A Neural Network Model to study Non-Communicable Diseases in India

[1]Ms. Aakanksha Agarwala, [2]Dr. Manab Deka

[1]Department of Statistics, Gauhati University, Guwahati, Assam, India.
aakankshaagarwala@gmail.com
[2]Department of Statistics, Arya Vidyapeeth College, Guwahati, Assam, India.
manab.deka@avcollege.ac.in

Introduction

The rising epidemic of non-communicable diseases (NCDs) are the leading cause of death worldwide that affects the global health system, policies and socio-economic developments, thereby leading to increasing country level disparities. These diseases carry huge costs which traps people in poverty. The worst sufferers are the low and lower-middle-income countries who are hardly able to avail the services needed to diagnose and treat non-communicable diseases (NCDs). In 2016, NCDs were responsible for 41 million of the world's 57 million deaths (71%). The main NCD'S behind these deaths include cardio-vascular diseases, cancer, chronic respiratory diseases and diabetes. The risk factors being both behavioral (like harmful use of alcohol, physical inactivity, excessive salt/sodium intake and tobacco consumption) and metabolic risk factors (raised blood pressure, raised blood glucose, obesity). 75% of premature adult (aged 30-70) deaths were caused by NCD's which implies that NCD's do not target the older population alone. Thus, it's high time that the world should seek for prevention and treatment of these chronic diseases and thereby uplift the mental and physical well-being of the masses.

India, one of the largest economies in the world, also faces the NCD challenge. NCD's were estimated to account for 63% of total deaths in India in 2016. Despite substantial improvements in health indicators made in the past decade, the Indian healthcare system continues to contribute disproportionately to the global disease burden. The country is presently experiencing a phase of rapid health transition, wherein the mounting magnitude of NCDs is gaining prominence with substantial repercussion on health and economic productivity which implies that there is an urgent need to prioritize NCD control techniques.

Although researchers have attempted to study the significance of different socio-demographic characteristics, lifestyle behaviour and household status on different NCDs in India, however, technique such as neural network technique has yet not been used extensively

by researchers in India. On the above discussion on backdrop, the present study attempts to identify certain risk factors of non-communicable diseases (NCDs) on the basis of a nationally representative survey.

Data and Methodology

Present study is based on the nationally representative survey known as Indian Human Development Survey (IHDS-II), conducted during 2011-12 (IHDS-II). The IHDS is a nationally representative survey as it covers all states and union territories of India, with the exception of Andaman, Nicobar, and Lakshadweep islands. The survey covered 33 states and union territories, 42152 households, 384 districts, 1420 villages and 1042 urban blocks located in 276 towns and cities, across India. IHDS was jointly organized by researchers from the University of Maryland and the 145 National Council of Applied Economic Research (NCAER), New Delhi (Desai, Sonalde, et al., IHDS-II).

We propose to use Neural network technology to achieve our objectives. The term neural network applies to a loosely related family of models, characterized by a large parameter space and flexible structure, descending from studies of brain functioning. Specific definitions of neural networks are as varied as the fields in which they are used. While no single definition properly covers the entire family of models, let us consider the following description (Haykin, 1998):

A neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:

- Knowledge is acquired by the network through a learning process.
- Interneuron connection strengths known as synaptic weights are used to store the knowledge.

