Medicine®

OPEN

Blood-based FTIR-ATR spectroscopy coupled with extreme gradient boosting for the diagnosis of type 2 diabetes

A STARD compliant diagnosis research

Peiwen Guang, MD[a], Wendong Huang, MD[b], Liu Guo, MD[a], Xinhao Yang, MD[a], Furong Huang, PhD[a,*], Maoxun Yang, PhD[c,*], Wangrong Wen, MD[d], Li Li, MD[d]

Abstract

Timely diagnosis of type 2 diabetes and early intervention and treatment of it are important for controlling metabolic disorders, delaying and reducing complications, reducing mortality, and improving quality of life. Type 2 diabetes was diagnosed by Fourier transform mid-infrared (FTIR) attenuated total reflection (ATR) spectroscopy in combination with extreme gradient boosting (XGBoost). Whole blood FTIR-ATR spectra of 51 clinically diagnosed type 2 diabetes and 55 healthy volunteers were collected. For the complex composition of whole blood and much spectral noise, Savitzky–Golay smoothing was first applied to the FTIR-ATR spectrum. Then PCA was used to eliminate redundant data and got the best number of principle components. Finally, the XGBoost algorithm was used to discriminate the type 2 diabetes from healthy volunteers and the grid search algorithm was used to optimize the relevant parameters of the XGBoost model to improve the robustness and generalization ability of the model. The sensitivity of the optimal XGBoost model was 95.23% (20/21), the specificity was 96.00% (24/25), and the accuracy was 95.65% (44/46). The experimental results show that FTIR-ATR spectroscopy combined with XGBoost algorithm can diagnose type 2 diabetes quickly and accurately without reagents.

Abbreviations: ATR = attenuated total reflection, DM = diabetes mellitus, FPG = fasting plasma glucose, FTIR = Fourier transform mid-infrared, GBDT = Gradient Boosting Decision Tree, OGTT = oral glucose tolerance testing, PCA = principal component analysis, XGBoost = extreme gradient boosting.

Keywords: extreme gradient boosting, Fourier transform mid-infrared attenuated total reflection spectroscopy, type 2 diabetes, whole blood

1. Introduction

Diabetes mellitus (DM) is a group of metabolic disorders characterized by hyperglycemia resulting from decreased insulin secretion or insulin inaction.[1] As per the 2017 report of the International diabetes foundation, diabetes affects 425 million adults worldwide, with the total set to reach 629 million by 2045. An estimated 90% are affected by type 2 diabetes, which is largely preventable. One in 2 people with diabetes has not yet been diagnosed and so their diabetes is not controlled.[2] When diabetes is uncontrolled, it can have dire consequences for health and well-being and result in a number of serious complications such as cerebral hemorrhage, cerebral infarction, retinal disease, diabetes kidney disease, neurological and cardiovascular diseases.[3] Therefore, timely diagnosis of type 2 diabetes and early intervention and treatment of it are important for controlling metabolic disorders, delaying and reducing complications, reducing mortality, and improving quality of life.

Editor: Liang-Jun Yan.

WH is the co-first author.

This work was supported by the National Natural Science Foundation of China (61975069); Natural Science Foundation of Guangdong Province, China (2018A0303131000); Academician Workstation of Guangdong Province, China (2014B090905001); the key project of scientific and technological projects of Guang Zhou, China (201604040007 and 201604020168). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

[a] Department of Opto-Electronic Engineering, Jinan University, Guangzhou, [b] Department of Pharmacy, Maoming People's Hospital, Maoming, [c] Zhuhai Hopegenes Medical & Pharmaceutical Institute Co., Ltd, Hengqin New Area, Zhuhai, [d] First Affiliated Hospital of Jinan University, Guangzhou, Guangdong, China.

[*] Correspondence: Furong Huang, Department of Opto-Electronic Engineering, Jinan University, Guangzhou 510632, China (e-mail: furong_huang@163.com), Maoxun Yang, Zhuhai Hopegenes Medical & Pharmaceutical Institute Co., Ltd, Hengqin New Area, Zhuhai 519000, China (e-mail: yangmaoxun1980@163.com).

How to cite this article: Guang P, Huang W, Guo L, Yang X, Huang F, Yang M, Wen W, Li L. Blood-based FTIR-ATR Spectroscopy Coupled with Extreme Gradient Boosting for the Diagnosis of Type 2 Diabetes: A STARD conpliant diagnosis research. Medicine 2020;99:15(e19657).

