Diagnosis of Type 2 Diabetes and Pre-diabetes Using Machine Learning

Erika Severeyn[1(✉)], Sara Wong[2], Jesús Velásquez[1], Gilberto Perpiñán[3], Héctor Herrera[4], Miguel Altuve[5], and José Díaz[6]

[1] Departamento de Termodinámica y Transferencia de Calor, Universidad Simón Bolívar, Caracas, Venezuela
severeynerika@usb.ve
[2] Departamento de Electrónica y Circuitos, Universidad Simón Bolívar, Caracas, Venezuela
[3] Faculty of Electronic and Biomedical Engineering, Antonio Nariño University, Cartagena, Colombia
[4] Departamento de Tecnología de Procesos Biológicos y Bioquímicos, Universidad Simón Bolívar, Caracas, Venezuela
[5] Faculty of Electrical and Electronic Engineering, Pontifical Bolivarian University, Bucaramanga, Colombia
[6] Departamento de Ingeniería Eléctrica y Computación, Universidad Autónoma de Ciudad Juarez, Ciudad Juarez, Mexico

Abstract. The type-2 diabetes (T2D) is a multifactorial chronic disease that reduces the quality of lifestyle and produces the death of a large percentage of the population worldwide. Before the development of T2D a series of symptoms are presented even years before T2D diagnosis. This condition that appears before the development of T2D is called prediabetes. Prediabetes and T2D are diagnosed from the oral glucose tolerance test (OGTT). The OGTT consists in the measurement of glucose and insulin in five-time intervals, the first after 8 h of fasting (0 min) and the other four measurements after taking 75 g of oral glucose in 30-minutes intervals (30, 60, 90 and 120 min). Some parameters have been used to improve the efficiency in the diagnosis of prediabetes and T2D, for example: the area under the glucose (AUC_G) and insulin (AUC_I) curve during OGTT has been used as a parameter for the diagnosis of prediabetes, T2D and obesity. The aim of this study is to assess the *k*-means clustering algorithm in the classification of subjects with prediabetes and T2D using the AUC_G and AUC_I. A database of 188 subjects (male = 88 subjects, age = 42.11 ± 14.91 years old) with values of plasma glucose and insulin during OGTT was used. The *k*-means clustering performed for AUC_G presents acceptable results since the silhouette coefficient is above 0.6 in all cases. The findings in this study indicate that the *k*-means applied in the AUC_G classify subjects with T2D, prediabetes and control. Furthermore, it could even predict those subjects with high probabilities of developing T2D.

Keywords: *k*-means · Oral glucose tolerance test · Area under the insulin curve · Area under the glucose curve · Statistical analysis

C. A. González Díaz et al. (Eds.): CLAIB 2019, IFMBE Proceedings 75, pp. 792–802, 2020.
https://doi.org/10.1007/978-3-030-30648-9_105

1 Introduction

Type-2 diabetes (T2D) is a multifactorial chronic disease that reduces the quality of lifestyle and produces the death of a large percentage of the population worldwide [1]. The T2D imposes an enormous economic burden to the global health systems [2]. The total annual expenses of this disease range between US$ 141.6 million and 174 billion, and it is estimated that the health expenditure of T2D patients is at least twice the expenditure of healthy people [3–5].

The T2D has become entrenched in developing countries, and in very recent years, more than 80% of the deaths caused by this disease have occurred in low and middle-income countries. Moreover, their morbidity burden is estimated that it will increase worldwide and particularly in developing countries [6–8]. Currently, the global prevalence of T2D in people older than 18 years has increased from 4.7% (108 million people) in 1980 to 8.5% (400 million people) in 2017 and this increase has been faster in middle and low-income countries [8]. Even more, the global prevalence of T2D is projected to increase to more than 600 million affected persons by the year 2035, and the 70% of that global estimation will be placed in developing countries [8].

