2021 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON) | 978-1-6654-2686-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/IEMECON53809.2021.9689083

Prediction of Child Malnutrition using Machine Learning

Shubham Kar
Dept. of CSE
University of Engineering & Management, Jaipur, India
shubhamkar1999@gmail.com

Susmita Pratihar
Dept. of CSE
University of Engineering & Management, Jaipur, India
susmita.pratihar1409@gmail.com

Subhadip Nayak
Dept. of CSE
University of Engineering & Management, Jaipur, India
subhadipnayak56@gmail.com

Sauvik Bal
Dept. of CSE
University of Engineering & Management, Jaipur, India
sauvikbal@gmail.com

Gururaj H L
Dept. of CSE
Vidyavardhaka College of Engineering, Mysuru, India
gururaj1711@vvce.ac.in

Ravikumar V
Dept. of CSE
Vidyavardhaka College of Engineering, Mysuru, India
ravikumarv@vvce.ac.in

***Abstract*— Sometimes malnourished children fall into some serious health issues. And doctors are unable to find out the root causes of their illness, but they used to apply some practices which were not appropriate for every child. Children often die because of this reason. So, it is very dangerous for malnourished children. Along these lines, the fundamental point of our review is to anticipate hunger status of a 1 to 5 years more established kid in Asia by utilizing AI. Looked for ongoing examination papers (2010 - 2020) which identified our point and combined outcomes into a synopsis of what is and isn't known and tried to find out advantages and drawbacks. As explained in the introduction part, to do so, selected a suitable dataset from open source. And they went through many articles to know about Machine Learning Algorithms like their advantages and drawbacks. So, four widely used Machine Learning classifiers like Random Forest, Support Vector Machine, K-Nearest Neighbors, Logistic Regression have been considered to predict a good accuracy score of malnutrition status among under 5 years old. At last, they looked for the best algorithms according to their accuracy score. Based on various performances of Machine learning Algorithms, the best results were performed with Random Forest and Logistic Regression, which demonstrate an accuracy of 91.11 % and 89.88 %, Train accuracy of 1.000 and 0.847. Additionally, a most extreme discriminative ability appeared by Random Forest classification. Here they analyzed those 4 ML algorithms to find which one is performing best. Among them Random Forest and Logistic Regression performing very well. And in future they will do some beneficial work for malnourished children.**

Keywords*— *Child Malnutrition, Random Forest, SVM, Logistic Regression.

I. INTRODUCTION

Primarily, malnutrition is a disruption of the body's normal growth and development. Usually, children aged 0-5 years are malnourished in most cases. In a single word, malnutrition is the lack of nutrition in the body [1]. There are many different reasons behind child malnutrition. Lack of protein and vitamins in the diet is a major cause of malnutrition. In most cases, the baby will not be able to get the required protein because the amount of nutritious food like meat, egg, fish in the diet is low enough. Moreover, the population of India and Asia has increased significantly in the last few decades. The supply of cereals, fruits, fish, meat, and milk has declined, which is a major problem for the economically backward people. Often because of absence of cash they can't devour adequate protein and nutrients needed by the body. Children are also malnourished due to problems with food intake, digestion, absorption, and metabolism. Being infected with any complex disease like diarrhea, jaundice, typhoid etc. after birth is also one of the major causes of malnutrition. Malnutrition problems also depend on the rate of family education [2]. The number of malnourished children is increasing every year. Ailing health obstructs the psychological and actual development of youngsters, prompting an increment in the quantity of kids with gathered insight. Children also suffer from serious diseases like Kwashiorkor, Marashmas, Rickets, Pneumonia, Malaria, Measles, and even anemia. And most of the time die due to lack of proper treatment [3]. As of now, serious hunger is the reason for around 33% of all kids passing consistently. The number of deaths of girls is comparatively higher than that of boys. Essentially because of destitution and absence of training to get appropriate nutritious food the rate of disease among youngsters living in rustic and metropolitan ghetto regions is altogether higher than in metropolitan regions. The growth of the country will be measured not only based on resource and total capita, it will also be measured based on the health of the nation. The basic problem in India is malnutrition among the children. The main problem in India is malnutrition in children under 5 years of age. In the Statistics they show that the number of underweight children in India is the highest level and twice as high as in Africa [4]. The problem of malnutrition in India has been considered as a concentrated phenomenon. In emergent nations, around 2.3 million children between the ages of 6 and 60 die each year, mainly because of malnutrition. It presumes that the breadth of gentle underweight requests more mindfulness as a helpful sign of changing general medical issue among preschool youngsters in developing countries [5][6]. That recognize lack of healthy sustenance at a crude stage for the anticipating and usage of ideal intercession at the association for the wellbeing framework. As the author discussed in the introduction part, child Malnutrition will be a huge disaster for the children under 5 years. So, from our project they will be able to know about the main reason for malnutrition, most affected places, and others. and if those things are known to us also, they can handle the situation. Then they can save those affected children's lives. So, they were motivated for that reason..

