

Research Article

Improved CEEMDAN and PSO-SVR Modeling for Near-Infrared Noninvasive Glucose Detection

Xiaoli Li and Chengwei Li

School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin 150001, China

Correspondence should be addressed to Chengwei Li; lcw@hit.edu.cn

Received 14 April 2016; Revised 18 July 2016; Accepted 27 July 2016

Academic Editor: Thomas Desaive

Copyright © 2016 X. Li and C. Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Diabetes is a serious threat to human health. Thus, research on noninvasive blood glucose detection has become crucial locally and abroad. Near-infrared transmission spectroscopy has important applications in noninvasive glucose detection. Extracting useful information and selecting appropriate modeling methods can improve the robustness and accuracy of models for predicting blood glucose concentrations. Therefore, an improved signal reconstruction and calibration modeling method is proposed in this study. On the basis of improved complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and correlative coefficient, the sensitive intrinsic mode functions are selected to reconstruct spectroscopy signals for developing the calibration model using the support vector regression (SVR) method. The radial basis function kernel is selected for SVR, and three parameters, namely, insensitive loss coefficient ϵ , penalty parameter C , and width coefficient γ , are identified beforehand for the corresponding model. Particle swarm optimization (PSO) is employed to optimize the simultaneous selection of the three parameters. Results of the comparison experiments using PSO-SVR and partial least squares show that the proposed signal reconstitution method is feasible and can eliminate noise in spectroscopy signals. The prediction accuracy of model using PSO-SVR method is also found to be better than that of other methods for near-infrared noninvasive glucose detection.

1. Introduction

Diabetes is a chronic disease that poses a serious threat to human health. According to the International Diabetes Federation (IDF) in 2014, diabetes affects 387 million individuals around the world, and this figure is expected to increase to 592 million in 2035 [1]. The diabetes, cancer, and cardiovascular diseases are the main causes of death since 2005 [2]. At present, diabetes is treated by detecting blood glucose concentrations to adjust the dose of glucose-lowering drugs, thus controlling blood glucose levels to prevent and reduce the symptoms of diabetes and its complications [3]. The accurate detection of blood glucose concentrations is important in diabetes prevention and treatment.

Diabetes monitoring is usually carried out in hospitals or through self-monitoring [4]; diabetes monitoring commonly involves invasive detection, which uses high amounts of biochemical reagents, entails long testing times, and causes inevitable pain and inconvenience to patients. By contrast, noninvasive blood glucose detection [5–7] offers a number

of advantages, such as fast analysis speed, absence of trauma, low cost, and environmental friendliness. Noninvasive optical detection technology [8–10] is an important research topic in the area of noninvasive blood glucose detection. Since the 1970s, scientists have applied optics to determine the chemical composition of the human body. Noninvasive optical detection technologies include a variety of methods, such as near-infrared spectroscopy [11, 12], infrared spectroscopy [13], polarimetry [14], photoacoustics [15], Raman spectroscopy [16], and light-scattering coefficient method [17].

Near-infrared light, the wavelength of which varies from 780 nm to 2526 nm, is the electromagnetic wave between visible light and mid-infrared light that can penetrate the human skin and tissues. A good linear correlation exists between blood glucose concentrations and near-infrared spectrum absorption. In recent years, near-infrared spectroscopy measurement has been widely employed and has thus become fast-developing technology for analysis, particularly in medical applications [11, 18, 19]. The research into near-infrared spectroscopy combined with chemometrics is regarded as

an effective method for the noninvasive detection of blood glucose concentrations [11, 20].

Empirical mode decomposition (EMD), which is an adaptive time frequency data analysis method, is widely used in nonsteady and nonlinear systems [21]. However, mode mixing occurs in EMD. For example, different oscillations exist in the same intrinsic mode function (IMF), or similar oscillations exist in different IMFs. This problem is addressed with ensemble empirical mode decomposition (EEMD), which employs EMD to integrated signals with white Gaussian noise [22]. However, signals with added noise can produce varying numbers of IMFs, and reconstructed signals contain residual noise after decomposition. In complementary ensemble empirical mode decomposition (CEEMD), which can completely eliminate the residual noise in reconstructed signals [23], pairs of positive and negative noises are added to a signal to improve the efficiency of the original noise auxiliary method. EEMD or CEEMD will produce wrong ingredients components, and the IMFs obtained via decomposition may fail to meet the definition of IMF when parameter selection is ineffective. These limitations are resolved with another noise auxiliary algorithm, called CEEMDAN, which is used to achieve an accurate reconstruction of original signals and pure decomposed mode spectra [24]. The iterations of CEEMDAN are less than half of the iterations of EEMD. Moreover, CEEMDAN can accurately reconstruct original signals and recover the features of EMD that are lacking in EEMD. However, CEEMDAN still has some problems which need to be improved; for example, its modes contain some residual noise, and the signal information shows some spurious modes in the early stages of decomposition [25]. To overcome these two issues, the improved CEEMDAN method is applied in this paper to obtain modes with less noise and more physical meaning.

The EMD-based methods can decompose the signal into a series of IMFs which contain the noisy modes and information modes. Therefore, it can be powerful adaptive tool to extract the sensitive intrinsic mode functions to reconstruct the signal. The problems is how to select the sensitive mode to distinguish relevant IMFs and irrelevant IMFs in an efficient way. Reference [26] uses an analogue approach based on consecutive mean squared error (CMSE) criterion. The signal is reconstructed from the mode whose criterion is minimal. In [27], the authors propose an intuitive selected mode method by a new criterion based on Hausdorff distance (HD). Moreover, [28] introduces the mutual information (MI) to select the sensitive IMFs which can reflect the signal characteristics for signal reconstruction. In this paper, the correlative coefficient is used to select relevant IMFs to extract useful spectral information.

