ORIGINAL RESEARCH

AzoresDiab model: the risk prediction of type 2 diabetes in the Azores

AUTHORS

Duarte Pedro de Sousa Tavares[1] PhD, Assistant Professor *

Ana Filipa Jorge[2] MSc, MD

CORRESPONDENCE

*Asst Prof Duarte Pedro de Sousa Tavares duartepedro7@gmail.com

AFFILIATIONS

[1] Faculdade de Economia, Universidade do Algarve, Faro, Portugal; and Egas Moniz – Cooperativa de Ensino Superior, CRL, Almada, Portugal

[2] Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal

ABSTRACT:

Introduction: The incidence of type 2 diabetes mellitus is increasing in the Azores. The Azorean population has unique health-related characteristics that emphasize the necessity of developing a predictive model, namely the double insularity phenomena and consanguinity marriages. Therefore, the authors aimed to develop a model, the AzoresDiab model, that assesses the risk of type 2 diabetes for residents of the Azores.
Methods: The variables used for developing the model included the history of cardiovascular disease, hypertension, sex, body mass index, triacylglycerol level, glucose level, and age. This model was developed using binary logistic regression wherein the dependent variable was considered 0 if the patient had type 2 diabetes and 1 if the patient did not. The sample comprised 6834 individuals who were Azores residents, aged over 18 years and who were not missing values for the covariates under study; individuals were included regardless of whether they had a previous diagnosis of type 2 diabetes. Participants were considered to have type 2 diabetes if they had been previously diagnosed with type 2 diabetes or had been prescribed at least one antidiabetic drug listed in the norms of the Portuguese General Directorate of Health and the Portuguese Medical Association.
Results: This model showed an area under the receiver operating

characteristic curve of 0.863 based on internal validation performed with bootstrapping.

Conclusion: The AzoresDiab model exhibited excellent discrimination of patients with and without type 2 diabetes. The use of predictive risk models will enable the early implementation of disease prevention programs in medium- and high-risk individuals, and public health policies to prevent the onset of the disease in these populations.

FULL ARTICLE:

Introduction

Type 2 diabetes mellitus accounts for 90% of the incidence of diabetes and is considered by many as the 21st century epidemic disease[1] because of its increasing prevalence, inherent complications, associated deaths, and exorbitant costs to healthcare systems[2] such as the Portuguese National Health Service. There is a growing demand for prevention, early diagnosis, and effective treatment of this illness[3].

Diabetes affects an estimated 9.3% (463 million) of the worldwide adult population and contributes to almost 5 million deaths per year. The prevalence of diabetes has been increasing over the last three decades, and the disease is expected to affect about 700 million people by 2040[3]. Additionally, in 2010, diabetes caused 680 disability-adjusted life years (DALY) per 100 000 inhabitants worldwide, one of the highest DALY values registered[4].

The Azores, officially the Autonomous Region of the Azores, is an autonomous region of Portugal consisting of nine volcanic islands in the North Atlantic[5]. According to a 2014 Azores regional health inquiry, the prevalence of diabetes is 9.9% among the population between 20 and 74 years old residing in the Azores (9.8% of women and 10% of men)[6]. Furthermore, 66% of individuals with diabetes also have hypertension, while 22% and 34% are considered obese and pre-obese, respectively. In addition, a 2009 PREVADIAB (*Prevalência da Diabetes em Portugal*) study reported a prevalence of 14.3% of diabetes in the Azorean population: 9.2% with an established diagnosis and 5.1% with undiagnosed diabetes[6].

It is crucial to identify individuals with undiagnosed diabetes (who are already included in the established prevalence of diabetes) and those at risk of developing this disease, which will thus contribute to an incremental increase in the prevalence of diabetes in the near future. This model would justify the need for new prevention and early diagnosis strategies[7].

Type 2 diabetes diagnosis was, for many years, subjective and not systematized[8]. Currently, the diagnostic criteria for type 2 diabetes are based on fasting plasma glucose levels (≥126 mg/dL), 2-hour plasma glucose levels (≥200 mg/dL 2 hours after the ingestion of 75 g anhydrous glucose in an oral glucose tolerance test), glycated hemoglobin (≥6.5%), and random plasma glucose levels (≥200 mg/dL if associated with diabetes symptoms)[9].

