

**APPLYING DECISION TREE FOR DETECTION OF A RISK FACTORS FOR TYPE 2 DIABETES: A POPULATION BASED STUDY**Maryam Tayefi^{1,3}, Habibollah Esmaeily², Majid Ghayour-Mobarhan^{1,3}, Ali Reza Amirabadi zadeh⁴

1: Biochemistry of Nutrition Research Center, 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: Department of Modern Sciences and Technologies, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.

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

Correspondence:

Ali Reza Amirabadi Zadeh, Tel:+989105090341, E-mail: amirabadiza921@gmail.com

TYPE OF ARTICLE: CONFERENCE ABSTRACT

ABSTRACT

Introduction: The aim of current study was to create a prediction model using a data mining approach and 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 the MASHAD database. Moreover, the receiver operating characteristic (ROC) curve was applied.

Results: The prevalence rate of T2DM was 14% in our population. For the decision tree model, the accuracy, sensitivity, and specificity value for identifying the related factors with T2DM were 78.7%, 61.2%, and 83%, respectively. In addition, the area under the ROC curve (AUC) value for recognizing the risk factors associated with T2DM was 68%. The identified variables included family history of diabetes, triglycerides, systolic blood pressure, body mass index, hs-crp, education.

Conclusion: Our findings demonstrated that decision tree analysis, using routine demographic, clinical, and 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 substantially decrease the number of subjects needed for screening and recognition of subjects at high risk.

KEYWORDS: Data mining, Decision tree, Type 2 diabetes