Journal of the Korean Physical Society (2021) 78:373–378
https://doi.org/10.1007/s40042-021-00056-8

Online ISSN 1976-8524
Print ISSN 0374-4884

ORIGINAL PAPER

Machine learning for the diagnosis of early-stage diabetes using temporal glucose profiles

Woo Seok Lee[1] · **Junghyo Jo**[2] · **Taegeun Song**[3]

Received: 14 October 2020 / Revised: 30 October 2020 / Accepted: 30 October 2020 / Published online: 1 February 2021

Abstract
Machine learning shows remarkable success for recognizing patterns in data. Here, we apply machine learning (ML) for the diagnosis of early-stage diabetes, which is known as a challenging task in medicine. Blood glucose levels are tightly regulated by two counter-regulatory hormones, insulin and glucagon, and the failure of glucose homeostasis leads to a common metabolic disease, diabetes mellitus. It is a chronic disease that has a long latent period that complicates detection of the disease at an early stage. The vast majority of diabetes cases result from that diminished effectiveness of insulin action, and that insulin resistance modifies the temporal profile of blood glucose. Thus, we propose to use ML to detect subtle changes in the temporal pattern of the glucose concentration. Time series data on blood glucose with sufficient resolution is currently unavailable, so we confirm the proposal by using synthetic glucose profiles produced using a biophysical model that considers glucose regulation and hormone action. Multi-layered perceptrons, convolutional neural networks, and recurrent neural networks all identified the degree of insulin resistance with high accuracy above 85%.

Keywords Machine learning · Diagnosing diabetes · Insulin resistance

1 Introduction

Glucose homeostasis is essential to stably supply fuel to the brain [1, 2]. Blood glucose levels (BGLs) are tightly regulated by hormones from the endocrine pancreas (Fig. 1a), and the normal fasting glucose concentration is about 4 mM [3]. The American Diabetes Association Guideline defines hyperglycemia as $5.6 < \mathrm{BGL} < 7$ mM, severe hyperglycemic ($\mathrm{BGL} > 7.8$ mM average at 2 h fasting) is defined as diabetes mellitus (DM) [4]. This chronic disease causes long-term damage, dysfunction, and failure of diverse organs, resulting in complications.

DM is grouped into three categories based on the origin of the metabolic disorders [5]. Type-1 diabetes mellitus (T1DM) results from insufficient production of insulin due to the destruction of insulin-producing endocrine cells caused by autoimmunity. An artificial pancreas can help such patients. Type-2 diabetes mellitus (T2DM) is a result of diminished effectiveness of insulin action even though it is produced as normal. T2DM comprises 90% of the total DM patients. Gestational diabetes is a temporary condition in women who develop hyperglycemia during pregnancy.

Insulin resistance has been a key component of health monitoring for a few decades. The incidence of T2DM is closely related to obesity; in which people, about 60–75% of the population who have body mass index (BMI) ≤ 25 kg/m^2 avoids DM [6, 7]. Insulin resistance is of utmost importance in the pathogenesis of T2DM, hypertension, and coronary heart disease, including syndrome X [8].

Here, we evaluate machine-learning (ML) as a method to predict development of insulin resistance from the time series of BGLs. ML has been used to diagnose DM by considering various features of individuals, such as age, gender, BMI, waist circumference, smoking, job, hypertension, residential region (rural/urban), physical activity, and family history of diabetes [9]. Use of clustering algorithms (linear regression, random forest, *k*-nearest neighbors, and support vector machine) to evaluate those risk factors can predict whether or not subjects are diabetic. To date, most ML

✉ Taegeun Song
tsong@postech.ac.kr

[1] Center for Theoretical Physics of Complex Systems, Institute for Basic Science (IBS), Daejeon 34126, Korea

[2] Department of Physics Education and Center for Theoretical Physics, Seoul National University, Seoul 08826, Korea

[3] Department of Physics, Pohang University of Science and Technology, Pohang 37673, Korea