Identifying risk factors associated with a greater likelihood of developing type 2 diabetes is extremely important, given the existence of asymptomatic patients who fulfil the laboratory criteria, the epidemiological studies showing that type 2 diabetes may be present up to 10 years before diagnosis, and the several patients who have specific diabetes complications when the diagnosis is established[10].

Several risk factors for type 2 diabetes are already known: family history (parents or siblings) of type 2 diabetes, obesity, sedentary lifestyle, specific ethnicities, history of fasting plasma glucose levels >100 mg/dL, decreased glucose tolerance or glycated hemoglobin >5.7–6.4%, history of gestational diabetes or delivery of newborn weighing >4 kg, hypertension, high-density lipoprotein cholesterol levels <35 mg/dL or triacylglycerol (the main constituent of body fat in humans) levels >250 mg/dL, history of polycystic ovary syndrome or of acanthosis nigricans, and history of cardiovascular disease[7,9,11,12]. Subjective risk factors should also be considered: age, abdominal circumference (measured below the ribs), a diet rich in carbohydrates and saturated fats and poor in fruits and vegetables, smoking habit, low birth weight, low socioeconomic level, and regular medication with corticosteroids or antipsychotics.

It is estimated that the combined effect of these risk factors contributes to an 80% increase in type 2 diabetes risk[7]. All these risk factors can be used to establish risk predictor models for type 2 diabetes. There are various models designed for this purpose: the Finnish Diabetes Risk Score (FINDRISC)[13], which is recommended by the American Diabetes Association and is the most used model in Portugal; the Framingham Offspring Study (FOS)[14], which was once the most widely used model in the world; the Leicester Practice Risk Score (LPRC)[15], a European model used mainly in the UK and recommended by the National Institute for Health and Care Excellence; and the German Diabetes Risk Score (GDRS)[16], a model mainly used in Germany but also recommended by the American Diabetes Association. Nevertheless, most of the risk factors differ between models (see Discussion). For instance, only age, anthropometric factors and history of hypertension were common to all four international models.

It is crucial to evaluate all models to understand which are useful, applicable to the population, reliable, and effective. Usually, the evaluation of risk predictor models is based on their objectives, covariates, methodology, sample size, selection criteria, and model performance (evaluated by internal and/or external validation). A good risk predictor model should have well-defined objectives, proven validation, and good performance, reliability, and clinical utility[7].

Guidelines do not indicate the need for a risk predictor model.

Instead, in the future, it seems that new models will be developed that will enable each individual to estimate their own risk of developing type 2 diabetes and that will enable delineation of high-risk areas ('hotspots'); this will be of interest in the field of public health[17].

This study aimed to develop AzoresDiab, a type 2 diabetes risk predictor model specific to the Azorean population, which is highly affected by the insularity phenomenon and consanguinity marriages, and has developed unique genetic characteristics of health as expected by Rudan[18].

Methods

To develop the AzoresDiab model, the authors used anonymous patient-level data from 2013 and 2014 that related to primary care services in all the Azores islands provided by the entity that manages the health system in the Azores (Saudaçor SA). In total, data from 272 705 patients (total population) were supplied. A patient was defined as having type 2 diabetes if they had an International Classification of Primary Care (ICPC-2) code of T90 in the data or had at least one prescription of an antidiabetic drug listed in the Portuguese Guidelines for Diabetes Clinical Management[12].

As there is no consensus regarding the best methodology for allocating values to missing values in health[19], and because pediatric populations have specific and different clinical patterns and assumptions related to type 2 diabetes, the following exclusion criteria were established: patients missing covariate values and those aged <18 years. After these criteria were applied, 6834 individuals were eligible for inclusion in the development of the risk model.

The AzoresDiab model is based on binary logistic regression[20] (the dependent variable is 0 if the person does not have type 2 diabetes and 1 if the person has type 2 diabetes) to describe the individual probabilistic risk of type 2 diabetes development, as suggested by Iezzoni[19]. To avoid confounding effects, an α value of 0.05 was used as the reference for the independent variables[21], and these variables were also included in an interaction model[20]. The independent variables included in the AzoresDiab model were those available and collected by Saudaçor SA (Table 1), and their connection with type 2 diabetes was supported by the Kasper et al study[9] and by two review articles[7,22]. The model was internally validated by the bootstrapping method as described by Efron and Tibshirani[23], and the ability of the model to discriminate this binary scenario was calculated using the area under the curve (AUC)[24]. All statistical analysis was conducted using SPSS v23 (www.spss.com).

