

Diabetes prediction based on discrete and continuous mean amplitude of glycemic excursions using machine learning

by Artikel 9

Submission date: 17-Jan-2024 11:04AM (UTC+0700)

Submission ID: 2272323026

File name: Diabetes_prediction_based_on_discrete_and_continuous_mean.pdf (704.3K)

Word count: 5444

Character count: 29679

Diabetes prediction based on discrete and continuous mean amplitude of glycemic excursions using machine learning

Lailis Syafaah¹, Setio Basuki², Fauzi Dwi Setiawan Sumadi³, Amrul Faruq⁴, Mauridhi Hery Purnomo⁵

^{1,4}Department of Electrical Engineering, Faculty of Engineering, Universitas Muhammadiyah Malang, Indonesia

^{2,3}Department of Informatics Engineering, Faculty of Engineering, Universitas Muhammadiyah Malang, Indonesia

⁵Department of Electrical Engineering, Institute Teknologi Sepuluh Nopember, Indonesia

Article Info

Article history:

Received Jan 29, 2020

Revised Mar 22, 2020

Accepted May 3, 2020

Keywords:

Chronic hyperglycemia
Diabetes mellitus
Glycemic variability
Machine learning
Mean amplitude of glycemic excursion (MAGE)

ABSTRACT

Chronic hyperglycemia and acute glucose fluctuations are the two main factors that trigger complications in diabetes mellitus (DM). Continuous and sustainable observation of these factors is significant to be done to reduce the potential of cardiovascular problems in the future by minimizing the occurrence of glycemic variability (GV). At present, observations on GV are based on the mean amplitude of glycemic excursion (MAGE), which is measured based on continuous blood glucose data from patients using particular devices. This study aims to calculate the value of MAGE based on discrete blood glucose observations from 43 volunteer patients to predict the diabetes status of patients. Experiments were carried out by calculating MAGE values from original discrete data and continuous data obtained using Spline Interpolation. This study utilizes the machine learning algorithm, especially k-Nearest Neighbor with dynamic time wrapping (DTW) to measure the distance between time series data. From the classification test, discrete data and continuous data from the interpolation results show precisely the same accuracy value that is equal to 92.85%. Furthermore, there are variations in the MAGE value for each patient where the diabetes class has the most significant difference, followed by the pre-diabetes class, and the typical class.

⁹
This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Lailis Syafaah,
Department of Electrical Engineering,
Faculty of Engineering,
Universitas Muhammadiyah Malang,
246 Tlogomas, 65144, Malang, Indonesia.
Email: lailis@umm.ac.id

1. INTRODUCTION

Diabetes mellitus (DM) sufferer's prediction status is important since this disease has become a major issue in the world. Based on studies in the world population with age ranging from 20-79, people with diabetes in 2017 around 424.9 Million will reach 628.6 Million in 2045 [1]. Due to DM is very closely related to body metabolism, it is important to monitor the blood vessels function to guarantee that they work normally. Monitoring the blood vessel is also crucial to obtain the fluctuation of blood glucose levels [2]. Mean amplitude of glucose excursion (MAGE) as one of the glycaemic variability is a method for measuring the blood glucose fluctuations associated with body metabolism. The position of the MAGE, which is directly related to the blood vessels makes it appropriate to be used in predicting a person's diabetes status. To obtain the MAGE value, blood glucose fluctuations are observed continuously with continuous glucose monitoring (CGM) that transmits the blood glucose level every 5 minutes.

The idea of glycaemic variability (GV) boosted popularity as a tool for categorizing the unique properties of blood glucose (BG) in particularly of the major increased use and reliability of CGM systems.

Journal homepage: <http://beei.org>

Using classification based on impaired glucose tolerance (IGT), the Authors in [3] compared with type-2 diabetes (T2D) using CGM data. 62 subjects of time-series data were collected as a CGM dataset. It the ability to distinguish between the subjects affected by impaired glucose tolerance versus T2D. Moreover, using simple linear support vector machine classifier can distinguish IGT and T2D by reduced set of CGM-based glycaemic variability indices [4].

Recently, it is reported that glycaemic variability is related to the severity of coronary artery disease in patients with poorly measured type 2 diabetes and acute myocardial infarction [5]. Reference [6] proposed an easy metric that may be utilized by clinicians to quickly assess the glycaemic variability standing of patients and thereby determine those patients with continuing high levels of glycaemic variability. The new measurement the glycaemic fluctuation share (GVP), gave a quantitative measuring of glycaemic inconstancy over a given interim of your time by dissecting the length of the CGM worldily follow standardized to the period underneath examination.

Although previously published studies demonstrated that MAGE can be measured by observing blood glucose data continuously, generally reported as in [7-9]. However, the provision of CGM for personal and mass use is considered very expensive. In this study, blood samples were taken for several days using CGM. This results in blood glucose data that can be easily obtained in the community cannot be used as data for research related to MAGE [10]. To overcome this problem, our previous study [11] used discrete data interpolated for the measurement of MAGE. In that study, used Spline interpolation techniques for smoothing discrete data of 21 points were observed within three days. The interpolation technique is implemented by forming 864 interpolation points. This value is obtained from simulating the amount of data generated by CGM for three days. If CGM transmits data every 5 minutes, in 1 hour there will be 12 data, in 1 day there will be 288 data, and three days there will be 864 data. The size of the 3-day observation was taken from the study in [12]. From the experimental results, the linear spline technique produces the lowest RMSE value than the other techniques, namely quadratic and cubic spline. The issue that we tend to raise in this study is measuring MAGE costs based on discrete data to be used to predict a person's with diabetes status. This study revealed that discrete data is capable to predict the MAGE costs utilizing machine learning technique.

2. RESEARCH METHOD

To answer the predetermined research question, this research was conducted by the following number of workflows shown in Figure 1.