T2D is a progressive disease that even years before its diagnosis a series of symptoms are present. This condition that appears before the development of T2D is called prediabetes [9]. Prediabetes and T2D are diagnosed from the oral glucose tolerance test (OGTT). The OGTT consists in the measurement of glucose and insulin concentrations in five-time intervals (5-samples OGTT), the first measurement is taken after 8 h of fasting (0 min) and the other four measurements after taking 75 g of liquid glucose in 30-minute intervals (30, 60, 90 and 120 min) [10]. T2D is diagnosed according to the World Health Organization if fasting glucose is above 125 mg/dL and/or postprandial glucose (120 min of OGTT) is above 200 mg/dL [9]. Prediabetes is diagnosed according to the ADA if fasting glucose ranges between 100-125 mg/dL, which is known as impaired fasting glucose (IFG) and/or postprandial glucose range between 140–200 mg/dL, which is known as impaired glucose tolerance (IGT) [9].

Prediabetes can be controlled, even reversed through very simple treatments or changes in lifestyle; otherwise, when T2D is diagnosed it is not possible to reverse it. Furthermore, it can only be controlled with more expensive treatments that diminished the patient life quality [11]. Therefore, the diagnosis of the prediabetic diseases is crucial to prevent the development of diabetes and thus reduce public health spending [12]. On the other hand, it is known that early diagnosis of T2D can prevent the progression of the disease to other related conditions such as diabetic neuropathy, atherosclerosis, and diabetic foot [13].

There are studies that have used different methodologies to classify prediabetes, T2D and its related diseases (metabolic syndrome and insulin resistance). Some studies have used *k*-means to differentiate diabetic subjects from normal subjects, achieving accuracy classification above 80% [14, 15]. Other studies have designed indices that include anthropometric variables [16], biochemical variables [17] and heart rate variability parameters [18] to classify subjects with insulin resistance, prediabetes and metabolic syndrome, obtaining receiving operating characteristics curves with areas above 0.70 [19]. On the other hand, the area under the glucose (AUC_G) and insulin

(AUC_I) curves during OGTT have also been used to improve the efficiency in the diagnosis of prediabetes, T2D and obesity [20, 21].

The aim of this study is to assess the *k*-means clustering algorithm using AUC_G and AUC_I in the classification of subjects with prediabetes and T2D. A database of 188 subjects with values of glucose and insulin during OGTT was used. In the next section the methodological procedure will be explained. In section three and four, results and discussion will be shown. And finally, in section five, the conclusions and future works proposals will be presented.

2 Methodology

2.1 Database

At the Clinical Research Laboratory of the Venezuela Central University, 188 adults were enrolled (male = 88 subjects, age = 42.11 $\pm$ 14.91 years old), between the years 2010 and 2013. Each participant underwent the 5-samples OGTT. In the OGTT of five samples, insulin and glucose levels were measured in the five different blood samples: a sample in fasting of 8 h (G_0, I_0) and four others after oral intake of 75 g of glucose at 30 min (G_{30}, I_{30},), 60 min (G_{60}, I_{60}), 90 min (G_{90}, I_{90}),and 120 min (G_{120}, I_{120}).

Every participant was classified as T2D according to WHO criteria ($G_0 > 125$ and/or $G_{120} \geq 200$) [22], as prediabetic according to ADA criteria ($140 \leq G_{120} < 200$ and/or $100 \leq G_0 \leq 125$) [9] and as a control group if it does not meet the criteria for T2D or prediabetes. The characteristics of the database are shown in Table 1. The clinical protocol adhered to the principles of the Declaration of Helsinki and it was approved by the Bioethical Committee of the Medical Science Faculty of Venezuela Central University; all the subjects signed a written informed consent.

2.2 *k*-Means Clustering Method

In this investigation the *k*-means clustering algorithm is used to classify the one-dimensional observations [23]. *k*-means technique goal is the partition of the dataset into different clusters, so that observations in each cluster share some similar characteristics. The distance between observations and centroids were calculated by squared Euclidean distance and it was repeated 10 times to avoid local minima. The silhouette coefficient (SC) was used to evaluate the assignment of the dataset into the cluster [24].

Two experiments were carried out using the AUC_I, and two more experiments using the AUC_G. The AUC_G was calculated from the curve constructed with the glucose values of 30, 60 and 90 min of the OGTT. The glucose values of 0 min and 120 min from the OGTT are dismissed because they are dependent variables in the diagnosis of T2D and prediabetes. The AUC_I was calculated from the curve constructed with the insulin values of 0, 30, 60, 90 and 120 min of the OGTT.