Table 1: Clinical and sociodemographic variables under study[25,26]

Independent variable	Variable description
Cardiovascular disease history	Dummy (0=no history, 1=with history) History of cardiovascular disease means, at least, registration of one of a selection of ICPC-2 codes (K71, K74, K75, K70, K84, K83, K73, K79, K78, K80, K77, K90, K91, K92, K99, K94, K95, K96, B71, B99 or K88) or drug prescription history related to Chapter 3 of the INFARMED therapeutic book (except subsection 3.4) [ref. 25].
Hypertension history	Dummy (0=no history, 1=with history) A history of hypertension means, at least, registration of ICPC-2 code K86 or K87 or drug prescription history related to the subsection 3.4 of the INFARMED therapeutic book [ref. 25].
Sex	Dummy (0=male, 1=female)
Age	Numerical (years)
Body mass index	Numerical (kg/m^2)
Triacylglycerol level	Numerical (mg/dL)
Glucose level	Numerical (mg/dL) Calculation of the average value of glucose from the value of glycated hemoglobin, as suggested by Nathan et al [ref. 26].

ICPC-2, International Classification of Primary Care.

Ethics approval

The databases used are approved by the Protection Data National Commitee and the authors have only worked with anonymous data.

Results

The background characteristics of the study population are presented in Table 2.

The probability of an inhabitant of the Azores developing type 2 diabetes is based on the following expression:

$$\hat{P}(\text{having type 2 diabetes}) = \frac{e^a}{1 + e^a}$$

where a is given by the following equation:

$$a = -8.209 - 0.480 \times \text{Sex} + 0.065 \times \text{Age} + 0.043 \times \text{BMI} + 0.024 \times \text{Gli} - 0.031(\text{CVD} \times \text{Age}) + 0.030 \times$$
$$(\text{CVD} \times \text{Gli}) - 0.019 \times (\text{HTA} \times \text{Age}) - 0.040 \times (\text{HTA} \times \text{BMI}) + 0.023 \times (\text{Gli} \times \text{HTA}) + 0.002 \times (\text{Sex} \times \text{Tri})$$

where BMI is the body mass index, CVD is the history of

cardiovascular disease, Gli is the glucose level, HTA is the history of

hypertension, and Tri is the triacylglycerol level.

Following internal validation, the model described above had an AUC of 0.863, with a 95% confidence interval of 0.854–0.872; this shows excellent discrimination of individuals with and without

type 2 diabetes[20].

Table 3 shows the significance of the variables included in the model, and Figure 1 presents the receiver operating characteristic curve.

Table 2: Exploratory analysis results

Characteristic	n (%)
Diagnosis of type 2 diabetes	
No	2173 (31.8)
Yes	4661 (68.2)
Age (years)	
18–49	1077 (15.8)
50–64	2663 (39.0)
65–79	2531 (37.0)
≥80	563 (8.2)
Body mass index (kg/m²)	
<18.5	22 (0.3)
18.5–24.9	769 (11.2)
25.0–29.9	2364 (34.6)
≥30	3679 (53.8)
Sex	
Male	2827 (41.4)
Female	4007 (58.6)
Cardiovascular disease history	
Yes	4551 (66.6)
No	2283 (33.4)
Hypertension history	
Yes	5036 (73.7)
No	1798 (26.3)
Triacylglycerol level (mg/dL)	
Normal (<250)	5744 (84.0)
Abnormal (≥250)	1090 (16.0)
Glucose level (mg/dL)	
<110	1040 (15.2)
110–125	1659 (24.3)
≥126	4135 (60.5)

Table 3: The final model, with the significance of the included variables

Independent variables	β	Standard error	Odds ratio (e^β)	95% confidence interval	p-value
Sex × Tri	0.002	0.001	1.002	1.001–1.004	0.002
Sex	−0.480	0.115	0.619	0.393–0.845	<0.001
HTA × BMI	−0.040	0.010	0.960	0.940–0.981	<0.001
BMI	0.043	0.010	1.044	1.026–1.063	<0.001
HTA × age	−0.019	0.005	0.981	0.971–0.991	<0.001
CVD × age	−0.031	0.006	0.969	0.957–0.982	<0.001
Gli × HTA	0.023	0.003	1.023	1.017–1.029	<0.001
CVD × Gli	0.030	0.004	1.031	1.023–1.038	<0.001
Age	0.065	0.005	1.067	1.057- 1.078	<0.001
Gli	0.024	0.002	1.024	1.021–1.028	<0.001
(Constant)	−8.209	0.451	<0.001	–	<0.001