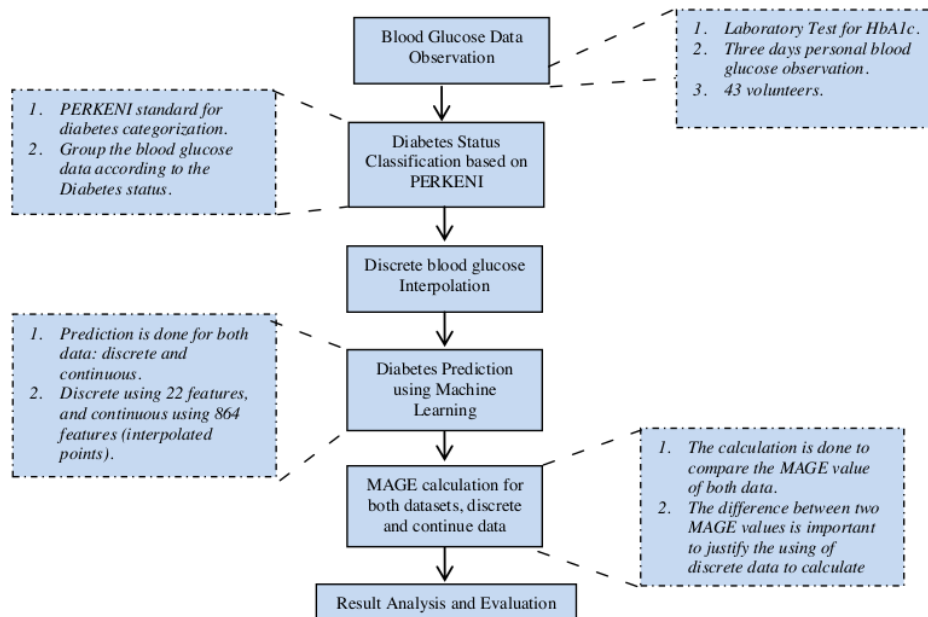


Figure 1. Process flow of diabetes prediction based on MAGE

2.1. Blood glucose data observation

In this study, blood glucose data was used in our previous study [11]. This data was obtained from 43 volunteer patients who were recorded using a monitoring card as shown in Table 1. In addition to recording blood glucose values based on a given scenario, all patients also recorded each food consumed. This is needed for further analysis of the trend of increasing and decreasing blood glucose values before and after eating. After expressing a willingness to volunteer, patients are also asked to do blood tests conducted at a health center or hospital. This test is intended to obtain fasting glucose and HbA1c values that are used to classify the diabetes status of each patient.

Table 1. Blood glucose monitoring card

Name / Age / Occupation		 / / / / / /						
HbA1c		 %											
Fasting Glucose		 (mg/dl)											
Day ... (1/2/3)								Before Breakfast	1 hour after Breakfast	Before Lunch	1 hour after Lunch	Before Dinner	1 hour after Dinner	Before Bedtime
Observation Time		
Blood Glucose Level	Hyper Level	>300
		260-299
		220-256
	Normal Level	180-219
		145-179
		101-144
Hype Level	80-100	
	50-79	
			<50	

Breakfast Menu:
Lunch Menu:
Dinner Menu:

2.2. Diabetes status classification based-on PERKENI

When starting to introduce the scenario of taking blood glucose to each patient, a brief interview was conducted to obtain information regarding diabetes status. This interview aims to get an even distribution of each category of diabetes status, namely Normal, Pre-diabetes, and Diabetes. However, trying to balance the number of patients in each category is not easy due to two factors: i) the patient is not willing to volunteer, and ii) the patient has never checked diabetes status before. Furthermore, data preparation is done to form a classification dataset from blood glucose observation data. As a reference, the process of labeling the dataset uses the *Perkumpulan Endokrinologi Indonesia* (PERKENI) criterion [13], which consists of three classes as previously mentioned namely Normal, Pre-diabetes, and Diabetes based on HbA1c values. Table 2 shows the PERKENI standard for classifying all three patients with diabetes status.

Table 2. PERKENI standard for diabetes classification

Parameters	HbA1c (%)	Fasting glucose (mg/dl)	Plasma glucose 2 hours after TTGO (mg/dls)
Diabetes	>=6.5	>=126	>=200
Pre-diabetes	5.7-6.4	100-125	140-199
Normal	<5.7	<100	<140

2.3. Blood glucose interpolation with linear spline

Linear spline interpolation (first-order spline) is used to find the value of a point by connecting ordered points using linear function. This technique is the simplest piece of a polynomial function. The resulting interpolation curve is relatively similar to the non-linear interpolation model. The curve contains a steep slope change in the data. The linear spline formula for the ordered data $x_0, x_1, x_2, ..$ and x_n as follows in (1):

$$\begin{aligned}
 f(x) &= f(x_0) + m_0(x - x_0) & x_0 \leq x \leq x_1 \\
 f(x) &= f(x_1) + m_1(x - x_1) & x_1 \leq x \leq x_2 \\
 &... & ... \\
 f(x) &= f(x_{n-1}) + m_{n-1}(x - x_{n-1}) & x_{n-1} \leq x \leq x_n
 \end{aligned}
 \tag{1}$$

with (slope) straight line m_i is stated in (2):

$$m_i = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} \quad (2)$$

This approach is used to interpolate 22 points on the observation of blood glucose to 864 points. The number of interpolation points is obtained assuming that in CGM sends data every 5 minutes, then in an hour, it will transmit 12 data. Then, there will be 288 data per day. Thus, within 3 days there will be 864 data.

2.4. Time series classification and dynamic time warping (DTW)

The classification was carried out to determine the diabetes status of patients based on observations for three days based on the k-Nearest Neighbor (kNN) algorithm. The classification performed on the two types of data namely discrete and continue is intended to compare again that the two types of data produce a classification accuracy that is nearly identical. In other words, discrete data can be used as a valid representation of continued observations. Because the dataset used is data in the form of sequences (time series), this research uses the dynamic time warping (DTW) algorithm as the similarity measurement.

