Medical Technologies Journal, Volume: 1, Issue: 4, October-December 2017, Pages:76-136. Doi : https://doi.org/10.26415/2572-004X-vol1iss4

Comparing three data mining algorithms for identifying associated risk factors of Type 2 Diabetes

Type of article: conference abstract

Maryam Tayefi1,3, Habibollah Esmaeily2, Majid Ghayour-Mobarhan1,3, Ali Reza Amirabadizadeh4*

1) Department of Modern Sciences and Technologies, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.

2) Department of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran.

3) Biochemistry of Nutrition Research Center, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.

4) Medical Toxicology and Drug Abuse Research Center (MTDRC), Birjand University of Medical Sciences

* amirabadiza921@gmail.com

Abstract

Introduction: Type 2 diabetes (T2DM) shows increasing prevalence and global health burden, causing a concern among health service providers and health administrators.. The current study is aimed at developing and comparing some statistical models that are useful in measuring or establishing such associations. The three particular statistical methods investigated in this study are artificial neural network (ANN), support vector machines (SVM) and multivariate logistic regression (MLR) using demographic, anthropometric and biochemical characteristics on a sample of 9528 individuals from Mashhad city.

Methods: The statistical methods involved in this study are also known as machine learning algorithms and require dividing the available data in to training and testing dataset. This study has randomly selected 70% cases (6654 cases) for training and reserved the remaining 30% (2874 cases) for testing. The three methods are compared with help of the receiver operating characteristic (ROC) curve.

Results: The prevalence rate of T2DM is 14% in our population. The ANN model has 78.7% , accuracy, 63.1% sensitivity and 81.2% specificity. Values of these three parameters are 76.8%, 64.5% and 78.9% respectively for SVM and 77.7%, 60.1% and 80.5%, respectively for MLR. The area under the ROC curve (AUC) is 0.71 for ANN, in SVM model was 0.73 for SVM, and 0.70 for MLR.

Conclusion: The overall conclusion is that ANN performs better than two models and can be used effectively to identify associated risk factors of T2DM.

Keywords: Artificial neural network, Support vector machine, Logistic regression method, Type 2 diabetes

1. Declaration of conflicts

This abstract is selected from the First International Congress of Diseases and Health Outcomes Registry and First National Congress of Medical Informatics, 14-17 February 2017, Mashhad, Iran

2. Authors' biography

No biography.

3. References

No references.