High-precision prediction of blood glucose concentration utilizing Fourier transform Raman spectroscopy and an ensemble machine learning algorithm

Shuai Song[1#], Qiaoyun Wang[1,2*#], Xin Zou[1], Zhigang Li[1], Zhenhe Ma[1], Daying Jiang[3], YongQing Fu[4], Qiang Liu[1,2]

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. Zhongyou BSS (Qinhuangdao) Petropipe Company limited, Qinhuangdao, 066004, China

4. Faculty of Engineering and Environment, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK

*** Corresponding author, *E-mail address:*** wangqiaoyun@neuq.edu.cn (Q.Y. Wang).

Shuai Song and Qiaoyun Wang contributes equally to this work.

Abstract

Raman spectroscopy has gained popularity in analyzing blood glucose levels due to its non-invasive identification and minimal interference from water. However, the challenge lies in how to accurately predict blood glucose concentrations in human blood using Raman spectroscopy. This paper researches a novel integrated machine learning algorithm called Bagging-ABC-ELM. The optimal input weights and biases of extreme learning machine (ELM) model are obtained by artificial bee colony (ABC) algorithm. The bagging algorithm is used to obtain a better the stability of the model and higher performance than ELM algorithm. The results show that the mean value of coefficient of determination is 0.9928, and root mean square error is 0.1928. Compared to other regression models, the Bagging-ABC-ELM model exhibited superior prediction accuracy, robustness, and generalization capability. The Bagging-ABC-ELM model presents a promising alternative for analyzing blood glucose levels in clinical and research settings.

Keywords: Raman spectroscopy; Blood glucose; Bagging algorithm; Artificial Bee Colony algorithm; Extreme Learning Machine

1. Introduction

In the forthcoming decades, there will be a substantial escalation in the global population of individuals afflicted by diabetes[1]. Diabetes is a metabolic disease that manifests as high blood glucose levels, leading to the degeneration of vital tissues and organ systems, such as the eyes, kidneys, and heart[2, 3]. Monitoring the concentration level of blood glucose in clinical treatment is one of the most effective ways to delay or prevent complications diseases of diabetes. The traditional finger-pick testing is invasive, and painful, and this method will increase the risk of infection[4]. So, there is a highly demand of blood glucose monitoring method which is non-invasive, sensitive, robust and continuous[5].

Optical techniques have emerged as the preferred approach for non-invasive and accurate detection[6-9]. Raman spectroscopy[10-12] has gained significant attention for its robust qualitative measurement capability and high resistance to water interference, making it an effective non-invasive technique for blood glucose concentration detection[13-15]. Several studies have proposed Raman spectroscopy as a non-invasive method for detecting blood glucose concentration[6, 16]. Due to the complexity of the raw spectrum obtained through Raman spectroscopy detection, chemometric approaches are required to obtain the most relevant dateset from complex Raman spectral data.

Chemometric approaches can effectively utilize the information in Raman spectroscopy to determine target component concentrations, resulting in improved prediction accuracy and model robustness[17]. Machine learning algorithms have garnered extensive utilization in the quantitative analysis of Raman spectroscopy. The main models used in Raman spectroscopy quantitative analysis are convolutional neural network (CNN)[18, 19], support vector machine (SVM)[20], extreme learning machine (ELM)[21, 22], and partial least squares regression (PLSR)[23]. The learning process of ELM depends on the randomization of input weights and biases to achieve its objectives. It has a faster training speed compared to traditional methods[24]. ELM has found wide applications in diverse fields, including fault diagnosis[25], medical

diagnosis[26], image processing[27], and pattern classification[28]. The ELM model is an excellent choice for spectroscopy analysis because of good convergence, strong generalization capability, and simple structure[24]. For example, Wang et al. investigated the octane number prediction of three different ELM models, and the results revealed that the accuracy and stability of OS-ELM and SaDE-ELM are higher than the traditional ELM model[29]. Xiao et al. developed a coal recognition model that combines CNN with ELM，and the model demonstrated remarkable performance with an accuracy rate of 97%, enabling rapid and accurate identification of different types of coal[30]. Wang et al. proposed an MC-LASSO-ELM model to predict the hemoglobin concentration in blood with infrared spectroscopy technology, and the results demonstrated the model's superior performance[31]. Zhang et al. combined Raman spectroscopy with the ELM model for cheese product identification, with an average identification accuracy rate of 98%, providing a new research method for identifying similar samples[32].

This study introduced a novel model, named "Bagging-ABC-ELM", that utilized Raman spectroscopy for predicting glucose concentration in whole blood. ELM was chosen as the base model and optimized using the ABC algorithm to enhance its accuracy by optimizing the input weights and biases. The Bagging algorithm was introduced to improve the stability of the ultimate model. Compared with PLSR, SVR, ELM, Bagging-ELM, and ABC-ELM models, the Bagging-ABC-ELM model has a better performance. The mean values of R^2 and RMSEP for the proposed model were 0.9928 and 0.1928, respectively, indicating that the combination of Raman spectroscopy with the Bagging-ABC-ELM model holds great promise as an effective approach for quantitative analysis of blood glucose concentration.

2. Materials and Methods

2.1. Experiments

For this research, a comprehensive set of 106 blood samples was obtained in the First Hospital of Qinhuangdao, China. The sample group consisted of 60 individuals

who were deemed healthy volunteers and 46 individuals diagnosed with diabetes based on clinical evaluation. All the volunteers and patients in the study were between the ages of 29 and 87, comprising 51 males and 55 females. The blood glucose concentrations of the samples ranged from 4.12 mmol/L to 16.32 mmol/L. The collected samples were then separated into two groups. The first group was used for the clinical standard testing of blood glucose concentration, and the clinical standard method used at the hospital was employed to obtain the blood glucose concentration values which were subsequently used as the standard values for the study. The second group of samples was scanned using Raman spectroscopy to obtain the whole blood Raman spectrum that were used for training the Bagging-ABC-ELM prediction model. The specific experimental steps are depicted in **Fig. 1**.

The output power of 1064 nm excitation radiation is 90 mW in the experiment. Germanium detector is cooled by liquid nitrogen. The spectrum range is from 400 cm^{-1} to 4000 cm^{-1}. Each sample is scanned at a spectral resolution of 6 cm^{-1} and the times of scanning is 64. In order to improve the accuracy, three measurements were performed on every sample, and the ultimate spectrum was derived by averaging the three measurements. The OPUS 7.0 software was used for data acquisition.

