

## ORIGINAL RESEARCH

# Data-driven modeling based on kernel extreme learning machine for sugarcane juice clarification

Yanmei Meng<sup>1</sup> | Shuangshuang Yu<sup>1</sup> | Hui Wang<sup>1</sup> | Johnny Qin<sup>2</sup>  | Yanpeng Xie<sup>1</sup>

<sup>1</sup>College of Mechanical Engineering, Guangxi University, Nanning, China

<sup>2</sup>Energy, Commonwealth Scientific and Industrial Research Organisation, Pullenvale, Queensland, Australia

**Correspondence**

Yanmei Meng, College of Mechanical Engineering, Guangxi University, Nanning, China.

Email: gxu\_mengyun@163.com and

Johnny Qin, Energy, Commonwealth Scientific and Industrial Research Organisation, Pullenvale, QLD, Australia. Email: Johnny.Qin@csiro.au

**Funding information**

National Natural Science Foundation of China, Grant/Award Number: 61763001

**Abstract**

Clarification of sugarcane juice is an important operation in the production process of sugar industry. The gravity purity and the color value of juice are the two most important evaluation indexes in the cane sugar production using the sulphitation clarification method. However, in the actual operation, the measurement of these two indexes is usually obtained by offline experimental titration, which makes it impossible to timely adjust the system indicators. A data-driven modeling based on kernel extreme learning machine is proposed to predict the gravity purity of juice and the color value of clear juice. The model parameters are optimized by particle swarm optimization. Experiments are conducted to verify the effectiveness and superiority of the modeling method. Compared with BP neural network, radial basis neural network, and support vector machine, the model has a good performance, which proves the reliability of the model.

**KEYWORDS**

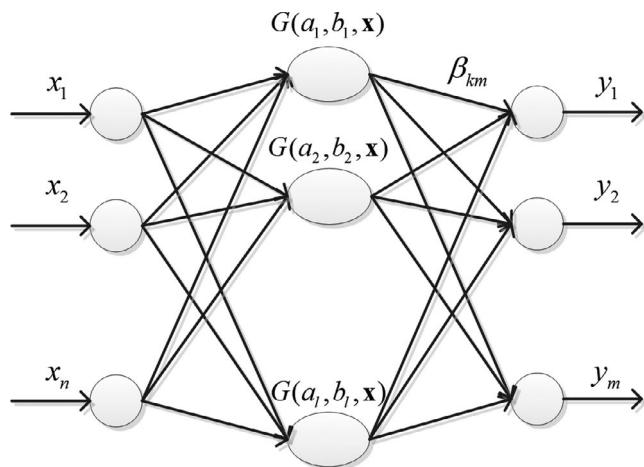
color value, extreme learning machine, gravity purity, particle swarm optimization, sugarcane juice clarification

## 1 | INTRODUCTION

Sugarcane juice clarification involves complex physical and chemical reactions. It is desirable to establish a mathematical model to analyze and control the clarification process, so that optimal final product be achieved. In control of complex industrial processes, there are generally two types of mathematical modeling approaches: modeling based on mechanism and modeling based on data-driven. Abderafi and Bounahmidi (1999) used the adapted Peng–Robinson equation of state to estimate the boiling temperatures of industrial beet and sugar cane juices over a wide range of dry substance content. Jourani and Bounahmidi (2002) explained the reaction process of calcium phosphate in the first stage by combining the growth rate equation of crystal with the dissolution equation through the kinetic method. Mirsaedghazi et al. (2010) proposed a mathematical

modeling of mass transfer in the concentration polarization layer of flat-sheet membranes during clarification of pomegranate juice. Cheng et al. (2011) used the data regression method to establish the quantitative relationship between calcium salt cations and acidic anions and constructed a mathematical model of calcium salt content and PH value. Hamerski, Silva, Corazza, Ndiaye, and Aquino (2012) presented a study of sugarcane juice carbonation and the evaluation of variable effects such as pH, carbonation time, and temperature on industrially relevant parameters for the quality of sugarcane juice. Three different batches of sugarcane juice were evaluated using a complete two-level factorial design with central point performed in triplicate.