Table 1. Glucose and insulin values of the OGTT, AUC_I and AUC_G for prediabetic, diabetic and control groups.

Variables	Control[e] Male = 33, n = 87	Prediabetic Male = 51, n = 85 IGT = 35, IFG = 72 IGT and IFG = 22	T2D Male = 4, n = 16	p-value[a]
Age [years]	35.78 ± 13.02[f] 18.00–75.00[g] 33.05–38.52[h]	46.32 ± 14.69 20.00–78.00 43.20–49.44	54.19 ± 10.24 42.00–72.00 49.17-59.20	<0.01[b,c,d]
G_0 [mg/dl]	90.74 ± 6.23 75.00–99.00 89.43–92.04	104.18 ± 7.45 77.00–119.00 102.59–105.76	134.25 ± 22.17 99.00–187.00 123.39–145.11	<0.01[b,c,d]
G_{30} [mg/dl]	134.10 ± 24.97 85.00–202.00 128.86–139.35	163.54 ± 27.22 100.00–230.00 157.75–169.33	214.81 ± 48.63 135.00–312.00 190.98–238.64	<0.01[b,c,d]
G_{60} [mg/dl]	125.54 ± 31.54 73.00–227.00 118.91–132.17	167.19 ± 39.45 80.00–256.00 158.80–175.58	243.44 ± 53.62 136.00–349.00 217.16–269.71	<0.01[b,c,d]
G_{90} [mg/dl]	111.67 ± 27.45 58.00–198.00 105.90–117.44	149.75 ± 37.47 71.00–245.00 141.79–157.72	243.03 ± 53.63 114.00–326.00 216.75–269.31	<0.01[b,c,d]
G_{120} [mg/dl]	100.03 ± 20.21 51.00–139.00 95.79–104.28	133.52 ± 29.46 72.00–194.00 127.25–139.78	231.69 ± 54.47 105.00–346.00 205.00–258.38	<0.01[b,c,d]
I_0 [μUI/ml]	6.76 ± 8.18 2.00–55.00 5.04–8.48	12.50 ± 20.36 2.00–154.00 8.17–16.82	15.70 ± 15.50 2.00–57.90 8.10–23.29	<0.01[b,c,d]
I_{30} [μUI/ml]	65.38 ± 55.50 12.00–293.00 53.72–77.04	80.15 ± 56.93 16.80–300.00 68.04–92.25	63.38 ± 58.96 15.00–233.00 34.48–92.27	<0.01[b,d]
I_{60} [μUI/ml]	72.22 ± 67.32 5.00–300.00 58.07–86.37	99.60 ± 69.67 10.00–300.00 84.79–114.41	87.14 ± 73.39 26.00–300.00 51.18–123.11	<0.01[b,c,d]
I_{90} [μUI/ml]	64.43 ± 64.74 8.79–300.00 50.83–78.03	96.32 ± 66.42 7.00–300.00 82.20–110.44	94.57 ± 74.13 26.00–300.00 58.25–130.89	<0.01[b,c,d]
I_{120} [μUI/ml]	53.74 ± 54.59 3.76–300.00 42.27–65.22	86.88 ± 62.35 7.00–300,00 73.63–100.14	100.87 ± 88.02 20.00–300.00 57.74–144.00	<0.01[b,c,d]
AUC_I [(μUI/ml) min]	6968.43 ± 6056.84 1248.60–31780.50 5695.68–8241.17	9772.55 ± 6137.39 2055.00–31087.50 8467.79–11077.31	9101.12 ± 7424.06 2385.00–30358.50 5463.33–12738.91	<0.01[b,d]
AUC_G^i [(mg/ml) min]	7452.76 ± 1591.46 4425.00–12435.00 7118.34–7787.18	9715.06 ± 1967.07 5175.00–13905.00 9296.88–10133.24	14170.78 ± 2956.88 7815.00–19650.00 12721.91–15619.65	<0.01[b,c,d]

[a]Statistically significant difference if p-value < 0.01.
[b]Statistically significant differences between control and prediabetic group.
[c]Statistically significant differences between control and diabetic group.
[d]Statistically significant differences between prediabetic and diabetic group.
[e]Control subjects are those who do not belong to any of the groups with pathology.
[f]Average and standard deviation.
[g]Maximum and minimum value.
[h]95% confidence interval.
[i]The AUC_G was calculated from the glucose values of 30, 60 and 90 min from OGTT since the values of glucose of 0 and 120 min from OGTT are dependent variables of prediabetes and T2D diagnosis.

