An Extreme Learning Machine optimized by Differential Evolution and Artificial Bee Colony for Predicting the Concentration of Whole Blood with Fourier Transform Raman Spectroscopy

Qiaoyun Wang[1,2*], Shuai Song[1], Lei Li[1], Da Wen[1], Peng Shan[1], Zhigang Li[1], YongQing Fu[3]

1. College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning Province,110819, China

2. Hebei Key Laboratory of Micro-Nano Precision Optical Sensing and Measurement Technology, Qinhuangdao, 066004, China

3. Faculty of Engineering & Environment, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK

*** Correspondence:** wangqiaoyun@neuq.edu.cn

Abstract: Raman spectroscopy, with its advantages of non-contact nature, rapid detection, and minimum water interference, is promising for non-invasive blood detection or diagnosis in clinic applications. However, there is a critical issue that how to accurately analyze blood composition by Raman spectroscopy. In this study, we apply extreme learning machine (ELM) algorithm and a multivariate calibration regression model to analyze the results from Raman spectroscopy and determine the component's concentrations in blood samples, including glucose, cholesterol, and triglyceride. Self-adaption differential evolution artificial bee colony (SADEABC) algorithm was further applied to increase the data's accuracy and robustness. The obtained data for coefficient of determination, root mean square error of calibration, root mean square error of prediction, and relative percent deviation, were 0.9822, 0.3993, 0.3827, and 6.6679 for glucose, 0.9786, 0.2104, 0.2088 and 5.9533 for cholesterol, and 0.9921, 0.2744, 0.3433 and 10.5075 for triglyceride, respectively.

Results demonstrated that the model based on SADEABC-ELM show much better prediction data than those models based on the ELM and ABC-ELM.

Keywords: Extreme Learning Machine; Raman spectroscopy; Artificial Bee Colony algorithm; Self-Adaption Differential Evolution; blood detection

1.Introduction

Along with rapid economic development and significantly improvement of people's living and diet qualities, there is an increasing issue on the rising cases of cardiovascular and diabetic disorders [1]. Therefore, the top priority issue is to detect the disease in time, and effectively treat as soon as possible. In clinical practice, analysis concentrations or levels of relevant components in the blood samples is one of the most commonly methods to detect these diseases in timely [2]. However, these invasive detection methods have disadvantages including consumption of a large amount of blood samples, high risk of infection, time-consuming, and needs of different chemical reagents to analyze different substances [3, 4]. Therefore, minimally invasive or non-invasive blood Spectral detection has recently attracted extensive attention because of its safety, non-polluting, and rapidity [5].

Spectral detection technologies can be applied to identify the chemical compositions and relative contents of substances simultaneously based on the spectrum obtained from the tested platform. Raman spectroscopy can achieve non-invasive and rapid detection for trace amount, and provide crucial information on compositions and structures of blood samples. The unique feature of high resistance to water interference using the Raman spectroscopy analysis makes it one of the preferred spectroscopic techniques for analyzing blood samples [6-8].

Combining Raman spectroscopy techniques with chemometric methods enables the opportunities for quantitative analysis of blood components [7], which not only relies on detection methods to obtain valid data, but also needs good regression analysis

methods. For this purpose, there are a few frequently used regression methods developed, including principal component analysis (PCA) [9], partial least squares regression (PLSR) [10], support vector machine (SVM) [11], and extreme learning machine (ELM) [12]. In recent years, convolutional neural network (CNN) has also been increasingly used in Raman spectroscopy. Mozaffari used 1D-CNN to train on millions of Raman spectra from standard available reference databases. The experimental results verified that 1D-CNN could identify one pure unknown Raman instance from thousands of classes with a high accuracy [13]. Wu developed a CNN model to identify the quality of honey by Raman spectroscopy of honey [14]. Wang have used the one-dimensional shallow CNN structure combined with elastic nets with elastic nets to predict the glucose concentration [15, 16]. Compared with other algorithm, ELM was learned without iteration, and it is easy to reach global optimal solution than other methods. Due to its superior convergence, less optimization constrains, good generalization capability and simplicity [17], the ELM algorithm is a good choice for quantitative analysis with Raman spectroscopy.

ELM has been widely applied for solving various classification and regression problems in many fields such as computer vision [17-19], bioinformatics [20, 21] and environmental science [22, 23]. This paper is focused on the ELM as the core algorithm of the quantitative regression model. There are many studies for applying the ELM model in the field of spectral analysis. For example, Chen et al. proposed an ensemble ELM algorithm (EELM), and showed that EELM outperformed the classical PLSR [24]. Tan et al. used near-infrared spectroscopy to identify detergent powder brands and constructed an ensemble ELM model to classify different detergent powder brands, achieving an accuracy of the model 100% [25]. Chu et al. used laser-induced breakdown spectroscopy to identify nasopharyngeal carcinoma, and combined the random forest (RF) method and ELM to build an RF-ELM model, and the recognition accuracy of the model reached 98.330% [26]. Wang et al. proposed a partial least squares regression residual extreme learning machine (PLSRR-ELM) model for rapid determination of research octane number (RON) of blended gasoline, and showed that the model performed better than other models [27].

In this study, we choose ELM as the fundamental model to enhance the algorithm and boost model's robustness and prediction accuracy. For quantitative analysis of blood glucose, cholesterol, and triglyceride concentrations in the blood sample, we develop an ELM model optimized by the Self-Adaption Differential Evolution Artificial Bee Colony (SADEABC) algorithm. At first, the theory of the SADEABC-ELM is given and then the SADEABC-ELM model is used to analysis of the Raman spectrum of whole blood. The results show that the combination of the SADEABC-ELM model with Raman spectroscopy is a promising method for blood components quantitative analysis.