There are few studies on the mechanism model of the sugarcane juice clarification process. This is because the cane juice clarification process is a large time delay, multivariable coupling, and nonlinear



**FIGURE 1** Network structure of extreme learning machine

process. It is extremely difficult to establish a complete mechanism model describing the clarification process which involves various complex physical and chemical reactions. Therefore, more and more researches turn to the process based on data-driven approach, without the need for the mechanism description of the clarification process. By relying on the online and offline data of the monitoring system, a model predicting the development of the clarification process can be obtained through mathematical processing. Lin and Yang (2009) established an Elman network model which improved the dual heuristic dynamic programming to predict the neutralized pH value and the purified pH value of sugarcane juice. Song, Wu, Lin, and Liu (2012) used a generalized dynamic fuzzy neural network to predict the color value and alkalinity during the carbonation clarification process of sugarcane juice and obtained a satisfactory result. Sartori et al. (2017) proposed artificial neural network (Lambda NN) models to predict the effects of different variables on sugarcane juice color removal and sucrose content.

The above-mentioned models of the clarification process focused on pH prediction, while more critical process parameters such as the gravity purity have not been involved. Moreover, these models are mainly based on the gradient descent method to update the model parameters (Al-Batah, Mat Isa, Zamli, & Azizli, 2010). While the generalization performance of these models is good, there are some problems, such as slow training speed and easy to fall into the local optimum, which limits the application and development of the model (Kaya & Uyar, 2013; Mohammed, Minhas, Jonathan Wu, & Sid-Ahmed, 2011).

In order to solve the problems of traditional learning machine of slow training speed and easy to fall into local optima, Huang, Zhu, and Siew (2006) proposed a new learning method, namely extreme learning machine (ELM). The ELM learning method has the advantages of less training parameters, very fast speed, and good generalization performance. Many researchers have applied ELM with different variants to solve different industrial problems. Wong, Wong, Vong, and Cheung (2015) used kernel-based ELM and cuckoo search to model and optimize the performance of biodiesel engine.

Farias et al. (2014) used extreme learning machine and bat algorithms to monitor product quality and provide fast and reliable product quality assessment of key process variables in second-generation ethanol production. Mohammadi et al. (2015) proposed an extreme learning machine (ELM)-based model for prediction of daily dew point temperature, and the model enjoys much greater prediction capability than SVM and ANN. Sadgrove, Falzon, Miron, and Lamb (2017) presented a color feature extreme learning machine (CF-ELM) for fast object detection in pastoral landscapes, which takes three color inputs instead of the standard grayscale input.

Among the different applications of ELM, the kernel-based ELM proved to have similar generalization performance to SVM while maintaining a much faster learning speed (Uçar & Özalp, 2017). Therefore, the kernel-based ELM is employed in our research to tackle the large time delay and strong coupling problem in the sugarcane juice clarification process where establishing of mechanism model is difficult.

In our data-driven model based on kernel extreme learning machine, four easy-to-measure variables in the sugarcane juice clarification process are selected as input, including the flow rate of the mixed juice, the intensity of sulfitation, the neutralization PH value, and the preliming PH value. Two difficult-to-measure variables are chosen as output, including the gravity purity of juice and the color value of clear juice. The parameters of the model are optimized by particle swarm optimization, and the effectiveness of the model is verified by experiment. To further evaluate the model performance on accuracy and time consuming, the results predicted from this model are also compared with those from other models such as BP, RBF and SVM.

## 2 | KERNEL EXTREME LEARNING MACHINE AND DATA-DRIVEN MODELING

### 2.1 | Kernel extreme learning machine

The support vector machine will not fall into the local minimum point during the learning process which makes it more generalized than BP in training feedforward neural networks, and the fitting degree of the test data set is more reliable (Huang, 2014; Wang, Zheng, Yoon, & Ko, 2018). However, the support vector machine has limited applicability to the field of system modeling. For a complex system, it may be necessary to build multiple parallel networks, which will cause a long modeling period. In order to overcome the shortcomings of neural networks and support vector machine, inspired by biological learning, Huang et al. (2006) proposed a new learning method, namely extreme learning machine (ELM). Unlike the traditional neural network methods which usually are time consumed and easy to get overfitting results, the ELM does not need to tune parameters for its hidden layers, resulting in a faster training speed and an improved generalization performance (Huang, 2014; Lu, Du, Liu, Xia, & Yeap, 2017).

