# Deep-learning-based characterization of glucose biomarkers to identify type 2 diabetes, prediabetes, and healthy individuals

**Sina Akbarian**
Klick Applied Sciences
Klick Inc.
175 Bloor Street East, Toronto, Canada
sakbarian@klick.com

**Qayam Jetha**
Klick Applied Sciences
Klick Inc.
175 Bloor Street East, Toronto, Canada
qjetha@klick.com

**Jouhyun Jeon**
Klick Applied Sciences
Klick Inc.
175 Bloor Street East, Toronto, Canada
cjeon@klick.com

## Abstract

Type 2 Diabetes (T2D) is a common chronic disease that can lead to serious comorbidities. Prediabetes is a state of increased health risk that is defined by abnormal glucose homeostasis and is strongly associated with the development of T2D and diabetic complications. Novel diagnostic or screening tools are required to identify T2D and prediabetic patients. In this study, we developed a predictive model that uses continuous glucose monitoring (CGM) signals to classify individuals as T2D, prediabetic, or healthy. We tested different durations of CGM signals to determine the minimum length of time required to achieve a reliable prediction of diabetic outcomes. We found that 12 hours of CGM signals were sufficient to achieve a classifier with a high degree of accuracy. The performance of the 12-hour model was equivalent to the performance of a model using the full period of CGM signals. The 12-hour model achieved AUCs of 0.83, 0.69, and 0.77 to identify T2D, prediabetes, and healthy individuals, respectively. The overall AUC of the 12-hour ensemble model was 0.86. Our findings propose a new application of currently available CGM systems to identify T2D and prediabetes based on only a short-time series of glucose profiles.

## 1 Introduction

Diabetes is a chronic metabolic disease characterized by elevated blood glucose (BG) levels due to insufficient insulin production or decreased insulin sensitivity [1]. Diabetes has become a major public health problem worldwide with an estimated 463 million people diagnosed. T2D accounts for approximately 90% of the diagnosed population [2]. T2D can lead to serious damage to the heart, eyes, kidneys, and nerves if patients fail to receive appropriate treatment on time [1]. Prediabetes is a state of increased health risk that is defined by abnormal glucose homeostasis, such as impaired fasting glucose or impaired glucose tolerance [3]. It is increasingly recognized as a critical metabolic state, since prediabetic individuals are at increased risk of developing T2D and diabetic complications. The global prevalence of prediabetes was estimated at 7.3% of the adult population in 2017 [4]. It also has been reported that about 90% of prediabetic U.S adults were unaware of their condition [5].

Workshop on Learning from Time Series for Health, 36th Conference on Neural Information Processing Systems (NeurIPS 2022).

When potential T2D is detected early, patients may delay or prevent progression to diabetes. Disease prevention is significantly less expensive than the treatment of T2D and diabetic complications. Therefore, developing a new test to identify potential patients with T2D or prediabetes is essential for improving patient outcomes and quality of life, reducing the medical burden, and enhancing overall population health. Here, we propose a novel predictive model that can classify T2D patients, prediabetic patients, and healthy individuals using continuously reported blood glucose levels. To do this, we recruited 436 participants (172 T2D, 87 prediabetes and 177 healthy individuals) and measured the blood glucose levels of participants as they went about their daily lives. Next, we investigated the effect of different durations of CGM signals on the identification of patient types, and found the minimum duration of CGM signals required to build a reliable predictive model.

## 2 Related Work

A CGM measures blood glucose levels every 15 minutes for a maximum of 14 days. In addition to blood glucose levels, CGM systems provide alerts if hypo- or hyper-glycemic events, which can lead to serious medical complications [6, 7], are detected. CGM systems also provide information on glucose variability, such as the degree of BG level fluctuation, variance from average BG levels, and the frequency of variations [8].

Researchers have explored various applications of CGM data by integrating machine-learning algorithms such as XGBoost [9], convolutional neural network [10], and recurrent neural network [11, 12] to predict future BG levels in 30-60 minutes, hypoglycemia events [12, 13], and hyperglycemia events [13] in diabetic patients. CGM data, also, has been used to predict the level of glycated hemoglobin (HbA1c), whose elevated level is significantly associated with the risk of developing diabetic complications [14].

