

Applying decision tree for detection of a low risk population for type 2 diabetes: A population based study

Type of article: conference abstract

Habibollah Esmaeily¹, Maryam Tayefi^{2,3}, Hassan Doosti¹, Majid Ghayour-Mobarhan^{2,3}, Ali Reza Amirabadizadeh^{4,*}

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

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

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

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

* Email: amirabadiza921@gmail.com.

Abstract

Introduction: The aim of current study was to create a prediction model using data mining approach, decision tree technique, to identify low risk individuals for incidence of Type 2 diabetes (T2DM), using the Mashhad Stroke and Heart Atherosclerotic Disorders (MASHAD) Study program.

Methods: a prediction model was developed using classification by the decision tree method on 9528 subjects recruited from MASHAD database. Moreover, the receiver operating characteristic (ROC) curve was applied.

Results: The prevalence rate of T2DM was ~14% in our population. For decision tree model, the accuracy, sensitivity, and specificity value for identifying the related factors with T2DM were 78.7%, 47.8% and 83%, respectively. In addition, the area under the ROC curve (AUC) value for recognizing the risk factors associated with T2DM was 0.64. Moreover, we found that subjects with family history of T2DM, age \geq 48, SBP \geq 130, DBP \geq 81, HDL \geq 29, LDL \geq 148 and occupation=other have more than 59% chance of this disorder, while the chance of T2DM in subjects without history with TG \geq 184, age \geq 48 and hs-CRP \geq 2.2, have approximately 51% chance.

Conclusion: Our findings demonstrated that decision tree analysis, using routine demographic, clinical, anthropometric and biochemical measurements, which combined with other risk score models, could create a simple strategy to predict individuals at low risk for type 2 diabetes in order to decrease substantially the number of subjects needing for screening and recognition of subject at high risk.

Keywords: Data mining, Decision tree, 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.