Original Article

Adherence to Insulin Treatment in Participants with Type 2 Diabetes: Comparison of Logistic Regression and Conditional Tree and Forests to Determine the Effective Factors

Alireza Mirahmadizadeh[1], MD, PhD; Sadaf Sahraian[2], MD; Hamed Delam[3], MSc; Mozhgan Seif[4], PhD

[1]Non-Communicable Diseases Research Center, Department of Epidemiology, School of Health, Shiraz University of Medical Sciences, Shiraz, Iran;
[2]Division of Neuroradiology, The Russell H. Morgan Department of Radiological Sciences, Baltimore, MD, USA;
[3]Student Research Committee, Larestan University of Medical Sciences, Larestan, Iran;
[4]Department of Epidemiology, School of Health, Shiraz University of Medical Sciences, Shiraz, Iran

Correspondence:
Mozhgan Seif, PhD;
Department of Epidemiology, School of Health, Razi St., P. O. Box: 71536-75541, Shiraz, Iran
Tel: +98 9374322827
Email: m_seif@sums.ac.ir
Received: 12 July 2019
Revised: 20 August 2019
Accepted: 19 September 2019

Abstract

Background: Type 2 diabetes is the most prevalent chronic disease in the world. Timely and appropriate control can significantly reduce the burdens and costs of this disease. Although insulin injection is the most efficient method to control type 2 diabetes, patients avoid this method for unknown reasons. The main aim of the present study is to determine the factors influential in non-adherence to insulin using tools and models that have not been applied in this field so far.

Methods: The tendency to insulin injection in 457 patients with type 2 diabetes was investigated in this cross-sectional study using the classic logistic regression and new learning algorithms, including conditional tree, conditional forest, and random forest. Different fits were compared so that the best model can be determined to identify the factors in non-adherence to insulin.

Results: Although random forest had the highest accuracy among the fitted models, all the methods had a relative consensus that having life insurance, academic education, and insulin injection experience in immediate family members increase the tendency to accept insulin therapy. Our results also showed that younger patients and those who were committed to a specific diet better approved insulin therapy.

Conclusion: The reasons for non-adherence to insulin can be summarized in economic and psychological aspects. Therefore, the health system policies are recommended to address economic issues and also raise public awareness about this treatment method.

Please cite this article as: Mirahmadizadeh AR, Sahraian S, Delam H, Seif M. Adherence to Insulin Treatment in Participants with Type 2 Diabetes: Comparison of Logistic Regression and Conditional Tree and Forests to Determine the Effective Factors. J Health Sci Surveillance Sys. 2019;7(4):171-177.

Keywords: Type 2 diabetes, Insulin, Decision trees, Forests

Introduction

Diabetes mellitus is a metabolic disorder due to the lack of insulin production or due to the lack of response to the produced insulin, classified as Type 1 diabetes and Type 2 diabetes, respectively.[1, 2] The prevalence of diabetes is ascending at an alarming rate, and is known as one of the costliest chronic diseases worldwide.[3-5] In Iran, the prevalence of Type 2 diabetes, as the most common type of diabetes,[6] has reached over 16% in some cities.[3] However, early detection and treatment of type 2 diabetes can reduce 80% of its side effects.[4] Although controlling method is different for different people,[7] insulin injection is considered a great transformation[8] such that some studies have introduced this treatment as the best available one for controlling diabetes.[9] Despite the higher efficiency and effectiveness of this treatment, especially when began timely or early,[10] insulin injection is prescribed as the last treatment recommended due to patients' non-adherence to insulin.[9, 11]

Many studies have investigated the patients' unwillingness to insulin injection.[12] In a review article in 2004, it was stated that only one-third of the young patients accepted insulin injections.[5] In another study on 502 diabetic patients in the United States, it was reported that more than half of them have refrained from insulin injection.[13] In contrast, this rate was said to be 60% in the last update around the southern areas of Iran.[14] In another study on type 1 diabetic patients, it was shown that the avoidance depended on the age; so that by increasing it, the patient's willingness to inject this drug increases. It was the only study in which the decision tree was used to identify the influential factors in non-adherence to insulin.[15]

A decision tree is one of the machine learning algorithms that has been highly practical in the field of diabetes. Using a conditional tree, Huang predicted nephropathy based on the genetic and clinical characteristics of type 2 diabetic patients.[16] Using the information of the patients visiting seventeen treatment clinics, Leung fitted three different types of conditional trees and random forests to predict the diabetic kidney side effects.[17] Tapak, who compared the classic statistical models such as logistic regression and new learning techniques such as random forest, states that learning algorithms have predicted diabetes with higher accuracy. Accordingly, she recommends applying these algorithms to predict other diseases.[18] It is noteworthy that, in the field of other conditions rather than diabetes, different trees have provided more accurate predictions in comparison to classical models such as survival ones.[19, 20] All these studies highlight that there is a need to compare methods to identify the best fitting, as more fitness leads to more accurate identification of affective factors.