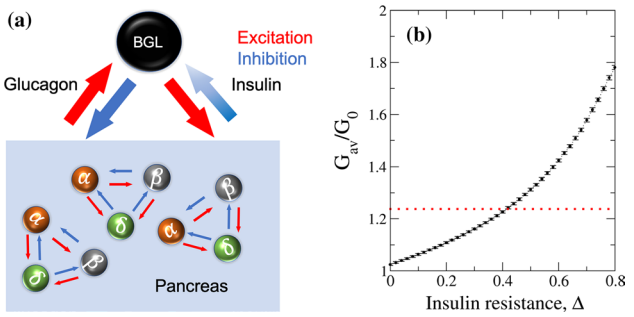

Fig. 1 **a** Schematic diagram of glucose homeostasis. Blood glucose levels (BGLs) are regulated by insulin and glucagon secreted from endocrine systems. In the pancreas (blue boxed area), each endocrine system consists of three cell types that interact with one another. The blue graded arrow represents the diminished action of insulin, which is insulin resistance. **b** Two-hour averaged blood glucose concentration depending on the degrees of insulin resistance. The red dotted line represents the standard hyperglycemic threshold of $7.8\,\mathrm{mM}/G_0 \approx 1.24$, where $G_0 = 6.3$ mM is the normal glucose concentration. Insulin resistance $\Delta > 0.4$ leads to hyperglycemia (color figure online)

applications to DM have focused on finding biomarkers [9, 10]. In this paper, we provide a novel insight to detect DM development by extracting the increment of insulin resistance, a critical factor of T2DM, from the time trend of the BGL.

This idea has not been explored yet because time series data on BGLs with sufficient temporal resolution are currently not available. BGLs are regulated by two counter-regulatory hormones, insulin and glucagon, secreted in a pulsatile manner with a 5 to 10-min period. The signal of fluctuating BGLs can be regarded as an outcome of the balanced response to insulin and glucagon. Successful probing of the signal of fluctuating BGLs requires a temporal resolution that is fine enough to detect the response of the pulsatile hormones with a shorter time interval than the hormone pulses. The time resolution of the current state-of-the-art continuous glucose monitoring sensor has reached 5 min, which is only comparable to the period of hormone pulses. Therefore, to test our proposal, we use synthetic data for glucose profiles produced by using a biophysical model [11, 12] that considers both glucose regulation and hormone action.

This paper is organized as follows: In Sect. 2, we briefly introduce the biophysical model that produces time series data for BGLs. In Sect. 3, we explain the machine learning methods that we use in this study. In Sect. 4, we provide discuss the results.

2 Data preparation of glucose time traces

To produce the data on glucose profiles with a dependence on insulin resistance, we adopt a biophysical model that describes glucose regulation by endocrine systems [11, 12]. Because of the importance of the metabolic disease, diabetes, many biophysical models exist in this field. Some models describe how glucose stimulates insulin secretion at cellular or organ levels [13–15] while other models describe how glucose and insulin regulate each other [16–19]. Unlike the one-way response model or the hormone-level description, the biophysical model we adopt formulates the closed loop between glucose regulation and the fluctuations of endocrine systems.

The human pancreas has a few million islets, endocrine systems, and each islet consists of α, β, and δ cells. Insulin secreted by β cells decreases BGLs whereas glucagon secreted by α cells increases BGLs. Somatostatin secreted by δ cells does not directly regulate BGLs, but the three endocrine cell types interact with one another. The interaction signs between α, β, and δ cells are very special (Fig. 1a). Depending on the glucose concentration, the endocrine cells show biological rhythms with active/silent phases that lead to corresponding hormone secretion. The biophysical model describes the rhythmic cellular activities responding to glucose stimuli as phase oscillators modulated by the environment [12, 20]. The model also considers interactions among endocrine cells within islets; these interactions correspond to the couplings in the oscillator model. The model was used to explain the entrainment of insulin secretion by alternating glucose stimuli in experiments [19, 21].