2. Materials and methods

2.1 FT-Raman spectroscopy and acquisition conditions

The blood samples used in this experiment were collected from the First Hospital of Qinhuangdao City, Hebei Province. Total 106 blood samples (including healthy volunteers and clinically diagnosed diabetic patients) were collected from 106 individual patients. Each of blood sample (about 5 mL) were divided into two groups. The first group (about 4mL) were measured according to the clinical standard method of the First Hospital of Qinhuangdao as standard values. The other group (about 1mL) were immediately stored in a refrigerator at 4°C for Raman spectroscopy analysis.

Raman spectroscopy at 1064 nm excitation radiation (Nd: YAG laser with the maximum output power of 500mW) were recorded with a FT-Raman Bruker MultiRAM spectrometer (Bruker Optics, Germany) equipped with a high-sensitivity Germanium diode detector operating at liquid nitrogen temperature. The spectrum range is 400–4000 cm^{-1}. All the Raman spectroscopy were recorded with 64 scans at a spectral resolution of 6 cm^{-1} under the laser power of 90mW. To ensure the accuracy of the experiment, the spectral data was acquired three times for each sample and then averaged as the raw spectral data. Spectral data were acquired using OPUS 7.0 software.

2.2 Extreme Learning Machine

ELM is a single-hidden-layer structure, whose input weights and hidden-layer

biases are generated randomly [28]. Compared with the other algorithms, ELM has better learning accuracy and speed [24]. **Fig.1** depicts its structure, showing the input, hidden and output layers. For a given set of *N* samples $\{(x_i, y_i)|i \in [1, n]\}$, the input vector and the output vector of the network can be written as $x_i = [x_{i1}, x_{i2}, \cdots, x_{il}]^T \in R^l$ and $y_i = [y_{i1}, y_{i2}, \cdots, y_{il}]^T \in R^l$. x_i is the Raman spectrum of ith sample, and y_i is the concentration of ith sample. The number of hidden layer nodes is *l*, and the SLFN with activation function $g(x)$ can be denoted as:

$$y_i = \sum_{j=1}^{l} \beta_j g(\omega_j \cdot x_i + b_j), i = 1, \cdots, N \tag{1}$$

Where $\omega_j = [\omega_{j1}, \omega_{j2}, \cdots, \omega_{jl}]^T$ is the weight vector between the input layer and the jth node of the hidden layer, $\beta_j = [\beta_{j1}, \beta_{j2}, \cdots, \beta_{jm}]^T$ is the output weight vector between the node of jth hidden layer and the nodes of output layer, and b_j is the deviation of the jth hidden node [29].

The formula of what can be written as follows:

$$\mathbf{H\beta} = \mathbf{Y} \tag{2}$$

where $\mathbf{H}$ is the output matrix of the hidden layer of the network, and $\boldsymbol{\beta} = [\beta_1\ \beta_2\ \cdots\ \beta_l]_{l\times m}^T$ is the output weight matrix, and $\boldsymbol{Y} = [y_1\ y_2\ \cdots\ y_m]_{l\times m}^T$ is the expected concentration matrix. If the network structure parameters are remained unchanged and the function $g(x)$ is infinitely differentiable, the output weight matrix $\boldsymbol{\beta}$ can be solved by Eq. (3):

$$\min_{\boldsymbol{\beta}} \left\| \mathbf{H\beta} - \mathbf{Y} \right\| \tag{3}$$

the solution of $\boldsymbol{\beta}$ can be obtained using Eq.(4):

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^{+}\mathbf{Y} \tag{4}$$

where $\mathbf{H}^{+}$ is the Moore-Penrose generalized inverse of $\mathbf{H}$.

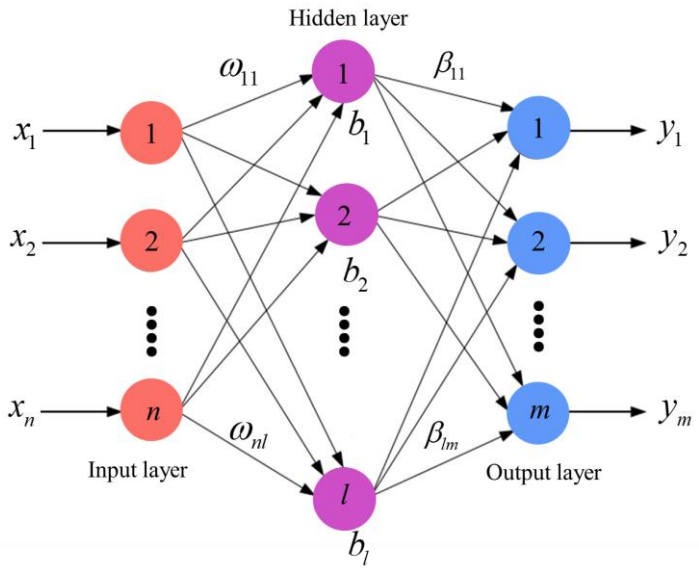

Fig. 1 The illustration of ELM

2.3 SADEABC Algorithm

Over the past decade, there are extensive research on the parameter optimization algorithms of the random initialized weights and biases in ELM, such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Artificial Bee Colony Algorithm (ABC). Compared with other algorithms, ABC algorithm just has fewer parameters and is easy implementation [30].