Despite the endless number of studies on diabetes, unfortunately, the rate of non-adherence to insulin is still high, and this has led to irreversible consequences on the patients' life and an excess burden on the healthcare system. While according to the records, no studies have used learning algorithms, including conditional tree, conditional forest, and random forest so far to determine factors of non-adherence to insulin in type 2 diabetic patients. Therefore, to identify the factors and causes of non-adherence to insulin, the learning algorithms, and logistic regression was fitted and compared so that by determining the most accurate fitness, the most practical design can be prepared in line with insulin adherence.

Materials and Methods

Study Population

The present cross-sectional study is based on the records of 457 type 2 diabetic patients that were collected using convenience sampling among the visitors to 12 clinical centers for diabetes care in Shiraz, located in the south of Iran during Jan-July 2017. The patient inclusion criteria were as follows:

1) The patient should be at least 30 years old.

2) The patient's HbAlc level should be at least 7.5% (58 mmol/mol).

3) Safe drugs for controlling diabetes should be used at the maximum dose.

4) No insulin should have been used before; however, the doctor should have prescribed it to the patient.

5) At least one of the diabetes side effects (including any kidney and cardiovascular diseases and diabetic foot ulcer) should have been experienced.

Experience of at least one side effect would cause more certainty that insulin is a vital treatment for patients. Pregnancy and/or having severe mental illnesses and the patient's lack of consent to participate in this study were the exclusion criteria. It must be stated that the Ethics Committee approved this study of Shiraz University of Medical Sciences (code: IR.SUMS.REC.1395.S1084).

Measured Variables

A total number of 36 independent variables including the demographic and clinical records of the patients were measured. Some of them are as follow: age, gender, marital status, insurance coverage, education (including the illiterate, elementary school, secondary school and academic levels), family background of insulin injection, the period the patient is suggested to receive insulin (years), background of visiting a nutrition consultant and the nutrition status (including two classes of regular diet and diabetic diet).

Statistical Analysis

This study applied backward stepwise logistic regression in addition to learning algorithms, including conditional tree, conditional forest, and random forest.

The conditional tree is the most evolved and unbiased type of decision tree. This algorithm split the patients' population according to practical factors and consequently forms homogeneous subsets of patients. It is worth considering that this type of tree provides Bonferroni adjustment along with the automatic selection of influential variables.

An ensemble of trees forms a forest, which reduces the variance of parameter estimation, resulting from averaging over all trees.[21] There are two common types of forests; conditional and random. The conditional forests consist of conditional trees. This forest reduces the variable selection bias in the way that a stepwise algorithm is applied to form homogeneous subsets.

At the first step, the most relevant variable would be detected and then at the second step the best cut-point would be selected along the detected variable. Whereas, random forest selects the split variable and its cut-point simultaneously, with a greedy search on the all possible variables and all cut-points.[22]

Similar to all learning algorithms, trees and forests are susceptible to overfitting. Therefore, 10-fold cross-validation was used to prevent overestimation in reporting the accuracy of algorithms.[21] The goodness of fit indices for comparing models via cross-validation included: sensitivity, specificity, accuracy, the area under ROC, and Brier score. R software was used for analysis, and the significance level was set at 0.05.

Results

Of all 457 patients with type 2 diabetes who were prescribed to receive subcutaneous insulin infusion only 182 (38.3%) accepted insulin therapy. The median of patients' age was 54 years, and the mean±SD duration of diagnosed diabetes type 2 was 3.4±2.5 years. The t-test revealed that the younger the patients, the more acceptance of insulin therapy (P=0.014).

Overall, 23.63% of the patients held insurance, 21.01% had a family history of insulin usage and 16.85% of college education. According to the results of chi-square tests, the association between having mentioned factors and insulin therapy acceptance was statistically significant (all P<0.001).