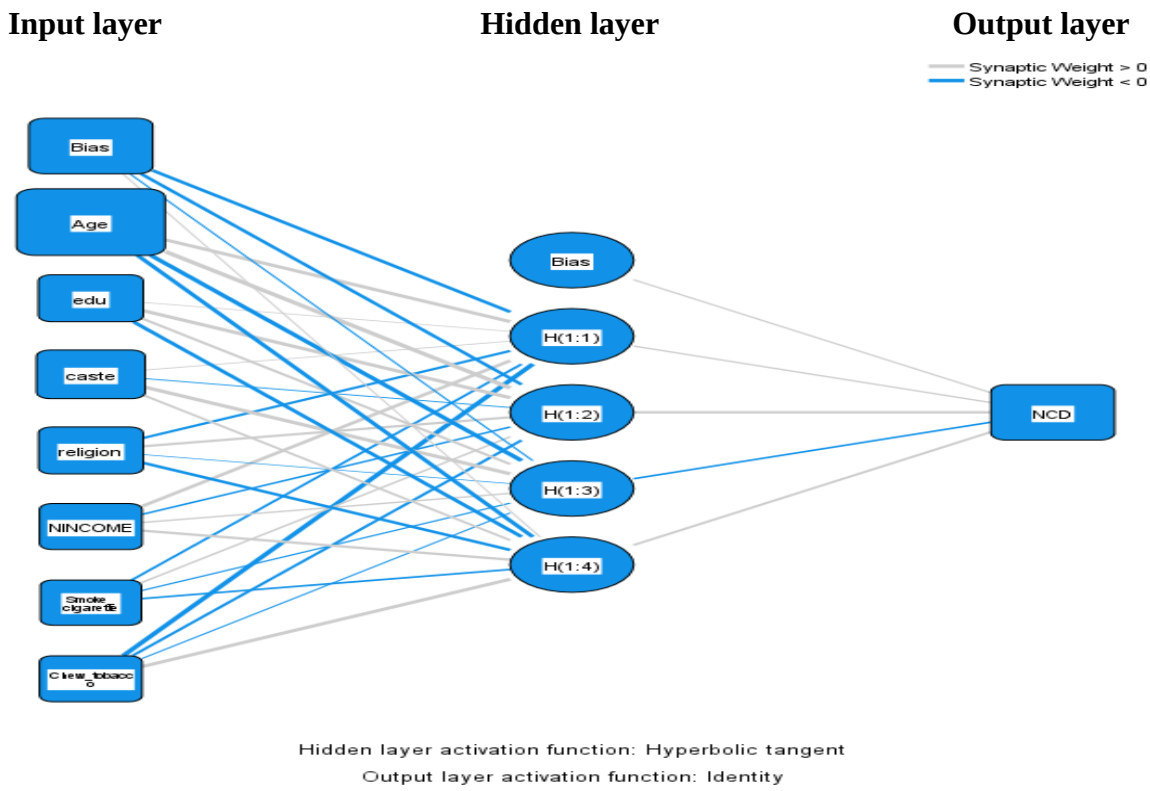
Neural networks are the preferred tool for many predictive data mining applications because of their power, flexibility, and ease of use. Predictive neural networks are particularly useful in applications where the underlying process is complex. Neural networks used in predictive applications, such as the multilayer perceptron (MLP), are supervised in the sense that the model-predicted results can be compared against known values of the target variables. The Multilayer Perceptron (MLP) procedure produces a predictive model for one or more dependent (target) variables based on the values of the predictor variables.

As IHDS II is a huge dataset and containing about 100 variables, we wanted to use only those covariates for predicting NCD, which have some significant impact on NCD. For this purpose, we used logistic regression technique to identify the covariates which significantly affect the dependent variable NCD. The results suggested that socio-demographic factors like age, level of education, caste, religion, income quantiles, smoking cigarettes and chewing tobacco/gutkha have significant impact on NCDs. As such, we have used these variables as covariates in neural network set up.

Moreover, the Independent variable importance analysis, of neural network technique, performs a sensitivity analysis, which computes the importance of each predictor in determining the neural network. The analysis is based on the combined training and testing samples or only on the training sample if there is no testing sample. This creates a table displaying importance for each predictor.

Results and Discussion

The analysis was carried out on 204569 respondents across India, of which about 3.3% were afflicted by one or other non-communicable diseases. We present below the network diagram comprising of Input layer, Hidden layer and Output layer obtained through neural network analysis.



This structure is known as a feedforward architecture because the connections in the network flow forward from the input layer to the output layer without any feedback loops. In this figure:

- The input layer contains the predictors.
- The hidden layer contains unobservable nodes, or units. The value of each hidden unit is some function of the predictors; the exact form of the function depends in part upon the network type and in part upon user-controllable specifications.
- The output layer contains the responses. Since the dependent variable NCD is a categorical variable with two categories, it is recoded as two indicator variables. Each output unit is some function of the hidden units. Again, the exact form of the function depends in part on the network type and in part on user-controllable specifications.

Moreover, the Independent variable importance analysis which computes the importance of each predictor in determining the neural network, has identified age as the most important covariate with weight 0.649 (normalized importance 100%), followed by caste, religion, income etc.

Conclusion

The study has identified certain socio-demographic characteristics, lifestyle behaviour, household status and contextual characteristics such as age, education level, caste, religion, income, smoking habit and chewing of tobacco as having significant effect on NCDs. Moreover, the Independent variable importance analysis has identified age as the most important covariate among all covariates. Our present study is an attempt to fill the void that exits when it comes to studying about the prevalence of NCDs with help of sophisticated technique like neural network.