The use of DTW is intended to measure the similarity of two or more sequences that are unequally spaced time points. More specifically, points measured by distance (as a classification attribute) are not always at the same time (non-linear). For instance, observations of taking blood glucose at point seven in discrete data are not always done at the same time. Thus, the similarity measure cannot be done based on Euclidean distance alone. Thus, the classification of both discrete and continuous data is performed using the kNN algorithm with DTW to measure the distance between sequences based on the best alignment.

The principle works for DTW technique as described in [14]: given time series sequences t and r with length m and n as it is illustrated in Figure 2. DTW algorithm is used to search for mapping the path $\{(p1, q1), (p2, q2), \dots, (pk, qk)\}$ with aiming to minimize the distance on $\sum_{i=1}^k |t(pi) - r(i)|$ with restriction: (1) $(p1, q1) = (1,1)$, $(pk, qk) = (m, n)$, (2) for every node (i, j) on path, there is a restriction that $(i - 1, j)$, $(i, j - 1)$, $(i - 1, j - 1)$. Thus, to obtain the optimum distance value is done based on forward DP:

- The optimum-value function is defined as $D(i, j)$ as DTW-distance between $t(1:i)$ and $r(1:j)$
- Recursion Function as stated in (3):

$$D(i, j) = |t(i) - r(j)| + \min \begin{cases} D(i - 1, j) \\ D(i - 1, j - 1) \\ D(i, j - 1) \end{cases} \quad (3)$$

Initial condition $D(1,1) = |t(1) - r(1)|$

- Final answer: $D(m, n)$

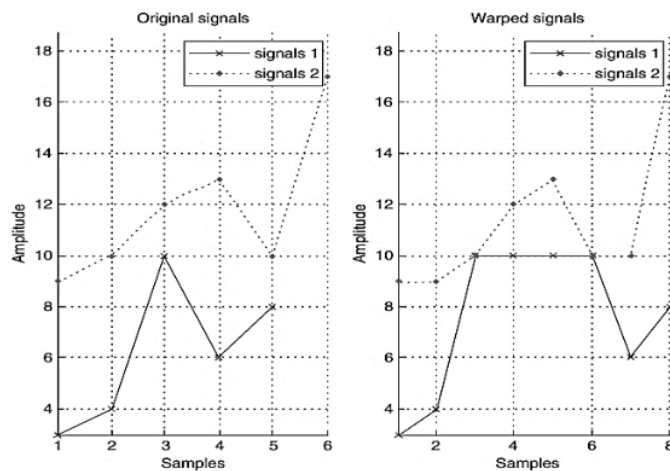


Figure 2. Original signal and warped signal with DTW [15]

2.5. Calculation of MAGE

Several studies have shown the bad effects of sustained chronic hyperglycemia that results in excessive protein glycation and activates oxidative stress [16]. Role variability glucose (glycemic) is less documented, but the value of the average fluctuation of glucose will activate oxidative stress. Thus, it is recommended that the treatment strategies for diabetes should be directed to reduce to the minimum of the various components of Disglychemic. Glycemic variability is an important parameter that is used to resolve potential clinical problems in patients with diabetes. It is known that glycemic variability produces oxidative stress and potentially contribute to the development of macro and microvascular complications [12, 17]. Currently, the best measurement for assessing glycemic variability is by mean amplitude of glycemic excursion (MAGE).

However, MAGE not in routine clinical use. Routine measurement of glycemic variability clinical causes an important measure of overall glucose control. It predictors to the risk of complications of diabetes that are not detected by glycosylated hemoglobin levels (HbA1c) [18]. Glycated hemoglobin levels are regularly measured to monitor and evaluate the glycemic control of diabetics [19]. Good influence also reported by the DCCT (Diabetes Control Complications Trial) and by the UKPDS (the United Kingdom Prospective Diabetes Study) 1998: Reduction of microvascular complications contribute to the reduction of glycemic excursion (glycemic variable). Assessment of glycemic control in diabetic patients should include three parameters, describes as "glucose triad" [17]. Furthermore, referring to [20-22] these parameters are hemoglobin A1c (HbA1c), fasting plasma glucose (FPG), and postprandial glucose (PPG) as respectively as illustrated in Figure 3.

MAGE is a general size of the volatility of blood glucose levels, an indication level of diabetes control. Mage is usually used with a blood glucose monitoring system continuously (CGM). However, there were published medical studies that use MAGE algorithm by setting the data smaller, typically 7-10 observation per day for 2-3 days [12]. The MAGE value is calculated by dividing λ (the series of blood glucose levels) by x as a number of observations.

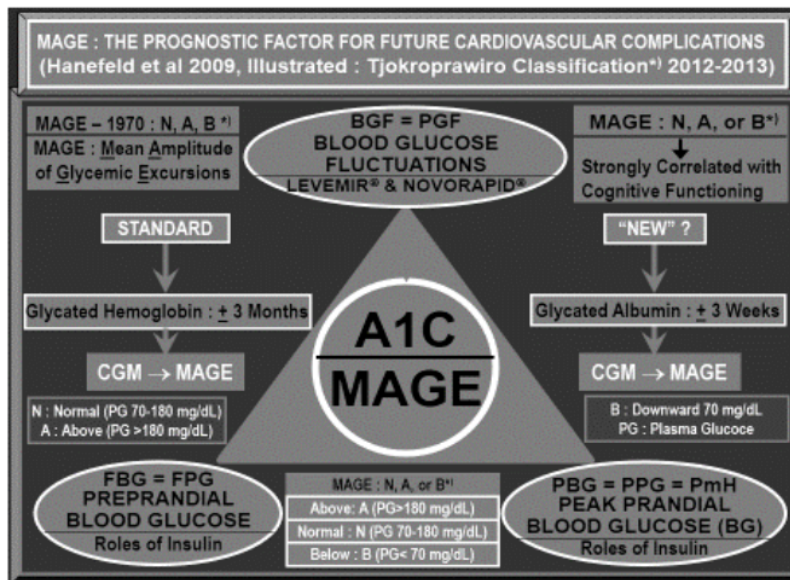


Figure 3. The illustration of three glycaemia parameters [20, 23]

Two aspects are important in the calculation of MAGE: first, not considering the MAGE excursion significant frequency, just average amplitude. Second, not considering the MAGE excursion glycemic which outside the normal range. MAGE only counts a high average amplitude of glucose that exceeds the standard deviation for a particular day (only include the value of the peak to peak on the bottom or vice versa) as shown in Figure 4, according to [12].