Chemometrics, which was proposed by Bhattacharjee in 1994 [29], employs a multivariate statistical analysis of calibration methods and computing technologies to calculate the sample content of each component combined with the near-infrared spectrum. Common linear chemometrics modeling methods include multiple linear regression, principal components regression, and partial least squares (PLS) regression. Examples of nonlinear modeling methods include artificial

neural networks and support vector regression (SVR). Generally, modeling is the process of selecting parameters and methods. SVR can obtain the global optimal solution in spectrum detection and convert linear regression to nonlinear regression, as well as kernel function to the linear mapping of high-dimensional space. The basic principle of SVR, which is a regression method developed from support vector machine, is to map the original data to high-dimensional feature space through nonlinear mapping and to establish a regression model in this space. Applying SVR in near-infrared spectrum quantitative analysis modeling produces a good effect. The commonly used kernel functions include linear kernel functions, polynomial kernel functions, radial basis kernel functions, and sigmoid kernel functions. Many researches and experiments demonstrate that radial basis kernel functions are preferable options if previous knowledge is insufficient. The particle swarm optimization- (PSO-) SVR method is proposed to select ε , C , and γ simultaneously. The results show the satisfactory learning precision and generalization ability with PSO-SVR.

The paper is organized as follows. Section 2 provides a description of the spectrum reconstruction method based on improved CEEMDAN and the PSO-SVR model. The CEEMDAN algorithm, improved CEEMDAN algorithm, correlative coefficient, PSO, and SVR are also introduced. Section 3 presents the near-infrared spectrum experiments on glucose solutions and the results of different modeling methods. Section 4 presents the conclusion of the study.

2. Methods

2.1. CEEMDAN Algorithm. The basis of CEEMDAN is EEMD. Thus, the decomposition theory of the EEMD method is described first [24].

- (1) Set $x^i(n) = x(n) + w^i(n)$, where $w^i(n)$ ($i = 1, \dots, I$) is a different white Gaussian noise.
- (2) The modes $\text{IMF}_k^i(n)$ of each $x^i(n)$ ($i = 1, \dots, I$) can be obtained by EMD, where $k = 1, \dots, K$ representatives modes.
- (3) The k th mode of $x(n)$ is set to $\overline{\text{IMF}}_k$, and the corresponding average of IMF_k^i is

$$\overline{\text{IMF}}_k(n) = \frac{1}{I} \sum_{i=1}^I \text{IMF}_k^i(n). \quad (1)$$

In EEMD, each independently decomposed $x^i(n)$ produces residue $r_k^i(n) = r_{k-1}^i(n) - \text{IMF}_k^i(n)$. However, the decomposition modes are called $\overline{\text{IMF}}_k$, and the first residue is $r_1(n) = x(n) - \overline{\text{IMF}}_1(n)$, where $\overline{\text{IMF}}_1(n)$ is obtained by employing EEMD. $\overline{\text{IMF}}_2(n)$ is the mean value of the result. $r_1(n)$ with a different given noise is decomposed by EMD. The next residue is $r_2(n) = r_1(n) - \overline{\text{IMF}}_2(n)$. Other modes continue this process until the stop condition is met.

Operator $E_j(\cdot)$ is the j th mode of a given signal decomposed by EMD, where w^j is the white noise with the mean value of zero and the variance of one.

If $x(n)$ is the signal, then the steps of CEEMDAN are described as follows:

- (1) The signal $x(n) + \varepsilon_0 w^i(n)$ is decomposed by EMD I times to obtain the first mode: $\widehat{\text{IMF}}_1(n) = (1/I) \sum_{i=1}^I \text{IMF}_1^i(n) = \overline{\text{IMF}}_1(n)$.
- (2) When $k = 1$, the first residue $r_1(n) = x(n) - \widehat{\text{IMF}}_1(n)$ is calculated.
- (3) $r_1(n) + \varepsilon_1 E_1(w^i(n))$ ($i = 1, \dots, I$) is decomposed until the first EMD mode is obtained. The second mode is then calculated:

$$\widehat{\text{IMF}}_2(n) = \frac{1}{I} \sum_{i=1}^I E_1(r_1(n) + \varepsilon_1 E_1(w^i(n))). \quad (2)$$

- (4) When $k = 2, \dots, K$, the k th residue $r_k(n) = r_{k-1}(n) - \widehat{\text{IMF}}_k(n)$ is calculated.
- (5) $r_k(n) + \varepsilon_k E_k(w^i(n))$ ($i = 1, \dots, I$) is decomposed until the first EMD mode is obtained. The $(k+1)$ th mode is then defined:

$$\widehat{\text{IMF}}_{(k+1)}(n) = \frac{1}{I} \sum_{i=1}^I E_1(r_k(n) + \varepsilon_k E_k(w^i(n))). \quad (3)$$

- (6) Steps (4)–(6) are repeated until the obtained residue can not be decomposed; that is, the residue has a maximum of one extreme at most. The final residue meets $R(n) = x(n) - \sum_{k=1}^K \widehat{\text{IMF}}_k$, where K is the total mode number. Thus the expression of signal $x(n)$ is

$$x(n) = \sum_{k=1}^K \widehat{\text{IMF}}_k + R(n). \quad (4)$$

2.2. Improved CEEMDAN Algorithm. According to [25], the improved CEEMDAN algorithm is described based on CEEMDAN as follows:

- (1) For $x^i(n) = x(n) + \beta_0 E_1(w^i(n))$, calculate the local means of I realizations by EMD to obtain the first residue, $r_1 = \langle M(x^i(n)) \rangle$, where $\beta_0 = \varepsilon_0 \text{std}(x) / \text{std}(E_1(w^i(n)))$ and $M(\cdot)$ is the operation which produces the local mean of the signal.
- (2) When $k = 1$, calculate the first mode: $\widehat{\text{IMF}}_1(n) = x(n) - r_1$.
- (3) Estimate the second residue as the average of local means of the realizations $r_1(n) + \beta_1 E_2(w^i(n))$; then the second mode is defined

$$\widehat{\text{IMF}}_2(n) = r_1 - r_2 = r_1 - \langle M(r_1 + \beta_1 E_2(w^i(n))) \rangle. \quad (5)$$

- (4) Calculate the k th residue and k th mode ($k = 3, \dots, K$):

$$r_k = \langle M(r_{k-1} + \beta_{k-1} E_k(w^i(n))) \rangle, \quad (6)$$

$$\widehat{\text{IMF}}_k(n) = r_{k-1} - r_k. \quad (7)$$

- (5) Repeat (4) for the next k .