Received: 13 October 2019 / Received in final form: 17 February 2020 / Accepted: 17 February 2020

http://dx.doi.org/10.1097/MD.0000000000019657

Currently, the clinical diagnoses of type 2 diabetes are mainly carried out by fasting plasma glucose (FPG) and blood glucose 2 hours after oral glucose tolerance testing (OGTT).[4] However, these clinical diagnosis methods have the disadvantages of complicated procedures, time-consuming and high cost, and are not suitable for the screening of type 2 diabetes in large-scale populations.[5] To solve this problem, researchers have tried many new methods to diagnose type 2 diabetes. Tong et al[6] diagnosed diabetes by analyzing the content of acetone in human exhalation. Although the method is simple and time-saving, in practical applications, the acetone content is low and the analysis results are greatly affected by other gases. Kong et al[7] explored the clinical significance of 7 diabetes-related serum microRNAs during the pathogenesis of type 2 diabetes and found the expression levels of all 7 miRNAs of type 2 diabetes were significantly up-regulated compared with healthy person. Only 70.6% of type 2 diabetes subjects (12/17) were recognized by canonical discriminant function, the sensitivity is low and serum miRNAs were determined by real-time Reverse Transcription-Polymerase Chain Reaction which is expensive. There are also various methods for the determination of glucose have been reported, such as colorimetry,[8–10] fluorescence,[11–13] electrochemistry,[14,15] chemiluminescence,[16,17] capillary electrophoresis,[18] surface-enhanced Raman scattering,[19] and surface plasmon resonance.[20] However, in many cases, the complicated material modification, intrinsic toxicity, and turn-off sensing mode cause inevitable disadvantages, including operational complexity, false results, and unsatisfactory sensitivity, for type 2 diabetes diagnosis.[21] Therefore, it is still a challenge to explore a simple, non-reagent, rapid, and accurate method for the diagnosis of type 2 diabetes.

Vibrational spectroscopy has been widely used to discriminate and classify normal and pathological populations using cells, tissues, or biofluids.[22] FTIR-ATR is an excellent vibrational spectroscopic technique for the analysis of biofluids (e.g., blood) due to its rapidity and ease of translation to the clinical environment, that is, Fourier transform mid-infrared attenuated total reflection (FTIR-ATR) requires no sample preparation when analyzing blood. In the past few years, FTIR-ATR spectroscopy have been used for diagnosing a variety of diseases.[23] Hands et al[24] report the application of FTIR-ATR spectroscopy for stratified serum spectroscopic diagnostics capable of diagnosing at brain tumor. Paraskevaidi et al[25] diagnose Alzheimer disease using FTIR-ATR spectroscopy from blood. Lima et al[26] use FTIR-ATR spectroscopy coupled with variable selected techniques on plasma or serum specimens as an alternative approach for early detection of ovarian cancer.

Due to complicated massive spectral data and multidimensional analyses, a fast and accurate multivariate statistical method is required to be developed for the applications of the FTIR-ATR.[27,28] Extreme gradient boosting (XGBoost) is an efficient implementation of the Gradient Boosting Decision Tree (GBDT) algorithm, originally proposed by Dr Chen of the University of Washington.[29] It is used in Kaggle competition and has attracted wide attention because of its superior efficiency and high prediction accuracy. Liu et al[30] used visible near-infrared shortwave infrared spectroscopy combined with XGBoost to quantitative assessment of soil properties.

Classification of type 2 diabetes through FTIR-ATR spectra using the XGBoost algorithms has never been studied. In this paper, FTIR-ATR spectroscopy based on human whole blood samples was used to diagnose type 2 diabetes. Savitzky–Golay smoothing and principal component analysis (PCA) was used to eliminate spectral noise and extract principal component. The diagnosis model was established by XGBoost algorithm. The model was optimized and a simple operation, reagent-free, fast, and accurate method for diagnosis of type 2 diabetes was proposed.

2. Materials and methods

2.1. Experimental apparatus

A Vertex 70 transform infrared spectrometer produced by Bruker was used with an attenuated total emission sample measurement attachment manufactured by Specac. The ATR sample cell was made of a ZnSe crystal, with a 45° incidence angle and 3 reflections. Each measured spectrum was obtained from 4500 to 600 cm^{-1} with a spectral resolution of 2 cm^{-1}, the beam splitter was KBr and the number of scans was 32. The laboratory temperature was 24 ± 1°C and the relative humidity was 41%.

2.2. Experimental samples

The whole blood samples were collected from 113 volunteers (62 women and 51 men) in the Endocrinology Department of the First Affiliated Hospital of Jinan University, of which 51 cases were newly diagnosed with type 2 diabetes (28 women and 23 men). The average age of patients with type 2 diabetes and healthy volunteers was 49 ± 12 and 44 ± 10 respectively. Blood samples were collected in the fast manner and stored in 4°C refrigerator. All samples came from the same race and social economic background. This study has been approved by ethics committee of the hospital and all respondents have been informed about the research program.

Baseline correction and dark background subtraction were performed with the Application Programming Interface function of the spectrometer when the spectrum was collected. During spectral collection, the test tube was thoroughly shaken, and 0.075 mL samples were placed in the sample plot. The FTIR absorption spectra of samples were obtained through ATR. The spectrum was collected 3 times for each sample, and the average spectrum was calculated as the sample spectrum.

