pp. 75-108, 2016.

[3] A. M. L. Tabunda, J. R. G. Albert and I. Angeles-Agdeppa, "Results of an impact evaluation study on DepED's School-Based Feeding Program (No. 2016-05)," PIDS Discussion Paper Series, 2016.

[4] T. Panch, P. Szolovits and R. Atun, "Artificial intelligence, machine learning and health systems," Journal of global health, vol. 8, no. 2, 2016.

[5] H. Ashrafian and A. Darzi, "Transforming health policy through machine learning," PLoS Medicine, vol. 15, no. 11, e1002692, 2018.

[6] A. L. Beam and I.S. Kohane, "Big data and machine learning in health care. Jama, vol. 319, no. 13, pp. 1317-1318, 2018.

[7] A. Rajkomar, E. Oren, K. Chen, A. M.. Dai, N. Hajaj, M. Hardt and J. Dean, "Scalable and accurate deep learning with electronic health records," NPJ digital medicine, vol. 1, no. 1, pp. 18, 2018

[8] C. Farrar, and K. Worden, "Structural health monitoring: a machine learning perspective," John Wiley & Sons, 2012.

[9] N. Dukhi, R. Sewpaul, M. D. Sekgala and O. O. Awe,. "Artificial intelligence approach for analyzing anaemia prevalence in children and adolescents in BRICS countries: a review," Current Research in Nutrition and Food Science Journal, vol. 9, no. 1, pp. 01-10, 2021.

[10] A. Talukder and B. Ahammed, "Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh," Nutrition, vol. 78, pp. 110861, 2020.

[11] 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.

[12] K. Viswanath and K. Bond "Social determinants and nutrition: reflections on the role of communication'" Journal of nutrition education and behavior, vol. 39, no. 2, pp. S20-S24, 2007.

[13] Y. Ma, B. C. Olendzki, S. L. Pagoto, T.G. Hurley, R. P. Magner, I. S. Ockene and J. R. Hébert, "Number of 24-hour diet recalls needed to estimate energy intake," Annals of epidemiology, vol. 19, no. 8, pp. 553-559, 2009.

[14] M. C. A. Ieiri, S. Kosaka, E. Tomitsuka and M. Umezaki, "Factors affecting undernutrition among school children in Cebu, Philippines," Ecology of Food and Nutrition, vol. 60, no. 2, pp. 182-197, 2021.

[15] S. M. Krebs-Smith, H. Smiciklas-Wright, H. A. Guthrie and and J. Krebs-Smith,"The effects of variety in food choices on dietary quality," Journal of the American dietetic association, vol. 87, no. 7, pp. 897-903, 1987.

[16] J. Coates, A. Swindale and P. Bilinsky, "Household Food Insecurity Access Scale (HFIAS) for measurement of food access: indicator guide: version 3," 2007.

[17] M. M. Manore, "Exercise and the Institute of Medicine recommendations for nutrition," Current sports medicine reports, vol. 4, no. 4, pp. 193-198, 2005.

[18] E. Aurino and V. Morrow, "Food prices were high, and the dal became watery," Mixed-method evidence on household food insecurity and children's diets in India. World Development, vol. 111, pp. 211-224, 2018.

[19] R. P. Salo, "Filipinos' Right to Food: Violated," Ateneo LJ, vol. 57, pp. 638, 2012.

[20] T. Ooka, H. Johno, K. Nakamoto Y. Yoda, H. Yokomichi and Z. Yamagata, "Random forest approach for determining risk prediction and predictive factors of type 2 diabetes: large-scale health check-up data in Japan," BMJ Nutrition, Prevention & Health, vol. 4, no. 1, pp. 140, 2021

[21] L. Breiman "Random forests," Machine Learning, vol. 45, pp. 5–32, 2001.

[22] B. E. Boser, I. M Guyon and V. N. Vapnik, "A training algorithm for optimal margin classifiers," In Proceedings of the fifth annual workshop on Computational learning theory, pp. 144-152, 1992.