The diabetic diet was followed by 20% of the patients, while 38.51% of all patients experienced medical consultation with nutritionists. However, the chi-square tests revealed that patients who consulted with nutritionists and who observed the diabetic diet were more willing to accept insulin therapy (P<0.001).

Figure 1 depicts the conditional tree grown with all observations. The tree displays that having insurance, family history of insulin infusion and age were respectively the three most affecting factors to form homogenous classes of patients regarding their insulin acceptance.

The normalized variable importance is displayed in Figure 2, estimated by conditional 2(a) and random 2(b) forests, respectively. Generally, both forests identified almost the same variables. Both specified that having insurance and a higher level of education as the most important factors; so that 61% and 53% of insulin acceptance could be predicted by these two factors, respectively, using conditional and random forests. Other affecting factors were having a family history of insulin usage, adherence to a diabetic diet, consultation with nutritionists, duration of insulin suggestion by physicians, and age, in descending order.

Table 1 displays the results of the stepwise logistic regression to identify factors that significantly affect insulin acceptance. Accordingly, the family history of insulin usage and having insurance were the greatest determinants; each of them increased the odds of acceptance by almost four times. The results also showed that diabetic patients who adhere to the diabetic diet would accept insulin therapy 2.5 times more than others. Furthermore, an increase in academic grade would increment this chance by 34%. However, every one-year physician advisement for accepting insulin would cause a 19% increase in the odds of acceptance.

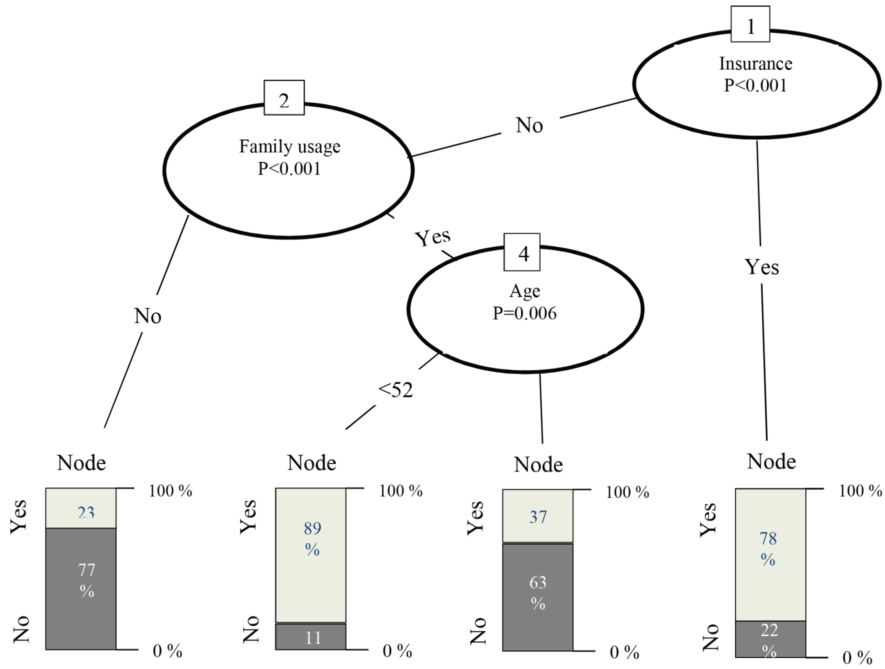

Figure 1: Conditional Inference Tree for classifying patients with type 2 diabetes to produce homogeneous subsets of Insulin Compliance.