2.3. Split of sample set

All whole blood samples were split into train and test sets, and the ratio was 3:2. The train set consisted of 67 blood samples, among which 37 from healthy volunteers and 30 from type 2 diabetes. The test set of 46 blood samples included 25 healthy volunteers blood samples and 21 type 2 diabetes blood samples (Table 1).

Table 1

Split of train set and test set.

Sample	Total sample	Type 2 diabetes	Healthy volunteers
Total sample	113	51	62
Train set	67	30	37
Test set	46	21	25

2.4. Extreme gradient boosting

The XGBoost algorithm is a class of lifting algorithms composes of a series of base classifiers. The principle is to divide the original data set into multiple sub-data sets. Each sub-data set is randomly assigned to the base classifier for prediction, the results of base classifier are calculated according to a certain weight, and the final result is the accumulation of weak classifier predictions. The base classifier of this paper is CART tree. Selecting the regression tree is based on our original experience.

Selecting CART as the basis function of the model, then the result of the Mth prediction for a single CART is:

$$f_m(x) = T(\mathrm{X}, \theta_m)$$

Thus, the basis function has been determined, where T represents the decision tree, m represents the number of base classifiers, and θ represents the path of the decision tree. The final prediction result is the previous prediction result plus the current decision tree, and the error term can be expressed as:

$$L(\widehat{y,}y) = L(y, f_{m-1}[x]) + T(\mathrm{X}, \theta_m)$$

$L(\widehat{y,}y)$ is the sum of the difference value between the true value y_i and predicted value$\widehat{y}\,i$. At this point, the error loss has been quantified. The Gini coefficient, pruning, and depth of control trees are important tools for CART classification. Actually, the variance and bias of the model are controlled by the above methods, making the model more capable of generalizing and fitting. For example, the structural complexity function can be defined by the number of leaf nodes T and the L2 square of the Leaf Score:

$$\phi(\theta) = \gamma T + \frac{1}{2}\lambda \sum\nolimits_{J=1}^{T} W_j^2$$

Where γ represents the complexity coefficient of the control tree, which is equivalent to the pruning of the tree of the XGBoost model; and λ represents how much proportion is used to change the regular terms, which is equivalent to give a penalty to the complex model, preventing the model from over-fitting. The comprehensive of deviation function and variance function can give the following objective functions:

$$Obj(\theta) = \sum_i l(y_t + y_i) + \gamma T + \frac{1}{2}\lambda \sum\nolimits_{j=1}^{T} W_j^2$$

The predecessors' method for the Gradient Boosting Decision Tree algorithm is to repeatedly calculate the error of the objective function so that the error becomes smaller and smaller, which is often referred to as the gradient descent algorithm. According to this method, as long as we get weak classifiers in this way each time, and then add each weak classifier, the result of the final model must be optimal.

The formula for gradient descent is as follows:

$$-\left[\frac{\partial L(y, f[x_i])}{\partial f(x_i)}\right] f(x) = f_{m-1}(x)$$

Since the number of base classifiers used by the XGBoost algorithm is large, we need a more general algorithm to achieve gradient descent. The inventor of the XGBoost algorithm uses Taylor second-order expansion instead of the original first-order derivative, making the algorithm more universality.

The objective function after adding Taylor second-order expansion:

$$Obj(\theta) = \sum\nolimits_{i=1}^{n} l([y_i, y_i{}^{m-1}] + f_m[x_i]) + \phi(f_m)$$

In the formula, n represents the number of used samples, m represents the number of current iterations, f_m indicating the error of the current iteration.

Taylor expansion

$$f(x + \Delta x) \cong f(x) + f'(x)\Delta x + \frac{1}{2} f''(x)\Delta x^2$$

Definition:

$$gi = \partial y(m-1) l(yi, y^{(m-1)}), hj = \partial^2 y(m-1)$$

Substituting in the formula:

$$Obj_m \cong \sum\nolimits_{i=1}^{n} (l[y_i, y_l{}^{m-1}] + g_i f_m[x_i] + \frac{1}{2} h_i f_m{}^2[x_i]) + \phi(f_m)$$

2.5. Evaluation index of model parameter

The performance of the model and the choice of optimal parameters need to be measured by appropriate evaluation indicators. The sensitivity, specificity, and accuracy are the most commonly used evaluation indicators in the classification problem. Sensitivity (SEN) (Equation (1)) and specificity (SPC) (Equation (2)) show the ability of the model to correctly classify true positive as positive and true negative as negative respectively, where tp, tn, fp, and fn are true positive, true negative, false positive, and false negative respectively. The accuracy (Equation (3)) is the sum of the correctly classified divided by the total number of classes. In this paper, accuracy is regarded as the selection criterion of the optimal parameter, and the performance of the model uses sensitivity, specificity, and accuracy as evaluation index.

$$\text{Sensitivity} = \frac{\text{tp}}{(\text{tp} + \text{fn})} \tag{1}$$

$$\text{Specificity} = \frac{\text{tn}}{(\text{tn} + \text{fp})} \tag{2}$$

$$\text{Accuracy} = \left(\frac{(\text{tp} + \text{tn})}{(\text{tp} + \text{tn} + \text{fn} + \text{fp})}\right) \tag{3}$$

2.6. Ethics approval and consent to participate

The study has been approved by the ethics committee of Jinan University. And written-informed consent was obtained from each participant.

