SCIENTIFIC REP☼RTS

A Novel Classification Indicator of Type 1 and Type 2 Diabetes in China

Yannian Wang[1], Shanshan Liu[1], Ruoxi Chen[1], Zhongning Chen[2], Jinlei Yuan[3] & Quanzhong Li[4]

Because of the differences of treatment, it is extremely important to classify the types of diabetes, especially for the diagnosis made by clinician. In this study, we proposed a novel scheme calculating an indicator of classifying diabetes, which contains two stages: the first is a model of feature extraction, 17 features are automatically extracted from the curve of glucose concentration acquired by continuous glucose monitoring system (CGM); the second is a model of diabetes parameter regression based on an ensemble learning algorithm named double-Class AdaBoost. 1050 curves of glucose concentration of type 1 and type 2 diabetics were acquired at the Department of Endocrinology in People's Hospital of Zhengzhou University China, and an upper threshold μ was set to 7 mmol/L, 8 mmol/L, 9 mmol/L, 10 mmo/L, and 11 mmol/L respectively according to the guideline of WHO. The experiments show that the coincidence rate of our scheme and clinical diagnosis is 90.3%. The novel indicator extends the criteria in diagnosing types of diabetes and provides doctors with a scalar to classify diabetes of type 1 and type 2.

Diabetes mellitus (DM) is a chronic metabolic disease caused by deficiency or diminished effectiveness of endogenous insulin. Poor glucose control can lead to complications in multiple organs resulting in increased rates of morbidity and mortality[1]. According to International Diabetes Federation (IDF), in the world, there are about 415 million patients suffering from diabetes in 2015[2], and this number is growing.

The efficacy of treatment in preventing diabetes complications has been confirmed by Diabetes Control and Complications Trial[3]. Specifically, the success[4,5] in Continuous Glucose Monitoring system (CGM), an invasive device is used to measure and record patients' glucose concentration every 5 minutes. Recently, the CGM has been introduced in the prediction of glucose concentration[6,7]. We consider it is worth investigating the classification according to CGM signal as tools for the management of DM[8,9].

According to the pathogeny of diabetes, there are 4 types of DM. Type 1 diabetes and Type 2 diabetes are the main categories. Clinically, the type is usually determined by tests, such as fasting plasma insulin (FINS), insulin releasing test (INS), C-Peptide test, insulin autoantibodies (IAA) and islet cell autoantibodies (ICA). Some of the tests are only temporary and incomplete to diagnose diabetes for the limitation of cognition. The scheme proposed in this paper tries to provide an effective and a supplementary indicator for diabetes classification, which would have benefits of perfecting the framework, raising the precision, and offering a convenient and intelligent method of classifying diabetes.

Classification is one of the hottest issues in data mining. Various classification algorithms have been introduced in many fields, such as Sound recognition[10], Bitcoin fraud[11], and Tomato plant disease[12]. In this study, a novel scheme calculating the indicator of classifying diabetes is proposed and it consists of two stages: the first is feature extraction, in which 17 features[13] are automatically extracted from the curves of glucose concentration using statistics methods, and the second is diabetes parameter regression based on an ensemble learning algorithm named double-Class AdaBoost. The scheme can give an intelligent and precise method to diagnose the type of diabetes.

The Scheme. Based on Adaboost and its variant, a diabetes classification indicator is proposed, and its processing steps are described below:

1) Utilize CGM, collect curves of diabetic glucose concentration.