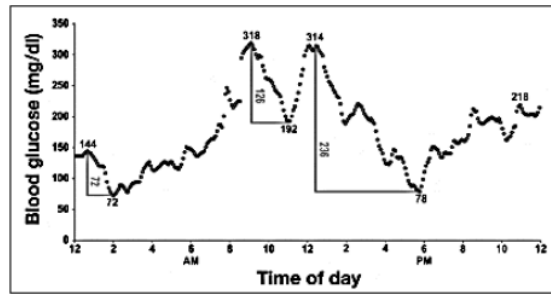


Figure 4. Example of MAGE calculation for 24 hours with CGM and SD=63 [12]

3. RESULTS AND DISCUSSION

3.1. Blood glucose observation result

Patients and the accompanying personnel recorded the results of blood glucose monitoring. Table 3 represents discrete blood sampling observation results from 43 volunteers (patients). The main difficulties in this observation are that the patients forgot to record the blood glucose level according to the suggested time and the inconsistency of the patients to take the blood sampling. If these conditions appear, then the observation should be repeated on the next day. Therefore, the role of personnel who monitor this observation becomes critical.

Screening value of fasting glucose is intended to categorize blood glucose data into three namely normal, pre-diabetes, and diabetes. The results of this stage are used for classification using a machine learning algorithm. Next, interpolation is done to obtain continuous values from discrete blood glucose data. The algorithm used is Linear Spline which in previous studies produced the smallest RMSE value [11]. Each patient's blood glucose data which amounts to 22 points will be interpolated to 864. In Figure 4, one example of the interpolation results is represented in graphical form.

Table 3. Three days blood glucose observation result form 43 volunteers and diabetes classification

Observation	Patient 1		Patient 2		Patient ...		Patient 43	
	Time in Minute	Blood Glucose	Time in Minute	Blood Glucose	Time in Minute	Blood Glucose	Time in Minute	Blood Glucose
HbA1c	5.6		5.8		...		9.7	
Fasting Glucose	
12	0	100	0	77	0	152
Day 1	1	326	58	335	112	...	364	252
	2	398	135	448	108	...	595	182
	3	753	106	662	121	...	950	262
	4	858	135	783	91	...	1096	139
	5	1105	89	1063	85	...	1187	161
Day 2	6	1191	149	1206	100	...	1320	141
	7	1368	103	1313	86	...	1543	156
	8	1805	87	1910	141	...	1893	214
	9	1878	104	2054	84	...	2047	181
	10	2184	81	2176	141	...	2302	220
Day 3	11	2262	131	2307	65	...	2429	156
	12	2558	121	2470	112	...	2590	199
	13	2819	106	2490	65	...	2724	169
	14	2826	121	2762	83	...	2946	164
	15	3213	92	3325	137	...	3225	220
	16	3319	126	3482	90	...	3359	153
	17	3596	81	3603	91	...	3837	188
	18	3660	125	3775	85	...	4000	172
	19	4037	104	3934	107	...	4082	239
	20	4177	130	4103	98	...	4220	200
	21	4369	111	4187	77	...	4330	152
Class	Normal		Pre-Diabetes				Diabetes	

Initial classification results were obtained in the form of 7 normal patients, 5 pre-diabetic patients, and 31 diabetes patients as shown in Figure 5(a). For classification-based prediction purposes, of course,

the blood glucose data obtained include imbalanced data. Such data distribution will not be able to form a fair classification model and cause classification bias. Thus, we use the synthetic minority over-sampling technique (SMOTE) algorithm to balance the distribution of data in each class [24]. Figure Figure 5(b) shows the distribution of new blood glucose data after the implementation of the SMOTE algorithm in which each class consists of 22 data.

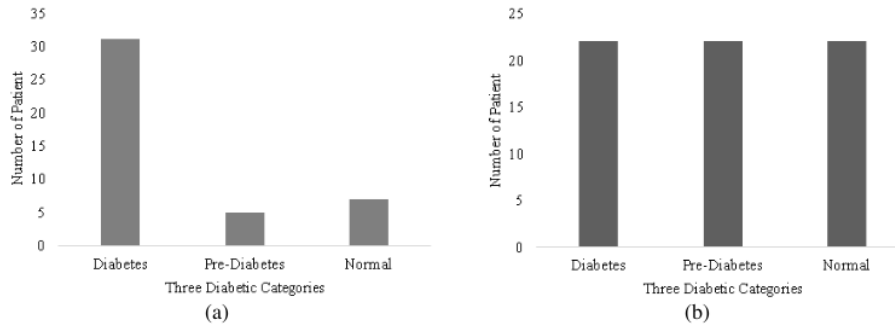


Figure 5. Blood glucose data, (a) Imbalance discrete, (b) Balanced using SMOTE

In the next stage, all blood glucose data were interpolated based on the linear Spline technique. Points generated from interpolation amounted to 864 according to the observations taken in every 5 minutes in a period of 3 days. Figure 6(a) shows the time series model of discrete observation data of blood glucose in one patient and in Figure 6(b) is the result of the interpolation of blood glucose data.