2.3. Correlative Coefficient. The correlative coefficient is widely applied in almost all areas of science and technology. The correlative coefficient is a dimensionless index used in multivariate statistics to represent the statistical relationship between two groups of variables. Its value ranges from -1 to 1 , and it is divided into three classes, namely, positive correlation, irrelevant correlation, and negative correlation. Generally, certain processing in the computation is necessary to combine the negative correlation with the positive correlation. The value of the correlative coefficient ranges from 0 to 1 , and a high value indicates a strong correlation. After setting the two groups of variables, namely, x, y , the correlative coefficient ρ_{xy} is

$$\rho_{xy} = \frac{\text{cov}(x, y)}{\sqrt{\text{cov}(x, x) \cdot \text{cov}(y, y)}}, \quad (8)$$

where $\text{cov}(x, y) = (1/(n-1)) \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$, $\text{cov}(x, x) = (1/(n-1)) \sum_{i=1}^n (x_i - \bar{x})^2$, $\text{cov}(y, y) = (1/(n-1)) \sum_{i=1}^n (y_i - \bar{y})^2$, $\bar{x} = (1/n) \sum_{i=1}^n x_i$, and $\bar{y} = (1/n) \sum_{i=1}^n y_i$.

Thus, the correlative coefficient can be expressed as

$$\rho_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (9)$$

2.4. Signal Reconstitution Method. Signal characteristics are not evident because of the overlapped hydrogen absorption peaks in the near-infrared spectrum. Moreover, the modeling result using the original spectroscopy data is inferior, and the accuracy is not high. Therefore, removing useless components can produce satisfactory predictions and simplifies the model. According to improved CEEMDAN and the correlation coefficient, the signal reconstitution method can be concluded by employing the following steps.

- (1) The original signal is decomposed into IMF_i ($i = 1, 2, \dots, n$) by using the improved CEEMDAN algorithm, and n is the number of IMFs.
- (2) All the correlative coefficient value between IMF_i and the original signal is calculated using formula (9). The sensitive IMFs are selected according to the correlative coefficient threshold [30], which is shown in formula (10).

$$\mu_h = \frac{\max(\mu_i)}{10 \times \max(\mu_i) - 3} \quad (i = 1, 2, \dots, n). \quad (10)$$

In the formula above, μ_i represents the correlative coefficient between IMF_i and the original signal, and the maximum number of correlative coefficient is denoted by $\max(\mu_i)$.

If the correlative coefficient value between IMF_i and the original signal is larger than μ_h , then the relevant IMF is maintained as the sensitive mode. Otherwise, the relevant IMF is removed as a false component.

- (3) The sensitive IMFs are selected to reconstruct the signal for modeling.

2.5. PSO-SVR Modeling Method. The PSO algorithm is a type of parallel global search strategy that is based on population. It is easy to implement, and its concept is relatively simple; in PSO, many parameters no longer require adjustments. PSO exhibits a fast convergence speed and the capability of dealing with high-dimensional problems.

The speed-position model is used in the PSO algorithm. In the D -dimension solution space, the position of the i th particle in the group is $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})^T$, and the velocity ratio is $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})^T$. The individual extreme value at the current time is p_{ibest} , and the global extreme value is g_{best} . In each iteration process, the particles adjust the position and velocity of the current time by tracking the individual extreme value and global extreme value and state in the previous time. The iterative formula is shown as follows:

$$\begin{aligned} v_i(k+1) &= \omega * v_i(k) + c_1 * \text{rand}() \\ &\quad * (p_{ibest} - x_i(k)) + c_2 * \text{rand}() \\ &\quad * (g_{best} - x_i(k)), \end{aligned} \quad (11)$$

$$X_i(k+1) = X_i(k) + V_i(k+1),$$

where $V(k), V(k+1), X(k), X(k+1)$ are the velocity and position at the current moment and next moment, respectively; $\text{rand}()$ is the random number within $[0, 1]$, and c_1, c_2 are the learning factors which are usually equal to two. ω is the weighting factor that should automatically decrease with algorithm iteration to accelerate convergence speed; it is generally defined as

$$\omega = \omega_{\min} + \frac{(\text{iter}_{\max} - \text{iter}) * (\omega_{\max} - \omega_{\min})}{\text{iter}_{\max}}, \quad (12)$$

where ω_{\max} and ω_{\min} are the maximum and minimum weighting factors, respectively, iter is the current iteration number, and iter_{\max} is the total iteration number.

For the sample data set $\{x_i, y_i\}$ ($i = 1, 2, \dots, n$, $x_i \in R^d$, $y_i \in R$), the regression function obtained by SVR fitting is

$$\begin{aligned} f(x) &= w \cdot \theta(x) + b \\ &= \sum_{i=1}^n (\hat{a} - a_i) (\varphi(x_i) \cdot \varphi(x)) + b^* \\ &= \sum_{i=1}^n (\hat{a} - a_i) k(x_i, x) + b^*, \end{aligned} \quad (13)$$

where a_i and \hat{a} are Lagrangian operators and b^* is the threshold. Consider the following:

$$\begin{aligned} K(x_i, y_i) \\ = \exp\left(-\frac{|x_i - x_j|^2}{2\gamma^2}\right) \end{aligned} \quad \text{is the kernel function.} \quad (14)$$

For PSO-SVR, the position and velocity of each particle are determined by 3D parameters (ε, C, γ) . The mean square

