Identifying Determining Factors of Low Birth Weight in India using Categorical Principal Component and Artificial Neural Network Model: An Insight from NFHS-5

Alokananda Ghosh[1]
[1] Department of Geography, Tehatta Sadananda Mahavidyalaya, Tehatta, Purba Bardhaman 713122, West Bengal, India, E-mail: alokanandaghosh04@gmail.com

1. Theoretical focus

Low birth weight (LBW) is defined as weight at birth of less than 2500 grams (5.5 pounds) (UNICEF & WHO, 2019). Every year, more than 20 million babies with LBW are born worldwide (Tessema et al., 2021). It is considered to be a significant public health problem as on an average 15% to 20% of all birth around the world are LBW (Anil et al., 2020). 96.5% of LBW births were reported from low and middle-income countries (LMICs) and, especially, in the most vulnerable populations (Zaveri et al., 2020). Globally 20 million births are LBW in a year, 90% of which occur in LMICs. India alone accounts for 40% of all LBW babies born in LMICs (Sharma et al., 2021; Ledinger et al., 2024). Nearly one in six women in India gives birth to newborns having LBW and the problem more prevalent in northern and western parts of the country (Girotra et al., 2023).

The first strength of the study is that to our knowledge, it is the sole study in India which has tried to identify the determining factors of LBW using both Categorical Principal Component (CATPCA) and Artificial Neural Network (ANN) models on NFHS-5 data set. Given the data availability, the study has considered four sets of putative predictive factors from the published source of literatures to identify the strong determinants of LBW in India (Chaudhuri et al., 2023; Jana, 2023; Girotra et al., 2023). These sets are- (1) The socio-demographic and economic conditional factors, (2) the prenatal conditions and factors of utilization of services and child factors (3) maternal complications and morbidity factors, and (4) the maternal nutritional environment.

2. Data and methods

2.1. Data

The present study utilizes data on LBW in India from the fifth round of the National Family Health Survey (NFHS 5, 2019–21), a nationally representative cross-sectional survey under the stewardship of Ministry of Health and Family Welfare (MoHFW), Government of India. A stratified, two-stage sample design is adopted in the NFHS-5 with primary sampling units being census enumeration blocks (CEBs) in urban areas and villages in rural areas. The NFHS-5 sample was designed to provide accurate estimates of all key indicators at the national, state, and district levels (for all 707 districts in India as of March 31, 2017). To achieve this, a sample size of around 636,699 households was determined. The survey collected information of 724,115 women in the reproductive age group (15–49 years), and 176,843 birth weight data of living children. Of these children (most recent birth), a total of 1637 (0.9%) births were multiple birth and 175,206 were (i.e., last-born) singleton births. Among these children who had very low birth weight i.e. <1500 grams (1450, 0.8%) and LBW i.e. <2500 grams (24916, 14.2%) were included in the study. Data on birth weight was collected using two methods: maternal recall and birth cards.

2.2. Outcome variable

Mother delivering (both vaginal and caesarian delivery) singleton live-born baby with birth weight less than 2500 grams without any congenital anomalies (UNICEF & WHO, 2019) and very low birth weight (VLBW) is less than 1,500 grams (Ledinger et al., 2024). The NFHS collected data on birth weight using the following questions: Was (name of the child) weighed at birth? How much did (name of the child) weigh? The information was reported in two ways; first, the mother's recall of her baby's weight, and second, reported with the help of any card of their baby's weight (IIPS and ICF, 2022). The outcome variables are dichotomous, '1' signifying the child had VLBW and '2' denoting the child had LBW.

2.3. Exposure variables

The correlates of adverse birth outcome were selected from published literatures and availability of variables in the NFHS-5 data set. The study has evaluated the association of socio-economic condition, prenatal factors and service utilization, maternal morbidity and obstetric complications, child factors and maternal nutritional factors with the birth outcomes.

2.4. Statistical analysis

2.4.1. Categorical principal component analysis and dimension reduction

The varimax rotation method adopted here and the factors of dimension 1 and 2 with loading value $\geq \pm 0.5$ of each CATPCA model have been considered as strong determinants of the final ANN predictive model. Therefore, total eight object score of each dimension (two dimensions from each model) of CATPCA were considered to construct the ANN model (Figure 1, Table 1).

