

# Using an artificial neural network to identify the population at risk of type 2 diabetes mellitus in the United States

Yue Fei, Bernard Man Yung Cheung  
Department of Medicine, University of Hong Kong, Hong Kong



**HKU  
Med**

LKS Faculty of Medicine  
Department of Medicine  
香港大學內科學系

## Introduction

- Prevention of type 2 diabetes mellitus (T2DM) is now believed to be possible.
- Artificial neural network (ANN) has emerged as a useful tool for disease prediction and classification.
- We aimed to construct an ANN to classify the risk of T2DM in a general population and to identify novel predictors of T2DM in order to help to prevent its development.

## Methods

- The demographic, clinical, and questionnaire data on the participants aged  $\geq 20$  years in the US National Health and Nutrition Examination Survey 2007-2010 were used to develop a multilayer perceptron ANN model in R.
- The diagnosis of T2DM was based on self-reported history.
- The analysis excluded pregnant women and participants diagnosed with type 1 diabetes mellitus.
- We randomly selected the same number of people without T2DM as those with T2DM as controls to build the ANN model.

## Results

- In the final dataset, there were 576 T2DM and 576 control participants for model establishment.
- A three-layer multilayer perceptron ANN model composed of an input layer of 28 variables, a hidden layer with 3 neurons, and an output layer representing the probability of T2DM, was finally established (Figure 1).
- Our ANN model achieved a specificity of 0.744, a sensitivity of 0.724, an accuracy of 0.734 (95% CI 0.679-0.784), and an area under the curve of 0.733 in T2DM classification (Table 1).
- In decreasing order of importance, waist circumference, age, race, LDL level, diet quality, close relatives having diabetes mellitus, education level, calcium level, marital status, and hypertension were identified as the 10 leading predictors of T2DM in the model (Figure 2).