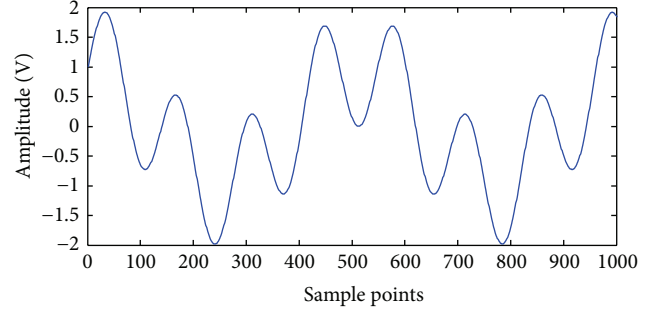


FIGURE 1: Waveform of pure signal.

error (MSE), which can directly reflect the regression performance of SVR, is used as the fitness function:

$$\text{MSE} = \left(\sum_{i=1}^n \frac{(\hat{y} - y_i)}{n} \right)^{1/2}, \quad (15)$$

where \hat{y} is the estimated value of a new sample.

The steps for the optimal selection of parameters (ε, C, γ) in PSO-SVR are described as follows.

- (1) The particle swarm ε, C, γ are initialized. Group size m is determined, the maximum and minimum weighting factors of algorithms $\omega_{\max}, \omega_{\min}$ are identified, and the maximum iteration number iter_{\max} is set.
- (2) The individual extreme value p_{ibest} of each particle is set as the current position. The fitness of each particle is set using the fitness function, namely, formulas (13) and (15). The individual extreme value corresponds to the particle with the best fitness as the global extreme value g_{best} .
- (3) On the basis of steps (1)–(3) for iteration calculation, the position and speed of particle are updated.
- (4) The fitness of each particle is evaluated using formulas (13) and (15).
- (5) If the fitness of each particle is better than the corresponding fitness p_{ibest} , then p_{ibest} is updated. Otherwise, the original value is retained.
- (6) If the updated p_{ibest} of each particle is better than the global extreme value g_{best} , the g_{best} is updated. Otherwise, the original value is retained.
- (7) If the maximum iteration is reached or if the solution does not change, the iteration is stopped. Otherwise, the process returns to step (3).

3. Experimental Results and Discussion

3.1. Simulation Signal Reconstruction Experiments. Consider the original signal $y(t) = \cos(4\pi t) + \sin(15\pi t)$. The length of the data is 1024, as shown in Figure 1. White Gaussian noise is added to the original signal with the input signal to noise ratio (SNR) fixed at 5 dB. Noisy signal $\nu(t)$ (Figure 2) is decomposed into eight modes. Figure 3 indicates that the sixth and seventh modes are two components of a pure signal.

TABLE 1: Correlative coefficients between each mode and noisy signal.

Mode	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8
Correlative coefficient	-0.0125	-0.0096	0.0054	0.0073	0.0579	0.7215	0.7429	0.0592

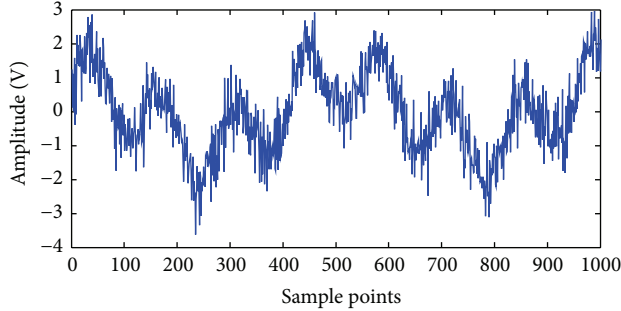


FIGURE 2: Waveform of noisy signal.

In the proposed method, the correlative coefficients between each IMF and noisy signal $y(t)$ are calculated, and the threshold is obtained with formula (7). The reconstructed signal is the sum of IMFs with correlative coefficients larger than the threshold. For noisy signal $y(t)$, the correlative coefficients of IMF6 and IMF7 are larger than the threshold (0.16721), which is shown in Table 1. The IMFs are arranged from high to low frequency, and the noise is often concentrated around the first IMFs. The first three modes should be removed to reconstruct the signal regardless of whether the correlative coefficients are larger than the threshold in the proposed method. The reconstructed signal is presented in Figure 4.

The simulated signal $y(t)$ is reconstructed to effectively reconstruct the spectral signal, and the reconstruction results for EMD, EEMD, CEEMD, CEEMDAN, and improved CEEMDAN are compared. In the experiments, the ensemble number and white Gaussian noise are fixed (the ensemble number is 100, and the SNR of white noise is 5 dB). The reconstruction effect is evaluated by introducing the SNR, root mean square error (RMSE), and correlative coefficient into this method. The SNR and RMSE are defined as

$$\text{SNR} = 10 \log_{10} \left(\frac{\sum_{t=1}^N (y(t))^2}{\sum_{t=1}^N (\bar{y}(t) - y(t))^2} \right), \quad (16)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N (\bar{y}(t) - y(t))^2},$$

where $\bar{y}(t)$ is the reconstructed signal and $y(t)$ is the original signal. Figure 5 demonstrates that EMD and improved CEEMDAN produce high SNR and small RMSE values. Hence, the reconstruction errors of these methods are smaller. However, EMD exhibits serious mode mixing. Moreover, the correlative coefficient values of these five methods are approximate, and the reconstruction effect can not be identified (Figure 5). Combining the SNR, RMSE, and correlative coefficients results (Table 2), the improved CEEMDAN method with the correlative coefficient exhibits a strong

TABLE 2: Values of SNR, RMSE, and correlative coefficient for reconstruction signal.

Methods	SNR	RMSE	Correlative coefficient
EMD	16.6707	0.1533	0.9896
EEMD	13.0117	0.2336	0.9754
CEEMD	14.2667	0.2021	0.9824
CEEMDAN	16.2156	0.1615	0.9888
Improved CEEMDAN	18.1517	0.1292	0.9924

TABLE 3: Values of SNR, RMSE, and correlative coefficient for different methods.