Recently, stratification of different patient types using CGM data has been investigated. From the analysis of CGM profiles, subsets of glycemic variability indices that distinguished healthy individuals from subjects with diabetes [15] and distinguished individuals with impaired glucose tolerance from T2D patients [16, 17] have been proposed. Through a multilevel clustering approach, four subgroups of T2D patients showing distinct CGM profiles and physiological characteristics were proposed [18]. Detrended fluctuation function has been applied to the CGM data to classify T1D and T2D patients [19]. Nevertheless, several challenges have remained. Many studies used a small number of samples which restricted applying machine- or deep-learning algorithms to build a classifier, and achieve reliable performance to identify T2D, prediabetes, and healthy individuals. Furthermore, it still remains unknown what time duration of CGM signals can capture unique glucose profiles for patients with T2D, prediabetes, and healthy individuals (i.e. how many hours of CGM recordings are sufficient to classify patient types).

## 3 Methods

### 3.1 Data

#### 3.1.1 Collecting CGM signals and demographic information

In this study, 568 participants were recruited from 4 states in India. The study was performed in accordance with relevant guidelines and regulations, and informed consent was obtained from all participants prior to study entry. Participants' blood glucose levels were measured using a FreeStyle Libre Pro glucose monitoring device (Abbott Diabetes Care). Participants were requested to report sex and age, and measured height, weight, and body mass index (BMI) at the first day of wearing a device. HbA1c concentration was measured at the last day of wearing a device. On average, each participant wore a CGM device for 12 days (288 hours), and we considered 288 hours as the full-time duration of CGM signals. 132 participants failed to measure BG levels due to a malfunctioning CGM device or failed to collect demographic information and were therefore removed from our dataset.

#### 3.1.2 Defining T2D, prediabetes and healthy individuals

T2D patients have HbA1c $> 6.5\%$. Their disease states are confirmed by medical doctors. Prediabetes patients have HbA1c concentration between $5.5\%$ and $6.5\%$. Healthy individuals have HbA1c $< 5.5\%$.

Table 1: Patient types and demographic characteristics

| | T2D | Prediabetes | Healthy individual |
|---|---|---|---|
| Number | 172 | 87 | 177 |
| Sex (Female / Male) | 50 / 122 | 29 / 58 | 69 / 108 |
| Age (years) | $51 \pm 9$ | $40 \pm 12$ | $30 \pm 10$ |
| BMI | $27.42 \pm 5.08$ | $27.47 \pm 4.80$ | $27.32 \pm 5.41$ |
| Height (cm) | $164 \pm 8.9$ | $164 \pm 7.4$ | $162 \pm 9.8$ |
| Weight (Kg) | $73.45 \pm 13.01$ | $73.61 \pm 12.32$ | $71.35 \pm 12.92$ |
| HbA1c (%) | $8.59 \pm 1.66$ | $5.85 \pm 0.28$ | $4.94 \pm 0.48$ |

* Average value and standard deviation are reported.

HbA1c level of 6.5% is considered as a symptom of hyperglycemia or diabetes [20, 21]. HbA1c level of 5.5% is related to the development of diabetes [22] and cardiovascular disease [23]. It is also associated with increased all-cause mortality of patients with acute ischemic stroke [24].

Of the 436 participants used for analysis, 172 were T2D, 87 were prediabetes and 177 were healthy individuals. Since T2D tends to be occurred more often in older adults [25], we observed T2D patients are relatively older than prediabetic patients or healthy individuals. A summary of patient types and demographic characteristics is shown in Table 1.