The two experiments using AUC_I and AUC_G were made as follow:

i. First experiment: The number of clusters used was set to k = 2, in order to see if the algorithm is able to group the observations into subjects with and without T2D.
ii. Second Experiment: The number of clusters used was set to k = 3 to group observations according to T2D, prediabetes and control group.

To evaluate the performance of *k*-means algorithm clustering to group T2D, prediabetic and control subjects, the accuracy (ACC), precision (P) and recall (R) were computed in each experiment [25].

2.3 Statistical Analysis

Two nonparametric statistical tests were used in this study:

i. The Kruskal-Wallis nonparametric statistical test was used to find significant differences between groups of three or more variables.
ii. The Mann-Whitney U test was used as post-hoc to determine the differences between groups of two.

A p value less than or equal to 5% was considered to be statistically significant [26]. The data in the text and in the tables are presented as values of mean and standard deviation, minimum and maximum values and 95% confidence interval.

Table 2. Best clustering results for k = 2, k = 3, accuracy, precision, recall and SC using AUC_I.

AUC_I		SC	T2D (n = 16)	Prediabetic (n = 85)	Control (n = 87)
k = 2 ACC = 0.79[g] P = 0.49 R = 0.50	Cluster 1 (n = 162)	0.71 ± 0.19	14[a] 87.5%[b]	70[c] 82.4%[d]	78[e] 89.7%[f]
	Cluster 2 (n = 26)		2 12.5%	15 17.6%	9 10.3%
k = 3 ACC = 0.55 P = 0.44 R = 0.42	Cluster 1 (n = 135)	0.64 ± 0.18	11 68.8%	53 62.4%	71 81.6%
	Cluster 2 (n = 12)		2 12.5%	6 7.1%	4 4.6%
	Cluster 3 (n = 41)		3 18.8%	26 30.6%	12 13.8%

[a]Number of subjects belongs to the T2D group and classify to a respective cluster.
[b]Percentage of subjects belongs to the T2D group and classify to a respective cluster.
[c]Number of subjects belongs to the prediabetic group and classify to a respective cluster.
[d]Percentage of subjects belongs to the prediabetic group and classify to a respective cluster.
[e]Number of subjects belongs to the control group and classify to a respective cluster.
[f]Percentage of subjects belongs to the control group and classify to a respective cluster.
[g]To construct the confusion matrix for the calculation of accuracy, precision and recall, the assignment of each cluster was made as follow: In the case of k = 2, the cluster 1 and cluster 2 were assigned as a predicted control-prediabetic and cluster 2 predicted T2D group, respectively. In the case of k = 3, the cluster 2, cluster 3 and cluster 1 were assigned as s predicted T2D, prediabetics and control groups, respectively.

Table 3. Best clustering results for k = 2, k = 3, accuracy, precision, recall and SC using AUC_G.

AUC_G		SC	T2D (n = 16)	Prediabetic (n = 85)	Control (n = 87)
k = 2 ACC = 0.76[g] P = 0.62 R = 0.83	Cluster 1 (n = 60)	0.60 ± 0.18	15[a] 93.8%[b]	39[c] 45.9%[d]	6[e] 6.9%[f]
	Cluster 2 (n = 128)		1 6.3%	46 54.1%	81 93.1%
k = 3 ACC = 0.70 P = 0.65 R = 0.75	Cluster 1 (n = 78)	0.62 ± 0.19	1 6.3%	55 64.7%	22 25.3%
	Cluster 2 (n = 80)		1 6.3%	16 18.8%	63 72.4%
	Cluster 3 (n = 30)		14 87.5%	14 16.5%	2 2.3%