ABC algorithm is a brand-new intelligent optimization algorithm which takes its cues from feeding strategies of bee colonies [31]. In the ABC algorithm, there are three different categories of bees that are used to search for the greatest food sources, i.e., employed bees, onlooker bees, and scout bees. Employed and onlooker bees accelerate algorithm convergence, while scout bees manage the global search. However, the disadvantages of ABC optimization are easily under-fitting solutions, lacked stability and robustness [30]. The Self-adaption differential evolution (SADE) algorithm is used to find the optimal solution of ABC algorithm. The specific steps for SADE-ABC can be listed as follows:

Step 1: Initialization. The parameters of ABC (e.g., the number of bees, the maximum number of cycles, and the maximum number of solution constants, etc.) are created and SN feasible solutions (nectar sources) are generated randomly. The feasible solutions are initialized as follows:

$$\mathbf{X}_i^j = \mathbf{X}_{\min}^j + r * (\mathbf{X}_{\max}^j + \mathbf{X}_{\min}^j) \tag{5}$$

where $i = 1, 2, \cdots, N$, $j = 1, 2, \cdots, D$, D is the dimension size, r is a random number between 0 and 1, $\mathbf{X}_i$ is the ith nectar source, and $\mathbf{X}_{\max}^j$ and $\mathbf{X}_{\min}^j$ are the maximum and minimum values of the nectar source position within the optimization range, respectively.

Step 2: Employed bee search phase. Bees are used to search near the current nectar source and a new solution is generated according to Eq. (6):

$$v_{ij} = x_{ij} + \phi_{ij} * (x_{ij} + x_{kj}) \tag{6}$$

where $i, k \in \{1, 2, \cdots, SN\}$ and $j \in \{1, 2, \cdots, D\}$ are randomly selected, and $k \neq i$; ϕ_{ij} is randomly generated in [-1, 1]. A greedy algorithm is used here to select a better candidate solution to replace the original solution. The greedy selection method is utilized as follows:

$$\mathbf{X}_i = \begin{cases} \mathbf{V}_i, & \text{if } \mathbf{V}_i \text{ outperforms } \mathbf{X}_i \\ \mathbf{X}_i, & \text{otherwise} \end{cases} \tag{7}$$

Step 3: Onlooker bee search phase. Each onlooker bee uses the selection probability p_i to select a solution among those of all the employed bees. Then the onlooker bees search for new solutions and use the greedy algorithm to select optimal solutions. The formula for calculating the selection probability p_i is

$$p_i = \frac{fit_i}{\sum_{k=1}^{SN} fit_k} \tag{8}$$

where fit_i is the fitness value of $\mathbf{X}_i$, and it can be calculated by

$$fit_i = \begin{cases} \dfrac{1}{1 + fit_i} , & if \quad f_i \geq 0 \\ 1 + abs(f_i) , & if \quad f_i < 0 \end{cases} \tag{9}$$

where f_i is the objective function value of $\mathbf{X}_i$.

Step 4: Scout bee search phase. When the solution hits the upper limit without being updated, it is abandoned to exit the local optimum and a new solution is created

randomly based on Eq. (5). At the same time, the corresponding employed bee is converted into the scout bee.

By switching between the above three types of functional bees, the algorithm heuristically searches the optimal solution both locally and globally. However, the ABC has issues such as premature convergence and low search accuracy, so its models and structures are still needed for further improvement.

The ABC has been combined with Differential Evolution (DE) [32] Algorithm to search for the optimal solution [33, 34]. The DE algorithm maintains a population containing *NP* candidate solutions, denoted as $\mathbf{X}_j^G = \{x_{1,j}^G, x_{2,j}^G, \cdots, x_{D,j}^G\}$, where G is the number of population evolution. The population goes through processes such as mutation operation, crossover operation and selection operation[35]. If the end condition is not met, these processes will continuously be repeated.

The most common method used in mutation operations is:

$$\mathbf{V}_j^G = \mathbf{X}_{r1}^G + F * (\mathbf{X}_{r2}^G - \mathbf{X}_{r3}^G) \tag{10}$$

where the mutation rate F is a fixed value between 0 and 1, $r1, r2, r3 \in \{1, 2, \cdots, NP\}$ are randomly selected individuals, in which $r1 \neq r2 \neq r3$.

The crossover operation realizes the information fusion of the target vector $\mathbf{X}_j^G$ and the mutation vector $\mathbf{V}_j^G$, and a new vector $\mathbf{U}_j^G$, which is called the trial vector and can be produced by:

$$u_{i,j}^G = \begin{cases} v_{i,j}^G & , rand(0,1) \leq CR \\ x_{i,j}^G & , rand(0,1) > CR \end{cases} \tag{11}$$

where $i \in \{1, 2, \cdots, D\}$,and CR is a fixed value on [0,1].

The selection operation performs a greedy selection between $\mathbf{X}_j^G$ and $\mathbf{U}_j^G$, and the one with better fitness will be kept. This operation is performed as:

$$\mathbf{X}_j^{G+1} = \begin{cases} \mathbf{U}_j^G & , f(\mathbf{U}_j^G) \leq f(\mathbf{X}_j^G) \\ \mathbf{X}_j^G & , f(\mathbf{U}_j^G) > f(\mathbf{X}_j^G) \end{cases} \tag{12}$$

In the DE algorithm, the mutation rate and the crossover rate are fixed values,

which are artificially set based on experience. They are remained unchanged in the entire algorithm. The concept of Self-Adaptation Differential Evolution (SADE) has been introduced to reduce the dependence on these fixed parameters, and the mutation rate and crossover rate are adjusted by using the current global optimal fitness value and the average fitness value. The adaptive adjustment method is based on Eq. (13) and Eq. (14):

$$F = F_0\left(1-\frac{GBestfit - mean(fit)}{GBestfit}\right) \tag{13}$$

$$CR = CR_0\left(1-\frac{GBestfit - mean(fit)}{GBestfit}\right) \tag{14}$$

The adaptive principle is summarized as follows. When the global optimal value of the current fitness value is significantly different from the mean value, the values of the mutation rate and the crossover rate will decrease accordingly, therefore, the probability for the occurrence of random mutation at each solution's position is reduced. The reduced degree of exchange of position information between the two sets of solutions helps to reduce the step size of the optimization process and keep the current optimal states as much as possible. When the global optimal value of the current fitness value is less than the mean value, the mutation rate and the crossover rate are increased accordingly, thus increasing the probability of random mutation of each solution's position and improving the degree of exchange of position information between the two sets of solutions. Increasing the step size during the optimization process is also beneficial as this increases the likelihood that the algorithm will be updated and departed from the local optimum in the global context. This improved method will reduce the probability of bad evolution when a better solution is obtained.