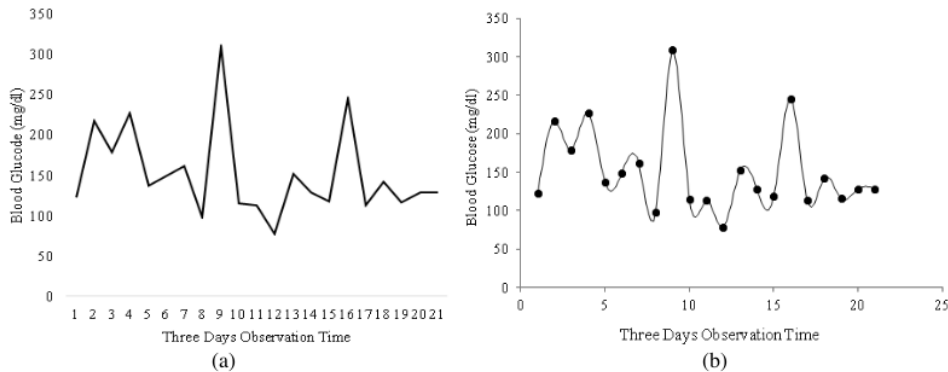


Figure 6. Blood glucose data, (a) Discrete, (b) Interpolated

3.2. Discrete and continuous classification comparison

Classification based prediction is done using kNN algorithm with DTW as a technique to calculate the distance between time series data. The data used for testing are data from the SMOTE data balancing results by distributing Normal, Pre-diabetes, and Diabetes data for each of the 22 data. Testing is done by splitting all data both discrete and continuous interpolation results to 80% (52 data) training data and 20% testing data (14 data). From the tests conducted, both discrete data and continuous data produce the same classification accuracy value, which is 92.85%. Furthermore, the confusion matrix, as shown in Table 4, is produced exactly where the test data for the Normal and Pre-diabetes categories can be classified correctly and the diabetes category contains 1 misclassification out of a total of 6 test data. Thus, from prediction testing using discrete data and continuous data, it can be described that discrete data with observations of patients for 3 days can represent observations that should be made continuously based on continuous glucose monitoring (CGM) taken every 5 minutes.

Table 4. Confusion matrix pregnancy risk prediction

	Normal	Pre-diabetes	Diabetes
Normal	3	0	0
Pre-diabetes	0	5	0
Diabetes	0	1	5

3.3. Discrete and continuous MAGE value comparison

In addition to classification-based testing, this study also examines the value of MAGE generated by both types of data, discrete and continuous. MAGE value is calculated by adding up all blood glucose values in all observations divided by the number of observations. In contrast to classification-based testing, MAGE values are not calculated using SMOTE balancing data. MAGE value is calculated using 43 original data from observations.

Table 5 shows the MAGE value calculation for both types of data. In general, it can be clearly seen that there is a difference in the value of MAGE for each patient in discrete and continuous data. In general, the difference in MAGE value is $0 < \text{Difference in MAGE} < 1$. However, in some patients with the Diabetes class, there is a difference in MAGE value that reaches a difference of two even a difference of three. In Figure 7 the average value of the MAGE value is shown in each class. In the blood glucose data used in this study, the largest MAGE difference is in the Diabetes class followed by the Pre-diabetes class, and the smallest Normal class. These results indicate that the higher the blood glucose value of the patient, the greater the difference in MAGE value. This can be caused because in the Pre-diabetes class and the Diabetes class, the blood pressure level tends to be more volatile than patients in the Normal class [25].

Table 5. Comparison of discrete and continuous MAGE value data

ID Patient	Class	Discrete MAGE	Continues MAGE
1	Diabetes	184.90	185.36
2	Diabetes	199.68	200.66
3	Diabetes	177.86	178.73
4	Diabetes	217.45	217.57
5	Pre-diabetes	97.81	98.41
6	Diabetes	222.50	222.05
7	Diabetes	178.90	179.81
8	Diabetes	195.68	195.03
9	Diabetes	244.09	243.26
10	Diabetes	323.54	326.40
11	Pre-diabetes	119.09	118.63
12	Diabetes	238.27	238.37
13	Diabetes	198.50	196.78
14	Normal	106.77	107.76
15	Diabetes	333.68	333.21
16	Diabetes	363.72	363.59
17	Diabetes	196.31	199.12
18	Pre-diabetes	133.40	134.76
19	Diabetes	386.04	386.35
20	Diabetes	400.59	403.66
21	Normal	117.77	118.28
22	Normal	115.90	117.30
23	Normal	92.72	92.96
24	Normal	115.81	115.96
25	Normal	100.45	100.48
26	Normal	105.68	106.10
27	Diabetes	152.68	153.72
28	Diabetes	282.22	283.88
29	Diabetes	151.54	151.90
30	Diabetes	142	142.06
31	Diabetes	179.63	179.47
32	Diabetes	256.95	257.06
33	Diabetes	151.36	151.99
34	Diabetes	198.95	198.50
35	Diabetes	165.90	166.59
36	Diabetes	400.59	403.66
37	Diabetes	233.63	233.11
38	Diabetes	227.54	228.33
39	Diabetes	176.22	175.80
40	Diabetes	308.68	309.95
41	Diabetes	282.22	283.88
42	Pre-diabetes	136.04	137.35
43	Pre-diabetes	100.54	101.07

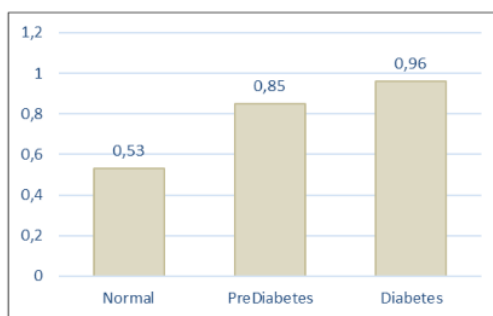


Figure 7. Average MAGE value between discrete data and continuous data

4. CONCLUSION

This study emphasizes to evaluate blood glucose data to predict the level of diabetes risk and calculate MAGE values. Discrete blood glucose values were compared with continuous interpolation data for the two tests. In this study used blood glucose data taken from 43 volunteer patients. Blood glucose data is taken 21 times in a span of 3 days. This observation resulted in three classes of diabetes risk level, namely the normal patients, 5 pre-diabetes patients, and 31 diabetes patients. For classification purposes, we use the SMOTE balancing technique which produces 22 data for each class.