[a]Number of subjects belongs to the T2D group and classify to a respective cluster.
[b]Percentage of subjects belongs to the T2D group and classify to a respective cluster.
[c]Number of subjects belongs to the prediabetic group and classify to a respective cluster.
[d]Percentage of subjects belongs to the prediabetic group and classify to a respective cluster.
[e]Number of subjects belongs to the control group and classify to a respective cluster.
[f]Percentage of subjects belongs to the control group and classify to a respectively cluster.
[g]To construct the confusion matrix for the calculation of accuracy, precision and recall, the assignment of each cluster were made as follow: In the case of k = 2, the cluster 2 and cluster 1 were assigned as a predicted control-prediabetic group and predicted T2D group, respectively. In the case of k = 3, the cluster 3, cluster 1 and cluster 2 were assigned as a predicted T2D, prediabetics and control groups, respectively.

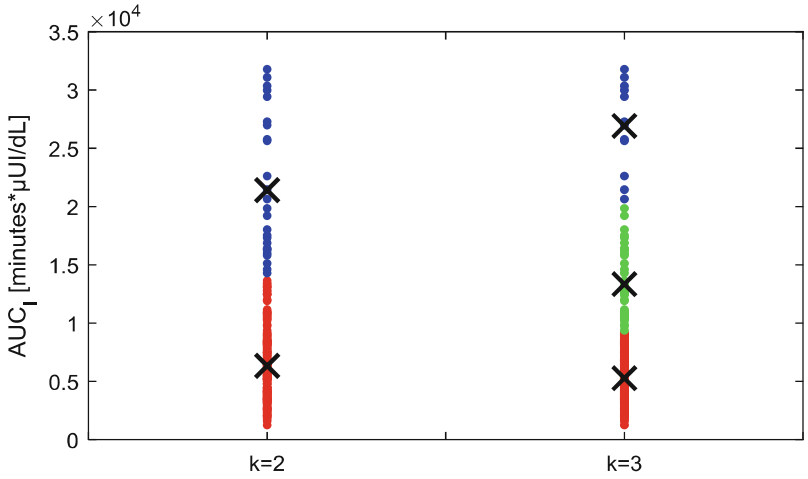

Fig. 1. Assignment of individuals (circles) to clusters for k = 2 and k = 3, using AUC_I. For k = 2, red circles belong to cluster 1 and blue circles to cluster 2, while for k = 3, red circles belong to cluster 1, blue circles to cluster 2 and green circles to cluster 3. Cluster centroids are represented by the character X.

3 Results

Table 1 reports the glucose and insulin values of the OGTT, the values of AUC_I and the modified version of AUC_G for the subjects with prediabetes, T2D and the control group. The database consists of 188 subjects, 46.6% belongs to the control group, 45.2% belongs to the prediabetic group and 8.5% endure T2D. Additionally, 25.9% of the subjects from the prediabetic group suffer from IGT and IFG in concomitance.

Tables 2 and 3 show the best clustering results of AUC_I and AUC_G, additionally, they report the silhouette coefficient, accuracy, precision and recall for k = 2 and k = 3. In the case of k = 2, to estimate the accuracy, precision and recall, the control and the prediabetics subjects were unified as one only group.

Figures 1 and 2 show the assignment of individuals to clusters for k = 2 and k = 3, using AUC_I and AUC_G, respectively, and the character X represents the centroids in each cluster.