2.4 Evaluation criterion of model

Several indicators, including the coefficient of determination (R^2), the root mean square error of calibration (RMSEC), the root mean square error of prediction (RMSEP), and the relative percent deviation (RPD), are often used to evaluate the performance of quantitative regression models and evaluate the established model. The R^2, RMSEC, RMSEP and RPD are defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_{i,p} - y_{i,p})^2}{\sum_{i=1}^{n}(\overline{y}_p - y_{i,p})^2} \tag{15}$$

$$\mathrm{RMSEC} = \sqrt{\frac{\sum_{i=1}^{m}(\hat{y}_{i,c} - y_{i,c})^2}{m}} \tag{16}$$

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_{i,p} - y_{i,p})^2}{n}} \tag{17}$$

$$\mathrm{RPD} = \sqrt{\frac{\sum_{i=1}^{n}(\overline{y}_p - y_{i,p})^2}{\sum_{i=1}^{n}(\hat{y}_{i,p} - y_{i,p})^2}} \tag{18}$$

where *m* and *n* are the numbers of the calibration and prediction samples, respectively, $y_{i,c}$ and $\hat{y}_{i,c}$ are the reference values and predicted values of the calibration samples, $y_{i,p}$, $\hat{y}_{i,p}$ and $\overline{y}_p$ are the reference values, predicted values and average value of the prediction samples, respectively.

3. Results and discussions

3.1 Sample data processing

The Raman spectral data of 106 groups of experimental blood samples were obtained, and examples of the obtained Raman spectra are shown in Fig.2. In the obtained Raman spectral data, there may be non-human interferences and external factors during transportation and storage [36], so it is necessary to perform preprocessing operations such as baseline correction and Savitzky-Golay denoising on the samples data before establishing a quantitative regression model.

To avoid excluding essential information related to the components in the special band interval in Raman spectra in the blood, the whole band interval is chosen to be modeled. The principal component analysis is introduced for dimensionality reduction of the experimental data [37]. In this study, the number of principal components selected after repeated experiments was 10, and the extracted data were proven to reflect the main feature information of the original data, thus avoiding complicated calculations.

The Duplex algorithm was employed to select sample for the training set and testing sets. The principle of Duplex algorithm is selected samples according to their mutual Euclidean distance. At first, two most distanced samples from the dataset T will be chosen, and then repetitively select samples with maximal distance to the previously

sampled examples. These methods can ensure a maximum coverage of T. And the numbers of training sets and testing sets were 65 and 41, respectively. Training sets were used to train the SADEABC-ELM model and produce the best performance. Finally, the data of the testing sets were input into the final model to obtain the final predicted concentration value.

3.2 Establishment and analysis of quantitative regression model

At first, we developed a quantitative regression model based on the ELM algorithm and optimized the parameters of the model by GA algorithm, PSO algorithm and ABC algorithm. The GA-ELM, PSO-ELM and ABC-ELM models were finally obtained, and then the performance of these three models was compared, and the comparison results are shown in **Table 1**. The results shown that the performance of the optimized model with intelligent optimization algorithms is better than the traditional ELM model. And the ABC-ELM model has higher accuracy, stability and shorter time than GA-ELM and PSO-ELM model. But the accuracy of ABC-ELM model not meet the requirements. So, the SADE algorithm is used to improve the accuracy of ABC-ELM model.

Table 1 Evaluation indexes of three algorithm optimization models

Models	R^2	RMSEC	RMSEP	RPD	NG	TIME（s）
ELM	0.8263	0.0041	0.1365	1.6384	——	——
GA-ELM	0.8429	0.0222	0.1328	1.5324	56	52.191
	0.8429	0.0222	0.1328	1.5324	90	87.3530
	0.8875	0.0090	0.1248	1.7802	320	312.711
	0.8875	0.0090	0.1248	1.7802	500	451.819
PSO-ELM	0.8767	0.0313	0.0788	2.8194	56	25.363
	0.8800	0.0296	0.0747	2.9735	90	47.560
	0.8800	0.0296	0.0747	2.9735	320	148.008
	08800	0.0296	0.0747	2.9735	500	213.912
ABC-ELM	**0.8942**	**0.0212**	**0.0748**	**2.9719**	**56**	**26.923**
	0.8942	0.0212	0.0748	2.9719	90	49.862
	0.8942	0.0212	0.0748	2.9719	320	146.543
	0.8942	0.0212	0.0748	2.9719	500	230.358

After preprocessing, sample selection and band selection of experimental data, quantitative regression models were established for glucose, cholesterol and triglyceride components in the blood samples for concentration detection, and the proposed and constructed SADEABC-ELM model was used for prediction. **Fig. 3** displays the obtained concentration forecast of the quantitative regression model of the blood glucose component parts. The concentration prediction of the quantitative regression model of cholesterol and triglyceride components in the blood samples are shown in **Fig. 4** and **Fig. 5**, respectively. The closer the data distribution is near to the straight-line of y=x, the better the prediction effect would be. Results shown in Figs. 3 to 5 reveal that data distribution using the SADEABC-ELM model is more concentrated on the diagonal line, while that using the ELM models more widely dispersed, which clearly prove the accuracy of SADEABC-ELM model.