The network structure of the extreme learning machine is shown in Figure 1.

Suppose a data sample set  $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$  has  $N$  samples, where  $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathbf{R}^n$  is the input matrix, indicating that there are  $n$  input variables, and  $\mathbf{y}_i = [y_{i1}, y_{i2}, \dots, y_{im}]^T \in \mathbf{R}^m$  is the output matrix, indicating that there are  $m$  output variables. According to the structure principle of feedforward neural network, the mathematical model of input and output with the number of  $l$  hidden layer neurons can be expressed by Equation (1).

$$f_N(\mathbf{x}_i) = \sum_{k=1}^l \beta G(x) = \sum_{k=1}^l \beta G(\langle \mathbf{a}_k, \mathbf{x}_i \rangle + b_k) \quad (1)$$

In the above formula,  $\beta$  is the network output weight;  $\mathbf{a}_k$  is the network input weight;  $\langle \mathbf{a}_k, \mathbf{x}_i \rangle$  is the inner product of  $\mathbf{a}_k$  and  $\mathbf{x}_i$ ; and  $b_k$  is the threshold of the  $k$ th hidden layer neuron. The activation function  $G(\bullet)$  of the hidden layer neurons can be any infinitely differentiable function such as sigmoid function, sine function, cosine function or compound function.

If the expression of Equation (1) is written as a matrix form consisting of  $N$  equations, it can be expressed as Equation (2).

$$\mathbf{H}\beta = \mathbf{Y} \quad (2)$$

where  $\mathbf{H}$  is the hidden layer output matrix.

$$\mathbf{H} = \mathbf{H}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_l, b_1, b_2, \dots, b_l, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$$

$$= \begin{bmatrix} G(\mathbf{a}_1 \cdot \mathbf{x}_1 + b_1) & G(\mathbf{a}_2 \cdot \mathbf{x}_1 + b_2) & \dots & G(\mathbf{a}_l \cdot \mathbf{x}_1 + b_l) \\ G(\mathbf{a}_1 \cdot \mathbf{x}_2 + b_1) & G(\mathbf{a}_2 \cdot \mathbf{x}_2 + b_2) & \dots & G(\mathbf{a}_l \cdot \mathbf{x}_2 + b_l) \\ \vdots & \vdots & \dots & \vdots \\ G(\mathbf{a}_1 \cdot \mathbf{x}_N + b_1) & G(\mathbf{a}_2 \cdot \mathbf{x}_N + b_2) & \dots & G(\mathbf{a}_l \cdot \mathbf{x}_N + b_l) \end{bmatrix}_{N \times l} \quad (3)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_l^T \end{bmatrix}_{l \times m}, \mathbf{Y} = \begin{bmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_N^T \end{bmatrix}_{N \times m} \quad (4)$$

Unlike the traditional feedforward neural network which needs to adjust all network parameters in the training process to get the optimality, Huang et al. demonstrated that ELM input weights and hidden layer neuron thresholds can be randomly initialized prior to training and remain unchanged during training. The weight vector  $\beta$  connected between the hidden layer and the output layer can be solved by Equation (5):

$$\left\| \mathbf{H}(\mathbf{a}_1, \dots, \mathbf{a}_l, b_1, \dots, b_l) \hat{\beta} - \mathbf{Y} \right\| = \min_{\beta} \left\| \mathbf{H}(\mathbf{a}_1, \dots, \mathbf{a}_l, b_1, \dots, b_l) \beta - \mathbf{Y} \right\| \quad (5)$$

The solution of the above formula is.

$$\hat{\beta} = \mathbf{H}^+ \mathbf{Y} \quad (6)$$

where  $\mathbf{H}^+$  is the Moore–Penrose generalized inverse of the hidden layer output matrix  $\mathbf{H}$ . It can be solved by various methods such as orthogonal projection, orthogonalization method, iterative method,

and singular value decomposition. When using Moore–Penrose generalized inverse to solve  $\mathbf{H}\beta = \mathbf{Y}$ , it is a least squares solution and is easily overfitting in the case of large samples. By introducing the concept of kernel function into the extreme learning machine (Huang, Zhou, Ding, & Zhang, 2012), the kernel extreme learning machine (KELM) is formed. This can effectively avoid the original randomness of ELM, achieving faster training, and better generalization performance (Huang, 2014; Jian et al., 2017)