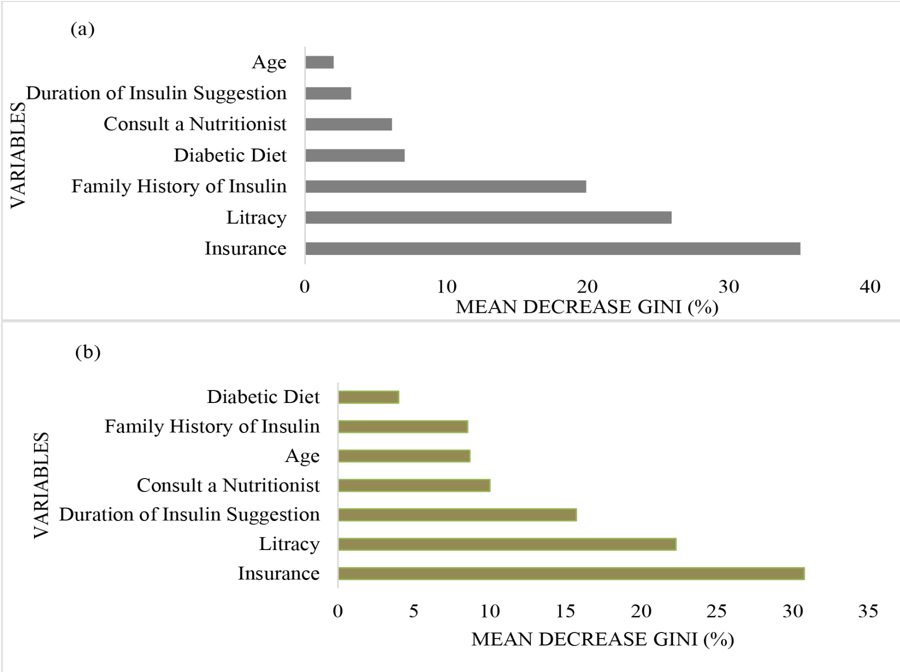

Figure 2: Normalized Variable Importance to Predict Insulin Compliance in patients with type 2 diabetes; respectively by Conditional (a) and Random (b) Forests

Table 1: Results of the logistic regression model for insulin acceptance affecting factors

	Coefficient	S.E.	OR†	95% CI for OR	
				Lower	Upper
Family History of Insulin Usage	1.393 **	0.32	4.02	2.18	7.59
Insurance	1.361 **	0.33	3.90	2.06	7.38
Diet for Diabetics	0.933 *	0.31	2.54	1.39	4.67
Literacy	0.300 **	0.08	1.34	1.15	1.59
Duration of insulin suggestion	0.178 *	0.07	1.19	1.05	1.38

*Significant at 1%; **Significant at 0.1%; †Odds Ratio

Table 2: Results of assessing the goodness of fit for different models in predicting the rate of insulin admission in patients with type 2 diabetes

	Sensitivity	Specificity	Accuracy	AUC†	Brier Score
Conditional tree	0.562	0.848	0.726	70.887	0.205
Conditional forests	0.491	0.901	0.723	82.129	0.177
Random forests	0.650	0.870	0.773	86.151	0.154
Logistic regression	0.561	0.843	0.721	78.254	0.190

†Area Under Curve

Table 2 shows the comparison of the models' goodness of fit according to different indices. For all models, the prediction of insulin acceptance was more valid than random guessing as the "Area Under Curve" was much more than 0.5. Conditional and random forests had the highest specificity; precisely, these two algorithms correctly recognized the patients avoiding insulin therapy. However, the random forest was the most sensitive model to identify patients accepting insulin therapy; therefore, this algorithm was the most accurate one. This forest also had the highest and lowest values, respectively, for the AUC and Brier score. Table 2 also shows that the conditioned forest had the second goodness of fits among others. The lower sensitivity of this ensemble algorithm is the consequence of its higher specificity. Both the sensitivity and specificity were almost the same for conditional tree and logistic regression; however, the AUC and Brier Score indicated the superiority of the logistic regression in comparison to the conditional tree.

Discussion

Different models were used in the present study to investigate the reasons for non-adherence to insulin among eligible type 2 diabetic patients. The applied models include a conditional tree, conditional forest, random forest, and backward stepwise logistic regression. Among the fitted models, although random forest has provided the best fitness, the overlap in the identified factors by different methods indicated that the results of the models were confirmed by each other. Research results suggest that although doctors' prescription is an important factor in insulin adherence, all the models have recognized insurance and patients' education more effective in insulin adherence. Other factors related to insulin adherence include: history of insulin injection with family members, age, time period since the first prescription, following a suitable diet and records of visiting nutrition consultants.