II. LITERATURE SURVEY

One of the major public health problems in Asia is malnutrition among children under 5 years of age. In Childhood After Adolescence and adulthood, there has been a big problem due to malnutrition, thus the topic is important in

a developing based country [7]. The major problem is discussing Childhood malnutrition children with a variety of health problems, and it is also late to recover the problems. In this way, Malnutrition is thus causing a major disease in developing countries. In this situation, India has already provided a large portion of the food for the malnourished children [8]. In this pandemic situation they will arrange some awareness camps to make people know the current situation and how to overcome this situation. Whenever they search for the malnutrition-based paper, they found GLIM. GLIM stands for The Global Leadership Initiative on Malnutrition. Basically, it is used to find out the state of nutrition [9]. Here Lack of proper blood tests to find out the nutritional status of children. These are comparatively effective in detecting malnutrition. When they compare the CONUT score, GLIM criteria, and SGA then they clearly see that detecting malnutrition [10]. Whenever they use GLIM, it gives the right value when the score can accurately predict the clinical cause of malnutrition [11]. Since malnutrition has become a major problem, everyone in the society, such as medical and healthcare workers need to pay attention to this situation [12]. Mortality due to malnutrition is on the rise, hospitals are running out of room for this, rehabilitation, physical problems may occur. firstly, find out the symptoms of malnutrition and see what to do next. subsequent nutritional intervention may benefit in addition to economic and reduce medical, healthcare costs. LMICs stands for low- and middle-income countries' studies of nutrition. It is generally based on malnutrition. According to the WHO Protein, energy, micronutrients such as insufficient supply of vitamins and frequent infections and diseases and their consequences. The conclusion is that as many as 2 billion people are suffering from malnutrition and so on [13]. It has been identified as the leading cause of death, disability, and morbidity [14]. After researching this paper, they saw that, Babies who receive poor nutrition within about 100 days of birth are later in the lower stages of education and when they become adult then their working rate is low. Malnourished youngsters likewise deal with the issue of being overweight, as cardiovascular sickness, diabetes, and disease, and of experiencing emotional wellness issues later in their life [15].in the present circumstance they will battle to make individuals mindful with regards to pandemic circumstances and make them reasonable. Whenever they reviewed some malnutrition papers, they saw that there are three main indicators of malnutrition, it's recognized internationally [16]. The three major indicators are Stunning, wasting, overweight and underweight. Wasting indicates that according to the height, weight is decreased. Children who suffer from this are so thin and their immune systems are not so strong, and it causes death. On the other hand, stunning refers to the fact that according to age height is low. Stunning also known as chronic malnutrition [17]. Those children who suffer from the wasting, they are also infected by the stunning. Also being overweight causes stunning, wasting, or stunning wasting both.

According to ICMR, ICMR stands for Indian Council of Medical Research it conveys that, children under the age of 5 die in India due to malnutrition. According to this in India child malnutrition death rate has dropped by about two thirds between 1990 to 2017. Nowadays child deaths are 68%. Above that point they discuss malnutrition indicators. Here they clearly see the birth weight is low, which is the main reason for the infection and health problem., it is the main issue for child growth as well as health problem., including underweight, wasting, and Stunning [18]. There are two aspects of malnutrition that they are now dealing with: these are the lack of food and second is nutrition which is the main issue. Recent published papers have been found to have an NCD effect about diet and all [19]. The burden of malnutrition is also assigned by obesity /overweight as well as conformity. This could be any place in the country, household, community, individual also city [20]. Here they briefly discuss the literature survey, they saw that malnutrition is a major problem. To escape the present circumstance, they need to make a few arrangements that will work, and the future will keep on utilizing this..