Replace the output matrix of the ELM with the corresponding kernel function, as shown in Equation (7):

$$\mathbf{H}\mathbf{H}^T(i, j) = K(x_i, x_j) \quad (7)$$

This leads to

$$\mathbf{H}\mathbf{H}^T = \mathbf{\Omega}_{\text{ELM}} = \begin{bmatrix} K(x_1, x_1) & \dots & K(x_1, x_j) \\ \vdots & \ddots & \vdots \\ K(x_N, x_1) & \dots & K(x_N, x_N) \end{bmatrix} \quad (8)$$

$$h(x)\mathbf{H}^T = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix} \quad (9)$$

Therefore, the output of KELM can be written as

$$f_N(x) = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_n) \end{bmatrix}^T \left( \frac{1}{C} + \mathbf{\Omega}_{\text{ELM}} \right)^{-1} \mathbf{Y} \quad (10)$$

where  $\mathbf{I}$  is the identity matrix,  $C$  is the penalty factor, and the Gaussian kernel function is chosen as the kernel function of the model.

## 2.2 | Data-driven modeling based on kernel extreme learning machine

### 2.2.1 | Input and output variables

There are many parameters that may have influence on the clarification of sugarcane juice. Table 1 lists the eight potential parameters that may have significant influence on the clarification process. To eliminate those parameters that have insignificant influence on the process so that to reduce the dimensionality of the data set required to be treated during the modeling, the principal component analysis (PCA) method is used. At the end, the parameters that have significant influence on the clarification process of the cane juice are extracted.

By calculating the size of the eigenvalues of each variable and the cumulative contribution rate to the clarification, it is able to determine the features to be extracted by the PCA feature dimension reduction. Table 2 shows the contribution rate statistics of each

**TABLE 1** Symbolic meaning table of parameter variable set for sugarcane juice clarification process

No.	Description
$n_1$	Mixed juice flow (t/hr)
$n_2$	Preliming valve opening (%)
$n_3$	Preliming PH value
$n_4$	Neutralization PH value
$n_5$	Clear juice PH value
$n_6$	Intensity of sulfitation (ml)
$n_7$	Floating color value
$n_8$	Clear juice brix (Bx)

variable calculated from a data set of 277 data samples, which were obtained by the experimental platform.

It can be seen from Table 2 that the cumulative contribution rate of the first four variables,  $n_1$ ,  $n_6$ ,  $n_4$ , and  $n_3$ , has already reached 86.19%, exceeding the usual requirement of 85% when using the cumulative variance contribution rate method (Niu, 2011). Therefore, the mixed juice flow ( $x_1$ ), the intensity of sulfitation ( $x_2$ ), the neutralization PH value ( $x_3$ ), and the preliming PH value ( $x_4$ ) are selected as input. Meanwhile, two unmeasurable parameters, the Gravity purity of juice ( $y_1$ ) and the Color value of clear juice ( $y_2$ ), are taken as output, as shown in Table 3.

### 2.2.2 | Parameter optimization

Particle swarm optimization (PSO) is a group intelligence global search optimization algorithm proposed by Kennedy in 1995 (Kennedy & Eberhart, 1995), which was inspired by the behavior of bird foraging groups. The particle swarm optimization algorithm has the characteristics of fast convergence, easy experimentation, and easy combination with other algorithms. It has been widely used in many fields, such as economic dispatch, robot application, signal processing, and image segmentation (Mahor, Prasad, & Rangnekar, 2009; Sengupta & Das, 2017; Suresh & Lal, 2017; Zhang, Gong, & Zhang, 2013). Therefore, the PSO is

**TABLE 2** Principal component contribution rate of parameters to sugarcane juice clarification process

No.	Contribution rate to the clarification process/%	Cumulative contribution rate/%
$n_1$	27.597	27.580
$n_6$	24.709	52.307
$n_4$	17.514	69.821
$n_3$	16.373	86.194
$n_5$	3.980	90.173
$n_2$	3.518	93.692
$n_7$	3.261	96.953
$n_8$	3.047	100

employed to optimize the parameters during the clarification process of the sugarcane juice.