BMI, body mass index. CVD, history of cardiovascular disease. Gli, glucose level. HTA, history of hypertension. Tri, triacylglycerol level.
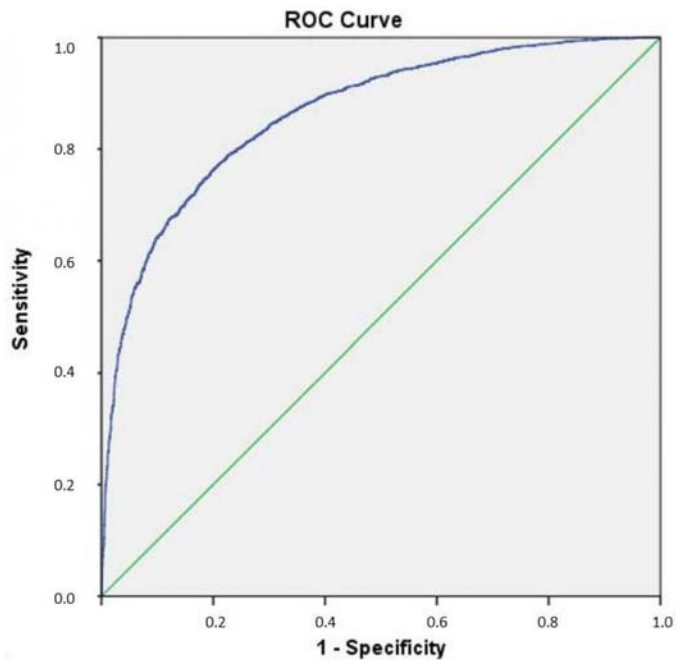
Figure 1: The model receiver operating characteristic curve.

Discussion

Early diagnosis of type 2 diabetes is crucial to prevent complications and create personalized diagnosis strategies for each individual, namely in remote and rural populations such as those in the Azores islands. With this in mind, the AzoresDiab model was developed to discriminate individuals with type 2 diabetes from those without type 2 diabetes; the model had an excellent AUC of 0.863.

The AzoresDiab model has similarities with and differences from four other international models used to predict type 2 diabetes: FINDRISC, FOS, LPRC and GDRS (Table 4).

All previous models compared the age, anthropometric factors, and history of hypertension. Nevertheless, the AzoresDiab is the most inclusive model because it includes all individuals aged 18 years or older. In contrast, FINDRISC[23], FOS[14] (45–64 years), GDRS[16] (35–65 years), and LPRS[15] (40–75 years) are more restricted models.

Changes in blood glucose level or oral glucose tolerance may be present without a definitive diagnosis of type 2 diabetes, indicating a prediabetic state that can be predicted by the LPRS model[15]. Considering this, the AzoresDiab model also includes blood glucose levels, which are also considered by the FINDRISC[13] and FOS[14] models.

A history of CVD is a leading predictor of disability and death among patients with type 2 diabetes[27]. Compared with the other four international models, the AzoresDiab model is the only one to use a history of CVD as a covariate. Using this covariate in the model is advantageous, namely in cases where type 2 diabetes has not yet been diagnosed, because it will increase the probability of predicting type 2 diabetes.

Larger samples are essential for constructing more robust models. The AzoresDiab model had the second-largest sample (6834), immediately behind the GDRS study (25 167), and followed by the LPRS[15] (6186), FINDRISC[16] (4746), and FOS[14] (3140) studies.

All four previous models are based on prospective cohort studies, whereas the AzoresDiab model uses binary logistic regression with retrospective information. The different objectives of the models justifies this discrepancy in methodologies. The AzoredDiab model aims to predict the probability of an individual having type 2 diabetes in a population with specific characteristics (remote and rural with double insularity phenomena and consanguinity marriages), while the others are more generalized and can be used for the population of the whole world.

Regarding the sample selection criteria, all the international models exclude individuals with a previous diagnosis of type 2 diabetes. However, these individuals were not excluded from the AzoresDiab model because one of the purposes of this model was that undiagnosed individuals could use it to confirm their risk of having type 2 diabetes.