III. METHODOLOGY

In this section, a detailed description of the dataset, data pre-processing and proposed methodologies on pre-processed data using various languages and algorithms has been elaborated.

A. Source of Data

The detailed datasets were collected. Then they checked all those datasets to select a suitable one. Finally raw data had been taken from open source. It had been designed to cover both the continents and subcontinent to malnourished children. Our dataset consists of different demographic e.g., birth record, Income scale, Height, Weight record across all over Asia. The dataset has 837 data together with 12 features which are used to analyse our model. Entities like Overweight, Underweight, Stunning is already present in this dataset. So, they skipped Anthropometric analysis.

B. Data Preprocessing

In phase 1, the dataset was not completely ready to fit and train our model because some null and duplicate values were present in our dataset. Therefore, at first all null values and duplicate values had been cleaned by using the most widely used Python library named “NumPy”. And they converted our worldwide dataset to only Asia’s dataset. And the Feature extraction process also had been done.

In phase 2, they simply created a dummy dataset of ‘World Bank’ attribute basis of its values (Low Income, Upper Middle Income, Lower Middle Income). And merged this new dataset with the previous one. After that, the “World Bank” attribute was deleted. And converted our dataset into a csv file with a proper arrangement of attributes.

C. Model Selection

From stage 2, the pre-processed information is currently ready to fit for our AI model. We subdivided our dataset into two parts. First one was for training and the other one was for prediction. ‘Train Test Split’ basically used to split data for training set and test set. After that they researched some suitable Machine Learning Algorithms for our Dataset which can help us to make a good ML model and predict good accuracy scores. So, they selected 4 ML algorithms like Random Forest, Support Vector Machine, Logistic Regression, K-nearest Neighbours based on their advantages and disadvantages. Then that had been imported to fit our model and to check accuracy (Number of correct predictions among total number of Predictions). At last, the comparison process had been done to declare the best algorithm.

Random Forest is not a biased algorithm, and it is also very stable. It functions admirably when you have both absolute and mathematical provisions and when information has missing qualities, or it has not been scaled well and functions admirably for huge datasets. Random Forest gives us high accuracy. Random Forest takes too much time to train our model. It is used to create many trees and then gather their outputs.

Support Vector Machine is a very clear algorithm which can give us a clear separation of data and classes and is much flexible to choosing the similar function. It can handle large types of dimensions. By using SVM in high dimension spaces it can classify the data very well. Support Vector Machine has low performance on the large dataset, and it cannot perform well when classes are not distinct.

Logistic Regression algorithm is a very explainable algorithm, its outputs are well calibrated and predicted probabilities. It's model training and prediction parts are very fast; its dimensions don't need scaling. It can perform very well with a small number of observations. It predicts the binary outcome from a given independent variable. In logistic regression, a high dimension dataset is used for overfitting the model, but it gives wrong output.

K neighbors' classifier is quite simple because it requires just two parameters. So, it is easy to implement. KNN could learn nonlinear decision boundaries. That's why it can separate the data with the help of nonlinear functions. It works very fast because it takes very less time to train data. It is not so efficient. If there are high dimensions in our dataset it does not work well. It does not handle categorical features so well.

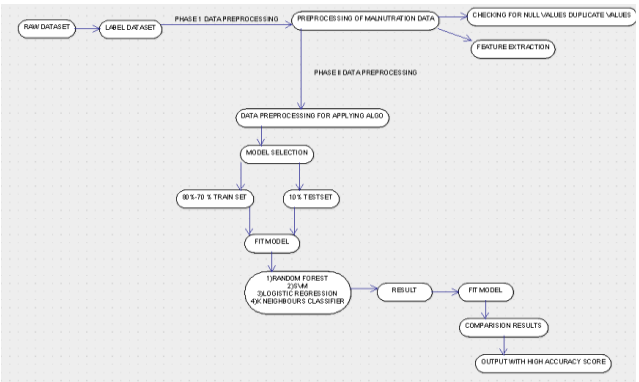

Fig. 1. Model Framework

IV. Results

Among those four algorithms they have implemented Random Forest, besides Random Forest another three Machine Learning Algorithms have been used sequentially. Like Support Vector Machine, Logistic Regression, K-Nearest Neighbors were also applied by us. RF approach showed the best result with accuracy of 91.11% with test size of 0.2% and n-estimators = 300.