It is proven from the conducted experiments, the prediction accuracy using the DTW-based kNN algorithm produces an identical value that is equal to 92.85%. Testing the calculation of MAGE values in both types of data shows the difference where the largest difference is experienced in the class of diabetes followed by the Pre-diabetes class and the Normal class. This can be caused due to patients who are included in the class of Pre-diabetes and Diabetes have blood glucose values that are more volatile.

ACKNOWLEDGEMENTS

The authors wish to express their gratitude to the Faculty of Engineering, Universitas Muhammadiyah Malang (UMM) for supporting this study. To the Department Electrical and Department of Computer Engineering, Institute Teknologi Sepuluh Nopember (ITS), Surabaya for facilitating and laboratory support of this research. The first author acknowledges the Ministry of Research, Technology, and Higher Education, Indonesia, under the post-doctoral research scheme to undertaking this work.

REFERENCES

- [1] International Diabetes Federation, "IDF diabetes atlas," Brussels, Belgium: International Diabetes Federation, vol. 8, 2017.
- [2] F. F. Li, L-Y Fu, W-L. Zhang, X-F. Su, J-D. Wu, J. Sun, L. Ye, and J-H. Ma, "Blood glucose fluctuations in type 2 diabetes patients treated with multiple daily injections," *Journal of Diabetes Research*, vol. 2016, pp. 1-8, 2016.
- [3] E. Longato, G. Acciaroli, A. Facchinetti, L. Hakaste, T. Tuomi, A. Maran, and G. Sparacino, "Glycaemic variability-based classification of impaired glucose tolerance vs. type 2 diabetes using continuous glucose monitoring data," *Computers in Biology and Medicine*, vol. 96, pp. 141-146, 2018.
- [4] E. Longato, G. Acciaroli, A. Facchinetti, A. Maran, and G. Sparacino, "Simple linear support vector machine classifier can distinguish impaired glucose tolerance versus type 2 diabetes using a reduced set of CGM-based glycaemic variability indices," *Journal of Diabetes Science and Technology*, vol. 14, no. 2, pp. 297-302, 2020.
- [5] M. Benalia, M. Zeller, B. Mouhat, C. Guenancia, V. Yameogo, C. Grecoa, H. Yao, M. Maza, B. Vergès, and Y. Cottina, "Glycaemic variability is associated with severity of coronary artery disease in patients with poorly controlled type 2 diabetes and acute myocardial infarction," *Diabetes & Metabolism*, vol. 45, no 5, pp. 446-452, 2019.
- [6] T. A. Peyser, A. K. Balo, B. A. Buckingham, I. B. Hirsch, and A. Garcia, "Glycemic variability percentage: A novel method for assessing glycemic variability from continuous glucose monitor data," *Diabetes Technology & Therapeutics*, vol. 20, no. 1, pp. 6-16, 2018.
- [7] G. Marics, Z. Lendvai, C. Lódi, L. Koncz, D. Zakariás, G. Schuster, B. Mikos, C. Hermann, A. J. Szabó, and P. Tóth-Heyn, "Evaluation of an open access software for calculating glucose variability parameters of a continuous glucose monitoring system applied at pediatric intensive care unit," *BioMedical Engineering OnLine*, vol. 14, no. 37, pp. 1-13, 2015.
- [8] X. Yu, L. Lin, J. Shen, Z. Chen, J. Jian, B. Li, and S. X. Xin, "Calculating the mean amplitude of glycemic excursions from continuous glucose data using an open-code programmable algorithm based on the integer nonlinear method," *Computational and Mathematical Methods in Medicine*, pp. 1-9, 2018.

Study on diabetes prediction based on discrete and continuous mean amplitude of... (Lailis Syafaah)