3. Results

3.1. Spectral analysis

The comparison of the average FTIR-ATR spectrum of the normal and the type 2 diabetes samples are shown in Fig. 1. As

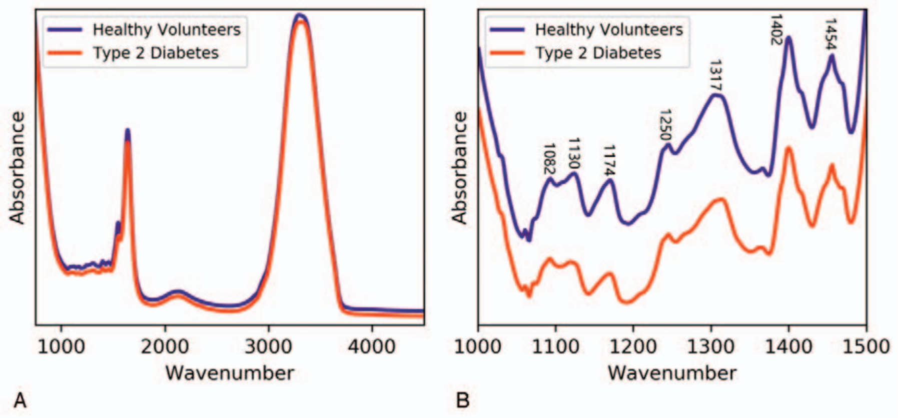

Figure 1. Comparison of the average FTIR-ATR spectra of 62 healthy volunteers blood and 51 type 2 diabetic blood. (A) Wavelength range of 700 to 4500 cm^{-1} and (B) wavelength range of 1000 to 1500 cm^{-1}. ATR=attenuated total reflection, FTIR=Fourier transform mid-infrared.

shown in Fig. 1A, the different pathologic groups are similar in spectral peak shape, in which the 2 main water-peak bands of 3000 to 4000 cm^{-1} and 1500 to 1800 cm^{-1} are the major absorption areas, but the difference in absorption intensity is significant. The vibrations and rotations of the various groups are overlapping due to the complex components of whole blood. It can be seen that there are many peaks in the range of 1000 to 1500 cm^{-1}, which is overlapping with the fingerprint area. The fingerprint region can be used to identify specific molecules, in this region, the vibrations of chemical bond are vulnerable to the effects from the adjacent chemical bond vibration, and minor structural changes may result in differences in this part of the spectra. Therefore, these bands are required to be separately displayed (Fig. 1B).

As shown in Fig. 1B, the average FTIR-ATR spectrum of the type 2 diabetes and healthy volunteers are similar in peak shape in the region of 1000 to1500 cm^{-1} while a significantly different in absorption intensity can be clearly seen. The assignments of the peaks in this region are listed in Table 2. In the average FTIR-ATR spectrum of patients with type 2 diabetes, the absorption intensities of those peaks were lower than that of healthy volunteers, which may be caused by more severe metabolism. Hence, the differences in the FTIR-ATR spectrum of the type 2 diabetes patients and healthy volunteers indicated that the FTIR-ATR spectroscopy can be effectively used for detection of type 2 diabetes.

Table 2

Major band positions observed from the region of 1000 to 1500 cm^{-1} along with their assignments.

Band (cm^{-1})	Assignment
1082	Symmetric vibration of phosphodiester bond
1130	Stretching vibration of C–O
1174	Stretching vibration of C–O in hydroxyl amino acid
1250	Antisymmetric vibration of phosphodiester bond
1317	Stretching vibration of C–N
1402	Bending vibration of OH, Symmetrical stretching vibration of O–C–C
1454	Scissor bending vibration of CH_2

3.2. Savitzky–Golay smoothing

The composition of the whole blood sample is very complicated, especially the anticoagulant is also added to make the purity of the sample not high, and various instrument noises are generated during the detection process. Therefore, it is necessary to perform smooth denoizing of the whole blood spectra. Savitzky–Golay smoothing is the most commonly used smoothing algorithm in spectral smoothing. This paper selects the optimal smoothing mode based on the best accuracy of XGBoost model. The polynomial has a search range of (1, 4) and a step size of 1; the search range of the window size is all odd values of (5, 25), and