In this study, we slightly modified the closed-loop model to consider insulin resistance. We use $\sigma \in \{\alpha, \beta, \delta\}$ to represent the three types of endocrine cells and n to indicate the islet index. The activity (or hormone secretion) of σ cells in the nth islet is denoted by amplitude $r_{n\sigma}$ and phase $\theta_{n\sigma}$. The dynamics of the interacting phase oscillators depends on the glucose level G:

$$\frac{\mathrm{d}r_{n\sigma}}{\mathrm{d}t} = [f_\sigma(G) - r_{n\sigma}^2]r_{n\sigma} + \sum_{\sigma' \neq \sigma} K_{\sigma\sigma'} r_{n\sigma'} \cos(\theta_{n\sigma'} - \theta_{n\sigma}) \tag{1}$$

$$\frac{\mathrm{d}\theta_{n\sigma}}{\mathrm{d}t} = \omega_{n\sigma} - \mu_\sigma(G)\cos(\theta_{n\sigma}) + \sum_{\sigma' \neq \sigma} K_{\sigma\sigma'} \frac{r_{n\sigma'}}{r_{n\sigma}} \sin(\theta_{n\sigma'} - \theta_{n\sigma}). \tag{2}$$

Here the model describes glucose-dependent amplitude modulations with sigmoidal functions of $f_\sigma(G)$ and phase modulations with linear functions of $\mu_\sigma(G)$ (see the Ref. [22] for their specific functional forms). The spontaneous oscillations of the cellular activities have angular frequencies,

$\omega_{n\sigma} \sim \mathcal{N}(\omega_0, 0.1)$, which follows a normal distribution with a mean value of $\omega_0 = 2\pi$ min^{-1} and a standard deviation of 0.1. The coupling signs between α, β, and δ cells follow experimental evidence: $K_{\alpha\beta} = K_{\beta\delta} = K_{\alpha\delta} = -1$ and $K_{\beta\alpha} = K_{\delta\alpha} = K_{\delta\beta} = 1$. Note that islet cells interact only within it; they do not interact with cells located in different islets.

The total amount of hormone secretion from all islets is then $\sum_n r_{n\sigma}(1 + \cos\theta_{n\sigma})$ for glucagon ($\sigma = \alpha$) and insulin ($\sigma = \beta$). The phase $\theta_{n\sigma} = 0$ shows maximal secretion whereas $\theta_{n\sigma} = \pi$ shows minimal secretion. Given the fact that insulin decreases the glucose concentration G, whereas glucagon increases G, the oscillator model of islets can make a closed loop for glucose regulation:

$$\frac{dG}{dt} = \frac{G_0}{N}\sum_{n=1}^{N} r_{n\alpha}(1 + \cos\theta_{n\alpha}) - \frac{G}{N}(1 - \Delta)\sum_{n=1}^{N} r_{n\beta}(1 + \cos\theta_{n\beta}). \quad (3)$$

Glucose clearance by insulin is proportional to the present glucose concentration G, unlike glucose production by glucagon. The multiplication of a constant G_0 in the glucose production term is included to consider the balance between the actions of glucagon and insulin at the normal glucose concentration G_0. In this study, we set $G_0 = 6.3$ mM. Equations (1)–(3) complete the closed-loop model for glucose regulation in the absence of external glucose stimuli [12]. To include the effect of insulin resistance, we introduce an auxiliary parameter Δ in the glucose clearance term. Δ is a reduction in the effectiveness of insulin action.

As the insulin-resistance parameter Δ increases, BGLs of G increase (Fig. 1b). In particular, beyond $\Delta = 0.4$, the 2-h averaged BGLs of $G_{\rm av}$ exceed 7.8 mM, so hyperglycemia is severe. Therefore, to reproduce early-stage diabetic conditions, we use five groups of $\Delta = \{0, 0.1, 0.2, 0.3, 0.4\}$. Given Δ, we numerically solved the coupled differential Eqs. (1)–(3) for $N = 200$ islets. Then we took 500 time steps (corresponding to 25 min) with a step size 0.05 min as a sample of glucose time traces. For each group of Δ, we prepared 2000 samples of the BGL time series for training and 200 samples for testing. Each sample included the group label of Δ, which has one-hot encoding (10000, 01000, 00100, 00010, 00001 for $\Delta = 0, 0.1, 0.2, 0.3$, and 0.4, respectively).