The fitness function is used to evaluate the pros and cons of the particles. It has a direct impact on the algorithm optimization results. During the optimization process, each particle in the group moves according to its own fitness function value, rather than a random flight. In each iteration, the optimal value of the individual particles is updated by comparing its historical optimal value with the current state optimal value.  $\mathbf{p}_i = (p_{i1}, p_{i2}, \dots, p_{iD})$  is used to denote the optimal position that individual particles can currently find, called "local optimum." After traversing each individual particle, the population global optimal value is updated by comparing the group history optimal value with the current state group optimal value.  $\mathbf{p}_g = (p_{g1}, p_{g2}, \dots, p_{gD})$  is used to indicate the optimal location that the group can currently find, called "global optimality."

The velocity and position iteration equations of the particle swarm algorithm are as follows:

$$v_{id}^{k+1} = \omega \cdot v_{id}^k + c_1 \cdot r_1 \cdot (p_{id}^k - x_{id}^k) + c_2 \cdot r_2 \cdot (p_{gd}^k - x_{id}^k) \quad (11)$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \quad (12)$$

where  $k$  is the number of algorithm iterations,  $d=1,2,\dots,D$  is the search solution in the D-dimensional space,  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{id})$  is the position vector of the particle  $i$  in the D-dimensional search space,  $\mathbf{v}_i = (v_{i1}, v_{i2}, \dots, v_{id})$  is the moving speed of the particle  $i$  position change,  $p_{id}^k$  is the optimal position of the individual particle after  $k$  iterations,  $p_{gd}^k$  is the optimal position of the group after  $k$  iterations.  $\omega$  is the inertia factor, and  $c_1, c_2$  are the acceleration factors, and  $r_1, r_2$  are two random numbers between (0, 1).

The performance of the KELM-based cane juice clarification process model will be affected by the penalty factor  $C$  and the kernel parameter  $\sigma$ . In this model, the prediction result RMSE of the sample data of the sugarcane juice clarification process is taken as the fitness function, and the optimal penalty factor  $C$  and kernel parameter  $\sigma$  can be obtained.

### 2.2.3 | Data-driven model

The specific steps of construction of the data-driven model based on KELM are shown in Figure 2. As mentioned previously, four

**TABLE 3** Input and output variable table

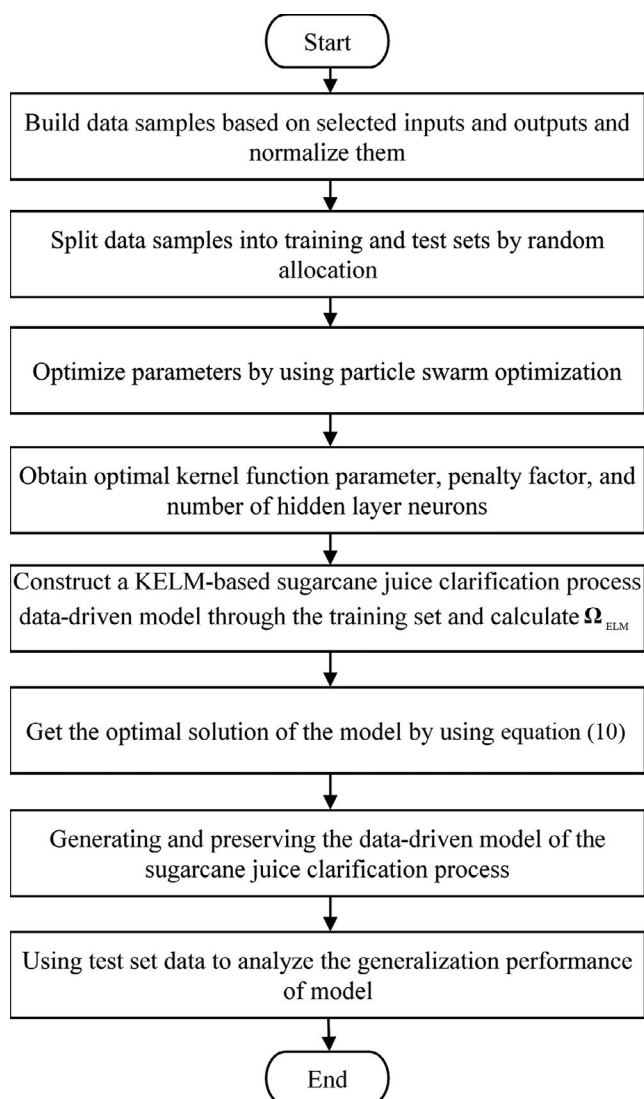
No.	Variables	Unit
$x_1$	Mixed juice flow	t/hr
$x_2$	Intensity of sulfitation	ml
$x_3$	Neutralization PH value	Null
$x_4$	Preliming PH value	Null
$y_1$	Gravity purity of juice	GP
$y_2$	Color value of clear juice	°St