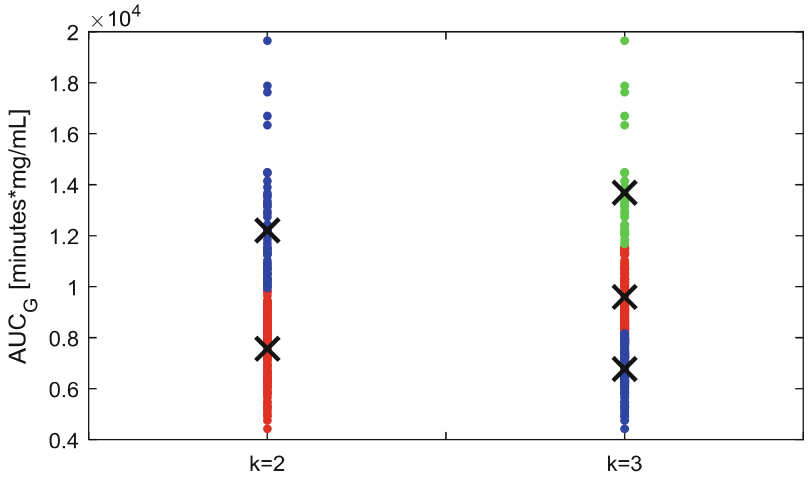

Fig. 2. Assignment of individuals (circles) to clusters for k = 2 and k = 3 using AUC_G. For k = 2, red circles belong to cluster 1 and blue circles to cluster 2, while for k = 3, red circles belong to cluster 1, blue circles to cluster 2 and green circles to cluster 3. Cluster centroids are represented by the character X.

4 Discussion

Table 1 shows significant statistical differences in the ages between all groups. The subjects with T2D show higher values of age than the prediabetic subjects, and these have a higher age than the control subjects. This could indicate that the prevalence of T2D as well as the risk of developing diabetes increase with age [27].

On the other hand, the AUC_I of the control and T2D subjects have not statistically significant differences, nevertheless, in the AUC_G, the T2D subjects have a higher AUC_G than the control subjects. All this indicates that control and T2D subjects have a similar insulin production; however, that is not reflected in the metabolism of glucose. This is consistent with researches that indicate the production of insulin does not ensure the correct metabolization of glucose [28]. In the same sense, it could be observed that prediabetic subjects have higher values of AUC_I compared to control subjects. It was also observed that the AUC_G is significantly higher in the prediabetic subjects compared to the control subjects. All this indicates that alterations in glucose absorption can be seen even in a prediabetic phase [29].

The *k*-means clustering performed for AUC_I and AUC_G presents acceptable results since the SC is above 0.6 in all cases. In Table 2, it can be seen that the *k*-means from the AUC_I values are not able to differentiate between prediabetic, T2D and the control subjects. In the case of k = 2, most of the T2D (87.5%), prediabetic (82.4%) and the control (89.7%) subjects are located in cluster 1. In the case of k = 3, most of the T2D (68.8%), prediabetic (62.4%) and the control (81.6%) subjects are located in cluster 2. On the other hand, it can be stated that although the AUC_I can identify the subjects without T2D, which explains the high accuracy in the case of k = 2 (ACC = 0.79), it cannot detect the subjects with T2D, which it is reflected in the low recall and precision values (P = 0.49, R = 0.50).All these indicate that the *k*-means method is not capable to differentiate between prediabetic, T2D and control with the AUC_I in this sample.

In Fig. 1, it can be observed that the subjects with higher AUC_I are in the cluster 1, and the subjects with lower AUC_I in the cluster 2, it is known that the higher AUC_I is related with the insulin resistance [30]. Insulin resistance is a condition that can appear in normal subjects, prediabetics subjects, or T2D subjects indistinctly. These facts suggest that the incapability of AUC_I *k*-means clustering to differentiate subjects with T2D and prediabetes could be due to the presence of insulin resistance in the three groups that bias the results [30].

Table 3 shows that AUC_G *k*-means clustering was able to classify the T2D, prediabetic and control subjects with accuracy and recall above 0.70 and precision above 0.60. Additionally, in the case of k = 2, in cluster 1 was located the 93.8% of the T2D subjects and 45.9% of prediabetic subjects; in cluster 2 was located the 93.1% and 54.1% of control and prediabetics subjects respectively. The prediabetic subjects located in cluster 1, 56% endure IFG and IGT in concomitance, also indicating that the *k*-means method was able to differentiate those subjects with high probabilities of developing T2D [30]. In the case of k = 3, it could be observed that 87.5% of the T2D subjects were located in cluster 3, 72.4% of the control subjects were located in cluster 2 and 64.7% of the prediabetic subjects were located in cluster 1. The 16.5% of prediabetics subjects located in cluster 3 suffer from IFG and IGT in concomitance; also they have the highest values of AUC_I and AUC_G. All these are in consistency with the findings with the k = 2 *k*-means clustering, where it was observed that the *k*-means method classifies as diabetic those subjects with high probabilities of developing T2D [31].

5 Conclusions

The findings in this study indicate that the *k*-means applied in the AUC_G classify subjects with T2D, prediabetes and the control group. Furthermore, it could even predict those subjects with high probabilities of developing T2D.