2.4.2. ANN model construction

A feed-forward ANN model framework was applied to construct final predictive model. The model structure comprised 3 layers. The input layer has the factors been chosen for model construction as neurons, and the output layer had 2 neurons (Very LBW and LBW events). This construction has been selected to evaluate the optimum number of neurons which has determined to be two, distributed in two hidden layers. Hence, the optimum network topology has been determined to be of form N-2-2-2 i.e., input layers carrying different predictors of different models, two neurons in each hidden layer, and two output layer. The sample randomized into a training set (70%) for model development and a test set (30%) for validation. The number of hidden layer neurons has been determined by 100 iterations.

2.4.3. Predictive model evaluation

The final ANN model performance has been enumerated through the approaches of discrimination, calibration, and reclassification (Jiang et al., 2023). The ROC plot is a popular way for discriminatory accuracy visualization of binary classification models and the area under curve (AUC) is a common measure of its exact evaluation. ANN begins operation with the normalization of the input data, which scales inputs and targets so that they fall in the range 1 to -1, corresponding to the highest and lowest values, respectively (Abidoye et al., 2018). The root mean squared errors (RMSE) of final predictive model has been calculated (Aryadoust & Goh, 2014).

3. Results

3.1.Descriptive statistics

Out of 176843 total birth, 175206 (99%) are singleton birth (mean age 26.93 years, minimum age 15yrs, SD - 5.148), amongst them 15.04% have experienced LBW in their recent pregnancy outcome. Maximum cases of the event have been observed among the rural poor. Considering socio-economic factors none of the exposure variables are significantly establishing the occurrence of the event, except caste (p= 0.000 <0.05). Full antenatal care (ANC) indicates the mother receiving at least 3+ANC visits, 2 tetanus toxoids (TT) taken, and 100 IFA consumed at the time of pregnancy, months or trimester of pregnancy registration (p= 0.000 <0.05) seemed significant in explaining the event of LBW and VLBW. Within the personal and child factor, age at marriage (p = 0.012 <0.05), age at first birth (p = 0.06 <0.05), parity or birth order (p = 0.048 <0.05), and birth spacing (p = 0.000 <0.05) are found to be significant. BMI is one of the significant influential exposure (p= 0.035<0.05), as a considerable percentage (22.09%) of LBW have been observed among the underweight respondent (<18.5). Mothers with gestational diabetes, hypertension, convulsion without fever (p = 0.000 <0.05), edema (p= 0.027<0.05) etc. are significantly influencing the birth outcome. Male neonates are more at risk of LBW than the female. Northern and western parts of the country are experiencing more fatal cases of LBW than the rest part. Uttar Pradesh is in most alarming situation constituting more than 16%

of the LBW neonates of India, followed by Madhya Pradesh (8%), Rajasthan (6.81%), Bihar (6,74%), Maharashtra (5.23%).

3.2.CATPCA models

A total of 14 exposure variables have been selected for model 2 (prenatal and child factors) which has explained more than 77% of the total variance. Adequate antenatal care (-0.523) and Parity (0.772) are the high loading factors of dimension 1 and age at marriage (0.658), age at first birth (0.626) and respondent's present age (0.856) are the dominating the third dimension. Model 4 (maternal nutritional conditions) is the next dominating CATPCA model with 15 selected exposure variables after model 2, which is explaining 73.5 % of the total variance. From the CATPCA analysis, it is evident that socio-economic factors (model 1) are not at all to be considered while predicting risk factors associated with low birth weight in India.

3.3. ANN predictive model

The loading factors with $\pm\geq 0.5$ dimension score of each components of each CATPCA model have been considered as predictive exposures of the final ANN model. Therefore, total eight object score of each dimension (two dimensions from each model) of CATPCA were considered to construct the ANN model. The RMSE is 0.226 and 0.231 in the training and testing stage respectively. Hence, the model can be a good proctor of the subject in concern. From the variable importance plot it is evident the main predictive factors are- edema and convulsion without fever from CATPCA model 3 (dimensions 2) and marital age, under age first birth and current age of respondent from CATPCA model 2 (dimension 2) (Figure 1, Table 1).