Methods	SNR	RMSE	Correlative coefficient
Improved CEEMDAN-CMSE	13.1254	0.2305	0.9817
Improved CEEMDAN-HD	16.8698	0.1498	0.9914
Improved CEEMDAN-MI	14.4521	0.1979	0.9835
Proposed method	18.2289	0.1289	0.9952

robustness to signal reconstruction and can extract useful signal information.

The experiment results from Figure 5 and Table 2 show that improved CEEMDAN method is better than that of the EMD, EEMD, CEEMD, and CEEMDAN. It is more suitable for the decomposition of nonstationary signal. For the signal $y(t)$, the proposed selected mode method based on correlative coefficient with improved CEEMDAN decomposition is compared with the methods in [26–28]. By the results shown as in Table 3, the correlative coefficient values of these four methods are approximate; however, the reconstruction effects of proposed method whose SNR is 18.2289 and RMSE is 0.1289 are superior to that of the other three methods. Therefore, the proposed method based on improved CEEMDAN has small reconstruction error and strong denoising ability.

3.2. Near-Infrared Spectrum Experiments

3.2.1. Near-Infrared Spectrum of the Glucose Solution. The instrument utilized for near-infrared spectroscopy is the Antaris II FT-NIR, produced by America Thermo Company. It is used to carry out full spectrum scanning transmission in the spectrum ranging from 833 nm to 2630 nm. The near-infrared spectroscopy of the 75 samples of 15 groups with different concentrations (range of 50–1000 mg/dL) of glucose solutions is selected to establish the spectrum calibration

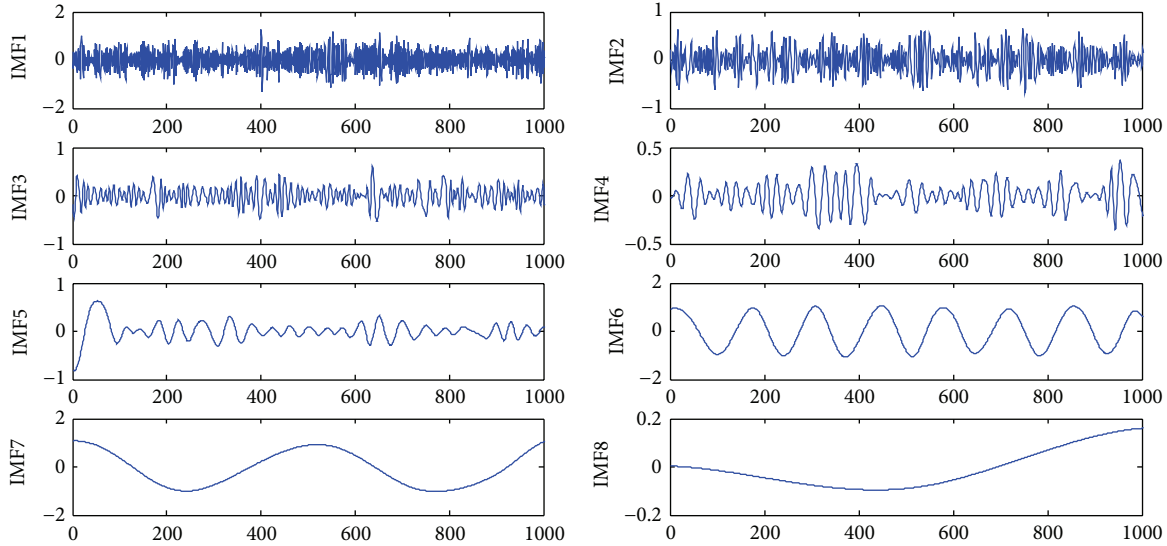


FIGURE 3: Waveform of IMFs.

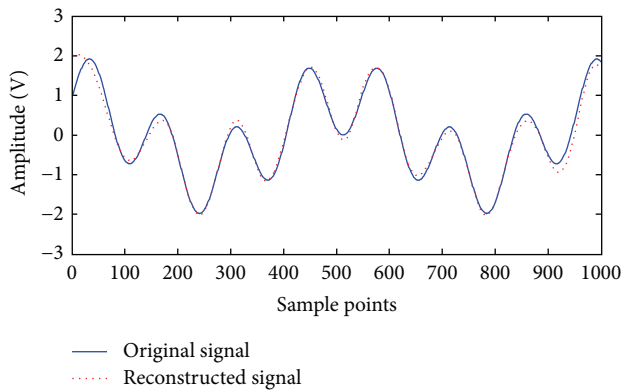


FIGURE 4: Waveform of reconstructed and original signal.

model. In the actual experiments, the transmission spectroscopy of the same sample is carried out several times to avoid the influence of instability of the near-infrared spectroscopy and to improve accuracy. Thus, in the modeling process, the average signal obtained from five measuring times for one solution sample is considered the training data. Some of the glucose solution spectra data are shown in Figure 6. The near-infrared spectrum of the samples is decomposed and reconstructed by employing improved CEEMDAN, which can eliminate the noise caused by some uncertain factors, such as spectrometer accuracy or test conditions, and can accurately provide spectral information for the calibration model.

3.2.2. Near-Infrared Spectrum Calibration Model Building. To maximize the sample data, we use cross validation until all the samples are tested once. The sample spectrum is collected independently. Thus, any individual can be chosen as the testing sample, and others are considered as training samples

to evaluate the modeling method. PSO-SVR only needs a small number of selected samples for training, particularly the data sets that are sensitive to noise, and can thus reflect the advantages of SVR and the intelligent optimization algorithm. To realize PSO-SVR, the constitution and prediction algorithms of SVR are embedded in the steps to calculate the fitness value in the PSO algorithm. Reference [31] provides the approximate scope of ε, C, γ as $\varepsilon = [0, 0.2]$, $C = [1, 10^8]$, and $\gamma = [0.01, 2]$; $\omega_{\max}, \omega_{\min}$, and m are generally taken as 0.9, 0.4, and 10, respectively. These initialization parameters can avoid choice blindness.