On the other hand, it was observed that the classifier does not work in the case of the AUC_I. This could be due to the fact that the T2D and the control subjects could produce comparable amounts of insulin, thus, the difference is not perceptible in the insulin values but in the values of glucose, because although they produce insulin, the correct metabolism of glucose does not occur. Another explanation is the presence of other associated diseases such as insulin resistance in all groups that could be biasing the results.

In this work, an unsupervised machine learning technique was used to identify subjects with T2D and prediabetes, other machine learning techniques such as neural networks and support vector machine will be explored in the future to assess the T2D diagnosis. Additionally, since the increment of the age has a direct proportional relationship with the prevalence of T2D, it would be interesting to investigate machine learning techniques using the age as a variable for the study.

Acknowledgments. This work was funded by the Research Vice-rectorate of the University Antonio Nariño, and the Research and Development Deanery of the Simón Bolívar University (DID) and Pontifical Bolivarian University.

References

1. Rowley, W.R., Bezold, C., Arikan, Y., Byrne, E., Krohe, S.: Diabetes 2030: insights from yesterday, today, and future trends. Popul. Health Manage. **20**(1), 6–12 (2017)
2. Nathanson, D., Sabale, U., Eriksson, J.W., Nyström, T., Norhammar, A., Olsson, U., Bodegård, J.: Healthcare cost development in a type 2 diabetes patient population on glucose-lowering drug treatment: a nationwide observational study 2006–2014. PharmacoEconomics-open **2**(4), 393–402 (2018)
3. Islam, S.M.S., Lechner, A., Ferrari, U., Laxy, M., Seissler, J., Brown, J., Holle, R.: Healthcare use and expenditure for diabetes in Bangladesh. BMJ Glob. Health **2**(1), e000033 (2017)
4. Eshwari, K., Kamath, V.G., Rao, C.R., Kamath, A.: Annual cost incurred for the management of type 2 diabetes mellitus—a community-based study from coastal Karnataka. Int. J. Diabet. Dev. Countries **39**(3), 590–595 (2019)
5. Karter, A.J., Parker, M.M., Solomon, M.D., Lyles, C.R., Adams, A.S., Moffet, H.H., Reed, M.E.: Effect of out-of-pocket cost on medication initiation, adherence, and persistence among patients with type 2 diabetes: the diabetes study of Northern California (DISTANCE). Health Serv. Res. **53**(2), 1227–1247 (2018)
6. Misra, A., Gopalan, H., Jayawardena, R., Hills, A.P., Soares, M., Reza-Albarrán, A.A., Ramaiya, K.L.: Diabetes in developing countries. J. Diabetes **11**(7), 522–539 (2019)
7. Narayan, K.V., Fleck, F.: The mysteries of type 2 diabetes in developing countries. Bull. World Health Organ. **94**, 241–242 (2016)