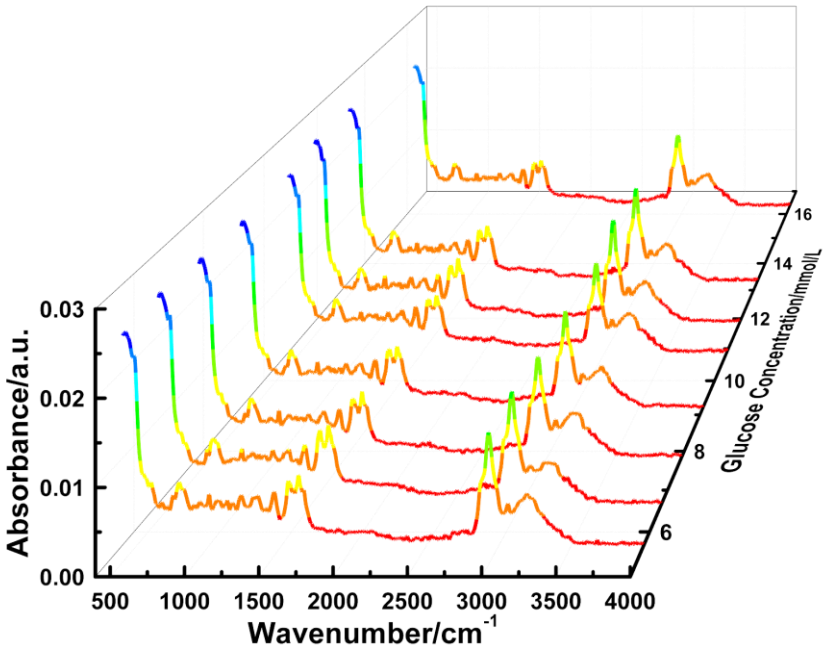

Fig. 2 Original Raman spectroscopy of blood samples with different glucose concentration

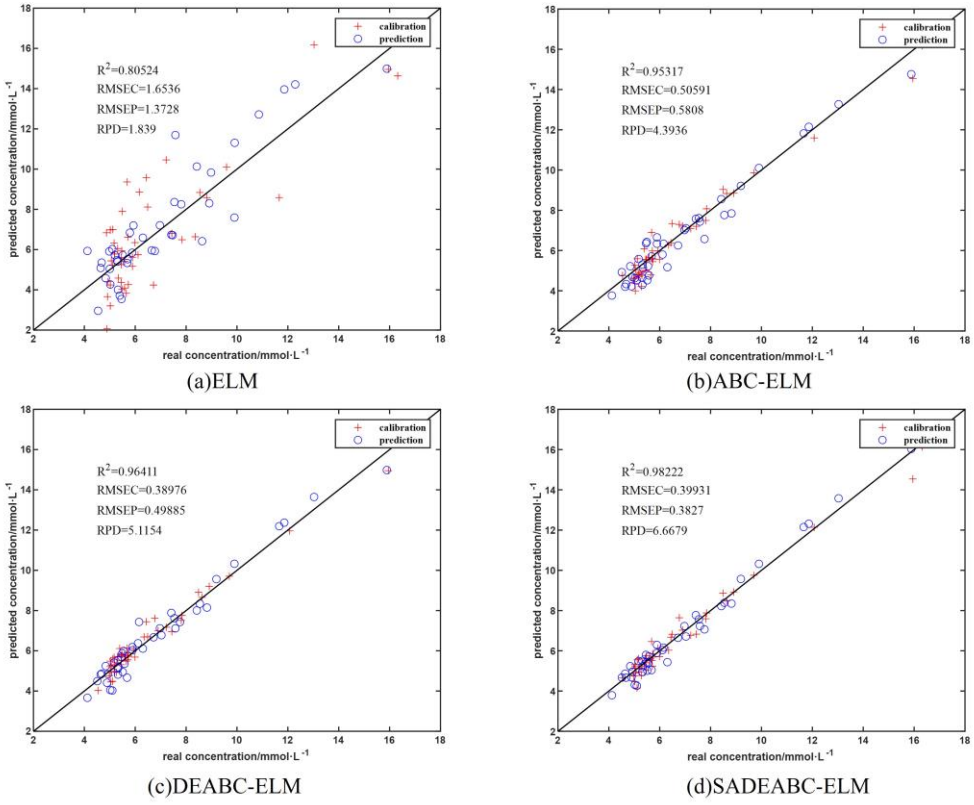

Fig. 3 Concentration prediction by quantitative regression model of blood glucose component in blood samples