PLS regression, a classical calibration method, can extract the latent variables associated with the dependent variable in the spectrum and establish the regression equation. In this work, the PLS is the comparison method for PSO-SVR. PSO-SVR and PLS are utilized to establish the calibration model between the spectral data and the real glucose concentration.

After establishing the calibration model, the results must be verified to confirm the reliability of the model. The correlation coefficient and the root mean square error of prediction (RMSEP) are employed to evaluate the correlation of the model. The correlation coefficient can describe the linear correlation degree between the spectral matrix X and the concentration matrix Y ; it is commonly denoted by R . when the R value is closer to one, the regression effect of model is satisfactory. Moreover, when the RMSEP is small, the prediction precision of model is high. The correlation coefficient and RMSEP are calculated as

$$R = \sqrt{1 - \frac{\sum (\hat{y}_i - y_i)^2}{\sum (\bar{y}_i - y_i)^2}}, \quad (17)$$

$$\text{RMSEP} = \sqrt{\frac{\sum (\hat{y}_i - y_i)^2}{n - 1}},$$

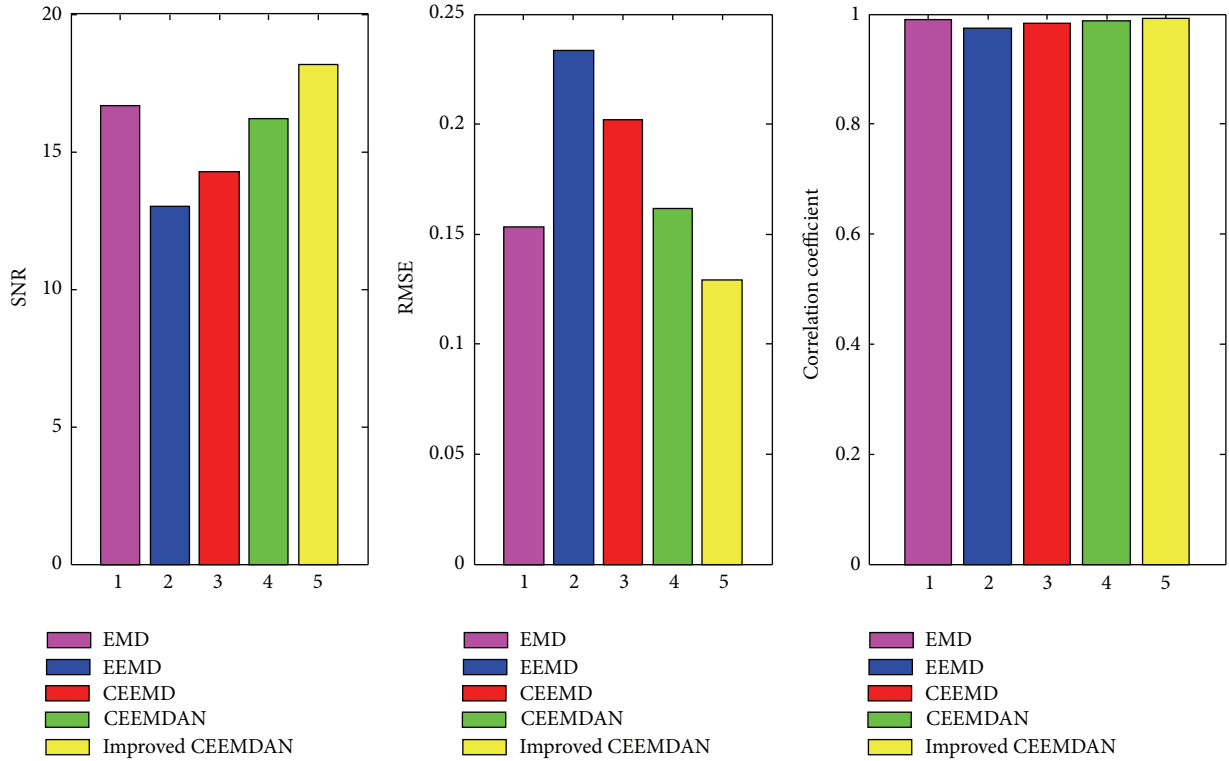


FIGURE 5: SNR, RMSE, and correlative coefficient for $y(t)$.

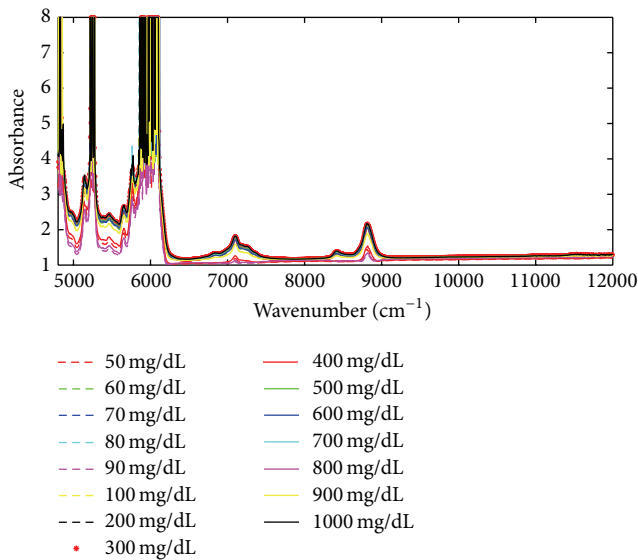


FIGURE 6: Some of the glucose solution spectrum.