In line with previous studies, the conditional tree developed in this study shows that patients with insurance will more probably accept insulin treatment.[23, 24] Of the patients without insurance, those with a background of insulin injection in their family, and the younger patients were more willing to take insulin. It might be claimed that the insulin therapy in immediate family members has reduced the fear of insulin injection and its side effects. Moreover, patients with this family history, have clearly observed the effectiveness of insulin therapy for their relatives. Therefore, believe more in this treatment. There are also previous studies which emphasized to create positive family interactions and cooperation regarding the insulin injection and adherence.[25]

As was mentioned earlier in our study, the conditional tree selected the age variable as one of the influential factors in insulin adherence. The estimated ratios in leaves show that young patients are more likely to risk the change of edible drugs by insulin injection. In agreement with the obtained results, other studies also show that by an increase in the age of the patients, their ability to follow treatments decreases.[26] The other point to be noted regarding the conditional tree is that the cut point on the branch (i.e., age variable) is found in such a way that the relationship between age and insulin adherence has revealed. The routine and classic method to determine the cut point to classify a variable is to use the mean/median of that variable.[27] If the conditional tree used the median of age to classify the patients, the significant relationship between age and insulin adherence was not revealed (P=0.17). Moreover, this discovered cut point (i.e., fifty years old) was already introduced by clinical experts as an ideal cut point to divide the type 2 diabetic patients.[28, 29] Age was also recognized as an influential variable through the only previous study in which used a conditional tree to identify factors of insulin resistance among type 1 diabetic patients.[15]

All the identified variables in the conditional tree and logistic regression are subsets of variables that have been identified by both forests as essential variables. Forests and logistic regression showed that the prescription and physicians' emphasis on insulin therapy has gradually persuaded the patients to receive the injection. Although, this acceptance is more probable among people who are willing to visit nutrition consultants and following specific diets. Previous studies have also shown that patients usually resist treatments at the beginning of the diagnosis;[30] however, as time passes and disease side effects appear, they finally believe the disease and undertake the treatment.[31] Furthermore, the obtained results consistent with other studies conducted on type 2 diabetic patients indicate that the patients' gender is not significantly valid on their willingness to take the treatment.[5, 32] It is noteworthy that similar to the obtained results from the same data, which was published by Mirahmadizadeh,[14] the variables including age, education, insurance, and following a proper diet have a significant relationship with the willingness to insulin adherence.

According to the estimated importance in both forests, patients' education would affect their insulin adherence. It is noteworthy that this effect was already proved in the study conducted on type 1 diabetic patients.[5] Furthermore, various studies have shown that educational programs lead to more comprehensive control over diabetes[33] and reduce the risk of catching hypoglycemia.[34] Although both forest algorithms have recognized education as an important variable, the conditional tree algorithm has not identified it significant, and in logistic regression, the other estimated odds ratios are more significant than the odds of education. However, it should be emphasised that the estimated Kappa coefficient between education (at academic and non-academic levels) and insurance variables was 0.52, which indicates a relatively good correlation.[35] Therefore, the underestimated education coefficient in logistic regression and not selecting this variable in tree growing might be related to this correlation. The correlation between the independent variables usually has not such a significant effect on the modeling and accuracy of forest predictions.[36] However previous studies have shown that the correlation between independent variables in logistic regression leads to inconsistency in the regression coefficients. Correlation also leads to overestimation of the variances and consequently the concealment of the significant relationships.[37]

Moreover, when a tree grows, initial splits would lead to the homogeneity of observations in each leaf, and this prevents the other correlated and similar variables from splitting. Many studies have addressed the bias and inconsistency of the tree estimates and its higher variance in comparison to forests.[21] In agreement with these results, it is estimated that in the study of Maroco (2011), which compared the learning machine algorithms including random forest and classic models including logistic regression, the accuracy of the random forest was significantly higher than the conditional tree and logistic regression.

No significant difference was observed between the conditional tree and logistic regression.[38] Apart from the study above, regarding diabetes, Leong showed that the prediction obtained from forests is more accurate than the conditional tree.[17] While in another study comparing these two algorithms, it was shown that although the prediction power of forests is greater than trees, the difference is not big enough to significantly recognize the fitness of forests much superior to trees.[39]

Although in this study random forests were the best algorithms, but other researches have concluded that introducing only one algorithm as the best one for all conditions and datasets is not possible.[38] Therefore the limitation of our study should be regarded that only four models were compared to determine the most appropriate methods to model the non-adherence to insulin therapy. As another limitation, it should be noted that, similar to all studies involving samples collected from medical centers, our study only included information from patients referring to these centers. Therefore the sample was biased in the way that it does not include patients who do not visit clinical centers.

Conclusion

Although random forest was recognized as the best-fitted algorithm, other methods did not also inferior to this method. Almost all the methods identified the same affective variables in the way that the reason for non-adherence to insulin could be summarized in two components:

1) The economic aspect, which is revealed by the insurance variable.

2) The psychological aspect, which is revealed by education and history of insulin injection in the family.