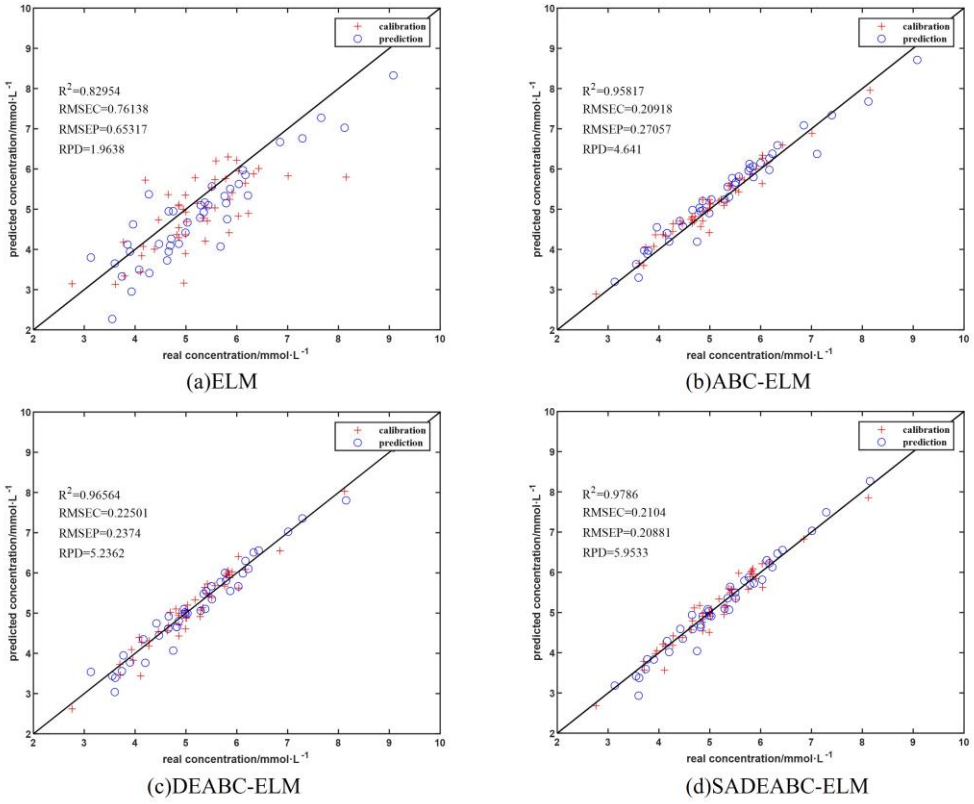

Fig. 4 Concentration prediction by quantitative regression model of cholesterol component in blood samples

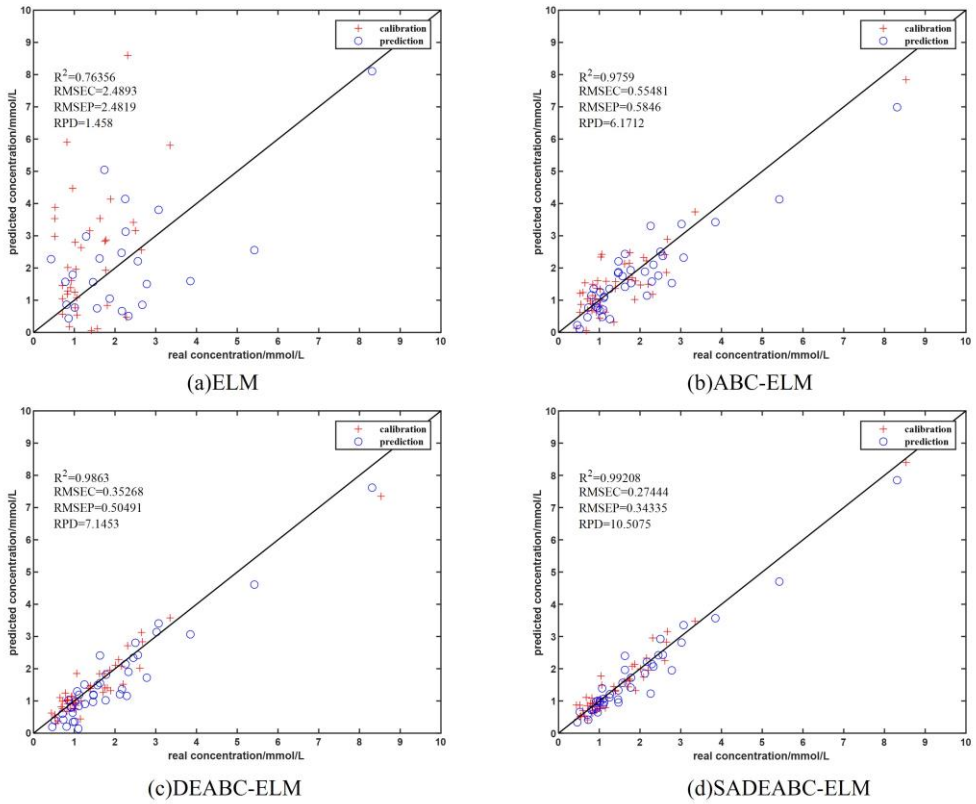

Fig. 5 Concentration prediction by quantitative regression model of triglyceride component in blood samples

The obtained parameters such as R^2, RMSEC, RMSEP and RPD for all the models and the evaluation indicators of ELM, ABC-ELM, DEABC-ELM and SADEABC-ELM are listed in **Table 2**. From the analysis results listed in **Table 2**, the SADEABC-ELM model shows the best performance compared with ELM, ABC-ELM and DEABC-ELM models. The correlation between the actual and predicted concentrations of the three blood components obtained from the ELM model shows the lowest value, and the model prediction is poor with a low accuracy. Both the ABC-ELM model and the DEABC-ELM model show improved prediction performance, but the the correlation between the actual and predicted concentrations of the three blood components using the SADEABC-ELM model obtained the highest value (all above 0.97). In the prediction of blood glucose concentration, the R^2 of the SADEABC-ELM model is 0.9822, which is an increase of 0.1770 compared with the ELM model, 0.0290 compared with the ABC-ELM model, and 0.0181 compared with the DEABC-ELM model. In the prediction of cholesterol concentration, the R^2 of the SADEABC-ELM model is 0.9786, which is an improvement of 0.1491 compared with the ELM model,

0.0204 compared with the ABC-ELM model, and 0.0130 compared with the DEABC-ELM model. In the prediction of triglyceride concentration, the R^2 of the SADEABC-ELM model is 0.9921, which is an improvement of 0.2285 compared with the ELM model, 0.0162 compared with the ABC-ELM model, and 0.0058 compared with the DEABC-ELM model.

Based on the above modeling analysis and data statistics of blood sample components, the performance indicators using the SADEABC-ELM model are more prominent in predicting the concentration of each component than the other ones, with a better strong prediction accuracy and model robustness. However, there is still potential for improvement in our proposed model's prediction accuracy. In follow-up studies, we are thinking about building many modules and integrating them to further increase the generalizability of the model using the concept of integration strategy in integrated learning.