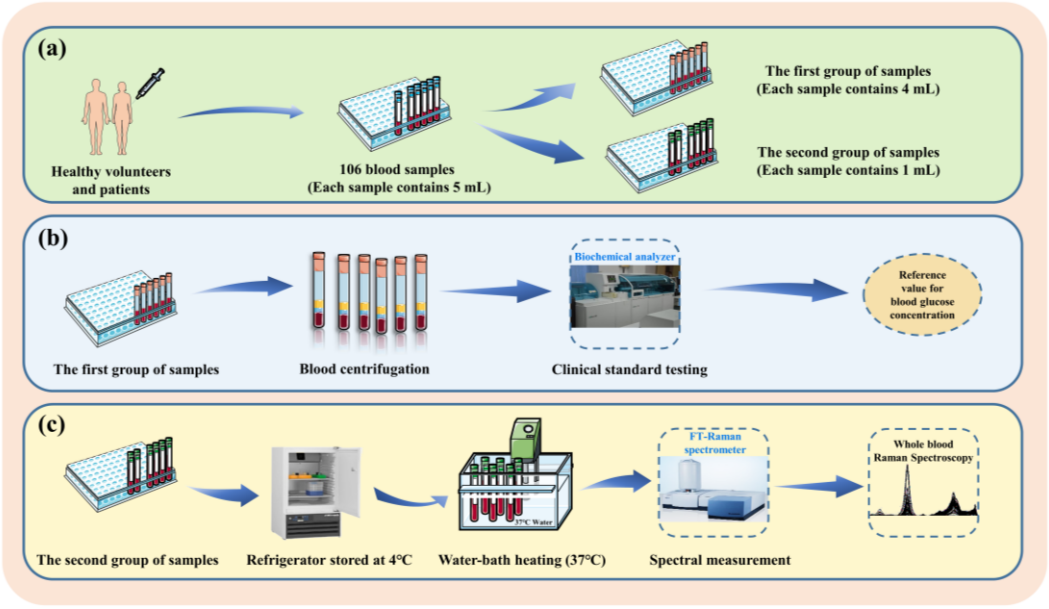

Fig. 1. The experimental process of this study.

2.2. Principle of ELM

The learning process of the ELM model depends on the randomization of input weights and biases[33]. It presents advantages over conventional methods such as BP and RBF networks in terms of training speed and generalization capability[34]. There are three layers in structure of ELM model, as depicted in **Fig. 2**. $\left\{\left(\boldsymbol{x}_j, \boldsymbol{t}_j\right), j=1,\cdots,N\right\}$ are the input samples, where N is the samples number, $\boldsymbol{x}_j=\left[x_{j1}, x_{j2},\cdots, x_{jn}\right]^T \in R^n$ is the input vectors of the jth sample and $\boldsymbol{t}_j=\left[t_{j1}, t_{j2},\cdots, t_{jm}\right]^T \in R^m$ is expected output vectors of the jth sample, respectively. The $\boldsymbol{t}_j$ can be obtained by:

$$\boldsymbol{t}_j=\sum_{i=1}^{l}\beta_i g\left(\boldsymbol{w}_i \cdot \boldsymbol{x}_j+b_i\right), j=1,\cdots,N \tag{1}$$

where $\boldsymbol{\beta}_i=\left[\beta_{i1}, \beta_{i2},\cdots, \beta_{im}\right]^T$ is the weight matrix for the hidden-to-output layer, $\boldsymbol{w}_i=\left[w_{i1}, w_{i2},\cdots, w_{in}\right]^T$ is the weight vector for the input-to-hidden- layer, the activation function in our experiments is $g\left(\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{b}\right)$, and b_i is used to represent the biases of the ith hidden node. Equation (1) can also be expressed as:

$$\boldsymbol{H}\boldsymbol{\beta}=\boldsymbol{T} \tag{2}$$

where $\boldsymbol{H}$ is the output matrix of the hidden layer, $\boldsymbol{H}$, $\boldsymbol{\beta}$ and $\boldsymbol{T}$ are represented as follows:

$$\boldsymbol{H}=\begin{bmatrix} g\left(\boldsymbol{w}_1 \cdot \boldsymbol{x}_1+b_1\right) & \cdots & g\left(\boldsymbol{w}_l \cdot \boldsymbol{x}_1+b_l\right) \\ \vdots & & \vdots \\ g\left(\boldsymbol{w}_1 \cdot \boldsymbol{x}_N+b_1\right) & \cdots & g\left(\boldsymbol{w}_l \cdot \boldsymbol{x}_N+b_l\right) \end{bmatrix}_{N\times l} \tag{3}$$

$$\boldsymbol{\beta}=\left[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2,\cdots, \boldsymbol{\beta}_l\right]^T \tag{4}$$

$$\boldsymbol{T}=\left[\boldsymbol{t}_1, \boldsymbol{t}_2,\cdots, \boldsymbol{t}_N\right]^T \tag{5}$$

Equation (6) yields the least squares solution for $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}}=\boldsymbol{H}^{+}\boldsymbol{T} \tag{6}$$

where $\boldsymbol{H}^{+}$ is the inverse matrix of $\boldsymbol{H}$.

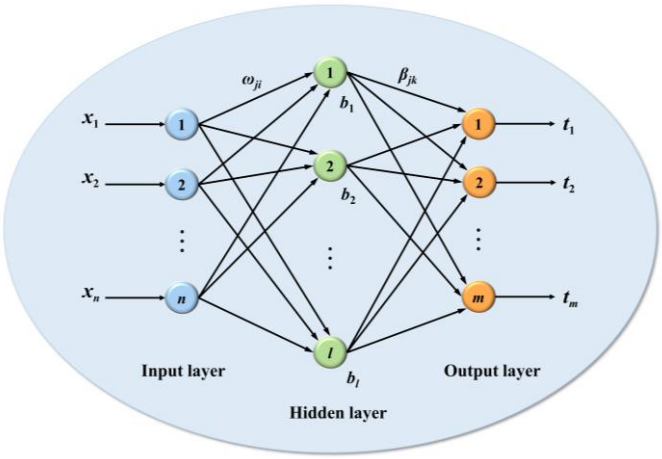

Fig. 2. The schematic diagram of the ELM network architecture.

2.3. Bagging-ABC-ELM algorithm

The complete prediction process is depicted in **Fig. 3**, which comprises the data pretreatment and prediction model building phases. Firstly, to achieve a more accurate model, operations such as preprocessing, dimensionality reduction and dataset partitioning of the whole blood Raman spectroscopy are required. Then, the Bagging-ABC-ELM model is proposed by combining ELM, ABC and bagging algorithms, and the final model is utilized to forecast the blood glucose concentration in human blood.