Identification of the insurance factor as the most effective variable shows that the most efficient approach in increasing the willingness to insulin adherence is to expand the insurance coverage and eliminate the economic problems in applying it. Moreover, less willingness to insulin adherence in patients with no backgrounds of insulin injection in the family indicated a social fear of insulin injection. This fear is even more intense in patients without academic education. Hence, educational programs to provide sufficient knowledge about diabetes and the insulin injection can be effective more than physicians' recommendations to adhere to insulin.

Acknowledgment

We appreciate Shiraz University of Medical Sciences, which funded this investigation as a part of Hamed Delam's dissertation (Code: 95/12448).

Conflict of Interest: None declared.

References

1 SOROUSH M, ABBASZADEH M, SOROOSH S. The Study of Bone Density in Type 2 Diabetic Patients and Comparison with Non-Diabetic Patients Referred to Two Hospitals in Tehran, Iran. *Biomed & Pharmacol J*. 2014; 7(2):585-9.

2 Deepika N, Poonkuzhali S. Design of Hybrid Classifier for Prediction of Diabetes through Feature Relevance Analysis. Int J Innov Res Sci Eng Technol. 2015; 2(10):788-93.

3 Larejani B, Zahedi F. Epidemiology of diabetes mellitus in Iran. J Diabetes Metab Disord. 2001; 1(1):1-8.

4 Mohamed EI, Linder R, Perriello G, Di ND, Pöppl S, De AL. Predicting Type 2 diabetes using an electronic nose-based artificial neural network analysis. Diabetes Nutr Metab. 2002; 15(4):215-21.

5 Farsaei S, Radfar M, Heydari Z, Abbasi F, Qorbani M. Insulin adherence in patients with diabetes: risk factors for injection omission. Prim Care Diabetes. 2014; 8(4):338-45.

6 Organization WH. Screening for type 2 diabetes: report of a World Health Organization and International Diabetes Federation meeting. 2003

7 Aljumah AA, Ahamad MG, Siddiqui MK. Application of data mining: Diabetes health care in young and old patients. J King Saud Univ Comp & Info Sci. 2013; 25(2):127-36.

8 Dall TM, Zhang Y, Chen YJ, Quick WW, Yang WG, Fogli J. The economic burden of diabetes. 2010; Health Aff, 29(2):297-303.

9 Meece J. Dispelling myths and removing barriers about insulin in type 2 diabetes. Diabetes Educ. 2006; 32(1):9-18.

10 Asche C, Bode B, Busk A, Nair S. The economic and clinical benefits of adequate insulin initiation and intensification in people with type 2 diabetes mellitus. Diabetes Obes Metab. 2012; 14(1):47-57.

11 Brod M, Kongsø JH, Lessard S, Christensen TL. Psychological insulin resistance: patient beliefs and implications for diabetes management. Qual Life Res. 2009; 18(1):23-32.

12 Morris AD, Boyle DI, McMahon AD, Greene SA, MacDonald TM, Newton RW, et al. Adherence to insulin treatment, glycaemic control, and ketoacidosis in insulin-dependent diabetes mellitus. The Lancet. 1997; 350(9090):1505-10.

13 Peyrot M, Rubin RR, Kruger DF, Travis LB. Correlates of insulin injection omission. Diabetes care. 2010; 33(2):240-5.

14 Mirahmadizadeh A, Delam H, Seif M, Banihashemi SA,

Tabatabaee H. Factors Affecting Insulin Compliance in Patients with Type 2 Diabetes in South Iran, 2017: We Are Faced with Insulin Phobia. Iran J Med Sci. 2018; 44(3):204-14.

15 Pinhas-Hamiel O, Hamiel U, Greenfield Y, Boyko V, Graph-Barel C, Rachmiel M, et al. Detecting intentional insulin omission for weight loss in girls with type 1 diabetes mellitus. Int J Eat Disord. 2013; 46(8):819-25.

16 Huang G-M, Huang K-Y, Lee T-Y, Weng JT-Y. An interpretable rule-based diagnostic classification of diabetic nephropathy among type 2 diabetes patients. BMC bioinformatics, BioMed Central. 2015.

17 Leung RK, Wang Y, Ma RC, Luk AO, Lam V, Ng M, et al. Using a multi-staged strategy based on machine learning and mathematical modeling to predict genotype-phenotype risk patterns in diabetic kidney disease: a prospective case–control cohort analysis. BMC Nephrol. 2013; 14(1):162-72.