Table 2 Evaluation indexes of blood sample components

Component	Regression method	R^2	RMSEC	RMSEP	RPD
Blood glucose	ELM	0.8052	1.6536	1.3728	1.8390
	ABC-ELM	0.9532	0.5059	0.5808	4.3936
	DEABC-ELM	0.9641	0.3998	0.4989	5.1154
	SADEABC-ELM	**0.9822**	**0.3993**	**0.3827**	**6.6679**
Cholesterol	ELM	0.8295	0.7614	0.6532	1.9638
	ABC-ELM	0.9582	0.2092	0.2706	4.6410
	DEABC-ELM	0.9656	0.2250	0.2374	5.2362
	SADEABC-ELM	**0.9786**	**0.2104**	**0.2088**	**5.9533**
Triglyceride	ELM	0.7636	2.4893	2.4819	1.4580
	ABC-ELM	0.9759	0.5548	0.5846	6.1712
	DEABC-ELM	0.9863	0.3527	0.5049	7.1453
	SADEABC-ELM	**0.9921**	**0.2744**	**0.3433**	**10.5075**

4. Conclusion

In this study, we used SADEAB algorithm to optimize the parameters in ELM and used it to predict the concentration of blood glucose, cholesterol, and triglyceride in human blood samples. According to the experimental results, the SADEABC-ELM model could accurately predict blood components' concentrations and showed a much higher prediction accuracy and robustness than those of ELM, ABC-ELM, and DEABC-ELM models. Compared with clinical blood tests, the Raman spectroscopy can predict the concentrations of blood glucose, cholesterol and triglyceride simultaneously with less blood samples. The technology combining the SADEABC-ELM model with Raman spectroscopy can achieve rapid, accurate, and no requirement for chemical reagents. Due to these advantages, the SADEABC-ELM model with Raman spectroscopy will be a promising method for quantitative analysis of blood components.

CRediT authorship contribution statement

Qiaoyun Wang: Funding acquisition, Investigation, Writing - review & editing, Resources, Software, Supervision, Project administration. Shuai Song: Validation, Writing - review & editing. Lei Li: Software, Validation. Da Wen: Validation. Peng Shan: Data curation, Formal analysis, Funding acquisition. Zhigang Li: Investigation, Methodology, Yongqing Fu: Funding acquisition, Investigation, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NFSC 11404054, 61601104), the Natural Science Foundation of Hebei Province (F2019501025, F2020501040), the Fundamental Research Funds for the Northeastern Universities (N2023006), and International Exchange Grant (IEC/NSFC/201078) through Royal Society and NFSC.

References

[1] D. Zhao, J. Liu, M. Wang, X. Zhang, M. Zhou, Epidemiology of cardiovascular disease in China: current features and implications, Nat Rev Cardiol, 16 (2019) 203-212.
[2] M.H. Sani, S. Khosroabadi, A novel design and analysis of high-sensitivity biosensor based on nano-cavity for detection of blood component, diabetes, cancer and glucose concentration, IEEE Sensors Journal, 20 (2020) 7161-7168.
[3] H. Ali, F. Bensaali, F. Jaber, Novel approach to non-invasive blood glucose monitoring based on transmittance and refraction of visible laser light, IEEE access, 5 (2017) 9163-9174.
[4] J. Yadav, A. Rani, V. Singh, B.M. Murari, Prospects and limitations of non-invasive blood glucose monitoring using near-infrared spectroscopy, Biomedical Signal Processing and Control, 18 (2015) 214-227.
[5] W. Villena Gonzales, A.T. Mobashsher, A. Abbosh, The progress of glucose monitoring-A review of invasive to minimally and non-invasive techniques, devices and sensors, Sensors, 19 (2019) 800.
[6] N.M. Ralbovsky, G.S. Fitzgerald, E.C. McNay, I.K. Lednev, Towards development of a novel screening method for identifying Alzheimer's disease risk: Raman spectroscopy of blood serum and machine learning, Spectrochim Acta A Mol Biomol Spectrosc, 254 (2021) 119603.
[7] Q. Wang, G. Wu, F. Pian, P. Shan, Z. Li, Z. Ma, Simultaneous detection of glucose, triglycerides, and total cholesterol in whole blood by Fourier-Transform Raman spectroscopy, Spectrochim Acta A Mol Biomol Spectrosc, 260 (2021) 119906.
[8] Y. Du, H. Huang, Y. Peng, J. Wang, Z. Gao, Rapid determination of Staphylococcus aureus enterotoxin B in milk using Raman spectroscopy and chemometric methods, J. Raman Spectrosc., 53 (2022) 709-714.
[9] A. Kaminska, M. Roman, A. Wrobel, A. Gala-Bladzinska, M.T.P. Malecki, C. Paluszkiewicz, E.L.P. Stepien, Raman spectroscopy of urinary extracellular vesicles to stratify patients with chronic kidney disease in type 2 diabetes, Nanomedicine, 39 (2022) 102468.
[10] M.A. Bakkar, H. Nawaz, M.I. Majeed, A. Naseem, A. Ditta, N. Rashid, S. Ali, J. Bajwa, S. Bashir, S. Ahmad, H. Hyat, K.S. Bukhari, F. Bonnier, Raman spectroscopy for the qualitative and quantitative analysis of solid dosage forms of Sitagliptin, Spectrochim Acta A Mol Biomol Spectrosc, 245 (2021) 118900.
[11] C. Robert, S.J. Fraser-Miller, W.T. Jessep, W.E. Bain, T.M. Hicks, J.F. Ward, C.R. Craigie, M. Loeffen, K.C. Gordon, Rapid discrimination of intact beef, venison and lamb meat using Raman spectroscopy, Food Chem, 343 (2021) 128441.
[12] H. Zhang, C. Chen, C. Ma, C. Chen, Z. Zhu, B. Yang, F. Chen, D. Jia, Y. Li, X. Lv, Feature Fusion Combined with Raman Spectroscopy for Early Diagnosis of Cervical Cancer, IEEE Photonics Journal, 13 (2021) 1-11.
[13] M. Hamed Mozaffari, L.L. Tay, Overfitting One-Dimensional convolutional neural networks for Raman spectra identification, Spectrochim Acta A Mol Biomol Spectrosc, 272 (2022) 120961.