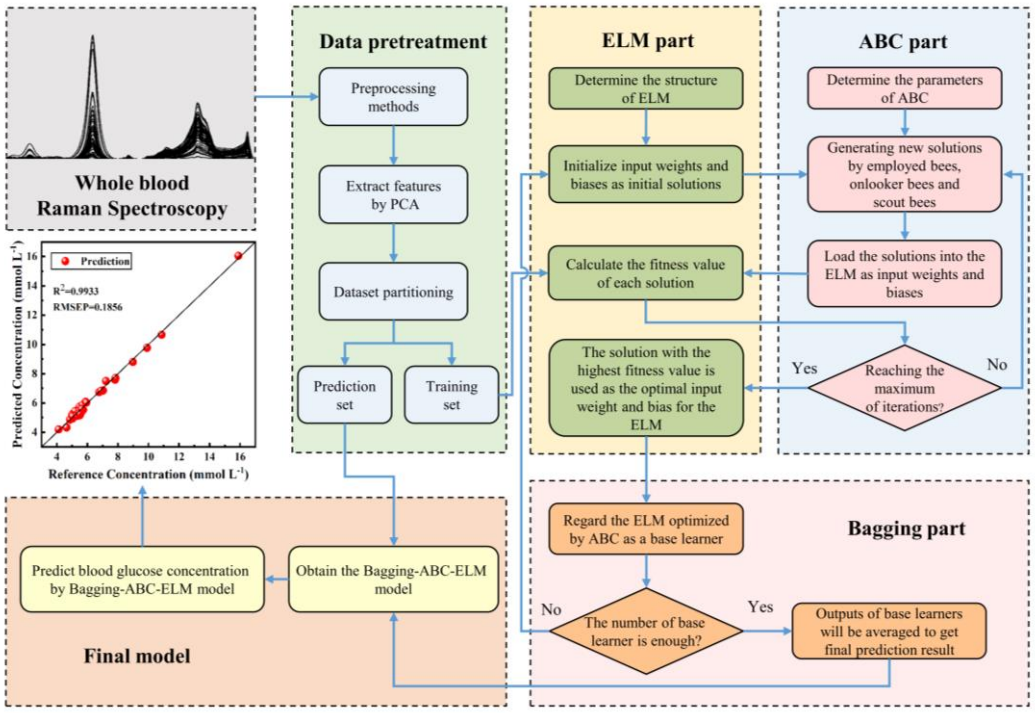

Fig. 3. The schematic of the prediction process.

The input weights and biases of the ELM algorithm are generated randomly, which will influent the performance of ELM model. Therefore, selecting a suitable optimization algorithm to optimize these parameters is crucial to achieving better results. Among various optimization algorithms, the ABC algorithm is preferred due to its simplicity, fewer control parameters, and efficacy in addressing multi-dimensional optimization problems[35, 36]. Hence, we adopted the ABC algorithm to optimize the ELM model.

In the ABC algorithm, the colony is divided into three groups: employed bees, onlooker bees, and scout bees[37]. During the optimization iteration, the three groups of bee exchange information with each other and cooperate to complete the optimization process. The ABC algorithm consists of the following stages:

(1) Initialization. Randomly generate *SN* nectar sources:

$$\boldsymbol{X}_i^j = \boldsymbol{X}_{\min}^j + rand(0,1) * (\boldsymbol{X}_{\max}^j - \boldsymbol{X}_{\min}^j) \tag{7}$$

where $i = 1, 2, \cdots, SN$, $j = 1, 2, \cdots, D$, $\boldsymbol{X}_i$ denotes the ith nectar source, while $\boldsymbol{X}_{\max}^j$ and $\boldsymbol{X}_{\min}^j$ are the upper and the lower bounds of the search space, respectively.

(2) Employed bee phase. The employed bees update iteratively near the initial nectar sources according to Equation (8) to obtain new sources:

$$v_{ij} = x_{ij} + \varphi_{ij} * (x_{ij} - x_{kj}) \tag{8}$$

where $i, k \in \{1, 2, \cdots, SN\}$ and $k \neq i$, $j \in \{1, 2, \cdots, D\}$, and $\varphi_{ij} \in [-1, 1]$, respectively. Then, the locally optimal source is retained through greedy selection. The greedy algorithm is shown below:

$$\boldsymbol{X}_i = \begin{cases} \boldsymbol{V}_i, & \text{if } \boldsymbol{V}_i \text{ outperforms } \boldsymbol{X}_i \\ \boldsymbol{X}_i, & \text{otherwise} \end{cases} \tag{9}$$

(3) Onlooker bee phase. Onlooker bees select nectar sources through a roulette wheel selection method, with Equation (10) as the roulette wheel selection formula. Then onlooker bees search near the selected nectar sources through Equation (8), and the locally optimal sources are retained by Equation (9) after each search.

$$p_i = fit_i / \sum_{k=1}^{SN} fit_k \tag{10}$$

$$fit_i = \begin{cases} 1/(1+f_i), & if \quad f_i \geq 0 \\ 1+abs(f_i), & if \quad f_i < 0 \end{cases} \tag{11}$$

where f_i is the objective function of $\boldsymbol{X}_i$, and fit_i is the fitness value of $\boldsymbol{X}_i$.

(4) Scout bee phase. Once an employed bee reaches a certain threshold of iterations without enhancing the fitness of a source, it transitions into a scout bee and abandons the current source. The scout bee then uses Equation (7) to find a new source.

Although the ABC-ELM model demonstrates superior predictive accuracy than the ELM model, its stability requires further enhancement. Compared to a single model, integration model can achieve more accurate and robust prediction[38]. Among various ensemble algorithms, bagging algorithm is widely adopted due to its strong interpretability and high performance[39]. Unlike boosting algorithm, bagging algorithm effectively enhances the anti-noise performance and avoids overfitting of the model[39, 40]. In this research, the bagging algorithm was introduced to enhance the model's generalization and robustness and to prevent overfitting. The Bagging-ABC-ELM model is constructed through the following steps:

Step (1): Utilize the ABC-ELM model as the base learner and employ the bootstrap approach to select *m* samples from the training samples, which formulates a subset.

Step (2): Repeat step (1) until the number of training subsets reaches *T*, where *T* is the number of base learners.

Step (3): The individual base learners are trained using their respective training subsets.

Step (4): When given a prediction set, each base learner makes an independent prediction, then the final prediction of the Bagging-ABC-ELM model is the average of all these individual predictions.

2.4. Model reliability

The performance of regression models is commonly used the coefficient of determination (R^2) and the root mean square error (RMSE). Following is a definition

of R^2 and RMSEP:

$$R^2 = 1 - \sum_{i=1}^{n} (y_{i,p} - y_{i,p})^2 / \sum_{i=1}^{n} (\bar{y}_p - y_{i,p})^2 \quad (12)$$

$$RMSEP = \sqrt{\sum_{i=1}^{n} (y_{i,p} - y_{i,p})^2 / n} \quad (13)$$

where *n* represents the of the prediction datasets, $y_{i,p}$ represents reference value of the ith sample in the prediction datasets, $y_{i,p}$ represents predicted value of the ith prediction sample and $\bar{y}_p$ represents average value of $y_{i,p}$. Typically, a model with an R^2 value closer to 1 and a smaller RMSEP value is considered to have better predictive performance[41].

3. Results and discussions

3.1. Spectral data treatment

Fig. 4 displays several original Raman spectra obtained from blood samples. Spectral data generated by the spectrometer may produce unwanted changes due to variations of factors such as sample condition, measurement mode, and environmental condition[42]. Therefore, in order to more accurately predict blood glucose concentrations, the original Raman spectral data needs to undergo preprocessing. In this study, the preprocessing operations we used were baseline correction and Savitzky-Golay denoising.