### 3.2 Predictive model development

#### 3.2.1 Preprocessing CGM signals and building predictive models

To examine the effect of CGM signal duration on the prediction of patient types, we generated 5 CGM datasets that represented different time durations (window sizes): 12, 24, 72, 168, and 288 hours. To make a balanced training dataset, a window was slided with a stride size of 150 minutes (T2D), 30 minutes (prediabetes), and 75 minutes (healthy individual). We only considered windows without missing values for further analysis. Next, the 1D CGM signal with a given window size was converted to a 2D spectrogram and fed into convolutional neural networks (CNNs). We used CGM signals from 70% of the participants to train and validate predictive models, while the remaining 30% of participants were used to test model performance. For the test set, we adopted a stride size of 150 minutes for all patient types. Adam was used to optimize a class-weighted cross-entropy loss. The categorical accuracy was used for selecting the best model based on validation. The initial value for the learning rate was 0.001, and 128 epochs were used to train the model with a batch size of 128-512 (depending on the input size). The total number of trainable parameters in this network were between 400,000 and 1,000,000 depending on the input size.

#### 3.2.2 Generating baseline models using demographic information

As a baseline model, we built a random forest (RF) and XGBoost classifiers using patients' demographic information, such as gender, age, height, weight, and BMI. Using grid-search, random forest parameters (the number of trees, depth of each tree, number of features selected at random, and class weight) and XGBoost parameters (number of trees, depth of each tree, and number of features to choose from at every split in a given tree) were optimized to have the highest categorical accuracy.

## 4 Results and Discussion

### 4.1 Performance of CGM-based models on patient type classification

We investigated the prediction ability of CGM signals to classify T2D, prediabetes, and healthy individuals. To examine whether the length of wearing a CGM device affected the prediction accuracy, we processed the CGM data into short-time (12, 24, 72, and 168 hour windows) and full-time (288 hour windows) BG readings. Processed datasets were fed into CNN models to identify distinct patient types. CGM-based models outperformed the baseline models (Table 2).

Both RF- and XGBoost-based baseline models had an overall balanced accuracy (BCC) of 0.48 (95% confidence interval (CI) = 0.41 to 0.54) for RF and BCC of 0.54 (95% CI = 0.47 to 0.60) for XGBoost.

Table 2: Performance of window-based predictions depending on the time duration of CGM signals

| | Overall performance (95% CI) | | Patient type-specific AUC (95% CI) | | |
|---|---|---|---|---|---|
| window size | BCC | AUC* | Healthy individual | Prediabetes | T2D |
| 12 hours | 0.67 (0.66-0.68) | 0.78 (0.773-0.784) | 0.77 (0.76-0.78) | 0.69 (0.68-0.70) | 0.83 (0.821-0.835) |
| 24 hours | 0.67 (0.66-0.68) | 0.79 (0.781-0.792) | 0.76 (0.75-0.77) | 0.66 (0.65-0.67) | 0.87 (0.868-0.880) |
| 72 hours | 0.68 (0.675-0.692) | 0.79 (0.785-0.796) | 0.77 (0.76-0.78) | 0.68 (0.67-0.69) | 0.86 (0.85-0.87) |
| 168 hours | 0.67 (0.66-0.68) | 0.80 (0.787-0.805) | 0.80 (0.79-0.81) | 0.62 (0.61-0.64) | 0.87 (0.86-0.88) |
| 288 hours | 0.67 (0.65-0.70) | 0.79 (0.78-0.82) | 0.81 (0.79-0.84) | 0.63 (0.59-0.67) | 0.86 (0.84-0.88) |
| Demographic info. only (RF) | 0.48 (0.41 - 0.54) | 0.65 (0.58-0.71) | 0.67 (0.57-0.77) | 0.49 (0.35-0.63) | 0.71 (0.62-0.81) |
| Demographic info. only (XGBoost) | 0.54 (0.47 - 0.60) | 0.70 (0.63-0.76) | 0.76 (0.67-0.85) | 0.52 (0.38-0.66) | 0.72 (0.63-0.82) |

* Weighted AUC was measured to correct imbalanced data.