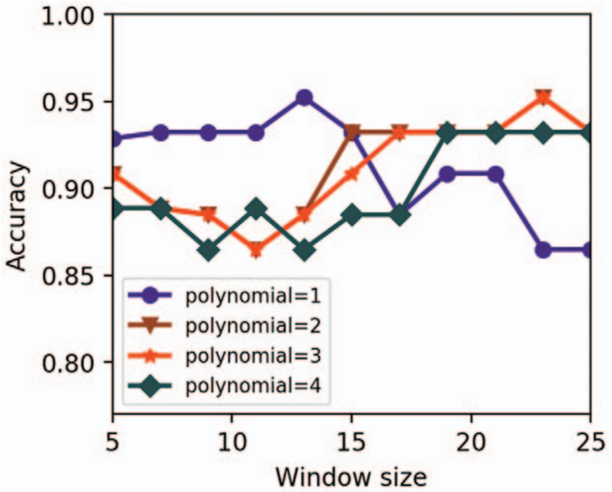

Figure 2. Optimization process for the best Savitzky–Golay smoothing mode.

Table 3

Optimization results for Savitzky–Golay smoothing mode.

Polynomial	Search range	Optimal window size	Optimal accuracy (%)
1	5:25	13	95.3
2	5:25	23	94.2
3	5:25	23	94.2
4	5:25	19	93.2

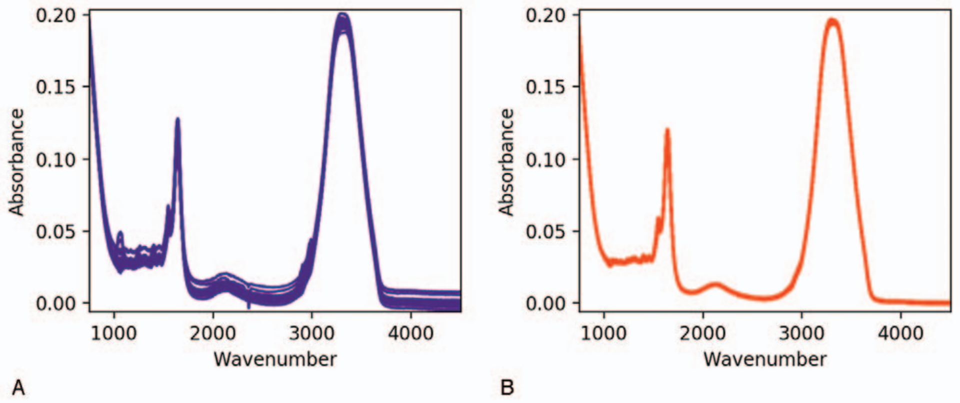

Figure 3. Comparison of (A) original FTIR-ATR Spectra and (B) Savitzky–Golay Smoothed FTIR-ATR Spectra. ATR = attenuated total reflection, FTIR = Fourier transform mid-infrared.

there are 44 Savitzky–Golay smoothing modes. Figure 2 is the optimization process of the Savitzky–Golay smoothing mode. The results of the optimization are shown in Table 3.

The results show that the Savitzky–Golay smoothing modes in the 2nd and 3rd order polynomial have the same processing effect. The Savitzky–Golay smoothing mode in the 1st order polynomial has the best processing effect, the best accuracy was 95.3%, which is 1.1% higher than the 2nd and 3rd order, and 2.1% higher than the 4th order. Therefore, this paper selects 1st order polynomial and window size of 13 points as the optimal Savitzky–Golay smoothing mode.

The original spectrum of the whole blood sample and Savitzky–Golay smoothed spectrum are shown in Fig. 3, it can be seen that the Savitzky–Golay smoothed spectrum has been significantly improved, indicating Savitzky–Golay smoothing has obvious effect on spectral denoising.

3.3. Principal component extraction

Although most of the noise interference of spectral data are eliminated by Savitzky–Golay smoothing, there is still a lot of redundant information. If they are all model inputs, the complexity of the model structure will be directly increased, affecting its performance and modeling time. Therefore, it is also necessary to extract principal components from the smoothed data to increase modeling efficiency and reduce model complexity. PCA has good performance in removing redundant data and extracting the principle components of data.

This paper used the best accuracy of XGBoost as a selection index to find the best number of principal components, which is similar to the selection of Savitzky–Golay smoothing mode. The search range of the principle component number is (1, 10) and the step size is 1. Figure 4 shows the optimization process for the optimal number of principal components.

It can be seen from Fig. 4 that when the number of principle components is 5, the best results of discrimination is achieved. At this time, the accuracy is 97.6%. Five principle components are extracted from 2022 features in the spectrum, which simplifies the modeling process and contributes to obtain a robust model.

3.4. XGBoost modeling

Constructing an XGBoost model is easy, but there are some difficulties in improving the performance of model. There are many parameters of the XGBoost algorithm. To increase the performance and generalization ability of the model, optimizing the model parameters is an indispensable step. This paper uses the grid search algorithm to optimize the relevant parameters, including n_estimators, learning_rate, min_child_weight, alpha, gamma, and subsample.