Appropriate feature selection algorithm will reduce training time, remove redundant components, and improve prediction performance of the model[43]. The original spectral data comprised 933 variables, with the majority of them being low intensity and uninformative. Therefore, in this research, we employed principal component analysis to extract key information from the spectral data, thereby simplifying the model input. We set the number of principal components to 24.

The dataset comprising 106 blood samples was partitioned into 42 training sets, 32 calibration sets, and 32 prediction sets employing a randomized partitioning

approach. The training sets were utilized for the establishment of the Bagging-ABC-ELM model, while the calibration sets played a crucial role in the selection of optimal hyper-parameters for the model. Subsequently, the prediction sets were employed to input data into the final Bagging-ABC-ELM model, generating predicted values.

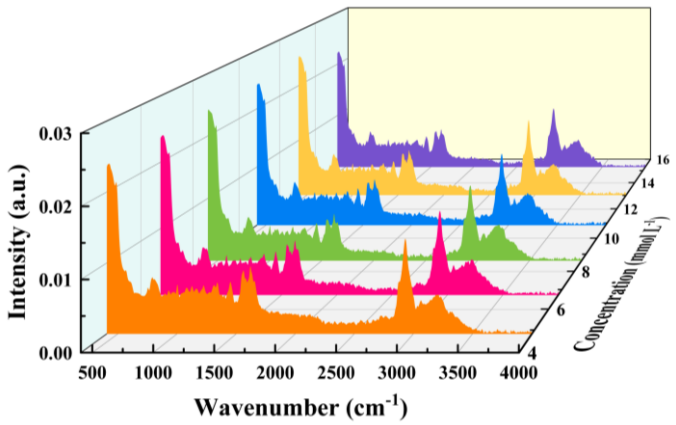

Fig. 4. The Raman spectroscopy data was obtained from blood samples with varying levels of glucose concentration.

3.2. Selection of regression model parameters

The characteristics of ELM model is significantly influenced by the hidden nodes (l) and the activation function. Bian et al. used MSR model for the parameter selection issue in ELM[44]. A higher MSR value indicates superior model performance. In this study, we selected sin, tanh, and sigmoid as activation functions, while setting the parameter l within the range of 1 to 40. For each combination of these two parameters, the ELM model was run 300 times, and the MSR value was calculated. **Fig. 5** displays the MSR values variation with these two parameters. The MSR values for each activation function initially increase with l, and then decrease as l increases. The maximum MSR values were achieved with 24, 25, and 25 hidden nodes for sin, tanh, and sigmoid activation functions, respectively. The ELM model exhibited the highest MSR value when the activation function was sigmoid and l was 25, demonstrating superior performance of the model. Consequently, we opted to use the sigmoid activation function and 25 hidden nodes.

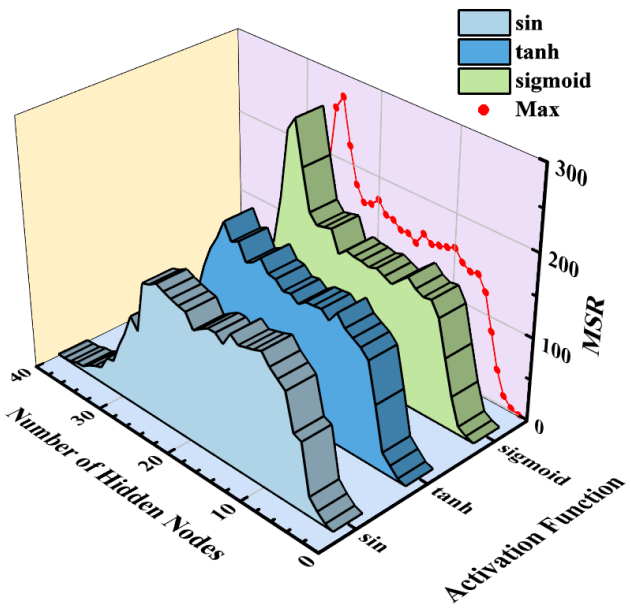

Fig. 5. MSR values variation with hidden nodes (l) and activation functions.

In the Bagging algorithm, the number of base learners (T) is an important parameter. When T is set too small, the model's performance may decline, while excessively large values of T can lead to longer runtimes. In this study, we set the selection range of T to be from 1 to 20 with an interval of 1. The repeat times of the Bagging-ABC-ELM model is 10 for each T, then calculate the RMSE mean and record the numerical fluctuations. The relationship between RMSE value and T is shown in **Fig. 6**. Firstly, the value of RMSE is decreased with the increased number of T, and then the fluctuation amplitude of RMSE value gradually narrows with the increased number of T. When T is greater than 15, the RMSE value is small and stable, the number of T is set 15 in our experiments.

The final Bagging-ABC-ELM model was established and used to predict blood glucose concentrations. **Table 1** summarizes the hyper-parameters used in this study.

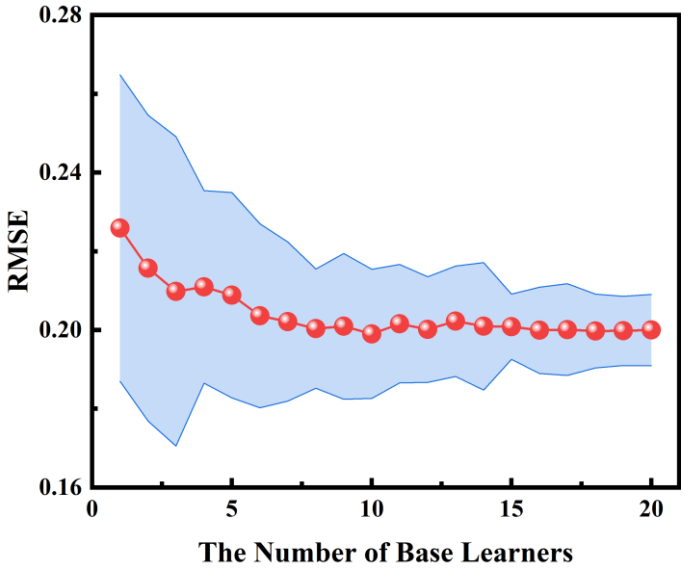

Fig. 6. Model performance with different *T*.

Table 1. The parameters of the Bagging-ABC-ELM model.

Parameters	Value/Type
Hidden nodes	25
Activation function	Sigmoid
Number of nectar sources	40
Iterations	100
Number of base learners	15
Training subset size	42

3.3. Establishment and evaluation of regression model

3.3.1. Performance of Bagging-ABC-ELM model

After conducting preprocessing, dimensionality reduction, and dataset partitioning on the experimental data, we established the Bagging-ABC-ELM model to predict the glucose concentrations in the blood. The model's effectiveness was assessed by comparing the predicted concentrations with the actual blood glucose concentrations, as shown in **Fig. 7**. A closer data distribution to the y=x line indicates higher prediction accuracy. The results displayed in **Fig. 7** demonstrate that the Bagging-ABC-ELM model performs exceptionally well, with the data distribution being concentrated along the y=x line. The R^2 and RMSEP values for the model are 0.9933 and 0.1856, respectively.

