Assessing Indicators of Chronic Child Undernutrition using Machine Learning Techniques

Kevin Mwenda[1], Guixing Wei[2]
Spatial Structures in the Social Sciences
Population Studies and Training Center, Brown University, Providence, RI 02912
[1]*e-mail address: kevin_mwenda@brown.edu*
[2]*e-mail address: guixing_wei@brown.edu*

Undernutrition remains a significant obstacle to children under five years old in Kenya. One in four children under five are stunted, meaning that they have height-for-age Z-scores (HAZ) less than minus two standard deviations below the median of a reference height-for-age standard. This paper employs machine learning techniques to identify significant socioeconomic, demographic, cultural, climatic and environmental indicators of child stunting from the Kenya Demographic and Health Survey (KDHS) 2014 dataset. Consequently, multivariate logistic regression is employed to determine and quantify the likelihood of any significant indicators in explaining stunting outcome. All in all, this paper suggests that machine learning techniques would be useful in elucidating significant indicators of child undernutrition that are in critical need of, and can respond well to, prioritized policy interventions.

Keywords: Undernutrition, Machine Learning; Logistic Regression; KDHS, Kenya

I. INTRODUCTION

Despite valiant efforts towards addressing food insecurity that have seen undernutrition decline globally, many sub-Saharan countries still struggle to feed their people. Undernutrition, a global health problem with effects particularly devastating in Sub-Saharan Africa, causes those affected to suffer from weakened immune systems, increased susceptibility to diseases, and higher mortality rates [1]. Chronic undernutrition remains a significant obstacle to children under five years old in Kenya, where approximately 1 in 4 children are stunted, meaning that they have height-for-age Z-scores less than minus two standard deviations below the median of a reference height-for-age standard [2]. Several studies have examined the extent of stunting prevalence by first identifying socioeconomic, demographic, cultural and climatic indicators using *a priori* knowledge or logical deduction [5-6] then conducted statistical analyses using methods such as descriptive & trend analyses and standard logistic regressions [5, 6-8] to model the probability of stunting among children under five. However, the usage of machine learning (ML) techniques to identify indicators of spatially heterogeneous phenomena such as stunting remains scant, despite the advent of greater computational power in the current "big data" era. In an effort towards addressing this knowledge gap, this paper aims to answer the following questions: Can ML techniques identify new explanatory variables (indicators) of child undernutrition? Further, how would these explanatory variables identified by ML be different from those found in extant literature? Lastly, how would these additional indicators contribute to child stunting outcomes? The objective of this paper is to employ ML techniques to identify significant socioeconomic, demographic, cultural, climatic and environmental determinants of child stunting from the Kenya Demographic and Health Survey (KDHS) 2014 dataset. Consequently, multivariate logistic regression is employed to determine and quantify the odds of any significant indicators in explaining stunting incidence.

II. DATA AND METHODS

To achieve our objectives, the first component of the paper is to identify and extract indicators associated with child stunting in the KDHS 2014 dataset [9] from existing literature and machine learning (ML). In their native form, potential candidates for socioeconomic and demographic indicators of child stunting binary outcome would consist of child factors (e.g. age of child, sex of child, birth-weight, type of birth and immunization); maternal factors (e.g. maternal age, education level, body mass index, birth interval, number of under-five children, head of household and wealth index). [4] Cultural, climatic and environmental factors would potentially include indicators such as poverty rate, latrine facility type, drinking water sources, precipitation [10] and Normalized Difference Vegetation Index (NDVI) [11] .

The second part of the paper would entail classification for all indicators using a Decision Tree Classifier which would internally use information gain calculated for all potential indicators of child stunting. Accuracy would be calculated based on the concept of cross-validation to reduce model variance to lead to the most significant indicators. At this stage, indicator selection will be a two-pronged approach, the first of which will entail selection of reduced indicators and the second of which will be based on meta-analysis of indicators identified from literature.