[1]School of Information Engineering, Zhengzhou University, 450001, Zhengzhou, Henan, P.R. China. [2]Department of Cardiology, University of Arkansas for Medical Sciences, 72205, Little Rock, AR, USA. [3]The Fifth Affiliated Hospital, Zhengzhou University, 450052, Zhengzhou, Henan, P.R. China. [4]Department of Endocrinology, People's Hospital of Zhengzhou University, 450003, Zhengzhou, Henan, P.R. China. Correspondence and requests for materials should be addressed to J.Y. (email: yuanjinliea@163.com) or Q.L. (email: loicdf@163.com)
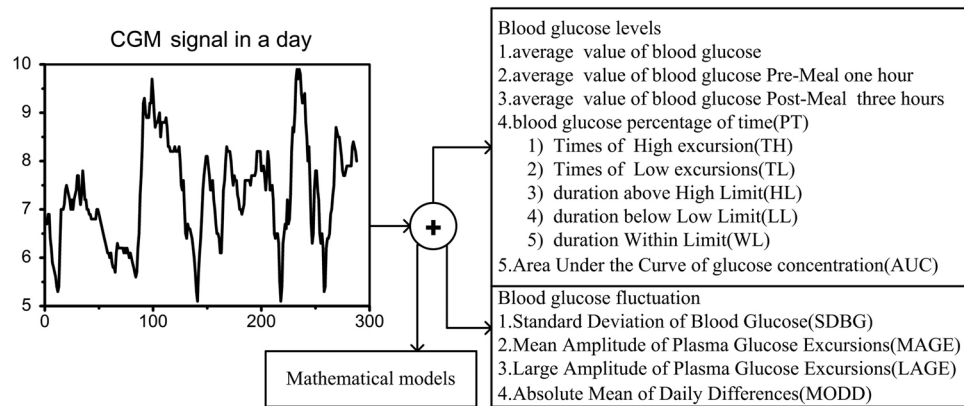
Figure 1. Feature extraction from CGM signal. The 17 features of blood glucose concentration and glucose fluctuation are extracted when CGM signal of one day is inputted to the Mathematical models.

2) Employ feature extraction model to achieve 17 features from the training curves of glucose concentration.
3) Build and train a classifier using the 17 features based on variants of AdaBoost.
4) Verify the classifier using the testing curves of glucose concentration.
5) Evaluate the indicator of the scheme to classify diabetes.

Methods

CGM. CGM is used in examination of how the blood glucose concentration reacts to insulin, exercise, food, and others. And it needs calibrating with traditional finger-stick measurements. A CGM acquires glucose concentration of patients on a continuous basis (every five minutes).

Feature Extraction. Feature extraction[14] is based on the morphological characteristics of signals to obtain the intrinsic features. Features usually possess some physical significance and could be extracted from complicated multi-component signals such as a time-series of glucose concentration. Hence, the feature extraction is taking a glucose concentration signal as input and gives the features as the output. The model of features extraction is illustrated in Figure 1.

The first feature is the average of blood glucose on the whole day; it can be calculated by Equation (1).

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n}x_i \tag{1}$$

where x_i is a discrete value of blood glucose concentration, n is the number of x_i in a day. The subsequent six features are also the average in different periods including pre-meal average and post-meal average of three meals. All those averages can be calculated by Equation (1).

SDBE this feature is Standard Deviation of Blood Glucose; it can be calculated by Equation (2).

$$SDBE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{2}$$

LAGE this feature is Large Amplitude of plasma Glucose Excursions; it can be calculated by Equation (3).

$$LAGE = x_{\max} - x_{\min} \tag{3}$$

where x_{max}, x_{min} are the maximum and minimum values of blood glucose in a day.

MODD this feature is absolute Mean Of Daily Differences, and it can be calculated by Equation (4).

$$MODD = \frac{1}{287}\sum_{i=1}^{288}|v_1(i) - v_2(i)| \tag{4}$$

where v_1, v_2 are arrays of glucose concentration with 288 values of one day, and they are the glucose concentrations of same diabetic in different day respectively.

Area Under the Curve of glucose concentration (AUC) indicates the area parceled by the glucose concentration-time curve and the threshold (upper threshold or lower threshold). The AUC should be calculated with two areas including under the curve of glucose concentration and the upper threshold, over the curve of glucose concentration and the lower threshold.

Mean amplitude of plasma glucose excursions (MAGE). This feature has been studied by many papers[15]. The MAGE can be calculated as follows: Step 1. Get all extreme points in the signal; Step 2. Find the first valid extreme point whose absolute differences of both adjacent extreme points are greater than the Standard Deviation of the

	Pre-meal			PT				
Average	Breakfast	Lunch	Dinner	TH	TL	HL	LL	WL
6.7	6.6	7.9	5.7	0	0	0	0	1
7.2	5.3	7.7	7.9	1	0.07	1	0.02	0.91
9.6	9.7	10.7	9.5	3	0.25	0	0	0.75

AUC	SDBG	MAGE	LAGE	MODD	Post-meal		
					Breakfast	Lunch	Dinner
0	0.7	0.8	3.4	0.725	6.6	7.3	6.4
0.1	1.8	−2.6	8.4	1.4327	9.8	7.3	8.9
0.5	1.6	−5.05	8	2.7753	10	12.5	8.6

Table 1. The features dataset (mmol/L). Each row was extracted from a curve of blood glucose concentration of diabetics.

signal; Step 3. Accumulate all differences of valid extreme point according to the left direction of the first valid extreme point in step 2; Step 4. MAGE is the average of sum counted in step 3.

$$MAGE = \frac{1}{n} \sum_{i=1,3,5,\dots}^{2n-1} ep_i - ep_{i+1}$$

(5)

where ep_i is the left adjacent extreme point of valid extreme point ep_{i+1}; n is the number of valid extreme point.