Table 3: Performance of ensemble predictions depending on the time duration of CGM signals

| | Overall performance (95% CI) | | Patient type-specific AUC (95% CI) | | |
|---|---|---|---|---|---|
| Ensemble of windows | BCC | Weighted AUC | Healthy individual | Prediabetes | T2D |
| 12 hours | 0.81 (0.74-0.86) | 0.86 (0.81-0.91) | 0.86 (0.80-0.93) | 0.84 (0.77-0.92) | 0.87 (0.80-0.93) |
| 24 hours | 0.75 (0.66-0.82) | 0.84 (0.79-0.90) | 0.84 (0.77-0.91) | 0.73 (0.62-0.84) | 0.90 (0.84-0.96) |
| 72 hours | 0.75 (0.65-0.83) | 0.83 (0.77-0.89) | 0.82 (0.74-0.90) | 0.76 (0.66-0.86) | 0.86 (0.80-0.93) |
| 168 hours | 0.68 (0.59-0.75) | 0.81 (0.75-0.86) | 0.81 (0.73-0.89) | 0.64 (0.52-0.77) | 0.88 (0.82-0.94) |
| 288 hours | 0.68 (0.60-0.76) | 0.80 (0.75-0.86) | 0.81 (0.73-0.89) | 0.63 (0.51-0.76) | 0.88 (0.81-0.94) |
| Demographic info. only (RF) | 0.48 (0.41 - 0.54) | 0.65 (0.58-0.71) | 0.67 (0.57-0.77) | 0.49 (0.35-0.63) | 0.71 (0.62-0.81) |
| Demographic info. only (XGBoost) | 0.54 (0.47 - 0.60) | 0.70 (0.63-0.76) | 0.76 (0.67-0.85) | 0.52 (0.38-0.66) | 0.72 (0.63-0.82) |

Baseline models showed the lowest performance to detect prediabetic patients (Area Under the Curve (AUC) = 0.49 for RF and 0.52 for XGBoost). CGM-based models improved overall BCC by 140% and AUC by 122%. CGM-based models also showed 1.21, 1.34, and 1.17 times higher prediction accuracy to identify T2D, prediabetes, and healthy individuals, respectively. Our findings imply that physiological indices could be weak indicators whereas CGM signals have the predictive power to classify different patient types.

We found that the 12-hour model achieved a competitive performance compared to the longer duration time-based models. The 12-hour model showed an overall BCC of 0.67 (95% CI = 0.66 to 0.68), which was equivalent to the performance of the model using the full-time CGM signals (Table 2). The 12-hour model had an overall AUC of 0.78 (95% CI = 0.773 to 0.784), which was similar to those of other models. The model also showed the highest accuracy to identify prediabetes (AUC = 0.69), and had high ability to predict T2D (AUC = 0.83) and healthy individuals (AUC = 0.77). Our results suggest that using only 4% of the entire duration of CGM signals is enough to capture unique BG characteristics of individual patient types.

### 4.2 Ensemble analysis and prediction of patient types

To examine whether ensemble predictions improved predictive performance, we assigned an overall individual patient type based on the majority of short-time predictions. Ensemble predictions showed higher accuracy (average 109% increase) and weighted AUC (average 105% increase) compared to window-based predictions. Interestingly, we found that predictive performance improved more when shorter-time CGM signals were combined. For example, the 12-hour ensemble model achieved 120% and 111% improvements in BCC and AUC compared to the 12-hour model. Meanwhile, we only found 101% - 103% improvements in BCC and AUC for the 168-hour and 288-hour ensemble models. The 12-hour ensemble model showed the highest overall BCC (0.81, 95% CI = 0.74-0.86) and weighted AUC (0.86, 95% CI = 0.81-0.91). Patient type-specific AUCs were 0.87, 0.84, and 0.86 for T2D, prediabetes, and healthy individuals, respectively (Table 3). We suspect that the short-time series of CGM signals contain information to identify distinct patient types, and their aggregation could enhance the predictive performance.