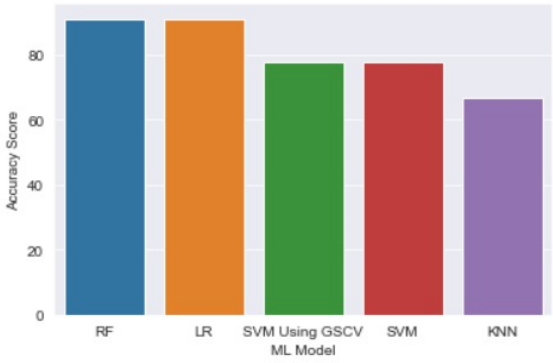

Fig. 2. The accuracy score of our Machine Learning algorithms

The Logistic Regression also gives the exact same result as compared to the Random Forest classifier (Figure 2) with an accuracy score of 91.11%. Training data and Testing data accuracy of Logistic Regression is 84.20% and 84.40% respectively whereas RF performs very well on training data and also testing data. Train and test accuracy of Rf is 99.40%, 91.10%. So that we can assume Random Forest performs better than Logistic Regression on train and test accuracy (Figure 3).

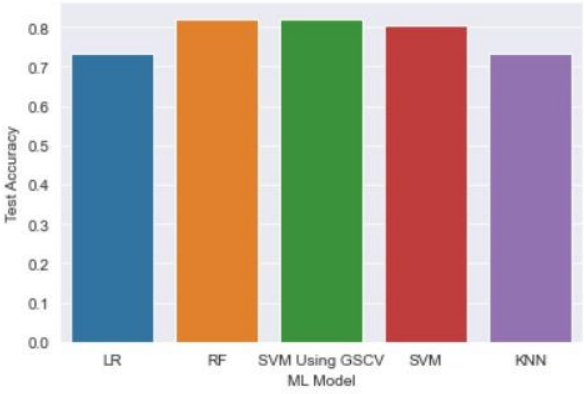

Fig. 3. Test accuracy score of our Machine Learning algorithms

The above graph (Figure 2, Figure 3) shows us that the SVM classifier performs moderately well with a 77.612% accuracy score in both the cases. Training data and Testing data accuracy was also very similar. KNN gives a poor accuracy score of 66.66%, 71.10% test accuracy. Whereas it performs really well to train data. Train accuracy of KNN is 99.40%.

V. Conclusion

In this study, we compared those four machine learning algorithms based on their accuracy score, train, and test accuracy. So that they can predict accurately whether a child is malnourished or not. Among those algorithms Random Forest and Logistic Regression performed very well with the most accurate prediction in Asia's dataset. In this research basically run our machine learning algorithms all over Asia. To foresee if a kid is malnourished. But in future we will do the same for India. So that they will be able to know India's current scenario about malnutrition. It is known that the approximate amount of the children who are suffering from this situation and they can find out some main reasons behind

it. Our main aim is to control the situation. So, they will arrange some campaigns to make people aware about this and to make them sensible. They will provide them proper guidance to overcome this situation. Also, in the future it will be a very good and valuable discovery to differentiate between children whether they are affected by malnutrition or not. And they will create an instrument which can identify malnourished children who are at higher risk. Thus, later on they will provide them with nutritional food, required medicines and other useful health supplements. In this way, our model will be valuable for controlling the situation and repeal of child malnutrition.

References

[1] Rahman A, Biswas SC. Nutritional status of under-5 children in Bangladesh. South Asian J Popul Health. 2009; 2(1):1–11.

[2] Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, de Onis M, et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet. 2013;382(9890):427–51. pmid:23746772.