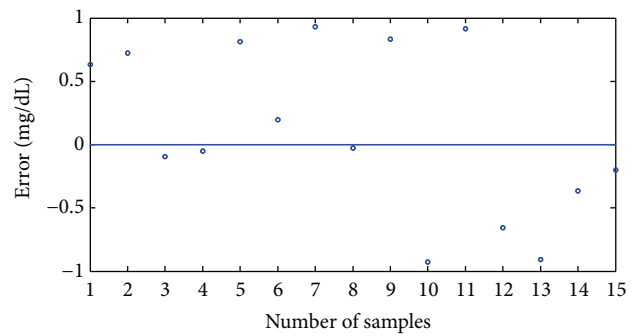


FIGURE 7: The errors between predicted values and true values.

TABLE 4: Values of R and RMSEP for calibration model.

Methods	R	RMSEP
PLS	0.9999825	0.9100
Improved CEEMDAN-PLS	0.9999985	0.6519
PSO-SVR	0.9999986	0.5560
Improved CEEMDAN-PSO-SVR	0.9999997	0.5352

where n is the sample quantity of the calibration set, y_i is the true value of the i th sample, \hat{y}_i is the predicted value of the i th sample, and $\bar{\hat{y}}$ is the average value of \hat{y}_i of all the samples in the calibration set.

Two types of calibration models introduced above are employed to predict the glucose concentrations in 15 glucose

solutions samples. The errors between the predicted values and the true values are then calculated. The results are shown in Figure 7. The predicted values and true values are provided in Figure 8. Table 4 shows that the PSO-SVR, which is based on improved CEEMDAN for predictions and whose R and RMSEP are 0.9999997 and 0.5352, respectively, is more

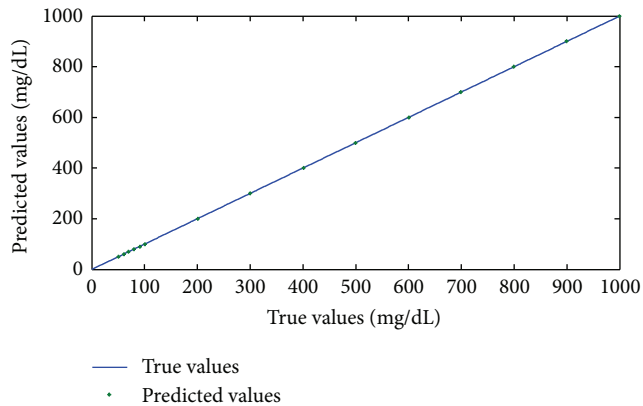


FIGURE 8: The predicted values and true values of cross validation.

robust and accurate than PLS, whose R and RMSEP are 0.9999825 and 0.91, respectively. The R and RMSEP values of the calibration model established with the spectrum data, which were reconstructed with improved CEEMDAN in the PLS and PSO-SVR methods, are better than those of the calibration model with the original spectrum data. Thus, PSO-SVR, which is based on improved CEEMDAN, exhibits a good performance in detecting the glucose concentrations. The excellent prediction results of PSO-SVR indicate that near-infrared spectroscopy technology can be employed to detect glucose concentrations.

4. Conclusion

The quantitative analysis of near-infrared spectroscopy data demonstrates the potential development in noninvasive blood glucose detection. PSO-SVR is an effective method to solve the regression problem on high-dimensional data matrixes. This study proposes the PSO-SVR modeling method that is based on the improved CEEMDAN algorithm. The proposed method can remove noise, extract useful information from near-infrared spectra, and optimize the parameters in SVR. The near-infrared spectrum analysis model established with the PSO-SVR based on improved CEEMDAN method is stable, accurate, and practicable and it exhibits a good predictive effect. The current experiments focus on glucose solutions. In the future, we will extend the experiments to human tissues.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

This work was supported by the “Fundamental Research Funds for the Central Universities” (Grant no. HIT.IBRSEM.201307) and Program for Harbin City Science and Technology Innovation Talents of Special Fund Project (Grant no. 2014RFXJ065).

References

- [1] M. Banerjee and M. Abbas, “Genetic studies of lifestyle diseases: type 2 diabetes mellitus and cancer,” *Journal of Experimental Zoology, India*, vol. 18, pp. 1–7, 2015.
- [2] A. Hussein and N. Abu, “Cardiovascular diseases, diabetes mellitus, and cancer in the occupied Palestinian territory,” *The Lancet*, vol. 373, no. 9668, pp. 1041–1049, 2009.
- [3] D. Ding, *Near infrared spectroscopy research and application in biomedicine [Ph.D. thesis]*, JiLin University, 2004.
- [4] R. N. Johnson and J. R. Baker, “Accuracy of devices used for self-monitoring of blood glucose,” *Annals of Clinical Biochemistry*, vol. 35, no. 1, pp. 68–74, 1998.
- [5] K. Song, U. Ha, S. Park, J. Bae, and H.-J. Yoo, “An impedance and multi-wavelength near-infrared spectroscopy IC for non-invasive blood glucose estimation,” *IEEE Journal of Solid-State Circuits*, vol. 50, no. 4, pp. 1025–1037, 2015.
- [6] H. C. Garcia, I. Papadopoulos-Kelidis, I. Sotiriou et al., “Glucose sensing in saline solutions using V-band waveguides,” in *Proceedings of the 4th International Conference on Wireless Mobile Communication and Healthcare (MOBIHEALTH '14)*, pp. 320–323, Athens, Greece, November 2014.
- [7] A. A. Anisimov and A. Grigorash, “Development of a system for the glucose level in the blood monitoring,” in *Proceedings of the IEEE NW Russia Young Researchers in Electrical and Electronic Engineering Conference (ElConRusNW '14)*, pp. 23–25, St. Petersburg, Russia, February 2014.
- [8] Y. Khawam, M. Ali, H. Shazada, S. Kanan, and H. Nashash, “Non-invasive blood glucose measurement using transmission spectroscopy,” in *Proceedings of the 1st International Conference on Communications, Signal Processing and Their Applications (ICCSPA '13)*, pp. 1–4, Sharjah, United Arab Emirates, February 2013.
- [9] J. Yadav, A. Rani, V. Singh, and B. M. Murari, “Near-infrared LED based non-invasive blood glucose sensor,” in *Proceedings of the 1st International Conference on Signal Processing and Integrated Networks (SPIN '14)*, pp. 591–594, February 2014.
- [10] K. A. U. Menon, D. Hemachandran, and A. T. Kunnath, “Voltage intensity based non-invasive blood glucose monitoring,” in *Proceedings of the 2013 4th International Conference on Computing, Communications and Networking Technologies (ICCCNT '13)*, pp. 1–5, IEEE, Tiruchengode, India, July 2013.
- [11] S. Ramasahayam, K. S. Haindavi, B. Kavala, and S. R. Chowdhury, “Non invasive estimation of blood glucose using near infrared spectroscopy and double regression analysis,” in *Proceedings of the 7th International Conference on Sensing Technology (ICST '13)*, pp. 627–631, Wellington, New Zealand, December 2013.
- [12] K. A. Unnikrishna Menon, D. Hemachandran, and T. K. Abhishek, “A survey on non-invasive blood glucose monitoring using NIR,” in *Proceedings of the 2nd International Conference on Communication and Signal Processing (ICCSP '13)*, pp. 1069–1072, IEEE, Melmaruvathur, India, April 2013.
- [13] Y.-J. Kim, S. Hahn, and G. Yoon, “Determination of glucose in whole blood samples by mid-infrared spectroscopy,” *Applied Optics*, vol. 42, no. 4, pp. 745–749, 2003.
- [14] B. D. Cameron, H. Gorde, and G. L. Cote, “Development of an optical polarimeter for in vivo glucose monitoring,” in *Proceedings of the 1999 Optical Diagnostics of Biological Fluids IV*, vol. 3566, pp. 43–49, January 1999.