The search range of the relevant parameters are set as follows: the search range of n_estimators is (1, 300) and the step size is 10; the search range of learning_rate is (0, 0.2), the step size is 0.01; the search range of min_child_weight is (1, 9), the step size is 1; the search range of alpha is (0, 10), the step size is 0.2; the search range of gamma is (0, 14), the step size is 1; the search range of subsample is (0, 1), the step size is 0.1. Figure 5 shows the GS optimization process for the parameters of XGBoost model. The results of the optimization are shown in Table 4.

According to Table 4, the optimal parameters of XGBoost model are n_estimators = 40, learning_rate = 0.05, min_child_weight = 1, alpha = 0.2, gamma = 0, subsample = 0.7. The optimal parameters were used in XGBoost model and the classification results of test sets were shown in Table 4.

As shown in Table 5, the optimized XGBoost model has good performance in identifying type 2 diabetes, and the sensitivity,

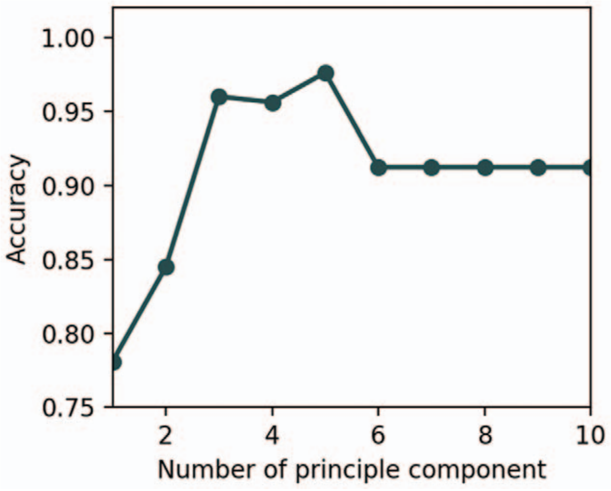

Figure 4. Optimization process for the best number of principal component.