- [9] G. Fritzsche, K.-D. Kohnert, P. Heinke, L. Vogt, and E. Salzsieder, "The Use of a computer program to calculate the mean amplitude of glycemic excursions," *Diabetes Technology & Therapeutics*, vol. 13, no. 3, pp. 319-325, 2011.
- [10] D. Rodbard, "Glucose variability: a review of clinical applications and research developments," *Diabetes Technology & Therapeutics*, vol. 20, no. S2, pp. S25-S215, 2018.
- [11] S. Lailis, Mauridhi Hery Purnomo, and S. Basuki, "Discrete mean amplitude of glycemic excursion (MAGE) measurement on diabetics with spline interpolation method," *International Journal on Electrical Engineering and Informatics*, vol. 10, no. 2, pp. 259-270, 2018.
- [12] C. R. Marling, J. H. Shubrook, S. J. Vernier, M. T. Wiley, and F. L. Schwartz, "Characterizing blood glucose variability using new metrics with continuous glucose monitoring data," *Journal of Diabetes Science and Technology*, vol. 5, no. 4, pp. 871-878, 2011.
- [13] S. A. Soelistijo, H. Novida, A. Rudijanto, P. Soewondo, K. Suastika, A. Manaf, H. Sanusi, D. Lindarto, A. Shahab, B. Pramono, Y. A. Langi, D. Purnamasari, N. N. Soetedjo, M. R. Saraswati, M. P. Dwipayana, A. Yuwono, L. Sasirani, Sugiarto, K. W. Sucipto, and H. Zufry, "Consensus on the management and prevention of type 2 diabetes mellitus in Indonesia," *PB. PERKENI*, Indonesia, 2015.
- [14] R. Jang, "Data clustering and pattern recognition," CS Dept., National Taiwan University, Taiwan, 2005. [Online]. Available at: <http://mirlab.org/jang/books/dcpr>. [Accessed: 22-Jan-2019].
- [15] S. I. Ao, "Applied time series analysis and innovative computing 59th ed," Springer Netherlands, 2010.
- [16] L. Monnier and C. Colette, "Glycemic variability should we and can we prevent it?," *Diabetes Care*, vol. 31, no. 2, pp. 150-154, 2008.
- [17] F. Zaccardi, P. Di Stefano, E. Busetto, M. O. Federici, A. Manto, F. Infusino, G. A. Lanza, D. Pitocco, and G. Ghirlanda, "Group of signs: A new method to evaluate glycemic variability," *Journal of diabetes science and technology*, vol. 2, no. 6, pp. 1061-1065, 2008.
- [18] S. Costantino, F. Paneni, R. Battista, L. Castello, G. Capretti, S. Chiandotto, L. Tanese, G. Russo, D. Pitocco, G. A. Lanza, M. Volpe, T. F. Lüscher, and F. Cosentino, "Impact of glycemic variability on chromatin remodeling, oxidative stress and endothelial dysfunction in type 2 diabetic patients with target HbA_{1c}," *Diabetes*, vol. 66, no. 9, pp. 2472-2482, 2017.
- [19] S. H. Alzahrani, M. Baig, M. M. Aashi, F. K. Al-shaibi, D. A. Alqarni, and W. H. Bakhamees, "Association between glycosylated hemoglobin (HbA_{1c}) and the lipid profile in patients with type 2 diabetes mellitus at a tertiary care hospital: A retrospective study," *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, vol. 12, pp. 1639-1644, 2019.
- [20] M. Hanefeld, S. R. Bornstein, and F. Pistrosch, "Shifting the disease management paradigm from glucose," *Diabetes Care*, vol. 32, no. 2, pp. 353-356, 2009.
- [21] A. T. Reutens, R. Hutchinson, T. V. Binh, C. Cockram, C. Deerochanawong, L.-T. Ho, Li. Ji, B. A. K. Khalid, A. P. S. Kong, M. A. Lim-Abraham, C.-E. Tan, A. Tjokroprawiro, K.-H. Yoon, P. Z. Zimmet, and J. E. Shaw, "The GIANT study, a cluster-randomised controlled trial of efficacy of education of doctors about type 2 diabetes mellitus management guidelines in primary care practice §," *Diabetes Research and Clinical Practice*, vol. 98, no. 1, pp. 38-45, 2012.
- [22] M. Hanefeld, E. Duetting, and P. Bramlage, "Cardiac implications of hypoglycaemia in patients with diabetes—a systematic review," *Cardiovascular Diabetology*, vol. 12, no. 1, p. 135, 2013.
- [23] M. Hanefeld, S. Sulk, M. Helbig, A. Thomas, and C. Köhler, "Differences in glycemic variability between normoglycemic and prediabetic subjects," *Journal of diabetes science and technology*, vol. 8, no. 2, pp. 286-290, 2014.
- [24] R. A. Rawlings, H. Shi, L.-H. Yuan, W. Brehm, R. Pop-Busui, and P. W. Nelson, "Translating glucose variability metrics into the clinic via Continuous glucose monitoring: A graphical user interface for diabetes evaluation (CGM-GUIDE®)," *Diabetes Technology & Therapeutics*, vol. 13, no. 2, pp. 1241-1248, 2011.
- [25] Y. Huang, C. Heng, J. Wei, X. Jing, X. Wang, G. Zhao, J. Hou, Q. Liu, and K. Jiao, "Influencing factors of glycemic variability in hospitalized type 2 diabetes patients with insulin therapy," *Medicine*, vol. 96, no. 36, 2017.

BIOGRAPHIES OF AUTHORS



Lailis Syafaah is an Associate Professor at the Department of Electrical Engineering, Faculty of Engineering, University of Muhammadiyah Malang (UMM). She received her Bachelor and Master Degrees from Electrical Engineering Department, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, in 1989 and 2002 respectively. In 2014, she received a Doctoral Degree from Medical Department, Universitas Brawijaya, Malang, Indonesia. Currently, she actively conducts research on Biomedical Engineering.



Setio Basuki was born in Indonesia, in 1984. He graduated from Institut Teknologi Telkom (IT Telkom), Indonesia, 2007, for his Bachelor's Degree in Informatics Engineering. He received a Master Degree in Informatics Engineering from Institut Teknologi Bandung (ITB), Indonesia, 2015. His research interests are the applications of Machine Learning and Natural Language Processing. Since 2009, he has worked as a lecturer in the Informatics Engineering Department, University of Muhammadiyah Malang (UMM), Indonesia.



Fauzi Dwi Setiawan Sumadi achieved his bachelor's degree in 2014 at Bandung Telkom University. He studied informatics engineering. Subsequently, in 2017 he graduated from the University of Queensland Australia for his master's degree program in computer science. Nowadays, he becomes one of the main lecturers in informatics at Universitas Muhammadiyah Malang and maintains his research in the implementation of artificial intelligence, software defined network, and distributed computer.



Amrul Faruq, is an electrical engineer and computer science enthusiast. He obtained bachelor and master's degree in Electrical Engineering in 2009 and 2013, from Universitas Muhammadiyah Malang and Universiti Teknologi Malaysia, respectively. Starting from 2018 onwards he is a PhD candidate in the Malaysia-Japan International Institute of Technology (MJIT), Universiti Teknologi Malaysia, Kuala Lumpur. His research interests about computational data science and optimization algorithms.



Mauridhi Hery Purnomo received his bachelor's degree from Electrical Eng., Institut Teknologi Sepuluh Nopember (ITS) in 1984, Master and Ph.D degree from Electrical Engineering Osaka City University (OCU) in 1989 and 1995 respectively. He is a Professor in Electrical Eng., ITS with research interest in artificial intelligent, deep learning, image and signal processing, renewable energy, online condition and monitoring system.