18 Tapak L, Mahjub H, Hamidi O, Poorolajal J. Real-data comparison of data mining methods in prediction of diabetes in iran. Healthc Inform Res. 2013; 19(3):177-85.

19 Safe M, Faradmal J, Mahjub H. A Comparison between Cure Model and recursive partitioning: a retrospective cohort study of Iranian female with breast cancer. Comput Math Methods Med, 2016.

20 Safe M, Faradmal J, Poorolajal J, Mahjub H. Model-based Recursive Partitioning for Survival of Iranian Female Breast Cancer Patients: Comparing with Parametric Survival Models. Iran J Public Health. 2017; 46(1):35-43.

21 Hastie T, Tibshirani R, Friedman J, Franklin J. The elements of statistical learning: data mining, inference and prediction. The Mathematical Intelligencer. 2005; 27(2):83-5.

22 Strobl C, Boulesteix A-L, Zeileis A, Hothorn T, editors. Bias in random forest variable importance measures. Workshop on Statistical Modelling of Complex Systems. 2006.

23 Hertz RP, Unger AN, Lustik MB. Adherence with pharmacotherapy for type 2 diabetes: a retrospective cohort study of adults with employer-sponsored health insurance. Clin Ther. 2005; 27(7):1064-73.

24 Kaplan RC, Bhalodkar NC, Brown EJ, White J, Brown DL. Race, ethnicity, and sociocultural characteristics predict noncompliance with lipid-lowering medications. Prev Med . 2004; 39(6):1249-55.

25 Anderson B, Vangsness L, Connell A, Butler D, Goebel-Fabbri A, Laffel L. Family conflict, adherence, and glycaemic control in youth with short duration type 1 diabetes. Diabet Med. 2002; 19(8):635-42.

26 Morris AD, Boyle DI, MacAlpine R, Emslie-Smith A, Jung RT, Newton RW, et al. The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. Bmj. 1997; 315(7107):524-8.

27 Strobl C, Kopf J, Zeileis A. Rasch trees: A new method for detecting differential item functioning in the Rasch model. Psychometrika. 2015; 80(2):289-316.

28 de Miguel-Yanes JM, Shrader P, Pencina MJ, Fox CS, Manning AK, Grant RW, et al. Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms. Diabetes care. 2011; 34(1):121-5.

29 Khattab M, Khader YS, Al-Khawaldeh A, Ajlouni K. Factors associated with poor glycemic control among patients with type 2 diabetes. Journal of diabetes and its complications. 2010; 24(2):84-9.

30 Caro JJ, Salas M, Speckman JL, Raggio G, Jackson JD. Persistence with treatment for hypertension in actual practice. CMAJ. 1999; 160(1):31-7.

31 Jin J, Sklar GE, Oh VMS, Li SC. Factors affecting therapeutic compliance: A review from the patient's perspective. Ther Clin Risk Manag. 2008; 4(1):269-86.

32 Donnan PT, MacDonald TM, Morris AD. Adherence to prescribed oral hypoglycaemic medication in a population of patients with Type 2 diabetes: a retrospective cohort study. Diabet Med. 2002; 19(4):279-84.

33 de Oliveira SMS, Guimarães DB, Reis JS. Illiteracy and diabetes: educational program for people with type 2 diabetes in the public health system. Diabetol Metab Syndr. 2015; 7(1):A177.

34 Hopkins D, Lawrence I, Mansell P, Thompson G, Amiel S, Campbell M, et al. Improved biomedical and psychological outcomes 1 year after structured education in flexible insulin therapy for people with type 1 diabetes. Diabetes Care. 2012; 35(8):1638-42.

35 Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions. 1st ed. John Wiley & Sons. 2013.

36 Boulesteix AL, Janitza S, Kruppa J, König IR. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdiscip Rev Data Min Knowl Discov. 2012; 2(6):493-507.

37 Onwukwe C, Aki I. On selection of best sensitive logistic estimator in the presence of collinearity. Am J Appl Math Stat. 2015; 3(1):7-11.

38 Maroco J, Silva D, Rodrigues A, Guerreiro M, Santana I, de Mendonça A. Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. BMC Res Notes. 2011; 4(1):299-313.

39 Safe M, Mahjub H, Faradmal J. A Comparative Study for Modelling the Survival of Breast Cancer Patients in the West of Iran. Glob J Health Sci. 2016; 9(2):215-22.