[3] Caulfield LE, de Onis M, Blossner M, Black RE. Undernutrition as an underlying cause of child deaths associated with diarrhea, pneumonia, malaria, and measles. Am J Clin Nutr. 2004;80:193–8.

[4] World Bank. India, Undernourished children: A call for reform and action. [Last accessed on 2014 Apr 05]. Available from: http://web.worldbank.org/WBSITE/EXTERNAL/COUNTRIES/SOUTHASIAEXT/0,contentMDK:20916955~pagePK:146736~piPK:146830~theSitePK:223547,00.html

[5] Bhagowalia P, Chen SE, Masters WA. Effects and determinants of mild underweight among preschool children across countries and over time. Econ Hum Biol. 2011;9:66–77. [PubMed] [Google Scholar]

[6] Global Nutrition Report Africa Nutrition Profile. Available online: https://globalnutritionreport.org/resources/nutrition-profiles/ (accessed on 6 August 2020).

[7] Vir SC, "Undernutrition in Children: Significance, Current situation, Determinants, Evidence Based Programme Principles and Intervention Package" in Public Health Nutrition in Developing Countries 2nd Edition, ed Sheila c. Vir, (Woodhead Publishers- India 2021) in press.

[8] Zhou L., Varadharajan V., Hitchens M. Achieving secure role-based access control on encrypted data in cloud storage.IEEE transactions on information forensics and security. 2013;8(12):1947-1960.

[9] Kirsch R, Matthews K, Williams V. Using global criteria to detect malnutrition: application in disease states. Nutr Clin Pract. 2020 Feb;35(1):85–97.

[10] Shimose S, Kawaguchi T, Iwamoto H, Tanaka M, Miyazaki K, Ono M, et al. Controlling nutritional status (CONUT) score is associated with overall survival in patients with unresectable hepatocellular carcinoma treated with lenvatinib: a multicenter cohort study. Nutrients. 2020 Apr;12(4):1076.

[11] Maeda K, Ishida Y, Nonogaki T, Mori N., "Reference body mass index values and the prevalence of malnutrition according to the global leadership initiative on malnutrition criteria". Clin Nutr. 2020 Jan;39(1):180–4.

[12] Crichton M.Craven D. Mackay H. Marx W de van der Schueren M.Marshall S., "A systematic review, meta-analysis and meta-regression of the prevalence of protein-energy malnutrition: associations with geographical region and sex. Age Ageing", 2019; 48: 38-48.

[13] Food and Agriculture Organization of the United Nations. The state of food insecurity in the world. Rome, Italy: FAO; 2011.

[14] World Health Organization. MDG1: eradicate extreme poverty and hunger. 2013. Available from: [cited 12 June 2015].

[15] Hoddinott, J., Alderman, H., Behrman, J. R., Haddad, L., & Horton, S. (2013). The economic rationale for investing in stunting reduction. Maternal & Child Nutrition, 9, 69–82.

[16] WHO (2020) Severe acute malnutrition. Nutrition topics. United Nations World Health Organization.

[17] Kennedy, E., Branca, F., Webb, P., Bhutta, Z., & Brown, R. (2015). Setting the scene: An overview of issues related to policies and programs for moderate and severe acute malnutrition. Food and Nutrition Bulletin, 36, S9–S14.

[18] Danaei G, Andrews KG, Sudfeld CR, Fink G, McCoy DC, Peet E, Sania A, Fawzi MCS, Ezzati M, Fawzi WW. Risk factors for childhood stunting in 137 developing countries: a comparative risk assessment analysis at global, regional, and country levels. PLoS Med. 2016;13(11):e1002164.

[19] World Health Organization, 2000. The management of nutrition in major emergencies. Geneva: WHO; De Onis M., Borghi E., Arimond M. et al., 2019. Prevalence thresholds for wasting, overweight and stunting in children under 5 years. Public Health Nutrition, 22:1, pp. 175–79, doi:10.1017/S1368980018002434

[20] The WHO defines 'double burden' as the "coexistence of undernutrition along with overweight and obesity, or diet related non communicable diseases, within individuals, households and populations, and across the lifecourse". For further information, see: The double burden of malnutrition, Lancet series, December 2019. Available at: www.thelancet.com/series/double-burden-malnutrition