Different groups of Δ have a clear feature in the time-averaged value $G_{\rm av} \equiv 1/\tau \sum_{t=1}^{\tau} G(t)$ of BGLs for the total time step $\tau = 500$. However, given real glucose time traces, one cannot judge whether the different $G_{\rm av}$ results from different degrees of insulin resistance or simply from individual variations. Therefore, to avoid this confusion, we focus on the temporal pattern itself rather than the shifted average level by using $G(t) - G_{\rm av}$. We produced different samples of $G(t) - G_{\rm av}$ by solving the model for different initial conditions or by randomly selecting different time windows from full-time traces (Fig. 2). The temporal patterns for different Δ are not apparent to the eye.

3 Pattern recognition of machine learning

Temporal pattern classification is one of the most challenging problems in ML [23] with a wide range of applications in human activity recognition [24], electroencephalogram (EEG) classification [25], and speech recognition [26, 27]. Here, we consider four different neural networks for each's ability to classify insulin resistance from the synthesized time traces of BGLs. First, we consider a shallow neural network (ShallowNet) as a control for comparison with more sophisticated network models. Second, we use a fully-connected deep neural network called a multilayer perceptron (MLP); it is the most basic structure for deep learning. Third, we use a convolutional neural network (CNN) because it has been very successful in recognizing spatial and temporal patterns [28, 29]. Finally, we also use a recurrent neural network (RNN) [30]; RNNs were originally specialized for temporal data by considering recurrent flows in a network.

Our task is a supervised learning with inputs of temporal glucose traces, $G(t) - G_{\rm av}$, and outputs of five labels Δ for insulin resistance. Thus, we assigned 500 nodes for the input layer, and 5 nodes for the output layer. We adopt the ReLU (Rectified Linear Unit) as a basic activation function except for the output layer [31]. For the output layer, we

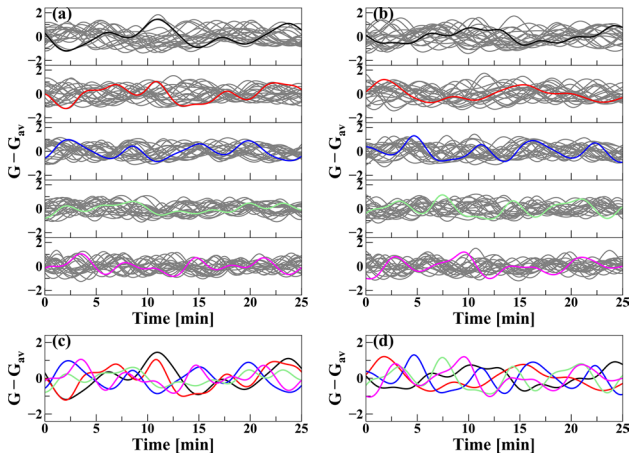

Fig. 2 Temporal glucose profiles under insulin resistance. Twenty-one samples were randomly selected **a** from 10,000 training data and **b** from 1000 test data. Among the gray time series samples, one is highlighted with colors. The five rows have different degrees of insulin resistance $\Delta = 0, 0.1, 0.2, 0.3, 0.4$ from top to bottom. For easy comparison, the colored samples are put together in **c** and **d** for the training and the test data, respectively (color figure online)

use a softmax function to obtain probabilistic predictions as $(p_1, p_2, p_3, p_4, p_5)$. For example, if $\mathrm{argmax}_k(p_k) = 1$, we conclude that the corresponding time trace is the first group, which has $\Delta = 0$. We use the Adaptive Moment Estimation (Adam) algorithm for the optimization of learning [32]. Now we specify the network structures that we used in this study.

ShallowNet The ShallowNet consists of two hidden layers of (1024, 256) nodes for each layer.

MLP The MLP has eight hidden layers of (256, 256, 512, 512, 512, 256, 128, 64) nodes, and every node in a layer is fully connected to every node in adjacent layers.