Diabetes prediction based on discrete and continuous mean amplitude of glycemic excursions using machine learning

ORIGINALITY REPORT

12%

SIMILARITY INDEX

9%

INTERNET SOURCES

9%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

- 1** www.slideshare.net 1%
Internet Source
- 2** Submitted to Cavite State University 1%
Student Paper
- 3** Zaccardi, F., P. Di Stefano, E. Busetto, M. O. Federici, A. Manto, F. Infusino, G. A. Lanza, D. Pitocco, and G. Ghirlanda. "Group of Signs: A New Method to Evaluate Glycemic Variability", Journal of Diabetes Science and Technology, 2008. 1%
Publication
- 4** web.sites.google.com 1%
Internet Source
- 5** www.liebertpub.com 1%
Internet Source
- 6** Lailis Syafaah, Desy Fatkhi Azizah, Inda Rusdia Sofiani, Merinda Lestandy, Amrul Faruq. "Self-Monitoring and Detection of Diabetes with art Toilet based on Image

Processing and K-Means Technique", 2020
IEEE International Conference on Automatic
Control and Intelligent Systems (I2CACIS),
2020

Publication

7 doaj.org 1 %
Internet Source

8 www.science.gov 1 %
Internet Source

9 Achmad Komarudin, Novendra Setyawan,
Leonardo Kamajaya, Mas Nurul Achmadiah,
Zulfatman Zulfatman. "Signature PSO: A novel
inertia weight adjustment using fuzzy
signature for LQR tuning", Bulletin of
Electrical Engineering and Informatics, 2021
Publication

10 liebertpub.com 1 %
Internet Source

11 Marling, C. R., J. H. Shubrook, S. J. Vernier, M.
T. Wiley, and F. L. Schwartz. "Characterizing
Blood Glucose Variability Using New Metrics
with Continuous Glucose Monitoring Data",
Journal of Diabetes Science and Technology,
2011.
Publication

12 d-scholarship.pitt.edu 1 %
Internet Source

13

helda.helsinki.fi

Internet Source

1 %

Exclude quotes On

Exclude matches < 1%

Exclude bibliography On



Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author: Artikel 9
Assignment title: Lailis Syfaah
Submission title: Diabetes prediction based on discrete and continuous mean...
File name: Diabetes_prediction_based_on_discrete_and_continuous_me...
File size: 704.3K
Page count: 11
Word count: 5,444
Character count: 29,679
Submission date: 17-Jan-2024 11:04AM (UTC+0700)
Submission ID: 2272323026

Bulletin of Electrical Engineering and Informatics
Vol. 9, No. 6, December 2020, pp. 2619–2629
ISSN: 2302-9285, DOI: 10.11591/eei.v9i6.2387

2619

Diabetes prediction based on discrete and continuous mean amplitude of glycemic excursions using machine learning

Lailis Syfaah¹, Setio Basuki², Fauzi Dwi Setiawan Sumodi³, Anrul Faruq⁴, Maarithi Hery Purmono⁵
¹Department of Electrical Engineering, Faculty of Engineering, Universitas Muhammadiyah Malang, Indonesia
²Department of Informatics Engineering, Faculty of Engineering, Universitas Muhammadiyah Malang, Indonesia
³Department of Electrical Engineering, Institute Teknologi Sepuluh Nopember, Indonesia

Article Info

Article history:
Received Jan 29, 2020
Revised Mar 22, 2020
Accepted May 3, 2020

Keywords:

Chronic hyperglycemia
Diabetes mellitus
Glycemic variability
Machine learning
Mean amplitude of glycemic excursions (MAGE)

ABSTRACT

Chronic hyperglycemia and acute glucose fluctuations are the two main factors that trigger complication in diabetes mellitus (DM). Continuous and sustainable observation of these factors is significant to be done to reduce the potential of cardiovascular problems in the future by minimizing the occurrence of glycemic variability (GV). At present, observations on GV are based on the mean amplitude of glycemic excursions (MAGE), which is measured based on continuous blood glucose data from patients using particular devices. This study aims to calculate the value of MAGE based on discrete blood glucose observations from 43 volunteer patients to predict the diabetes status of patients. Experiments were carried out by calculating MAGE values from original discrete data and continuous data obtained using Spline Interpolation. This study utilizes the machine learning algorithm, especially k-Nearest Neighbor with dynamic time wrapping (DTW) to measure the distance between time series data. From the classification test, discrete data and continuous data from the interpolation results show precisely the same accuracy value that is equal to 92.85%. Furthermore, there are variations in the MAGE value for each patient where the diabetes class has the most significant difference, followed by the pre-diabetes class, and the typical class.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Lailis Syfaah,
Department of Electrical Engineering,
Faculty of Engineering,
Universitas Muhammadiyah Malang,
246 Tlogomas, 65144, Malang, Indonesia.
Email: lailis@um.ac.id

1. INTRODUCTION

Diabetes mellitus (DM) sufferer's prediction status is important since this disease has become a major issue in the world. Based on studies in the world population with age ranging from 20-79, people with diabetes in 2017 around 424.9 Million will reach 628.6 Million in 2045 [1]. Due to DM is very closely related to body metabolism, it is important to monitor the blood vessels function to guarantee that they work normally. Monitoring the blood vessel is also crucial to obtain the fluctuation of blood glucose levels [2]. Mean amplitude of glucose excursion (MAGE) as one of the glycemic variability is a method for measuring the blood glucose fluctuations associated with body metabolism. The position of the MAGE, which is directly related to the blood vessels makes it appropriate to be used in predicting a person's diabetes status. To obtain the MAGE value, blood glucose fluctuations are observed continuously with continuous glucose monitoring (CGM) that transmits the blood glucose level every 5 minutes.

The idea of glycemic variability (GV) has boosted popularity as a tool for categorizing the unique properties of blood glucose (BG) in particular of the major increased use and reliability of CGM systems.

Journal homepage: <http://beei.org>