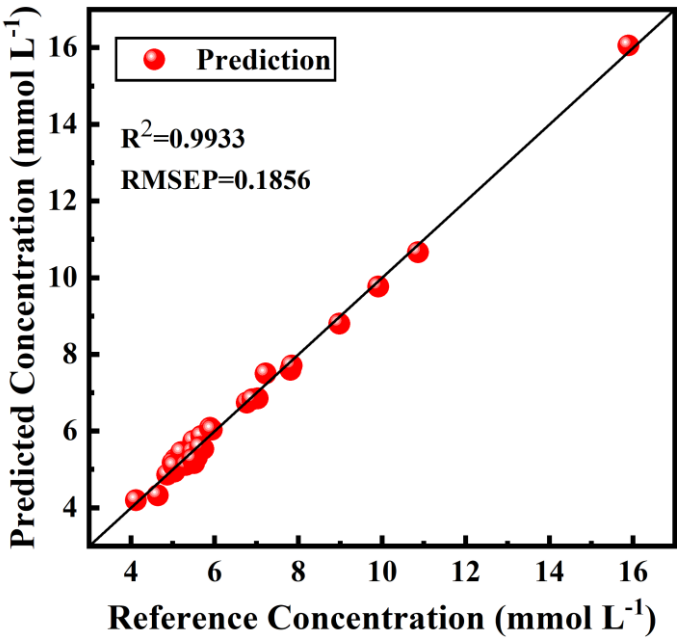

Fig. 7. Blood glucose concentration prediction by Bagging-ABC-ELM model.

The Clarke error grid is a commonly used tool for evaluating the clinical reliability of predicted blood glucose concentrations[45]. It is divided into five regions based on the severity of the clinical outcomes that could result. The predicted values in Region A are accurate and can be used to make correct clinical decisions. The predicted values in Region B have acceptable clinical errors, while the predicted values in Regions C, D, and E may lead to significant clinical mistakes. As can be seen from **Fig. 8**, the predicted blood glucose concentrations of the Bagging-ABC-ELM model established in this study are all distributed in Region A, demonstrating that the model accuracy is clinically acceptable.

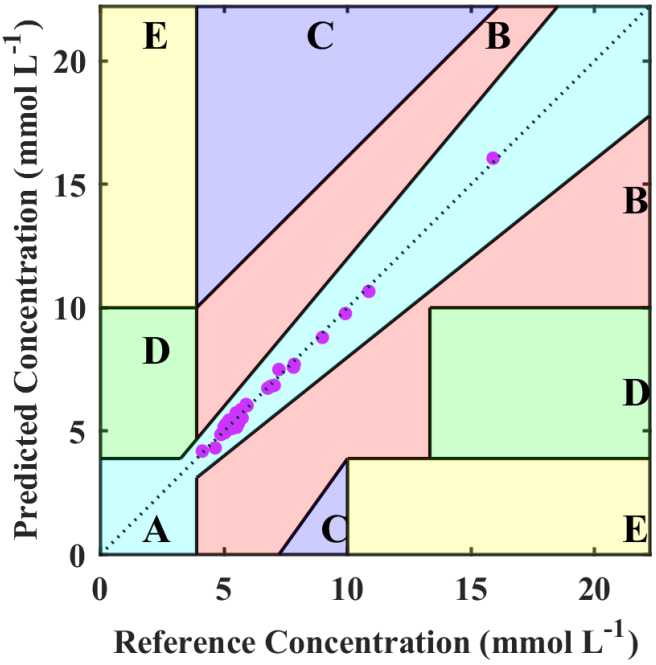

Fig. 8. Clarke error grid for blood glucose concentration prediction of Bagging-ABC-ELM model.

3.3.2. The comparisons of Bagging-ABC-ELM model and other ELM models

To demonstrate the superiority of the Bagging-ABC-ELM model, three other models (as shown in **Table 2**) were used as comparison models in this study. The blood glucose concentrations in blood samples were then predicted using the above four models. To reduce contingency, models were repeatedly run 30 times on the same dataset and the R^2 and RMSEP obtained each time are shown in **Fig. 9**. The means and standard deviations (SD) of the evaluation criterions for models are provided in **Table 2**. To facilitate a visual comparison of the performance of various models, boxplots of R^2 and RMSEP were generated and are presented in **Fig. 10**. In the boxplot, the evaluation criteria of the models are shown on the vertical axis, and each box represents a different method, with a smaller interquartile range (IQR) representing the higher stability of the model.

In **Fig. 10(a)**, the boxplot corresponding to the Bagging-ABC-ELM model is positioned at the highest point, indicating the highest R^2. The IQR of the Bagging-ABC-ELM model in the boxplot is 0.0004, while the IQRs of the other three models in the boxplots are 0.0058, 0.0014, and 0.0015, respectively. Similarly, in **Fig. 10(b)**, the

boxplot for the Bagging-ABC-ELM model is located at the lowest point, implying the lowest RMSEP. The IQR of the boxplot corresponding to the Bagging-ABC-ELM model is 0.0054, while the IQRs of the boxplots corresponding to the other three models are 0.0423, 0.0120, and 0.0199, respectively. Moreover, **Table 2** shows that the Bagging-ABC-ELM model has the highest R^2 mean and the lowest RMSEP mean, the results prove that the proposed Bagging-ABC-ELM model exhibits superior predictive accuracy compared to the other three models. And the Bagging-ABC-ELM model has the smallest SD of R^2 and RMSEP. It indicates that the proposed model is not only more accurate but also more stable than the other three models.

In summary, the Bagging-ABC-ELM model presented in this article has excellent performance in predicting blood glucose concentration in human blood. The results prove that this model outperforms ELM, Bagging-ELM, and ABC-ELM in terms of both prediction accuracy and stability.

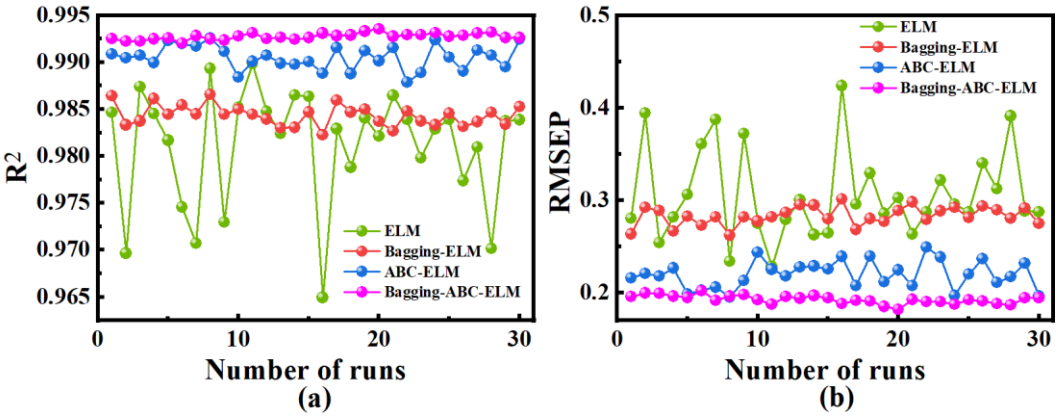

Fig. 9. Results obtained from Bagging-ABC-ELM models and the other three models after 30 runs. (a) R^2 and (b) RMSEP.