CNN The CNN for classification is usually composed of two parts. The first part is composed of convolution operations that extract features from input data. The second part takes the features extracted by the convolution layer and feeds them into the MLP for classification. The convolution layer consists of a set of trainable filters. Each filter convolves across the width and the height of input data and generates convolution outputs. The outputs can be interpreted as filtered input data. Therefore, optimizing the filter to extract relevant features from data is a crucial step for the CNN. If one uses many filters, one can extract multiple features. These convolution processes generate a multi-dimensional feature map, which becomes the input for the MLP that combines all the processed features and finally predicts the classification of the input data.

We prepared two different types of data encoding: 1-dimensional (1D) and 2-dimensional (2D) inputs. For 1D input, the feature map is generated by convolution only in temporal space. The CNN for a 1D input has five convolutional layers, with 100, 100, 200, 200 and 100 filters and filter sizes of 10, 10, 10, 3 and 3, respectively. A CNN is specialized for 2D image recognition, so we reshaped the 1D temporal data of (1×500) into 2D arrays of "images" such as (2×250), (5×100), or (10×50). Given a sequence $[x_1, x_2, \ldots, x_N]$, the reshaping of $(D \times M)$ changes the 1D sequence to $[[x_1, x_2, \ldots, x_M], [x_{M+1}, x_{M+2}, \ldots, x_{2M}], \ldots, [x_{(D-1)M+1}, x_{(D-1)M+2}, \ldots, x_{DM}]]$. This 2D reshaping may be able to capture internal structures such as periodicity in the original time traces. For the 2D input, the CNN also has five convolutional layers of 100, 100, 200, 200 and 100 filters, respectively, with the filter size (3, 3). Both 1D and 2D CNNs have the same fully connected MLP parts with four hidden layers with (200, 100, 50, 50) nodes.

RNN The RNN is designed to process sequential data. Arrays of input data are continuously fed into the RNN, and the activation of each node is transferred directly to connected nodes with recurrent flows. Thus, the RNN can naturally consider the order of time traces. Here, we use three types of RNNs: (i) vanilla RNN [30], (ii) long short-term memory (LSTM) [30, 33] and (iii) gated recurrent unit (GRU) [34]. Vanilla RNN is the simplest RNN, and LSTM and GRU consider the memory effect of temporal data. Vanilla RNN has three layers with 250 input, 200 hidden, and 50 output nodes for recurrent flows. For the vanilla RNN, we did not use the 1D input of (1×500) because it corresponds to a ShallowNet with a single hidden layer of 200 nodes. The LSTM and the GRU have gated memory cells, such as the LSTM unit or the GRU. As for the 2D CNN, we considered various input shapes to test memory effects in the LSTM and the GRU networks.

4 Results and discussion

For the learning, we used the Keras Python with a TensorFlow backend [35, 36]. The learning of this simple task did not take much computation time, usually less than a few tens of minutes. We examined the diagnosis accuracy (Table 1), which was measured by using the fraction of correct prediction of the degrees of insulin resistance among the tested glucose profiles. The overall accuracy ranged from 70 to 90%, with the MLP showing the best accuracy. The accuracy depended on the number of network parameters. The numbers of parameters (in millions) were approximately 1.99 (MLP), 0.99 (2D CNN), 0.87 (LSTM), 0.66 (GRU), and 0.26 (vanilla RNN).

In the analysis of the 2D reshaped data encoding, the 2D CNN showed invariant results for different reshaping whereas

Table 1 Benchmark results for various machine learning methods

Accuracy (%)	ShallowNet	MLP	CNN		RNN		
			(1D)	(2D)	(vanilla)	(LSTM)	(GRU)
$\Delta = 0.0$	92.5	89.0	79.0	84.0	88.0	82.0	85.0
$\Delta = 0.1$	70.5	80.5	63.5	78.0	70.5	81.0	80.0
$\Delta = 0.2$	84.5	89.5	65.0	89.5	76.0	83.5	81.5
$\Delta = 0.3$	94.0	94.5	73.0	89.5	80.5	84.5	89.0
$\Delta = 0.5$	95.0	96.5	84.5	91.5	92.0	99.0	97.5
Total	87.3	90.0	73.0	86.5	81.4	86.0	86.6