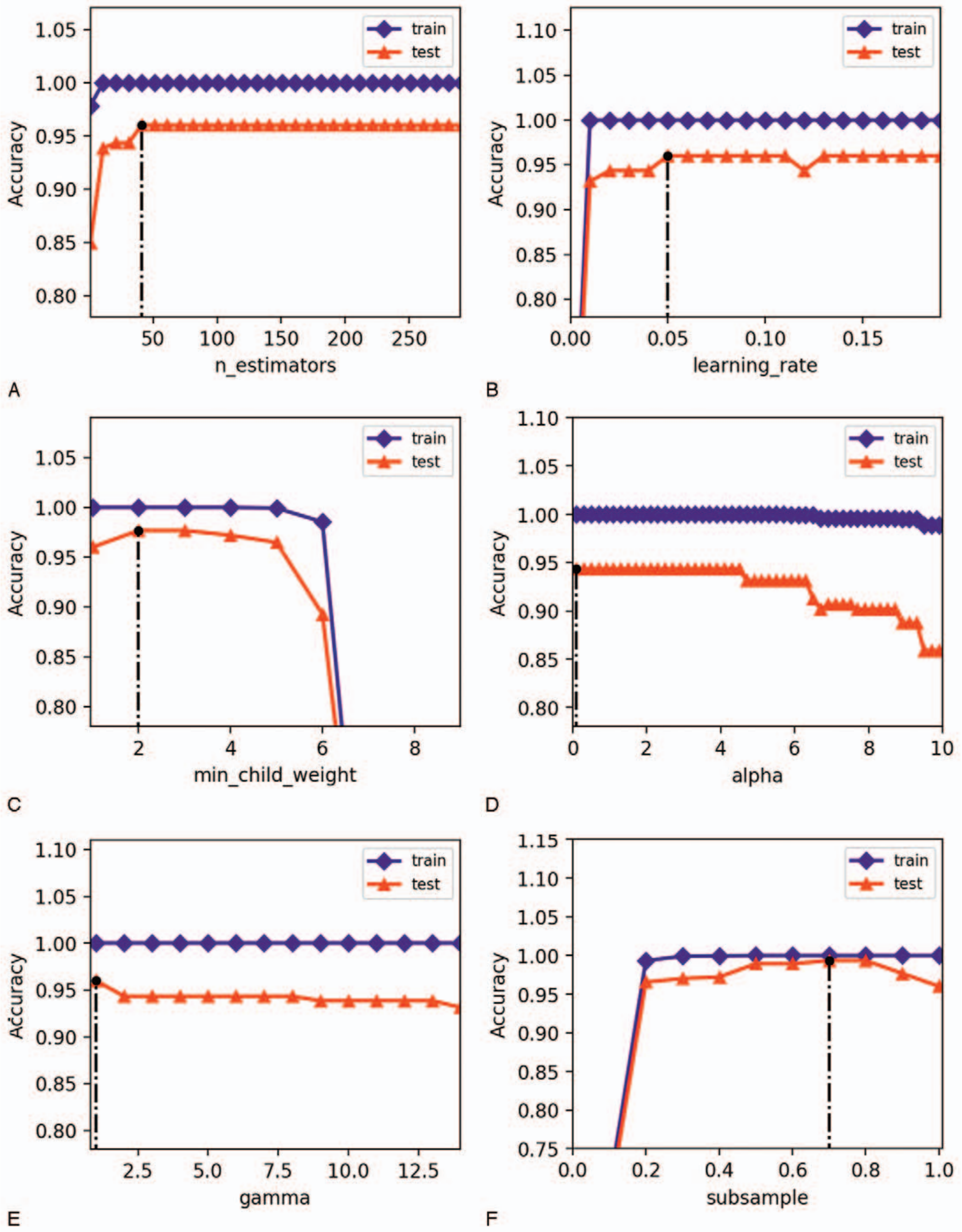

Figure 5. Optimization process for XGBoost model parameters. XGBoost = extreme gradient boosting.

Table 4

Optimization results of XGBoost model parameters.

Parameters	Step size	Search range	Optimal value	Accuracy	
				Train	Test
n_estimators	10	1:300	40	0.99	0.95
learning_rate	0.01	0:0.2	0.05	1	0.96
min_child_weight	1	1:9	1	1	0.96
alpha	0.2	0:10	0.2	1	0.96
gamma	1	0:14	0	1	0.96
subsample	0.1	0:1	0.7	0.99	0.96

XGBoost = extreme gradient boosting.

Table 5

Classification results of XGBoost model.

Sample	Sensitivity	Specificity	Accuracy
Whole blood	95.24% (20/21)	96.00% (24/25)	95.65% (44/46)

XGBoost = extreme gradient boosting.

specificity, and accuracy were 95.24%, 96.00%, and 95.65% respectively.

4. Discussion

This article applied the XGBoost algorithm to the data classification of FTIR-ATR spectral of whole blood samples in order to achieve rapid diagnosis of type 2 diabetes. For the problem that the purity of whole blood sample was not high and redundant information in spectra was too much, the Savitzky–Golay smoothing algorithm and PCA were used to preprocess the FTIR-ATR spectral data successively, eliminating the influence of most spectral noises and improving the modeling efficiency. In order to build a model with high accuracy, relevant parameters of XGBoost were optimizing and the diagnosis model of type 2 diabetes was established using the XGBoost algorithm combined with the whole blood FTIR-ATR spectral data processed by SG smoothing and principal component extraction. The results were encouraging and show the potential of the technique to diagnose of type 2 diabetes, and it may be used in the future as ancillary tools for clinical diagnostics.

Author contributions

Conceptualization: Peiwen Guang, Wendong Huang, Liu Guo, Xinhao Yang.
Formal analysis: Peiwen Guang, Wendong Huang.
Investigation: Peiwen Guang.
Resources: Wangrong Wen, Li Li.
Supervision: Wendong Huang, Wangrong Wen, Li Li.
Validation: Liu Guo, Xinhao Yang.
Writing – original draft: Peiwen Guang, Wendong Huang, Liu Guo, Xinhao Yang.
Writing – review & editing: Furong Huang, Maoxun Yang.
Furong Huang orcid: 0000-0001-8760-439X.

References

[1] Shanbhag VKL, Prasad KS. Graphene based sensors in the detection of glucose in saliva – a promising emerging modality to diagnose diabetes mellitus. Anal Methods 2016;8:6255–9.

[2] Meng RW, Liu N, Yu CQ, et al. Association between major depressive episode and risk of type 2 diabetes: a large prospective cohort study in Chinese adults. J Affect Disorders 2018;234:59–66.

[3] Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. Lancet 2017;389:2239–51.

[4] Meijnikman AS, De Block CEM, Dirinck E, et al. Not performing an OGTT results in significant underdiagnosis of (pre)diabetes in a high risk adult Caucasian population. Int J Obesity 2017;41:1615–20.

[5] Hulman A, Gujral UP, Narayan KMV, et al. Glucose patterns during the OGTT and risk of future diabetes in an urban Indian population: The CARRS study. Diabetes Res Clin PR 2017;126:192–7.

[6] Tong MM, Li J, Liu XW, et al. Analysis of acetone from exhalation based on information fusion technology. Acta Metrologica Sinica 2009;30:183–6.

[7] Kong L, Zhu J, Han WX, et al. Significance of serum microRNAs in pre-diabetes and newly diagnosed type 2 diabetes: a clinical study. Acta Diabetol 2011;48:61–9.

[8] Lin L, Song X, Chen Y, et al. Intrinsic peroxidase-like catalytic activity of nitrogen-doped graphene quantum dots and their application in the colorimetric detection of H2O2 and glucose. Anal Chim Acta 2015;869:89–95.