In the methods used to create the four international models and the AzoresDiab model, the discrepancies are not so notable in the diagnosis of type 2 diabetes as in the sample criteria: the five models are based on either the existence of a previously registered diagnosis of type 2 diabetes, the evidence of laboratory criteria for the diagnosis of type 2 diabetes, or the existence of ongoing antidiabetic medication. Laboratory evidence is the most objective criterion. However, AzoresDiab uses the registered diagnosis of type 2 diabetes or the prescription of at least one of the antidiabetic drugs referred to in the Portuguese Guidelines for Diabetes Clinical Management[12]. Most of these models are based on logistic regression, but it could also be interesting to study further the possibility of calculating the individual risk using the

Bayes theorem and machine learning[28-30].

The internally validated AUC in the present study of 0.863 was slightly above that of the other models (namely the FINDRISC[13] and the FOS[14] models), which obtained an internally validated AUC of 0.85. However, using the AzoresDiab model in the Azores still requires external validation before a proper comparison can be made with the other international models.

Besides the external validation, it is also important to note that the AzoresDiab model has other limitations, namely being restricted to the variables collected by Saudaçor SA. Risk factors such as physical activity, dietary habits (FINDRISC and GDRS), cholesterol levels (FOS) and family history of type 2 diabetes (FINDRISC, FOS and LPRS) were not available in the dataset used. Also, individuals were excluded who had data missing, and, therefore, the model may not represent the entire Azorean population. Nevertheless, as mentioned above, through AUC, the performance of the model was considered excellent in predicting type 2 diabetes even without other relevant risk factors reported in the literature and the other four models.

Despite the extensive creation of predictive risk models for type 2 diabetes, there are still some limitations in studies determining the impact of these models on type 2 diabetes. The premise of the AzoresDiab model is to meet the need for a model that can be used by any individual, even a layperson, to measure their risk of developing type 2 diabetes.

The use of predictive risk models will enable the early implementation of disease prevention programs in medium- and high-risk individuals. However, a high percentage of people are not interested in initiating prevention strategies despite having a high risk of developing the disease. Predictive risk models will also allow the delineation of groups of the population (local communities) with a medium and a high risk for type 2 diabetes; this will allow public health policies to be implemented to prevent the onset of the disease in these populations.

Table 4: Variables used by four international models and the AzoresDiab model with the respective area under the curve

Variable	Finnish Diabetes Risk Score	Framingham Offspring Study	Leicester Practice Risk Score	German Diabetes Risk Score	AzoresDiab model
Age	X	X	X	X	X
Sex		X	X		X
Ethnicity			X		
Anthropometric factors	X	X	X	X	
Family history of type 2 diabetes	X	X	X		
History of hypertension	X	X	X	X	
Hypertriglyceridemia		X			X
Blood glucose level	X	X			X
Cholesterol level			X		
Dietary habits	X			X	
Physical activity	X			X	
Smoking habit				X	
Cardiovascular disease history					X
Area under the curve	0.850	0.850	NR	NR	0.863

NR, not reported.

Conclusion

The objectives of AzoresDiab were to predict type 2 diabetes in any individual, while considering the epidemiological and geographical context in which it was developed – the remote and rural Azores islands. AzoresDiab also considers the most relevant risk factors for the development of type 2 diabetes. Using diagnostic data from Primary Health Care and registrations of prescriptions of antidiabetic drugs listed in the Portuguese Guidelines for Diabetes Clinical Management standards, the model achieved an internally validated AUC of 0.863, which shows its excellent discrimination power. In the near future, the next step will be the external validation of the AzoresDiab model to confirm its discriminatory power.

REFERENCES:

1 Tabish SA. Is diabetes becoming the biggest epidemic of the twenty-first century? *International Journal of Health Science* 2007; **1(2):** V-VIII.

2 Raposo JF. Diabetes: facts and numbers. 2016, 2017 e 2018. [In Portuguese]. *Revista Portuguesa de Diabetes* 2020; **15(1):** 19-27.

3 International Diabetes Foundation. *IDF diabetes atlas, 9th edn.* 266. 2019: 134-137. Available: web link (Accessed 11 July 2021).

4 Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; **380(9859):** 2197-2223. DOI link

5 National Printing House. Law no 2/2009, 12 January. *Diário da República* [official gazette of Portugal; in Portuguese]. **1(17).** Available: web link (Accessed 11 July 2021).

6 Direção Regional da Saúde. *2014's Regional Health Enquiry*. [In Portuguese]. 2015. Available: web link (Accessed 11 July 2021).