The accuracy for each group Δ is evaluated using 200 test data for each group. The total accuracy is evaluated using all 1000 test data

Table 2 Network performance for different 2-dimensional data encoding

Input shape	GRU	LSTM	2D CNN
(2×250)	86.6	86.0	85.3
(5×100)	84.7	84.5	85.7
(10×50)	82.2	84.0	86.5
(25×20)	79.5	83.3	84.2
(20×25)	80.7	83.9	85.4
(50×10)	79.8	82.0	85.5
(100×5)	74.6	77.5	82.3
(500×1)	74.5	73.0	–

$(D \times M)$ represents (segment dimension, time step in each segment)

the GRU and the LSTM showed diminished accuracies as the segment length was decreased (Table 2). This trend may be a result of the finite filter size for the CNN and the memory effect for the RNN. We also examined a different segmentation rule that change $[x_1, x_2, \ldots, x_N]$ to $[[x_1, x_2, \ldots, x_M], [x_2, x_3, \ldots, x_{M+1}], \ldots, [x_{N-M+1}, x_{N-M+2}, \ldots, x_N]]$, but it achieved a negligible increase (1–5%) in accuracy.

Real glucose profiles include intrinsic and measurement noise. Thus, to examine the noise effect, we added white noise to our synthesized glucose data and confirmed that our diagnosis was robust up to about a 10% fluctuation in the BGL.

In this study, we checked whether machine learning could detect the patterns of BGL under insulin resistance. The temporal change in the BGL results from the balanced response to the counter-regulatory hormones, insulin and glucagon. Thus, the ineffective action of insulin, called insulin resistance, should affect the BGL profile. Therefore, we simulated the glucose profiles under insulin resistance by using a biophysical model for glucose regulation and confirmed that a subtle change in the glucose profiles under insulin resistance could be recognized by using various machine-learning methods. This demonstrates great potential for the case of machine learning for the diagnosis of early-stage diabetes

A continuous-glucose-monitoring (CGM) system has been widely used for mainly T1DM as a closed-loop artificial pancreas with insulin pumps [37]. The recent development of a low-cost CGM system has revolutionized CGM usage towards wearable minimally-invasive CGM sensors [38, 39]. Such an efforts in conjunction with our proposal should guide the future direction of diabetes management. In addition to the difficulty in obtaining high-resolution glucose profiles, the high-accuracy of labels is another prerequisite for successful supervised learning.

Acknowledgements This research was supported by the POSTECH (T.S.) and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, NRF-2019R1F1A1052916, (J.J.) and NRF-2017R1D1A1B03034600 (T.S.).