- [15] Z. M. Zhao, *Pulsed photoacoustic techniques and glucose determination in human blood and tissue [Ph.D. thesis]*, University of Oulu, 2002.
- [16] A. J. Berger, T.-W. Koo, I. Itzkan, G. Horowitz, and M. S. Feld, "Multicomponent blood analysis by near-infrared Raman spectroscopy," *Applied Optics*, vol. 38, no. 13, pp. 2916–2926, 1999.
- [17] L. Heinemann, U. Kramer, H. Klotzer et al., "Noninvasive glucose measurement by monitoring of scattering coefficient during oral glucose tolerance tests," *Diabetes Technology and Therapeutics*, vol. 2, no. 2, pp. 211–220, 2000.
- [18] S.-W. Wang, X. Chen, W. Lu, L. Wang, Y. Wu, and Z. Wang, "Integrated optical filter arrays fabricated by using the combinatorial etching technique," *Optics Letters*, vol. 31, no. 3, pp. 332–334, 2006.
- [19] K. A. Unnikrishna Menon, D. Hemachandran, and T. K. Abhishek, "A survey on non-invasive blood glucose monitoring using NIR," in *Proceedings of the 2nd International Conference on Communications and Signal Processing (ICCSP '13)*, pp. 1069–1072, IEEE, Melmaruvathur, India, April 2013.
- [20] J. J. Burmeister, M. A. Arnold, and G. W. Small, "Noninvasive blood glucose measurements by near-infrared transmission spectroscopy across human tongues," *Diabetes Technology and Therapeutics*, vol. 2, no. 1, pp. 5–16, 2000.
- [21] Z.-Y. Zhang, G. Li, L. Lin, X.-Y. Cui, and B.-J. Zhang, "Application of EMD and SPA algorithm in the detection of benzoyl peroxide addition in flour by spectroscopy," *Spectroscopy and Spectral Analysis*, vol. 32, no. 10, pp. 2815–2819, 2012.
- [22] A.-H. Liao, C.-C. Shen, and P.-C. Li, "Contrast improvement by combining pulse inversion with EMD and EEMD," in *Proceedings of the IEEE International Ultrasonics Symposium (IUS '09)*, pp. 287–290, Rome, Italy, September 2009.
- [23] J.-R. Yeh, J.-S. Shieh, and N. E. Huang, "Complementary ensemble empirical mode decomposition: a novel noise enhanced data analysis method," *Advances in Adaptive Data Analysis*, vol. 2, no. 2, pp. 135–156, 2010.
- [24] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *Proceedings of the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '11)*, pp. 4144–4147, IEEE, Prague, Czech Republic, May 2011.
- [25] M. A. Colominas, G. Schlotthauer, and M. E. Torres, "Improved complete ensemble EMD: a suitable tool for biomedical signal processing," *Biomedical Signal Processing and Control*, vol. 14, no. 1, pp. 19–29, 2014.
- [26] A.-O. Boudraa and J.-C. Cexus, "EMD-based signal filtering," *IEEE Transactions on Instrumentation & Measurement*, vol. 56, no. 6, pp. 2196–2202, 2007.
- [27] A. Komaty, A. Boudraa, and D. Dare, "EMD-based filtering using the Hausdorff distance," in *Proceedings of the 12th IEEE International Symposium on Signal Processing and Information Technology (ISSPIT '12)*, pp. 292–297, Ho Chi Minh City, Vietnam, December 2012.
- [28] L. Han, C. Li, and H. Liu, "Feature extraction method of rolling bearing fault signal based on EEMD and cloud model characteristic entropy," *Entropy*, vol. 17, no. 10, pp. 6683–6697, 2015.
- [29] S. Bhattacharjee, "Chemometrics in analytical chemistry," in *Chemical Engineering*, pp. 1–13, 1994.
- [30] A. Ayenu-Prah and N. Attoh-Okine, "A criterion for selecting relevant intrinsic mode functions in empirical mode decomposition," *Advances in Adaptive Data Analysis*, vol. 2, no. 1, pp. 1–24, 2010.
- [31] B. Üstün, W. J. Melssen, M. Oudenhuijzen, and L. M. C. Buydens, "Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization," *Analytica Chimica Acta*, vol. 544, no. 1-2, pp. 292–305, 2005.