Blood glucose Percentage of Time (PT) includes two main aspects: times and the percentage of the time. The features relating to times, including Times of High excursion (TH), Times of Low excursion (TL), are the number of the extreme points of glucose concentration curve over threshold line in one day. The features relating to percentage of the time, including duration above High Limit (HL), duration below Low Limit (LL), duration Within Limit (WL), are the percentage of the time of glucose concentration curve over the threshold line one day.

In our research, a dataset was built to store the features, as shown in Table 1.

Usually, age is an important factor related to diabetes[16], and places heavy weight on the classification of type of diabetes, thus would cause under-fitting. Some other factors were not involved, such as exercise, food, and insulin or oral medicines, which are difficult to quantify as these factors are from different manufacturers and are difficult to homogenize. Furthermore the main purpose of our research is to provide an easy and approach available to diabetes diagnose.

AdaBoost. Boosting methods are iterative algorithms[17]. AdaBoost is a boosting method which united some simple "weak" classifiers to generate generalized models. It was proposed by Freund and Schapire to distinguish a binary classification[18], and later various AdaBoost variants such as Real Adaboost were proposed[19]. AdaBoost and its variants have contributed to various real-world applications, such as face detection[20] and human detection[21]. In our research, its variants Real Adaboost[19], Gentle AdaBoost[22], and Modest AdaBoost[23], were applied to the model of diabetes parameter regression in our scheme.

Diabetes parameter regression based on AdaBoost. Let s = {(g₁,y₁), (g₂,y₂), …, (gₘ,yₘ)} be a set of training samples with initial weights $D_1(g_i) = 1/m$, and m is the number of training data. Each g_i is a vector with 17 features which were extracted from CGM curve of glucose concentration, and each y_i is the label of g_i. In our research, the DM classification is a binary classification, so assuming label y_i equals 1 if the sample belongs to type 1 diabetes, and otherwise equals −1 when the sample belongs to type 2 diabetes.

Diabetes classification based on AdaBoost algorithm is described as follows:

Input: training dataset s = {(g₁,y₁), (g₂,y₂), …, (gₘ,yₘ)}, initialize data weights $D_1(g_i) = 1/m$, i = 1, …, m;

$$g_i \in G^{17}, y_i \in \{-1, 1\}$$

For $t = 1, 2, \dots T$:

Step 1: train weak classifier h_t using distribution D_t.
Step 2: Calculate the error of the weak classifier h_t: G → {−1,1}.

$$\varepsilon_t = \sum_{i=1}^{m} D_t(g_i)[y_i \neq h_j(g_i)]$$

Step 3: Calculate the weight $a_t = (1/2)ln((1-\varepsilon_t)/\varepsilon_t)$.
Step 4: Update data weight D_t and get new weights D_{t+1} by error.

$$D_{t+1} = \frac{D_t(g_i)\exp(-a_t y_i \, h_t(g_i))}{z_t}$$

where z_t is a normalization factor.

Output: final classifier: $H_{final}(g) = sign(\sum_{t=1}^{T} a_t h_t(g))$

	Predict P	Predict N
True P	TP	FN
True N	FP	TN

Table 2. Confusion matrix. In this table, TP is the number of true positives, TN the number of the true negatives, FP the number of false positives and FN the number of false negatives.