References

1. P.V. Röder, B. Wu, Y. Liu, W. Han, Exp. Mol. Med. **48**(3), e219 (2016)
2. L. von Bertalanffy, Science **111**, 239 (1950)
3. D.A. Lang, D.R. Matthews, J. Peto, R.C. Turner, New Engl. J. Med. **301**, 10237 (1979)
4. American Diabetes Association, Diabetes Care **41**(Suppl. 1), S13 (2018)
5. American Diabetes Association, Diabetes Care **42**(Suppl. 1), S13–S28 (2019)
6. J.C. Seidell, Br. J. Nutr. **83**(Suppl. 1), S5 (2000)
7. J.C. Seidell, Horm. Metab. Res. **21**, 155 (1998)
8. G.M. Reaven, Physiol. Rev. **75**(3), 473 (1995)
9. S. Larabi-Marie-Sainte, L. Aburahmah, R. Almohaini, T. Saba, Appl. Sci. **9**(21), 4604 (2019)
10. I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, I. Chouvarda, Comput. Struct. Biotechnol. J. **15**, 104 (2017)
11. D. Park, T. Song, D. Hoang et al., Sci. Rep. **7**, 1602 (2017)
12. T. Song, J. Jo, Phys. Biol. **16**, 051001 (2019)
13. G.M. Grodsky, J. Clin. Investig. **51**, 2047 (1972)
14. M. Komatsu, M. Takei, H. Ishii, Y. Sato, J. Diabetes Investig. **4**, 511 (2013)
15. I.J. Stamper, X. Wang, J. Theor. Biol. **318**, 210 (2013)
16. P. Palumbo, S. Ditlevsen, A. Bertuzzi, A.D. Gaetano, Math. Biolsci. **244**, 69 (2013)
17. I.M. Tolic, E. Mosekilde, J. Sturisa, J. Theor. Biol. **207**, 361 (2000)
18. J. Li, Y. Kuang, C.C. Mason, J. Theor. Biol. **242**, 722 (2006)
19. B. Lee, T. Song, K. Lee, J. Kim, S. Han, P.-O. Berggren et al., PLoS One **12**(2), e0172901 (2017)
20. T. Song, H. Kim, S.-W. Son, J. Jo, Phys. Rev. E **101**, 022613 (2020)
21. B. Lee, T. Song, K. Lee, J. Kim, P.-O. Berggren, S.H. Ryu et al., PLoS One **12**(8), e0183569 (2017)
22. We set the amplitude modulation functions $f_\alpha(G) = [1 - \tanh((G - G_0)/\delta G)]/2$, $f_\beta(G) = [1 + \tanh((G - G_0)/\delta G)]/2$, and $f_\delta(G) = [1 + \tanh((G - G_0 + \Delta G)/\delta G)]/4$ and the phase modulation functions $\mu_\alpha(G) = \mu \cdot (G - G_0)$ and $\mu_\beta(G) = \mu_\delta(G) = -\mu \cdot (G - G_0)$. Here, we used the parameters, $G_0 = 6.3$ mM, $\Delta G = 4.5$ mM, $\delta G = 1.8$ mM, and $\mu = 0.1$ mM^{-1}
23. P. Esling, C. Agon, ACM Comput. Surv. **45**(1), 1 (2012)
24. A. Bevilacqua, K. MacDonald, A. Rangarej, V. Widjaya, B. Caulfield, T. Kechadi, Human activity recognition with convolutional neural networks, in *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2018. Lecture Notes in Computer Science*, vol. 11053, ed. by U. Brefeld et al. (Springer, Cham, 2019)
25. A. Craik, Y. He, J.L. Contreras-Vidal, J. Neural Eng. **16**(3), 031001 (2019)
26. G. Hinton et al., IEEE Signal Process. Mag. **29**(6), 82 (2012)
27. A. Mohamed, G. Hinton, G. Penn, in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto* (2012), pp. 4273–4276
28. A. Krizhevsky, I. Sutskever, G. Hinton, *Advances in Neural Information Processing Systems* (NIPS 2012), pp. 1097–1105
29. H.I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, P.-A. Muller, Data. Min. Knowl. Discov. **33**, 917 (2019)
30. A. Sherstinsky (2018). arXiv:1808.03314
31. We used a tangent hyperbolic activation function for the LSTM given the input shape (100×5), because its long segment dimension caused the weight to explode when we used ReLU
32. D.P. Kingma, B. Jimmy (2014). arXiv:1412.6980

33. F.A. Gers, J. Schmidhuber, F. Cummins, in *Ninth International Conference on Artificial Neural Networks ICANN 99 (Conf. Publ. No. 470)* (1999)
34. K. Cho et al. (2014). arXiv:1406.1078
35. Chollet, François et al., Keras, 2015. Software available from https://keras.io
36. M. Abadi et al., TensorFlow: large-scale machine learning on heterogeneous systems (2015). Software available from https://www.tensorflow.org/
37. Z. Mian, K.L. Hermayer, A. Jenkins, Am. J. Med. Sci. **358**, 332 (2019)
38. M. Vettoretti, G. Cappon, G. Acciaroli, A. Facchinetti, G. Sparacino, J. Diabetes Sci. Technol. **12**, 1064 (2018)
39. G. Cappon, M. Vettoretti, G. Sparacino, A. Facchinetti, Diabetes Metab. J. **43**, 383 (2019)