[9] Dutta AK, Das S, Samanta S, et al. CuS nanoparticles as a mimic peroxidase for colorimetric estimation of human blood glucose level. Talanta 2013;107:361–7.

[10] Su L, Feng J, Zhou X, et al. Colorimetric detection of urine glucose based ZnFe2O4 magnetic nanoparticles. Anal Chem 2012;84:5753–8.

[11] Liu Z, Liu L, Sun M, et al. A novel and convenient near-infrared fluorescence “turn off–on” nanosensor for detection of glucose and fluoride anions. Biosens Bioelectron 2015;65:145–51.

[12] Karnati VV, Gao X, Gao S, et al. A glucose-selective fluorescence sensor based on boronicacid-diol recognition. Bioorg Med Chem Lett 2002;12:3373–7.

[13] Li YS, Du YD, Chen TM, et al. A novel immobilization multienzyme glucose fluorescence capillary biosensor. Biosens Bioelectron 2010;25:1382–8.

[14] Jia X, Hu G, Nitze F, et al. Synthesis of palladium/helical carbon nanofiber hybrid nanostructures and their application for hydrogen peroxide and glucose detection. ACS Appl Mater Inter 2013;5:12017–22.

[15] Noiphung J, Songjaroen T, Dungchai W, et al. Electrochemical detection of glucose from whole blood using paper-based microfluidic devices. Anal Chim Acta 2013;788:39–45.

[16] Yeh TY, Wang CI, Chang HT. Photoluminescent C-dots@ RGO for sensitive detection of hydrogen peroxide and glucose. Talanta 2013;115:718–23.

[17] Li N, Diao W, Han Y, et al. MnO2-modified persistent luminescence nanoparticles for detection and imaging of glutathione in living cells and in vivo. Chemistry 2014;20:16488–91.

[18] Wang X, Ma Y, Zhao M, et al. Determination of glucose in human stomach cancer cell extracts and single cells by capillary electrophoresis with a micro-biosensor. J Chromatogr A 2016;1469:128–34.

[19] Shafer-Peltier KE, Haynes CL, Glucksberg MR, et al. Toward a glucose biosensor based on surface-enhanced Raman scattering. J Am Chem Soc 2003;125:588–93.

[20] Li A, Guo ZY, Peng Q, et al. A saccharides sensor developed by symmetrical optical waveguide-based surface plasmon resonance. J Innov Opt Heal Sci 2015;8:1. doi:10.1142/S1793545815500030.

[21] Ma H, Liu X, Wang X, et al. Sensitive fluorescent light-up probe for enzymatic determination of glucose using carbon dots modified with MnO2 nanosheets. Microchim Acta 2017;184:177–85.

[22] Byoung-Gwon K, Eun-Mi J, Gyeong-Yeon K, et al. Analysis of methylmercury concentration in the blood of Koreans by using cold vapor atomic fluorescence spectrophotometry. Ann Lab Med 2012;32:31–7.

[23] Zohdi V, Whelan DR, Wood BR, et al. Importance of tissue preparation methods in FTIR Micro-Spectroscopical analysis of biological tissues: “Traps for New Users”. PLOS ONE 2015;10:e0116491.

[24] Hands JR, Clemens G, Stables R, et al. Brain tumour differentiation: rapid stratified serum diagnostics via attenuated total reflection Fourier-transform infrared spectroscopy. J Neurooncol 2016;127:463–72.

[25] Paraskevaidi M, Morais CLM, Lima KMG, et al. Differential diagnosis of Alzheimer’s disease using spectrochemical analysis of blood. Proc Natl Acad Sci USA 2017;114:7929–38.

[26] Lima KMG, Gajjar KB, Martin-Hirsch PL, et al. Segregation of ovarian cancer stage exploiting spectral biomarkers derived from blood plasma or serum analysis: ATR-FTIR spectroscopy coupled with variable selection methods. Biotechnol Prog 2015;31:832–9.

[27] Wongwattanakul M, Hahnvajanawong C, Seubwai W, et al. Potential prediction of patient survival and chemotherapeutic sensitivity in cholangiocarcinoma using FTIR microspectroscopy. J Mol Struct 2018;1166:416–21.

[28] Condurso C, Cincotta F, Tripodi G, et al. Characterization and ageing monitoring of Marsala dessert wines by a rapid FTIR-ATR method coupled with multivariate analysis. Eur Food Res Technol 2018;244:1073–81.

[29] Chen TQ, Guestrin C. Xgboost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International 2016; 785–794.

[30] Liu LF, Ji M, Dong YY, et al. Quantitative retrieval of organic soil properties from visible near-infrared shortwave infrared (Vis-NIR-SWIR) spectroscopy using fractal-based feature extraction. Remote Sens 2016;8:1035.