Diabetes parameter regression based on Variants algorithm of AdaBoost. Three variants algorithms of AdaBoost were used for diabetes parameter regression. Firstly, Real AdaBoost is a generalization of AdaBoost algorithm proposed by Schapire and Singer[19]. Its output is not binary, but a real number between $+1$ and -1. Real AdaBoost algorithm seems to Adaboost, except the steps 1–3 summarized as below:

For each weak classifier h_t

a. every value space of features is divided into several disjoint blocks $G_1, \ldots, G_n$
b. under the distribution D_t calculate

$$p_l^j(g) = p(g_i \in G_j, y_i = l) = \sum_{i=g=G_j, y_i=l} D_t(i) \, l \in 1, -1$$

c. set the output of h on each G_j as

$$h_t^j(g) \leftarrow \frac{1}{2} \log p_l^j/(1 - p_l^j) \in R \quad h_t(g_i) = h_t(g)$$

d. calculate the normalization factor

$$z = 2\sum_j \sqrt{p_{+1}^j p_{-1}^j} \quad z_t = \min z \quad h_t = \arg \min z$$

The second variants algorithm of AdaBoost named Modest AdaBoost which complete steps could be found in paper[23]. Gentle AdaBoost is the most efficient boosting algorithm and it has been used in Cascades object detection[24]. In each epoch, Gentle AdaBoost does a weighted regression based on least square. It means that the regression function $h_t(g)$ is fit by weighted least-squares of y_i to g_i.

Model Evaluation. In order to evaluate classification results, the present study applied two performance indicators: ACC (accuracy) and MCC (Matthews correlation coefficient). P and N represent the positive class and negative class respectively. T and F denote True and False respectively, as described in Table 2.

The ACC is as the formula

$$ACC = \frac{TP + TN}{P + N} \tag{6}$$

The MCC is as the formula

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{7}$$

Results
Patient Database. The diabetics were screened at ages ranging from 40 to 60 and the glucose concentration were acquired from the Department of Endocrinology in People's Hospital of Henan Province of China. There are 1050 samples of diabetes glucose concentration, and each sample is a curve with more than 864 values.

Experiment and analyses. To demonstrate the performance of proposed indicator, 300 of the 1050 samples were used as training set to construct a diabetes classifier while the other 750 were used as testing set to evaluate the classifier. Besides, 7 mmol/L, 8 mmol/L, 9 mmol/L, 10 mmol/L and 11 mmol/L were set to the upper threshold of glucose target range in the progress of feature extraction from the curves of glucose concentration monitored by CGM.

The Committee Report of diabetes expert of WHO diagnoses DM with fasting blood glucose concentration between 6.1 mmol/L and 6.9 mmol/L and plasma glucose of 11.1 mmol/L 2 hours post glucose-load (2 h PPG). This is the reason why 7 mmol/L, 8 mmol/L, 9 mmol/L, 10 mmo/L and 11 mmol/L were selected as the upper threshold when we extract the 17 features from glucose signal.

The models of Real AdaBoost, Modest AdaBoost and Gentle AdaBoost were applied to calculating the indicator of classification diabetes and the error rate was presented in Table 3. The error rate of Modest AdaBoost is 0.0970 when the upper threshold was set at 7 mmol/L and 8 mmol/L, which means that the coincidence rate of our scheme and clinical diagnosis is 90.3%.

After training 100 iterations, the three models of Real AdaBoost, Modest AdaBoost, and Gentle AdaBoost were to calculate the indicator of classifying diabetes. The test misjudging rate of indicator and clinical diagnosis illustrated in Figure 2. The upper thresholds of Figure 2 (a)–(e) were set at 7, 8, 9, 10 and 11 mmol/L respectively. It shows that when the upper limit was set at 7 mmol/L and 8 mmol/L the misjudging rate of three models were

Upper limit					
Error rate	7 mmol/L	8 mmol/L	9 mmol/L	10 mmol/L	11 mmol/L
Real AdaBoost	0.0921	0.0867	0.1491	0.1382	0.1057
Modest AdaBoost	0.0970	0.0970	0.1084	0.1220	0.1165
Gentle AdaBoost	0.0678	0.0786	0.1274	0.1030	0.1003