### 4.3 CGM signal and demographic characteristics of miss-classified participants

From the 12-hour ensemble prediction, we found that 23 participants were misclassified. To understand the cause of misclassification, we examined the changes in BG levels for these participants. We found that misclassified participants were likely to have unusual BG measurement patterns. For example, misclassified participants tended to report the same BG levels continuously over time. We counted the number of continuous blocks, which are a group of continuous timestamps with the

same BG level, and found that misclassified participants have continuous blocks 1.2 times more than correctly predicted participants (Mann-Whitney U test, P = 0.004). We suspect that misclassified participants generated 12-hour windows which were mainly composed of constant CGM signals, and these unusual BG measurements caused poor window-based predictions leading to the failure of ensemble prediction. No significant association was observed between misclassification and demographic indices: age (P = 0.36), height (P = 0.20), weight (P = 0.21), BMI (P = 0.38), and HbA1c (P = 0.27).

## 5 Conclusion

In this study, we demonstrated the potential application of CGM systems for the quick and accurate screening of diabetic outcomes. 12 hours of glucose profiles provide enough information to identify T2D, prediabetes, and healthy individuals, and are sufficient to achieve a classifier with a high degree of accuracy.

Based on our knowledge, our predictive model is the first CNN-based model which is trained and tested by over 400 participants' CGM signals, and classifies T2D, prediabetes and healthy individuals. Several studies have been conducted to identify patients with impaired glucose tolerance or diabetes [15–17, 19]. These studies collected CGM signals from a small number of participants ($\sim 100$ participants) and mainly focused on the classification of two patient groups (e.g. T2D vs impaired glucose tolerance or T1D vs T2D or healthy individual vs diabetic patients). Therefore, a direct comparison between our model and previous studies is challenging. We suggest that our predictive model could be benchmarked to other similar research studies.

Future developments should consider the imputation strategies for missing BG values to reflect the real-world behavior of the CGM device, and the application of CNN-based predictive models on larger holdout datasets with shorter durations of CGM signals (e.g. less than 12 hours) to evaluate the robust classification of patient types in different experimental settings.

## 6 Acknowledgements

Thanks to all of the participants, project members, supporters, and researchers at Klick Inc. for the successful development, implementation, and evaluation of this research. On behalf of all authors, the corresponding author states that there is no conflict of interest.

## References

[1] A. D. Association, "Diagnosis and classification of diabetes mellitus," *Diabetes care*, vol. 32, no. Supplement_1, pp. S62–S67, 2009.

[2] R. Goyal and I. Jialal, "Diabetes mellitus type 2," 2018.

[3] A. Tabák, C. Herder, W. Rathmann, E. Brunner, and M. Kivimäki, "Prediabetes: a high-risk state for diabetes development. lancet [internet]," 2012.

[4] U. Hostalek, "Global epidemiology of prediabetes-present and future perspectives," *Clinical diabetes and endocrinology*, vol. 5, no. 1, pp. 1–5, 2019.

[5] C. for Disease Control, Prevention, U. D. of Health, H. Services *et al.*, "Centers for disease control and prevention national diabetes statistics report, 2017," *Centers for Disease Control and Prevention, US Department of Health and Human Services: Atlanta, GA, USA*, 2017.

[6] A. Chawla, R. Chawla, and S. Jaggi, "Microvasular and macrovascular complications in diabetes mellitus: distinct or continuum?" *Indian journal of endocrinology and metabolism*, vol. 20, no. 4, p. 546, 2016.

[7] D. Aronson, "Hyperglycemia and the pathobiology of diabetic complications," *Cardiovascular diabetology: Clinical, metabolic and inflammatory facets*, vol. 45, pp. 1–16, 2008.

[8] B. Kovatchev and C. Cobelli, "Glucose variability: timing, risk analysis, and relationship to hypoglycemia in diabetes," *Diabetes Care*, vol. 39, no. 4, pp. 502–510, 2016.

[9] J. Jeon, P. J. Leimbigler, G. Baruah, M. H. Li, Y. Fossat, and A. J. Whitehead, "Predicting glycaemia in type 1 diabetes patients: experiments in feature engineering and data imputation," *Journal of healthcare informatics research*, vol. 4, no. 1, pp. 71–90, 2020.

[10] T. Zhu, K. Li, P. Herrero, J. Chen, and P. Georgiou, "A deep learning algorithm for personalized blood glucose prediction." in *KHD@ IJCAI*, 2018, pp. 64–78.