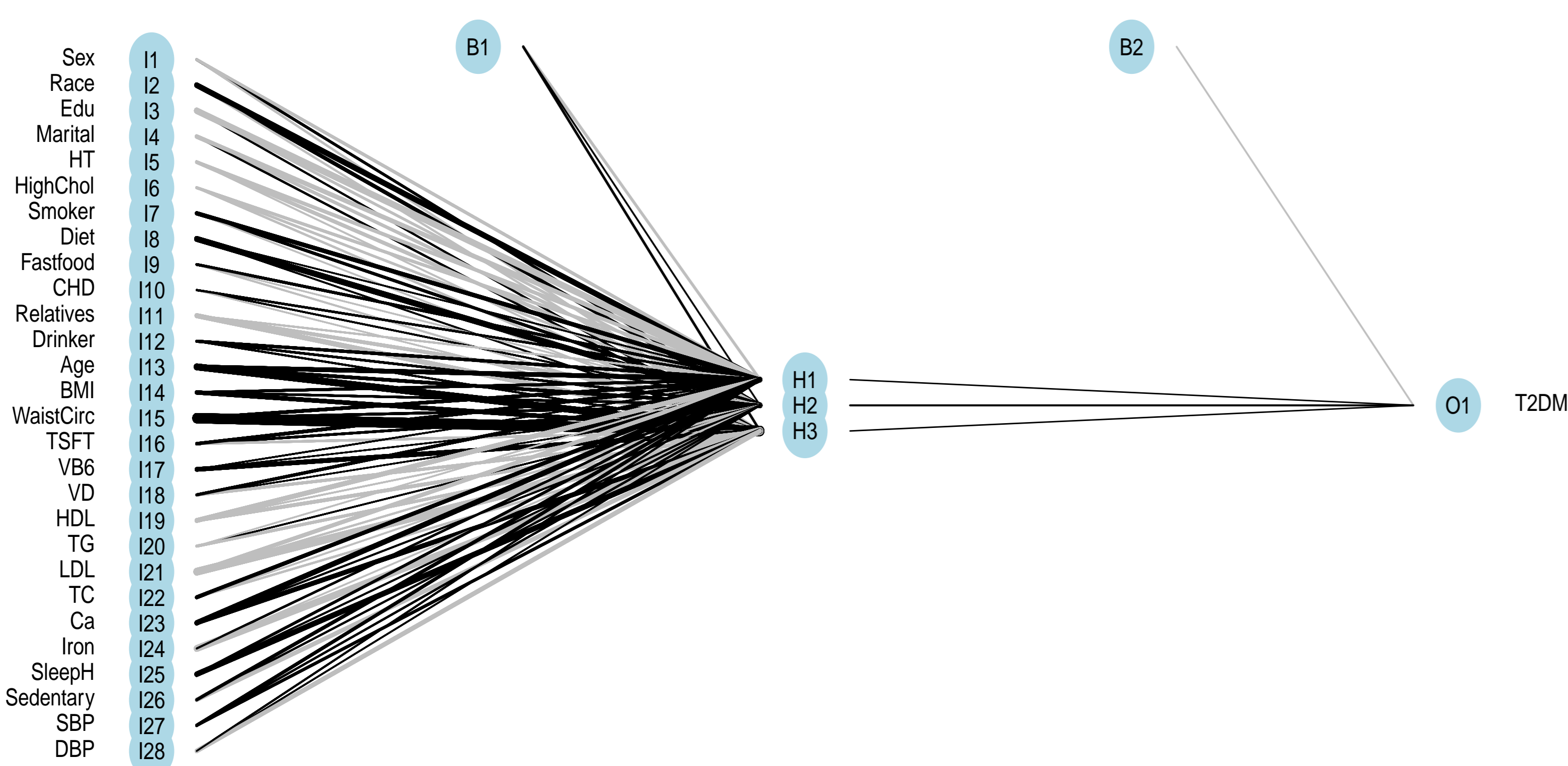


Figure 1. Architecture of the three-layer perceptron ANN

## Conclusion

- ANN is a novel method of identifying the risk of T2DM in the general population.
- The lead predictors associated with T2DM identified from our study can assist the medical experts to reach a better understanding of the patients with suspicious diabetes mellitus.
- The availability of big data allows rapid model construction for different populations that can help to improve prevention and intervention programs.

Table 1. Performance of the ANN model in T2DM classification

Parameters	Specificity	Sensitivity	Positive predictive value	Negative predictive value	Accuracy (95% CI)	Kappa	F1	AUC
Performance	0.744	0.724	0.708	0.759	0.734 (0.679-0.784)	0.461	0.741	0.733

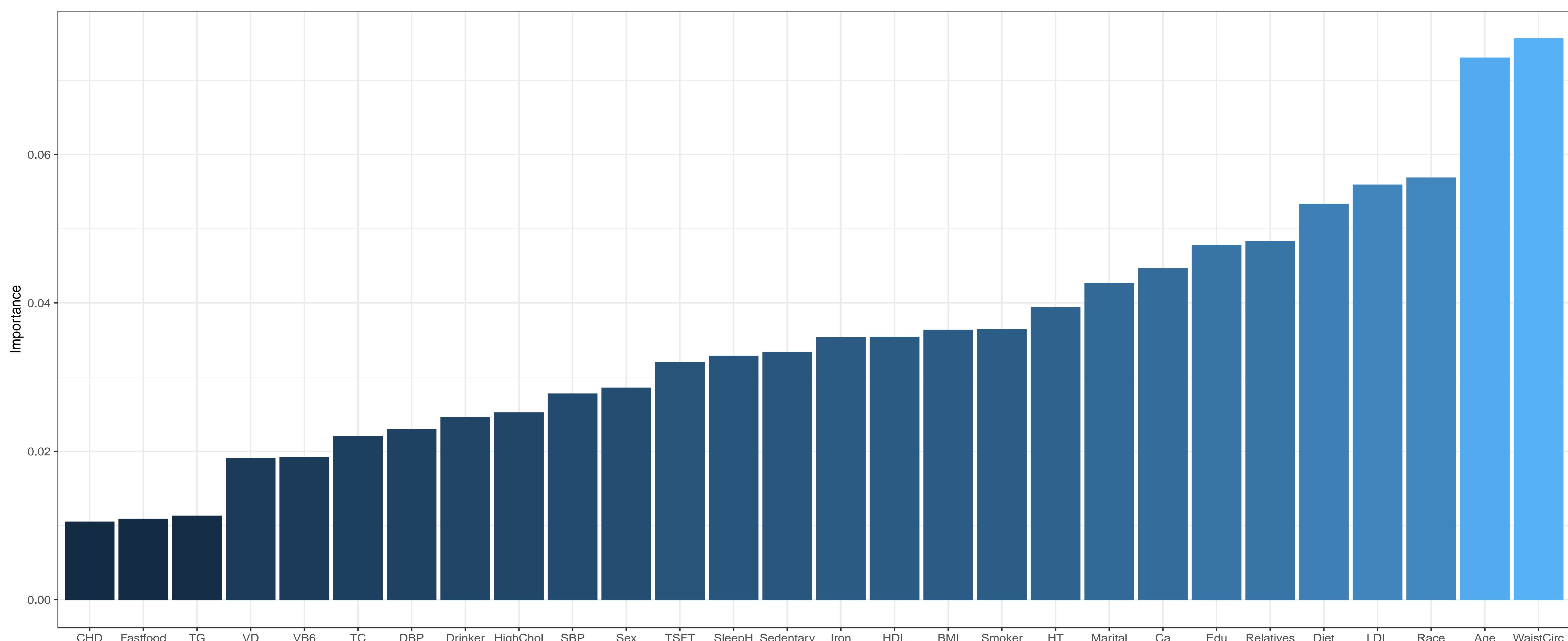


Figure 2. Variable importance on T2DM risk from ANN using Garson's algorithm

## References

1. International Diabetes Federation. 2017.
2. Cheung BMY, et al. *Am J Med*. 2009;122(5):443-53.
3. Contreras I, et al. *J Med Internet Res*. 2018;20(5):e10775.
4. Gulumbe SU, et al. *Machine Learning Research*. 2019;4(2):33-38.
5. Pradhan M, et al. *Int J Comput Sci Emerg Technol*. 2011;2(2):303-311.
6. Krogh A. *Nat Biotechnol*. 2008;26(2):195-7.