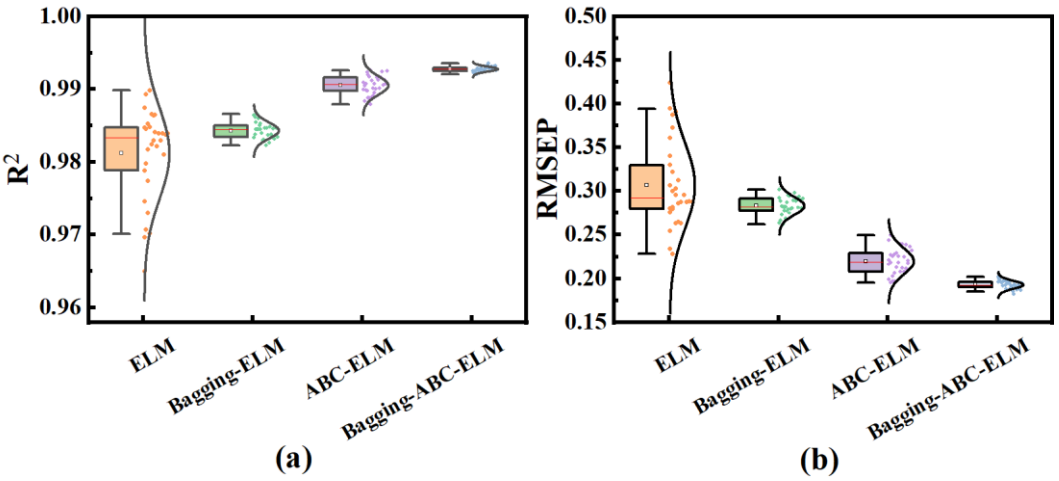

Fig. 10. Boxplots of results obtained from Bagging-ABC-ELM model and the other models after 30 runs. (a) R^2 and (b) RMSEP.

Table 2. The mean and standard deviation of R^2 and RMSEP obtained from each model after 30 runs.

Models	R^2		RMSEP	
	mean	SD	mean	SD
ELM	0.9812	0.0062	0.3065	0.0488
Bagging-ELM	0.9843	0.0011	0.2832	0.0101
ABC-ELM	0.9905	0.0013	0.2199	0.0148
Bagging-ABC-ELM	**0.9928**	**0.0003**	**0.1928**	**0.0046**

3.3.3. The comparisons of Bagging-ABC-ELM model and traditional models

To provide a comprehensive evaluation of the Bagging-ABC-ELM model, we also constructed PLSR and support vector regression (SVR) models as traditional methods for comparison in this study. PLSR is a widely used multivariate quantitative regression model in spectroscopy, while SVR is a method that uses SVM in regression problems and is an effective method for spectra regression. The three models were repeatedly run 30 times on the same dataset, and the R^2 and RMSEP obtained each time are presented in **Fig. 11**.

From **Fig. 11**, it shows that the there is no variation in PLSR model and SVR models results for each run. The R^2 and RMSEP of PLSR and SVR are 0.9775, 0.3397, and 0.9887, 0.2411, respectively. In contrast, the Bagging-ABC-ELM model outperformed the other two models in each run, with higher R^2 and lower RMSEP

values. Although the results of the Bagging-ABC-ELM model runs exhibit some fluctuation, the amplitude of the fluctuations is minimal, so it can be said that the model has high stability. The above analysis concludes that compared with conventional models, the Bagging-ABC-ELM model proposed in this paper performs better in predicting blood glucose concentration in human blood.

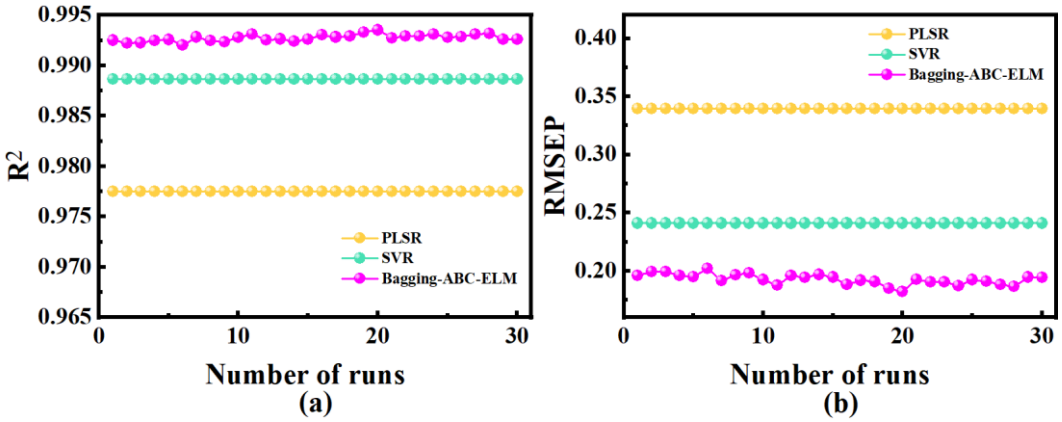

Fig. 11. Results obtained from PLSR, SVR, and Bagging-ABC-ELM models after 30 runs. (a) R^2 and (b) RMSEP.

4. Conclusions

This study proposes a novel predictive model called Bagging-ABC-ELM, which combines three distinct algorithms: bagging algorithm, ABC algorithm, and ELM algorithm. The model aims to predict blood glucose concentration in blood through the application of Raman spectroscopy technology. The ELM is employed as the fundamental framework for the model, and the optimization of the input weights and biases is obtained by the ABC algorithm. Subsequently, the bagging algorithm is incorporated into the model to improve its predictive accuracy and stability.

The experimental results demonstrate that the Bagging-ABC-ELM model has higher prediction accuracy, stability, and generalization ability compared to PLSR, SVR, ELM, Bagging-ELM, and ABC-ELM models. Specifically, the proposed model achieved a mean R^2 value of 0.9928, representing respective improvements of 0.0153, 0.0041, 0.0116, 0.0085, and 0.0023 compared to the aforementioned models. Additionally, the SD of R^2 for the proposed model was 0.0003, which demonstrated a

reduction of 95.16%, 72.73%, and 76.92% compared to the ELM, Bagging-ELM, and ABC-ELM models, respectively. Furthermore, the proposed model achieved a mean RMSEP of 0.1928, indicating respective reductions of 0.1469, 0.0483, 0.1137, 0.0904, and 0.0271 compared to the PLSR, SVR, ELM, Bagging-ELM, and ABC-ELM models, respectively. Additionally, the SD of RMSEP for the proposed model was 0.0046, demonstrating a reduction of 90.57%, 54.46%, and 68.92% compared to the ELM, Bagging-ELM, and ABC-ELM models, respectively.