parameters are selected as input, and two parameters are set as output. The Gaussian kernel function is chosen as the kernel function of the model.

### 3 | EXPERIMENTAL RESULTS AND ANALYSIS

#### 3.1 | Model performance index

In order to evaluate the performance of the data-driven model for sugarcane juice clarification, it is necessarily to define the model evaluation criteria. Suppose the actual value of the  $i$ -th test sample with  $m$  test samples is expressed as  $y_i$ , the predicted value of the corresponding  $i$ -th test sample is  $\hat{y}_i$ , and the mean value of the test sample is  $\bar{y} = \frac{\sum_{i=1}^m y_i}{m}$ . The performance of the data-driven model for the sugarcane juice clarification process was evaluated by using the indexes shown in Table 4 (Malik, 2005).



**FIGURE 2** Flow chart of constructing the data-driven model for sugarcane juice clarification process

**TABLE 4** The model performance indexes and their definitions

Indexes of model performance evaluation	Formula
Root mean square error: RMSE	$RMSE = \sqrt{\frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{m}}$
Mean absolute error: MAE	$MAE = \frac{\sum_{i=1}^m  y_i - \hat{y}_i }{m}$
Determination coefficient: $R^2$	$R^2 = \frac{\sum_{i=1}^m (y_i - \bar{y})^2}{\sum_{i=1}^m (y_i - \hat{y}_i)^2}$

#### 3.2 | Experimental results and analysis

##### 3.2.1 | Experimental platform

In order to verify the data-driven model of the sugarcane juice clarification process, a comprehensive experimental platform is developed. The platform mainly includes a sedimentation tank, some sugarcane juice tanks, an auxiliary part and a control valve part, as shown in Figure 3.

##### 3.2.2 | Experimental result

The computer operating environment provided in this paper is Inter(R) Corei5, 2.4GHz, 4G memory, Windows 7 operating system, and MATLAB 2014a is used as the running computing software. During the clarification process of a batch of sugarcane juice using the experimental platform, 277 sets of data samples are obtained, and the data samples are normalized and divided into training set and testing set by random allocation. The number of the training set and testing set are 200 and 77, respectively. The training set is mainly used for training the optimization of model parameters by the group intelligence algorithm and the construction of the



**FIGURE 3** Sugarcane juice clarification experimental platform (1, 3, 6—heaters; 2—sulfitor; 4—diffusion box; 5—sedimentation tank; 7, 8, 9—evaporators; 10—syrup tank; 11, 12—equal pressure tanks; 13—evaporating hot water tank; 14—clear juice tank; 15—flocculant tank; 16—lime milk tank; 17—neutralizer; 18—neutralizing juice tank; and 19—mixed juice tank)



**TABLE 5** Optimal parameter results of data-driven model for sugarcane juice clarification process

Optimization algorithm	Optimization time (s)	Root mean square error (RMSE)	Parameter optimization result	
			C	$\sigma$
PSO	38.4523	0.0562	112.76	71.38

data-driven model. The test set is used to test and analyze the data-driven model. Four variables including the mixed juice flow, intensity of sulfitation, neutralization PH value, and preliming PH value are selected as the input, and two production indexes including the gravity purity of juice and the color value of clear juice are used as output. Some parameters of the data-driven model, such as penalty factor C and kernel parameter  $\sigma$ , are optimized by PSO. The number of optimization iterations is set as 200, and the range of parameter optimization is set as [0, 1,024]. Using the root mean square error (RMSE) as fitness function, through a number of trial and error, the optimal number of hidden layer neurons is determined as 25. After the iteration is completed, the optimal parameter combination of the model is obtained. The optimization results are shown in Table 5.