Table 3. Comparison of error rate. The error rate of Real AdaBoost, Modest AdaBoost, and Gentle AdaBoost at 7, 8, 9, 10 and 11 mmol/L of the upper threshold of glucose target range.
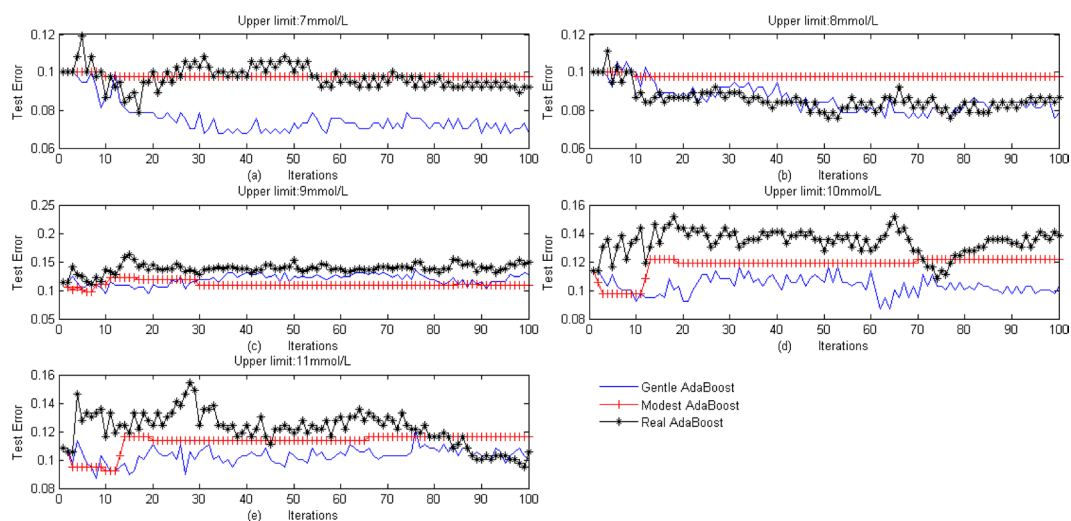

Figure 2. Comparison error of diabetes classification. Test error of the three ensemble methods, Real AdaBoost, Modest AdaBoost, Gentle AdaBoost, at 7, 8, 9, 10 and 11 mmol/L of the upper threshold of glucose target range. The y-coordinate of each point gives the test error rate, and the x-coordinate gives the times of iterations.

lower, and the misjudging rate of Model AdaBoost depicted by the line with the mark '|' is 0.0970. Furthermore, when the upper threshold was set at 10 mmol/L, the three models perform worst in diabetes classification. But when the upper threshold was set at 9 mmol/L or 11 mmol/L, the misjudging rate of Real AdaBoost is changing, and its largest error is greater than 0.12, therefore 9 mmol/L and 11 mmol/L are not suitable for regarding as the upper threshold. The value of upper threshold affects results of diabetes classification.

5-fold cross-validations were used to further demonstrate the accuracy of our scheme and seek out the best of upper threshold, after training 100 iterations, the indicator of classifying diabetes based on Real AdaBoost, Modest AdaBoost and Gentle AdaBoost were calculated and the test misjudging rate of indicator and clinical diagnosis illustrated in Figure 3. It shows that when threshold was set at 7 mmol/L or 8 mmol/L, the performance of our scheme is better, and only a few misjudging rates were above 0.1. It indicates that the coincidence rate of indicator calculated by our scheme and clinical diagnosis is better and the indicator is useful for doctors to diagnose diabetes.

The performance of our scheme was evaluated when the threshold was set at 7, 8, 9, 10 and 11 mmol/L respectively. The results are shown in Table 4. It shows that when threshold was set at 7 mmol/L or 8 mmol/L, the performance of our scheme is better.

Discussion

Due to the difference of epidemiology, etiology, pathogenesis and treatment of type 1 and type 2 DM, a knotty problem is how to effectively treat diabetes in clinic[25]. For a doctor, the reasonable solution is to classify the type of diabetes and suit the remedy to the case, so the diabetes can be in control. In fact, there are many clinical indicators to classify diabetes, such as the test results of Oral Glucose Tolerance Test (OGTT), INS, C-Peptide, IAA, ICA. The tests would contribute to providing guideline in treating diabetes, but the tests are incomplete and can't precisely reflect the heterogeneity of the Type 1 diabetes and Type 2 diabetes. Moreover, some of original symptoms about Type 2 diabetes have emerged on patients with Type 1 diabetes. At the moment, CGM can monitor the curve representing the fluctuation of glucose concentration in patients with type 1 and 2 diabetes[9,10], which is one of the most successful cases for diabetes controlling. In addition, the 17 features would be extracted from the curve of glucose concentration[13]. Those features can't directly diagnose the type of DM, but we attempt to build a novel scheme calculating the indicator of classifying DM by using those features.