In conclusion, this study proposes a novel approach to detect human blood glucose concentration, which integrates the Bagging-ABC-ELM model with Raman spectroscopy technology. The proposed approach holds significant potential for fast and precise determination of blood glucose levels, and provides new insights and ideas for the monitoring and management of diabetes and related diseases.

Acknowledgements

The work has been 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 Central Universities (2023GFZD002), and International Exchange Grant (IEC/NSFC/201078) through Royal Society and NFSC.

References:

[1] D. Li, N. Wu, Mechanism and application of exosomes in the wound healing process in diabetes mellitus, Diabetes Res Clin Pract 187 (2022) 109882. https://doi.org/10.1016/j.diabres.2022.109882.

[2] H. Wu, V. Norton, K. Cui, B. Zhu, S. Bhattacharjee, Y.W. Lu, B. Wang, D. Shan, S. Wong, Y. Dong, S.L. Chan, D. Cowan, J. Xu, D.R. Bielenberg, C. Zhou, H. Chen, Diabetes and Its Cardiovascular Complications: Comprehensive Network and Systematic Analyses, Front Cardiovasc Med 9 (2022) 841928. https://doi.org/10.3389/fcvm.2022.841928.

[3] J. Wei, J. Tian, C. Tang, X. Fang, R. Miao, H. Wu, X. Wang, X. Tong, The Influence of Different Types of Diabetes on Vascular Complications, J Diabetes Res 2022 (2022) 3448618. https://doi.org/10.1155/2022/3448618.

[4] L. Tang, S.J. Chang, C.J. Chen, J.T. Liu, Non-Invasive Blood Glucose Monitoring Technology: A Review, Sensors (Basel) 20(23) (2020). https://doi.org/10.3390/s20236925.

[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 (Basel) 19(4) (2019). https://doi.org/10.3390/s19040800.

[6] N. Li, H. Zang, H. Sun, X. Jiao, K. Wang, T.C. Liu, Y. Meng, A Noninvasive Accurate Measurement of Blood Glucose Levels with Raman Spectroscopy of Blood in Microvessels, Molecules 24(8) (2019). https://doi.org/10.3390/molecules24081500.

[7] K. Chen, N. Wang, M. Guo, X. Zhao, H. Qi, C. Li, G. Zhang, L. Xu, Detection of SF6 gas decomposition component H2S based on fiber-optic photoacoustic sensing, Sensors and Actuators B: Chemical 378 (2023). https://doi.org/10.1016/j.snb.2022.133174.

[8] S. Delbeck, T. Vahlsing, S. Leonhardt, G. Steiner, H.M. Heise, Non-invasive monitoring of blood glucose using optical methods for skin spectroscopy-opportunities and recent advances, Anal Bioanal Chem 411(1) (2019) 63-77. https://doi.org/10.1007/s00216-018-1395-x.

[9] Y. Ma, T. Liang, S. Qiao, X. Liu, Z. Lang, Highly Sensitive and Fast Hydrogen Detection Based on Light-Induced Thermoelastic Spectroscopy, Ultrafast Science 3 (2023). https://doi.org/10.34133/ultrafastscience.0024.

[10] J. Le, Y. Su, C. Tian, A.H. Kung, Y.R. Shen, A novel scheme for ultrashort terahertz pulse generation over a gapless wide spectral range: Raman-resonance-enhanced four-wave mixing, Light Sci Appl 12(1) (2023) 34. https://doi.org/10.1038/s41377-023-01071-z.

[11] Z. Zhang, T. Peng, X. Nie, G.S. Agarwal, M.O. Scully, Entangled photons enabled time-frequency-resolved coherent Raman spectroscopy and applications to electronic coherences at femtosecond scale, Light: Science & Applications 11(1) (2022). https://doi.org/10.1038/s41377-022-00953-y.

[12] W. Yang, F. Knorr, I. Latka, M. Vogt, G.O. Hofmann, J. Popp, I.W. Schie, Real-time molecular imaging of near-surface tissue using Raman spectroscopy, Light Sci Appl 11(1) (2022) 90. https://doi.org/10.1038/s41377-022-00773-0.

[13] H. Ding, D.J.J. Hu, X. Yu, X. Liu, Y. Zhu, G. Wang, Review on All-Fiber Online Raman Sensor with Hollow Core Microstructured Optical Fiber, Photonics 9(3) (2022). https://doi.org/10.3390/photonics9030134.
[14] Y. Xu, P. Zhong, A. Jiang, X. Shen, X. Li, Z. Xu, Y. Shen, Y. Sun, H. Lei, Raman spectroscopy coupled with chemometrics for food authentication: A review, TrAC Trends in Analytical Chemistry 131 (2020). https://doi.org/10.1016/j.trac.2020.116017.
[15] C. Zhang, S. Qiao, Y. Ma, Highly sensitive photoacoustic acetylene detection based on differential photoacoustic cell with retro-reflection-cavity, Photoacoustics 30 (2023) 100467. https://doi.org/10.1016/j.pacs.2023.100467.
[16] A. Pors, K.G. Rasmussen, R. Inglev, N. Jendrike, A. Philipps, A.G. Ranjan, V. Vestergaard, J.E. Henriksen, K. Norgaard, G. Freckmann, K.D. Hepp, M.C. Gerstenberg, A. Weber, Accurate Post-Calibration Predictions for Noninvasive Glucose Measurements in People Using Confocal Raman Spectroscopy, ACS Sens 8(3) (2023) 1272-1279. https://doi.org/10.1021/acssensors.2c02756.
[17] H.-P. Wang, P. Chen, J.-W. Dai, D. Liu, J.-Y. Li, Y.-P. Xu, X.-L. Chu, Recent advances of chemometric calibration methods in modern spectroscopy: Algorithms, strategy, and related issues, TrAC Trends in Analytical Chemistry 153 (2022). https://doi.org/10.1016/j.trac.2022.116648.
[18] 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, Spectrochim Acta A Mol Biomol Spectrosc 275 (2022) 121189. https://doi.org/10.1016/j.saa.2022.121189.
[19] 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, Spectrochim Acta A Mol Biomol Spectrosc 264 (2022) 120229. https://doi.org/10.1016/j.saa.2021.120229.
[20] S. Wang, S. Liu, Y. Yuan, J. Zhang, Z. Wang, X. Che, A novel CC-tSNE-SVR model for rapid determination of diesel fuel quality by near infrared spectroscopy, Infrared Physics & Technology 106 (2020). https://doi.org/10.1016/j.infrared.2020.103276.
[21] T. Ouyang, C. Wang, Z. Yu, R. Stach, B. Mizaikoff, G.-B. Huang, Q.-J. Wang, NOx Measurements in Vehicle Exhaust Using Advanced Deep ELM Networks, IEEE Transactions on Instrumentation and Measurement 70 (2021) 1-10. https://doi.org/10.1109/tim.2020.3013129.
[22] Q. Wang, S. Song, L. Li, D. Wen, P. Shan, Z. Li, Y. Fu, An extreme learning machine optimized by differential evolution and artificial bee colony for predicting the concentration of whole blood with Fourier Transform Raman spectroscopy, Spectrochim Acta A Mol Biomol Spectrosc 292 (2023) 122423. https://doi.org/10.1016/j.saa.2023.122423.
[23] C. Wang, Y. Sun, Y. Zhou, Y. Cui, W. Yao, H. Yu, Y. Guo, Y. Xie, Dynamic monitoring oxidation process of nut oils through Raman technology combined with PLSR and RF-PLSR model, Lwt 146 (2021). https://doi.org/10.1016/j.lwt.2021.111290.
[24] J. Wang, S. Lu, S.-H. Wang, Y.-D. Zhang, A review on extreme learning machine, Multimedia Tools and Applications 81(29) (2021) 41611-41660.