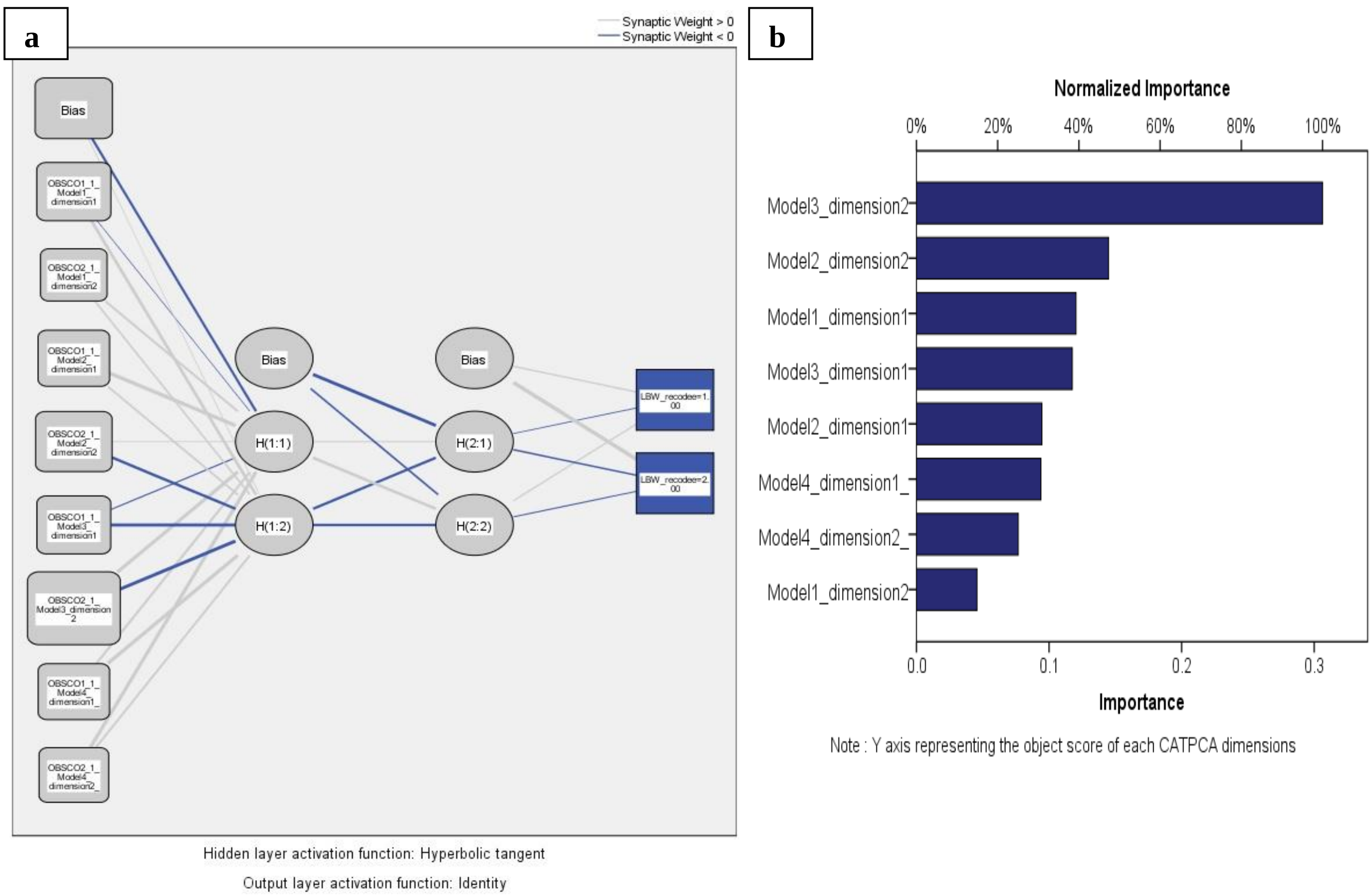

Figure 1: Structure of ANN model (a), Normalization Importance of Predictors (b)

Table 1: Loading Factors of 4 CATPCA models used in ANN model construction

Models	Component Loading (Loading factors were considered = ≥±0.5)			
	Dimension 1	**Loading Value**	**Dimension 2**	**Loading Value**
Model 1 (Socio-economic factors)	1. Types of place of residence	0.652	1. Religion	0.850
	2. Highest education level	-0.500	2. Caste	-0.640
	3. Wealth Index	-0.872		
	4. Availability of in-house toilet facility	0.667		
Model 2 (Prenatal and child factors)	1. Adequate antenatal care	-0.523	1. Age at marriage	0.658
	2. Parity	0.772	2. Age at first birth	0.626
			3. Respondent's present age	0.856
Model 3 (Maternal morbidity factors)	1. Currently has chronic respiratory disease	0.539	1. Currently has convulsion not from fever	0.799
	2. Currently has heart disease	0.642	2. Edema	0.800
	3. Currently has cancer	0.700		
	4. Currently has kidney problem			
Model 4 (Maternal nutritional factors)	1. Occasional consumption of eggs	0.907	1. Irregular consumption of pulses and beans	0.565
	2. Occasional consumption of fish	0.924	2. Irregular consumption of green leafy vegetables	0.533
	3. Occasional consumption of meat/chicken	0.940	3. Irregular consumption of fruits	0.629