We have constructed an effective scheme, which consists of feature extraction and classification. The experimental results show when the upper threshold μ is correctly set, the misjudging rate of classification is less than
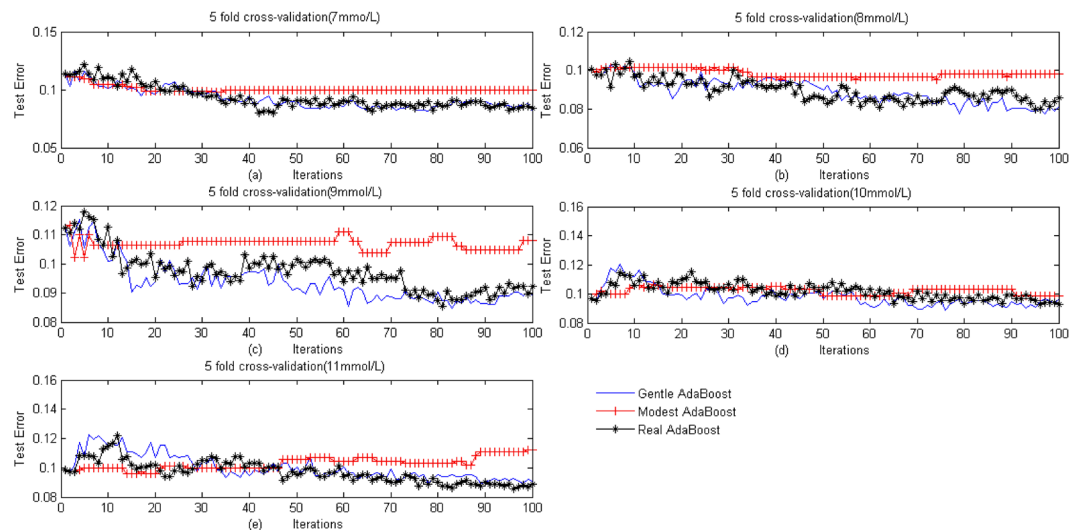
Figure 3. 5-fold cross-validation. The method of 5 fold cross-validationto validates the test error of Real AdaBoost, Model AdaBoost and Gentle AdaBoost at 7, 8, 9, 10 and 11 mmol/L of the upper limit of glucose target range. The *y*-coordinate of each point gives the test error rate, and the x-coordinate gives the times of iterations.

Upper limit	7 mmol/L	8 mmol/L	9 mmol/L	10 mmol/L	11 mmol/L
ACC	0.9038	0.9038	0.8428	0.8293	0.8293
MCC	0.5103	0.4736	0.0890	0.0872	0.0662

Table 4. Comparison of accuracy and Matthews correlation coefficient. The ACC and MCC of Modest AdaBoost at 7, 8, 9, 10 and 11 mmol/L of the upper threshold of glucose target range.

0.097, which suggests that the scheme achieves the best performance and the coincidence rate of our scheme and clinical diagnosis is up to 90.3%.

This experiment indicates that an indicator can be extracted from the curve of glucose concentration based on CGM and it is helpful for doctors to classify diabetes. In addition, more works should be considered, such as how to improve the precision of classifying diabetes, how to set a novel penalty to rectify the weight of diabetes samples according to the sampling distribution (D_t) of diabetes in the process of iteration, and our scheme should be validated whether it suffers data imbalance problems[26,27].