7 Collins GS, Mallet S, Omar O, Yu L. Developing risk prediction models for type 2 diabetes: a systematic review of methodology

and reporting. *BMC Medicine* 2011; **9:** 103-117. DOI link, PMid:21902820

8 Direção-Geral da Saúde. *Health Delegation Standard of Action 002/2011: Diagnosis and Classification of Diabetes Mellitus.* [In Portuguese]. 2011. Available: web link (Accessed 11 July 2021).

9 Kasper DL, Fauci AS, Hauser SL, Longo DL, Jameson JL, Loscalzo J (Eds). *Harrison's principles of internal medicine.* 19th edn. New York: McGraw-Hill Education, 2015.

10 American Diabetes Association. Classification and diagnosis of diabetes: Standards of Medical Care in Diabetes - 2021. *Diabetes Care* 2021; **44(January):** S15-S33. DOI link, PMid:33298413

11 World Health Organization Europe. *Diabetes - data and statistics.* Available: web link (Accessed 11 July 2021).

12 Direção-Geral de Saúde. *Health Delegation Standard of Action 052/2011: Pharmacological therapeutic in adults with type 2 diabetes mellitus.* [In Portuguese]. 2015. Available: web link (Accessed 11 July 2021).

13 Lindström J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* 2003; **26(3):** 725-731. DOI link, PMid:12610029

14 Wilson PWF, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB Sr. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Archives of Internal Medicine* 2007; **167(10):** 1068-1074. DOI link, PMid:17533210

15 Gray LJ, Taub NA, Khunti K, Gardiner E, Hiles S, Webb DR et al. The Leicester Risk Assessment Score for detecting undiagnosed type 2 diabetes and impaired glucose regulation for use in a multiethnic UK setting. *Diabetic Medicine* 2010; **27(8):** 887-895. DOI link, PMid:20653746

16 Schulze MB, Hoffmann K, Boeing H, Linseisen J, Rohrmann S, Möhlig M, et al. An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. *Diabetes Care* 2007; **30(3):** 510-515. DOI link, PMid:17327313

17 Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. *BMJ* 2011; **343:** d7163. DOI link, PMid:22123912

18 Rudan I. Health effects of human population isolation and admixture. *Croatian Medical Journal* 2006; **47(4):** 526-531.

19 Iezzoni LI (Ed.). *Risk adjustment for measuring health care outcomes.* 4th edn. Chicago: Health Administration Press, 2013.

20 Hosmer DW, Lemeshow S, Studivant RX. *Applied logistic regression.* 3rd edn. Hoboken: Wiley, 2013. DOI link

21 Fisher RA. *The design of experiments.* 2nd edn. London: Oliver and Boyd, 1937.

22 Mbanya V, Hussain A, Kengne AP. Application and applicability of non-invasive risk models for predicting undiagnosed prevalent diabetes in Africa: a systematic literature search. *Primary Care Diabetes* 2015; **9(5):** 317-329. DOI link, PMid:25975760

23 Efron B, Tibshirani RJ. *An introduction to the bootstrap.* 1st edn. London: Springer, 1993. DOI link

24 Swets J. Measuring the accuracy of diagnostic systems. *Science* 1988; **240:** 1285-1293. DOI link, PMid:3287615

25 INFARMED. *INFARMED Therapeutic Medical Record 10.* [In Portuguese]. 2011. Available: web link (Accessed 11 July 2021).

26 Nathan DM, Kuenen J, Borg R, Zheng H, Schoenfeld D, Heine RJ. Translating the A1C assay into estimated average glucose values. *Diabetes Care* 2008; **31(8):** 1473-1478. DOI link, PMid:18540046

27 Low Wang CC, Hess CN, Hiatt WR, Goldfine AB. Clinical update: cardiovascular disease in diabetes mellitus. *Circulation* 2016; **133(24):** 2459-2502. DOI link, PMid:27297342

28 Birjais R, Mourya AK, Chauhan R, Kaur H. Prediction and diagnosis of future diabetes risk: a machine learning approach. *SN Applied Sciences* 2019; **1(9):** 1-8. DOI link

29 Sisodia D, Sisodia DS. Prediction of diabetes using classification algorithms. *Procedia Computer Science* 2018; **132:** 1578-1585. DOI link

30 Agnal AS, Saraswathi E. Analyzing diabetic data using naive-Bayes Classifier. *European Journal of Molecular and Clinical Medicine* 2020; **7(4):** 2687-2698.