Variable Principal Normalization.

References

Abidoye, L. K., Mahdi, F. M., Idris, M. O., Alabi, O. O., & Wahab, A. A. (2018). ANN-derived equation and ITS application in the prediction of dielectric properties of pure and impure CO2. In Journal of Cleaner Production (Vol. 175). https://doi.org/10.1016/j.jclepro.2017.12.013

Anil, K. C., Basel, P. L., & Singh, S. (2020). Low birth weight and its associated risk factors: Health facility-based case-control study. PLoS ONE, 15(6 June), 1–10. https://doi.org/10.1371/journal.pone.0234907

Aryadoust, V., & Goh, C. C. M. (2014). CaMLA Working Papers. CaMLA Working Papers, 02(May), 1–39. https://doi.org/10.1016/j.cegh.2023.101450

Chaudhuri, S., Kumar, Y., Nirupama, A. Y., & Agiwal, V. (2023). Examining the prevalence and patterns of malnutrition among children aged 0–3 in India: Comparative insights from NFHS-1 to NFHS-5. Clinical Epidemiology and Global Health, 24(October), 101450.

Girotra, S., Mohan, N., Malik, M., Roy, S., & Basu, S. (2023). Prevalence and Determinants of Low Birth Weight in India: Findings From a Nationally Representative Cross-Sectional Survey (2019-21). Cureus, 15(3), 11–13. https://doi.org/10.7759/cureus.36717

Girotra, S., Mohan, N., Malik, M., Roy, S., & Basu, S. (2023). Prevalence and Determinants of Low Birth Weight in India: Findings From a Nationally Representative Cross-Sectional Survey (2019-21). Cureus, 15(3), 11–13. https://doi.org/10.7759/cureus.36717

International Institute for Population Sciences (IIPS) and ICF. (2022). Compendium of Fact Sheets, National Family Health Survey 5. Compendium of Fact Sheets, National Family Health Survey. National Family Health Survey (NFHS-5) 2019-21, 1–116.

Jana, A. (2023). Correlates of low birth weight and preterm birth in India. PLoS ONE, 18(8 AUGUST), 1–17. https://doi.org/10.1371/journal.pone.0287919

Jiang, H., Guo, J., Li, J., Li, C., Du, W., Canavese, F., Baker, C., Ying, H., & Hua, J. (2023). Artificial Neural Network Modeling to Predict Neonatal Metabolic Bone Disease in the Prenatal and Postnatal Periods. JAMA Network Open, 6(1), e2251849. https://doi.org/10.1001/jamanetworkopen.2022.51849

Ledinger, D., Nußbaumer-Streit, B., & Gartlehner, G. (2024). WHO Recommendations for Care of the Preterm or Low-Birth-Weight Infant. In Gesundheitswesen (Vol. 86, Issue 4). https://doi.org/10.1055/a-2251-5686

Sharma, S., Maheshwari, S., & Mehra, S. (2021). Association between Maternal Dietary Diversity and Low Birth Weight in Central India: A Case-Control Study. Journal of Nutrition and Metabolism, 2021. https://doi.org/10.1155/2021/6667608

Tessema ZT, Tamirat KS, Teshale AB, Tesema GA: Prevalence of low birth weight and its associated factor at birth in Sub-Saharan Africa: A generalized linear mixed model. PLoS One. 2021, 16:e0248417. 10.1371/journal.pone.0248417

UNICEF and WHO. (2019). Low Birthweight Estimates Levels and Trends 2000-2015. The Lancet Global Health, 7(7), e849–e860.

Zaveri A, Paul P, Saha J, et al.: Maternal determinants of low birth weight among Indian children: Evidence from the National Family Health Survey-4, 2015-16. PLoS One. 2020; 15(12 December 2020): 1–15.