[11] J. Iwasa, G. Myklebust, L. Engebretsen, R. Bahr, and T. Krosshaug, "Mechanisms for noncontact anterior cruciate ligament injuries," *Am J Sports Med*, vol. 38, no. 11, pp. 2218–2225, 2010.

[12] J. Martinsson, A. Schliep, B. Eliasson, C. Meijner, S. Persson, and O. Mogren, "Automatic blood glucose prediction with confidence using recurrent neural networks," in *Khd@ ijcai*, 2018.

[13] S. M. Pappada, B. D. Cameron, and P. M. Rosman, "Development of a neural network for prediction of glucose concentration in type 1 diabetes patients," *Journal of diabetes science and technology*, vol. 2, no. 5, pp. 792–801, 2008.

[14] S. I. Sherwani, H. A. Khan, A. Ekhzaimy, A. Masood, and M. K. Sakharkar, "Significance of hba1c test in diagnosis and prognosis of diabetic patients," *Biomarker insights*, vol. 11, pp. BMI–S38 440, 2016.

[15] G. Acciaroli, G. Sparacino, L. Hakaste, A. Facchinetti, G. M. Di Nunzio, A. Palombit, T. Tuomi, R. Gabriel, J. Aranda, S. Vega *et al.*, "Diabetes and prediabetes classification using glycemic variability indices from continuous glucose monitoring data," *Journal of diabetes science and technology*, vol. 12, no. 1, pp. 105–113, 2018.

[16] 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 glycemic variability indices," *Journal of diabetes science and technology*, vol. 14, no. 2, pp. 297–302, 2020.

[17] 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.

[18] R. Tao, X. Yu, J. Lu, Y. Shen, W. Lu, W. Zhu, Y. Bao, H. Li, and J. Zhou, "Multilevel clustering approach driven by continuous glucose monitoring data for further classification of type 2 diabetes," *BMJ Open Diabetes Research and Care*, vol. 9, no. 1, p. e001869, 2021.

[19] W. Liu, J. Chen, L. He, X. Cai, R. Zhang, S. Gong, X. Yang, J. Wang, X. Han, D. Shi *et al.*, "Flash glucose monitoring data analysed by detrended fluctuation function on beta-cell function and diabetes classification," *Diabetes, Obesity and Metabolism*, vol. 23, no. 3, pp. 774–781, 2021.

[20] G. Cavagnolli, J. Comerlato, C. Comerlato, P. Renz, J. Gross, and J. Camargo, "Hba1c measurement for the diagnosis of diabetes: is it enough?" *Diabetic medicine*, vol. 28, no. 1, pp. 31–35, 2011.

[21] T. Oshitari, N. Fujimoto, K. Hanawa, and E. Adachi-Usami, "Effect of chronic hyperglycemia on intraocular pressure in patients with diabetes," *American journal of ophthalmology*, vol. 143, no. 2, pp. 363–365, 2007.

[22] A. D. Association, "2. classification and diagnosis of diabetes: standards of medical care in diabetes—2018," *Diabetes care*, vol. 41, no. Supplement_1, pp. S13–S27, 2018.

[23] E. Selvin, M. W. Steffes, H. Zhu, K. Matsushita, L. Wagenknecht, J. Pankow, J. Coresh, and F. L. Brancati, "Glycated hemoglobin, diabetes, and cardiovascular risk in nondiabetic adults," *New England Journal of Medicine*, vol. 362, no. 9, pp. 800–811, 2010.

[24] S. Wu, C. Wang, Q. Jia, G. Liu, K. Hoff, X. Wang, A. Wang, C. Wang, X. Zhao, Y. Wang *et al.*, "Hba1c is associated with increased all-cause mortality in the first year after acute ischemic stroke," *Neurological research*, vol. 36, no. 5, pp. 444–452, 2014.

[25] M. A. B. Khan, M. J. Hashim, J. K. King, R. D. Govender, H. Mustafa, and J. Al Kaabi, "Epidemiology of type 2 diabetes–global burden of disease and forecasted trends," *Journal of epidemiology and global health*, vol. 10, no. 1, p. 107, 2020.