[14] X. Wu, B. Xu, R. Ma, Y. Niu, S. Gao, H. Liu, Y. Zhang, Identification and quantification of adulterated honey by Raman spectroscopy combined with convolutional neural network and chemometrics, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 274 (2022) 121133.
[15] F. Pian, Q. Wang, M. Wang, P. Shan, Z. Li, Z. Ma, A shallow convolutional neural network with elastic nets for blood glucose quantitative analysis using Raman spectroscopy, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 264 (2022) 120229.
[16] Q. Wang, F. Pian, M. Wang, S. Song, Z. Li, P. Shan, Z. Ma, Quantitative analysis of Raman spectra for glucose concentration in human blood using Gramian angular field and convolutional neural network, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 275 (2022) 121189.
[17] J. Wang, S. Lu, S.-H. Wang, Y.-D. Zhang, A review on extreme learning machine, Multimedia Tools and Applications, 81 (2022) 41611-41660.
[18] J. Sun, Y. Fu, S. Li, J. He, C. Xu, L. Tan, Sequential Human Activity Recognition Based on Deep Convolutional Network and Extreme Learning Machine Using Wearable Sensors, Journal of Sensors, 2018 (2018) 1-10.
[19] A. Pashaei, M. Ghatee, H. Sajedi, Convolution neural network joint with mixture of extreme learning machines for feature extraction and classification of accident images, Journal of Real-Time Image Processing, 17 (2019) 1051-1066.
[20] M. Rubiolo, D.H. Milone, G. Stegmayer, Extreme learning machines for reverse engineering of gene regulatory networks from expression time series, Bioinformatics, 34 (2018) 1253-1260.
[21] L.R. Ren, Y.L. Gao, J.X. Liu, J. Shang, C.H. Zheng, Correntropy induced loss based sparse robust graph regularized extreme learning machine for cancer classification, BMC Bioinformatics, 21 (2020) 445.
[22] Z.M. Yaseen, H. Faris, N. Al-Ansari, Hybridized Extreme Learning Machine Model with Salp Swarm Algorithm: A Novel Predictive Model for Hydrological Application, Complexity, 2020 (2020) 1-14.
[23] T. Hai, A. Sharafati, A. Mohammed, S.Q. Salih, R.C. Deo, N. Al-Ansari, Z.M. Yaseen, Global Solar Radiation Estimation and Climatic Variability Analysis Using Extreme Learning Machine Based Predictive Model, IEEE Access, 8 (2020) 12026-12042.
[24] H. Chen, C. Tan, Z. Lin, Ensemble of extreme learning machines for multivariate calibration of near-infrared spectroscopy, Spectrochim Acta A Mol Biomol Spectrosc, 229 (2020) 117982.
[25] C. Tan, H. Chen, Z. Lin, Brand classification of detergent powder using near-infrared spectroscopy and extreme learning machines, Microchem. J., 160 (2021) 105691.
[26] Y. Chu, T. Chen, F. Chen, Y. Tang, S. Tang, H. Jin, L. Guo, Y.f. Lu, X. Zeng, Discrimination of nasopharyngeal carcinoma serum using laser-induced breakdown spectroscopy combined with an extreme learning machine and random forest method, Journal of Analytical Atomic Spectrometry, 33 (2018) 2083-2088.
[27] H. Wang, X. Chu, P. Chen, J. Li, D. Liu, Y. Xu, Partial least squares regression

residual extreme learning machine (PLSRR-ELM) calibration algorithm applied in fast determination of gasoline octane number with near-infrared spectroscopy, Fuel, 309 (2022) 122224.
[28] G. Huang, An insight into extreme learning machines: random neurons, random features and kernels, Cognitive Computation, 6 (2014) 376-390.
[29] Y. Wang, R. Li, Y. Chen, Accurate elemental analysis of alloy samples with high repetition rate laser-ablation spark-induced breakdown spectroscopy coupled with particle swarm optimization-extreme learning machine, Spectrochimica Acta Part B: Atomic Spectroscopy, 177 (2021) 106077.
[30] H. Yu, L.J. Ming, R. Sumei, Z. Shuping, A hybrid model for financial time series forecasting—integration of EWT, ARIMA with the improved ABC optimized ELM, IEEE Access, 8 (2020) 84501-84518.
[31] J. Tang, G. Liu, Q. Pan, A review on representative swarm intelligence algorithms for solving optimization problems: Applications and trends, IEEE/CAA Journal of Automatica Sinica, 8 (2021) 1627-1643.
[32] M. Pant, H. Zaheer, L. Garcia-Hernandez, A. Abraham, Differential Evolution: A review of more than two decades of research, Eng Appl Artif Intel, 90 (2020) 103479.
[33] W. Gao, L. Huang, J. Wang, S. Liu, C. Qin, Enhanced artificial bee colony algorithm through differential evolution, Applied soft computing, 48 (2016) 137-150.
[34] H. Wang, W. Wang, S. Xiao, Z. Cui, M. Xu, X. Zhou, Improving artificial bee colony algorithm using a new neighborhood selection mechanism, Information Sciences, 527 (2020) 227-240.
[35] W. Deng, S. Shang, X. Cai, H. Zhao, Y. Song, J. Xu, An improved differential evolution algorithm and its application in optimization problem, Soft Computing, 25 (2021) 5277-5298.
[36] X. Zhang, J. Sun, P. Li, F. Zeng, H. Wang, Hyperspectral detection of salted sea cucumber adulteration using different spectral preprocessing techniques and SVM method, LWT, 152 (2021) 112295.
[37] Z. Tang, S. Wang, X. Chai, S. Cao, T. Ouyang, Y. Li, Auto-encoder-extreme learning machine model for boiler NOx emission concentration prediction, Energy, 256 (2022) 124552.