8. Dagogo-Jack, S.: Primary prevention of type 2 diabetes: an imperative for developing countries. In: Diabetes Mellitus in Developing Countries and Underserved Communities, pp. 7–31. Springer, Cham (2017)
9. American Diabetes Association2: Classification and diagnosis of diabetes: standards of medical care in diabetes—2018. Diabetes Care, **41**(Suppl. 1), S13-S27 (2018)
10. Kim, J.Y., Michaliszyn, S.F., Nasr, A., Lee, S., Tfayli, H., Hannon, T., Arslanian, S.: The shape of the glucose response curve during an oral glucose tolerance test heralds biomarkers of type 2 diabetes risk in obese youth. Diabetes Care **39**(8), 1431–1439 (2016)
11. Hays, L.M., Hoen, H.M., Slaven, J.E., Finch, E.A., Marrero, D.G., Saha, C., Ackermann, R.T.: Effects of a community-based lifestyle intervention on change in physical activity among economically disadvantaged adults with prediabetes. Am. J. Health Educ. **47**(5), 266–278 (2016)
12. Khan, T., Tsipas, S., Wozniak, G.: Medical care expenditures for individuals with prediabetes: the potential cost savings in reducing the risk of developing diabetes. Popul. Health Manage. **20**(5), 389–396 (2017)
13. Zheng, Y., Ley, S.H., Hu, F.B.: Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat. Rev. Endocrinol. **14**(2), 88 (2018)
14. Qu, H.Q., Li, Q., Rentfro, A.R., Fisher-Hoch, S.P., McCormick, J.B.: The definition of insulin resistance using HOMA-IR for Americans of Mexican descent using machine learning. PLoS One **6**(6), e21041 (2011)
15. Patil, B.M., Joshi, R.C., Toshniwal, D.: Hybrid prediction model for type-2 diabetic patients. Expert Syst. Appl. **37**(12), 8102–8108 (2010)
16. Velásquez, J., Severeyn, E., Herrera, H., Encalada, L., Wong, S.: Anthropometric index for insulin sensitivity assessment in older adults from Ecuadorian highlands. In: 12th International Symposium on Medical Information Processing and Analysis, vol. 10160, p. 101600S. International Society for Optics and Photonics, January 2017
17. Velásquez, J., Herrera, H., Encalada, L., Wong, S., Severeyn, E.: Análisis dimensional de variables antropométricas y bioquímicas para diagnosticar el síndrome metabólico. Maskana **8**, 57–67 (2017)
18. Velásquez, J., Severeyn, E., Herrera, H., Astudillo-Salinas, F., Wong, S.: Dimensional analysis of heart rate variability parameters for metabolic dysfunctions diagnosis. In: 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM), pp. 1–6, October 2017
19. Severeyn, E., Velásquez, J., Herrera, H., Wong, S.: Random sub-sampling cross validation for empirical correlation between heart rate variability, biochemical and anthropometrics parameters. In: Conference on Information Technologies and Communication of Ecuador, pp. 347–357. Springer, Cham (2018)
20. Potteiger, J.A., Jacobsen, D.J., Donnelly, J.E.: A comparison of methods for analyzing glucose and insulin areas under the curve following nine months of exercise in overweight adults. Int. J. Obesity **26**(1), 87 (2002)
21. Abdul-Ghani, M.A., Lyssenko, V., Tuomi, T., DeFronzo, R.A., Groop, L.: The shape of plasma glucose concentration curve during OGTT predicts future risk of type 2 diabetes. Diabetes/Metab. Res. Rev. **26**(4), 280–286 (2010)
22. World Health Organization: Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation (2006)
23. Hartigan, J.A., Wong, M.A.: Algorithm as 136: a kmeans clustering algorithm. J. Roy. Stat. Soc. Ser. C (Appl. Stat.) **28**(1), 100–108 (1979)
24. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)
25. Powers, D.: Evaluation: from precision, recall and F1-measure to ROC, informedness, markedness and correlation (2011)

26. Marusteri, M., Bacarea, V.: Comparing groups for statistical differences: how to choose the right statistical test? Biochem. Medica: Biochem. Medica **20**(1), 15–32 (2010)
27. Menke, A., Casagrande, S., Geiss, L., Cowie, C.C.: Prevalence of and trends in diabetes among adults in the United States, 1988–2012. JAMA **314**(10), 1021–1029 (2015)
28. Tangvarasittichai, S.: Oxidative stress, insulin resistance, dyslipidemia and type 2 diabetes mellitus. World J. Diabetes **6**(3), 456 (2015)
29. Burgeiro, A., Cerqueira, M., Varela-Rodríguez, B., Nunes, S., Neto, P., Pereira, F., Carvalho, E.: Glucose and lipid dysmetabolism in a rat model of prediabetes induced by a high-sucrose diet. Nutrients **9**(6), 638 (2017)
30. Vintimilla, C., Wong, S., Astudillo-Salinas, F., Encalada, L., Severeyn, E.: An aide diagnosis system based on *k*-means for insulin resistance assessment in eldery people from the Ecuadorian highlands. In: 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM), pp. 1–6, October 2017
31. Anjana, R.M., Rani, C.S.S., Deepa, M., Pradeepa, R., Sudha, V., Nair, H.D., Mohan, V.: Incidence of diabetes and prediabetes and predictors of progression among Asian Indians: 10-year follow-up of the Chennai urban rural epidemiology study (CURES). Diabetes Care **38**(8), 1441–1448 (2015)