References

1. Wilson, R.H., Foster, D. W., Kronenberg, H. N. & Larsen, P. R. William Textbook of Endocrinology, 9th ed. Philadelphia, PA: Saunders (1998).
2. Eyes on Diabetes. Word diabetes day. International Diabetes Federation. November 14 (2016).
3. Wang, Y., Wei, F., Sun, C. & Li, Q. The Research of Improved Grey GM (1, 1) Model to Predict the Postprandial Glucose in Type 2 Diabetes. Journal BioMed Research International 2016, ISSN: 2314-6133 (2016).
4. Cobelli, C. *et al*. Diabetes: models, signals, and control. *IEEE reviews in biomedical engineering* **2**, 54–96 (2009).
5. Tildesley, H. D. *et al*. A comparison of internet monitoring with continuous glucose monitoring in insulin-requiring type 2 diabetes mellitus. *Canadian journal of diabetes* **37.5**, 305–308 (2013).
6. Malik, S. *et al*. Gargling effect on salivary electrochemical parameters to predict blood glucose. Computational Techniques in Information and Communication Technologies (ICCTICT). New Delhi, India. 2016 International Conference on. IEEE. 11th−13th March (2016).
7. Eren-Oruklu, M. *et al*. Adaptive system identification for estimating future glucose concentrations and hypoglycemia alarms. *Automatica* **48.8**, 1892–1897 (2012).
8. Zhang, J. & Zhang, M. The Preliminary analysis of blood glucose fluctuation in type II diabetes. *Chinese Journal of Integrative Medicine on Cardio-/Cerebrovascular Disease* **11.4**, 500–501 (2013).
9. Wentholt, I. M. *et al*. Glucose fluctuations and activation of oxidative stress in patients with type 1 diabetes. *Diabetologia* **51.1**, 183–190 (2008).
10. McLoughlin, I. *et al*. Robust sound event classification using deep neural networks. IEEE/ACM Transactions on. *Audio, Speech, and Language Processing* **23.3**, 540–552 (2015).
11. Monamo, P., Marivate, V. & Twala, B. Unsupervised learning for robust Bitcoin fraud detection[C]//Information Security for South Africa (ISSA), 2016. IEEE, 129–134 (2016).
12. Sabrol, H. & Satish, K. Tomato plant disease classification in digital images using classification tree. Communication and Signal Processing (ICCSP). Melmaruvathur, TN, India. 2016 International Conference on. IEEE, 5th, 06–08 April (2016).
13. Jia, W. P. Clinical application guideline of dynamic blood glucose monitoring in china(version of 2009)[J]. *National Medical Journal of China* **89.48**, 3388–3392 (2009).
14. Hu, J. Summarization on feature dimensionally reduction. *Application Research Of Computers* **25.9**, 2601–2606 (2008).
15. Mo, Y. F. *et al*. Assessment index of glucose fluctuation—clinical significance and research progress of Average glucose fluctuations. *Chinese Journal of Diabetes Mellitus* **3.3**, 259–263 (2011).

16. Jiang, L & Peng, L. of Diabetes with II type Classification and features extraction using SVM, Science Technology and Engineering, 7.5, 721–726(2007).
17. Mining, Data. Practical machine learning tools and techniques, -Witten. Frank and Hall (1913).
18. Freund, Y., Schapire, R. & Abe, N. A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence* **14**(1612), 771–780 (1999).
19. Schapire, R. E. & Singer, Y. Improved boosting algorithms using confidence-rated predictions. *Machine learning* **37.3**, 297–336 (1999).
20. Guo, J.-M. *et al*. Complexity reduced face detection using probability-based face mask prefiltering and pixel-based hierarchical-feature adaboosting. *IEEE Signal Processing Letters* **18.8**, 447–450 (2011).
21. Xu, J. *et al*. Fast and accurate human detection using a cascade of boosted MS-LBP features. *IEEE Signal Processing Letters* **19.10**, 676–679 (2012).
22. Friedman, J., Hastie, T. & Tibshirani, R. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics* **28.2**, 337–407 (2000).
23. Vezhnevets, A. & Vezhnevets, V. Modest AdaBoost-teaching AdaBoost to generalize better. Graphicon. Vol. 12. No. 5 (2005).
24. Lienhart, R., Kuranov, A. & Pisarevsky, V. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. Joint Pattern Recognition Symposium. Springer Berlin Heidelberg (2003).
25. Li, W., Zheng, H., Bukuru, J. & De Kimpe, N. Natural medicines used in the traditional Chinese medical system for therapy of diabetes mellitus. *Journal of Ethnopharmacology* **92**, 1–21 (2004).
26. Provost, F. Machine learning from imbalanced data sets 101. Proceedings of the AAAI'2000 workshop on imbalanced data sets (2000).
27. Kotsiantis, S., Kanellopoulos, D. & Pintelas, P. Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering* **30.1**, 25–36 (2006).

Acknowledgements

Author Contributions

Conceived the idea: Y.N.W, S.S.L, Q.Z.L. Designed and did the experiments: S.S.L. Analyzed the data: Y.N.W, S.S.L. Wrote the manuscript: Y.N.W, S.S.L, J.L.Y. Revise the paper: S.S.L, R.X.C, Z.N.C. All authors read and approved the manuscript.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.