The third and final part of this paper would be to construct Logistic Regression models for child stunting outcome based on the identified reduced indicators from literature and ML, after which the odds of significant indicators will be analyzed and compared to garner further insights towards overall impact on child nutrition, health and well-being. The proposed model design is depicted in figure 1 below:

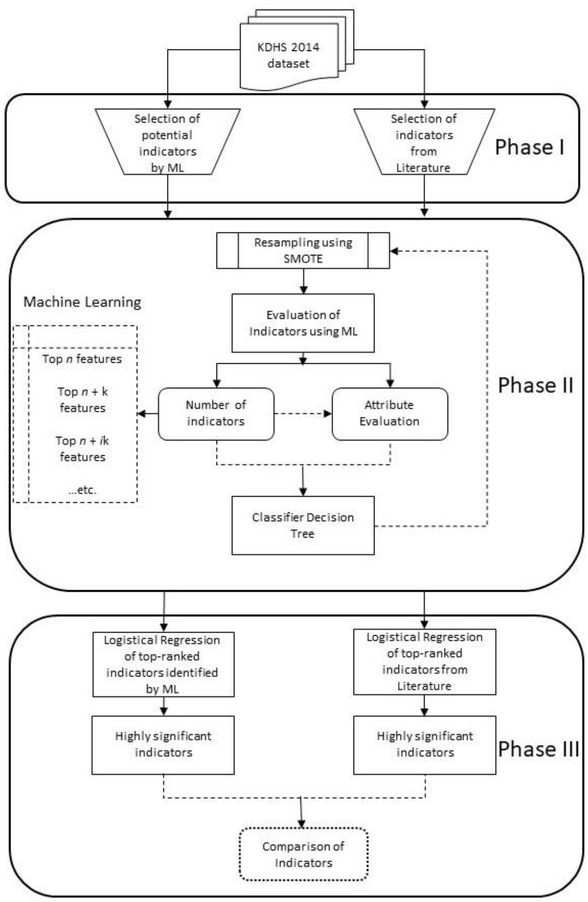

Fig 1. Proposed Model Design

III. EXPECTED FINDINGS

This paper will significantly improve understanding of the indicators behind this prevalent scourge of child undernutrition. To date, the vast majority of indicators of child undernutrition are solely identified and assessed by human deduction. This paper takes advantage of the availability of new sophisticated methods and tools that can improve analysis and prediction of health phenomena such as child malnutrition. Machine learning will be useful to select indicators of child undernutrition with reduced uncertainty, therefore enabling researchers, policy-makers and other stakeholders alike to discover the most effective policies and interventions to address this crisis.

REFERENCES

[1] Berkman DS, Lescano AG, Gilman RH, Lopez SL, Black MM. Effects of stunting, diarrhoeal disease, and parasitic infection during infancy on cognition in late childhood: a follow-up study. Lancet. 2002;359:564–71.

[2] WHO(World Health Organization). Physical status: the use and interpretation of anthropometry. Geneva: World Health Organization, 1995; 2006. p. 36. Technical report no.854.

[3] Adekanmbi VT, Uthman OA, Mudasiru OM. Exploring variations in childhood stunting in Nigeria using league table, control chart and spatial analysis. BMC Public Health. 2013;13:361.

[4] Haile D, Azage M, Mola T, Rainey R. Exploring spatial variations and factors associated with childhood stunting in Ethiopia: spatial and multilevel analysis. BMC Pediatrics. 2016; 16(9)

[5] Khare S, Kavyashree S, Gupta D, Jyotishi A. Investigation of Nutritional Status of Children based on Machine Learning Techniques using Indian Demographic and Health Survey Data. Procedia Computer Science. 2017;115: 338-49.

[6] Matanda DJ, Mittelmark MB, Kigaru DM. Child undernutrition in Kenya: trend analyses from 1993 to 2008–09. BMC Pediatrics. 2014; 14(5)

[7] Mengistu K, Alemu K, Destaw B. Prevalence of Malnutrition and Associated Factors Among Children Aged 6–59 Months at Hidabu Abote District, North Shewa, Oromia Regional State. Nutritional Disorders Therapy. 2013;T1:001.

[8] Tariku B, Mulugeta A, Tsadik M, Azene G. Prevalence and Risk Factors of Child Malnutrition in Community Based Nutrition Program Implementing and Nonimplementing Districts from South East Amhara, Ethiopia. Open Access Library J. 2014;1:1–17.

[9] KDHS, 2014: Final Report. http://www.issn.org (2014). Accessed 12 May 2017.

[10] Funk, Chris, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, James Rowland, Laura Harrison, Andrew Hoell & Joel Michaelsen. "The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes". Scientific Data 2, 150066

[11] López-Carr D, Mwenda KM, Pricope NG, Kyriakidis PC, Jankowska MM, Weeks J, Funk C, Husak G, Michaelsen J. Climate-Related Child Undernutrition in the Lake Victoria Basin: An Integrated Spatial Analysis of Health Surveys, NDVI, and Precipitation Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2016; 9(6): 2830 –5.