https://doi.org/10.1007/s11042-021-11007-7.
[25] P. Li, B. Anduv, X. Zhu, X. Jin, Z. Du, Diagnosis for the refrigerant undercharge fault of chiller using deep belief network enhanced extreme learning machine, Sustainable Energy Technologies and Assessments 55 (2023). https://doi.org/10.1016/j.seta.2022.102977.
[26] J. Xia, D. Yang, H. Zhou, Y. Chen, H. Zhang, T. Liu, A.A. Heidari, H. Chen, Z. Pan, Evolving kernel extreme learning machine for medical diagnosis via a disperse foraging sine cosine algorithm, Comput Biol Med 141 (2022) 105137. https://doi.org/10.1016/j.compbiomed.2021.105137.
[27] L. Armi, E. Abbasi, J. Zarepour-Ahmadabadi, Texture images classification using improved local quinary pattern and mixture of ELM-based experts, Neural Computing and Applications 34(24) (2021) 21583-21606. https://doi.org/10.1007/s00521-021-06454-0.
[28] B. Ahuja, V.P. Vishwakarma, Deterministic Multi-kernel based extreme learning machine for pattern classification, Expert Systems with Applications 183 (2021). https://doi.org/10.1016/j.eswa.2021.115308.
[29] X. Wang, K. Yang, J.H. Kalivas, Comparison of extreme learning machine models for gasoline octane number forecasting by near-infrared spectra analysis, Optik 200 (2020). https://doi.org/10.1016/j.ijleo.2019.163325.
[30] D. Xiao, T.T.G. Le, T.T. Doan, B.T. Le, Coal identification based on a deep network and reflectance spectroscopy, Spectrochim Acta A Mol Biomol Spectrosc 270 (2022) 120859. https://doi.org/10.1016/j.saa.2022.120859.
[31] K. Wang, X. Bian, M. Zheng, P. Liu, L. Lin, X. Tan, Rapid determination of hemoglobin concentration by a novel ensemble extreme learning machine method combined with near-infrared spectroscopy, Spectrochim Acta A Mol Biomol Spectrosc 263 (2021) 120138. https://doi.org/10.1016/j.saa.2021.120138.
[32] Z.-Y. Zhang, M.-Q. Jiang, H.-M. Xiong, Optimized identification of cheese products based on Raman spectroscopy and an extreme learning machine, New Journal of Chemistry 47(14) (2023) 6889-6894. https://doi.org/10.1039/d2nj06350f.
[33] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: Theory and applications, Neurocomputing 70(1-3) (2006) 489-501. https://doi.org/10.1016/j.neucom.2005.12.126.
[34] S. Ding, H. Zhao, Y. Zhang, X. Xu, R. Nie, Extreme learning machine: algorithm, theory and applications, Artificial Intelligence Review 44(1) (2013) 103-115. https://doi.org/10.1007/s10462-013-9405-z.
[35] 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. https://doi.org/10.1109/access.2020.2987547.
[36] D. Karaboga, B. Akay, A comparative study of Artificial Bee Colony algorithm, Applied Mathematics and Computation 214(1) (2009) 108-132. https://doi.org/10.1016/j.amc.2009.03.090.
[37] D. Karaboga, B. Basturk, A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm, Journal of Global Optimization 39(3) (2007) 459-471. https://doi.org/10.1007/s10898-007-9149-x.

[38] S. Faußer, F. Schwenker, Selective neural network ensembles in reinforcement learning: Taking the advantage of many agents, Neurocomputing 169 (2015) 350-357. https://doi.org/10.1016/j.neucom.2014.11.075.
[39] L. Yang, Y. Jiang, H. Liu, X. Yang, Dimensional Error Prediction of Grinding Process Based on Bagging–GA–ELM with Robust Analysis, Machines 11(1) (2022). https://doi.org/10.3390/machines11010032.
[40] J. Liao, J. Lv, Y. Shao, P. Li, X. Hu, Application of BP Neural Network Ensemble Model Based on Bagging Algorithm, International Journal of Machine Learning and Computing 9(2) (2019) 121-128. https://doi.org/10.18178/ijmlc.2019.9.2.775.
[41] X. Wu, S. Gao, Y. Niu, Z. Zhao, R. Ma, B. Xu, H. Liu, Y. Zhang, Quantitative analysis of blended corn-olive oil based on Raman spectroscopy and one-dimensional convolutional neural network, Food Chem 385 (2022) 132655. https://doi.org/10.1016/j.foodchem.2022.132655.
[42] P. Mishra, A. Biancolillo, J.M. Roger, F. Marini, D.N. Rutledge, New data preprocessing trends based on ensemble of multiple preprocessing techniques, TrAC Trends in Analytical Chemistry 132 (2020). https://doi.org/10.1016/j.trac.2020.116045.
[43] H.M. Abdulwahab, S. Ajitha, M.A.N. Saif, Feature selection techniques in the context of big data: taxonomy and analysis, Applied Intelligence 52(12) (2022) 13568-13613. https://doi.org/10.1007/s10489-021-03118-3.
[44] X.-H. Bian, S.-J. Li, M.-R. Fan, Y.-G. Guo, N. Chang, J.-J. Wang, Spectral quantitative analysis of complex samples based on the extreme learning machine, Analytical Methods 8(23) (2016) 4674-4679. https://doi.org/10.1039/c6ay00731g.
[45] S. Sengupta, A. Handoo, I. Haq, K. Dahiya, S. Mehta, M. Kaushik, Clarke Error Grid Analysis for Performance Evaluation of Glucometers in a Tertiary Care Referral Hospital, Indian J Clin Biochem 37(2) (2022) 199-205. https://doi.org/10.1007/s12291-021-00971-4.