After obtaining the optimal penalty factor and kernel function parameter, it is necessary to learn the training samples of the sugarcane juice clarification process and construct a data-driven model. The RMSE, MAE, and  $R^2$  are used to show the performance of the model. The training time represents the running time required by the CPU when the model is optimized using the training set to rebuild the model.

Figure 4a shows the predicted gravity purity including the measured actual values for the purpose of comparison. Prediction fits the measurement very well. The prediction error fluctuation is small. The maximum error is 0.10 (see Figure 4b), accounting 6.21% of the average gravity of 1.61 of the data sample.

Figure 5 shows the prediction result and error when the color value is the model output. It can be seen that the maximum error is 0.69, accounting 1.17% of the average (59) of the color values.

Table 6 shows the model performance of the sugarcane juice clarification process. As can be seen in Table 6, the training time of the data-driven model based on KELM is small (0.3325s), which indicates that the model has a high learning efficiency.

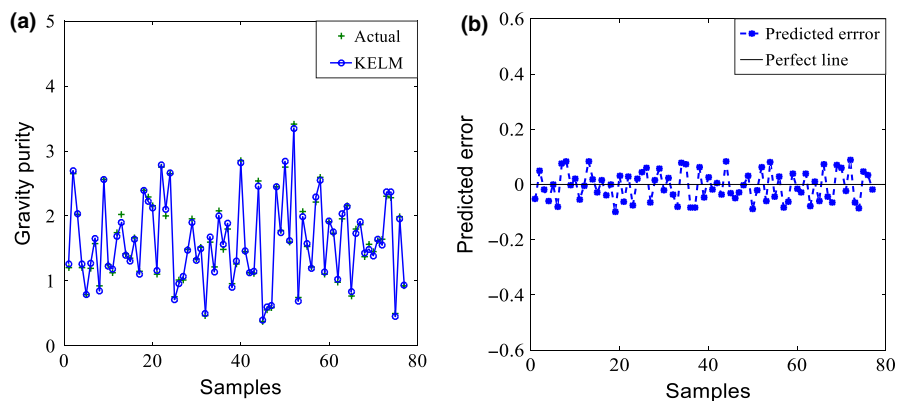
The performance evaluation indicators, RMSE, and MAE are relatively low values. The coefficient of determination  $R^2$  is 0.934 for predicting purity and 0.972 for predicting color value. Both are very close to 1, proving the excellent learning performance of the model.

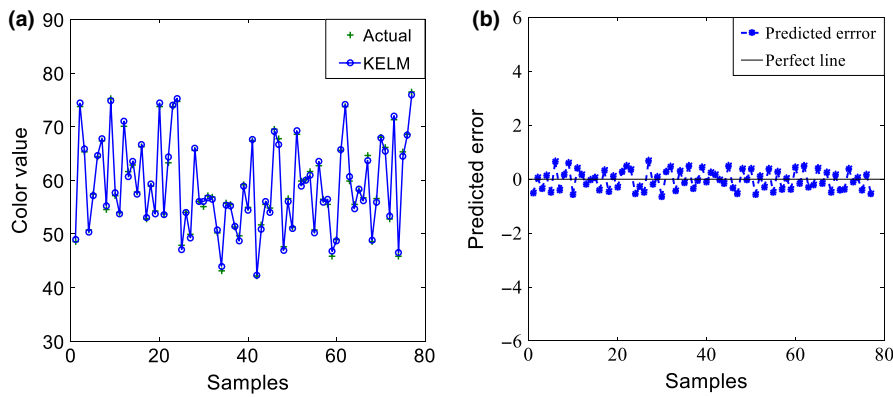
### 3.2.3 | Comparison with other learning methods

In order to verify the validity and superiority of the data-driven model based on KELM for the cane juice clarification process, a number of models with different learning methods including BP, RBF, and SVM have been run using the same data set of 277 data samples that was used in the KELM-based model to compare their performances. Figure 6 shows the performance comparison of the four models when the gravity purity of juice is chosen as output. Figure 7 shows the results of predicting the color value of clear juice. Tables 7 and 8 compare the performance indexes of different models in predicting gravity purity and color value, respectively.

It is seen from Figures 6 and 7 that all models perform better in predicting the gravity purity than in predicting the color value. While the prediction accuracy and generalization performance of both BP and RBF are substandard with  $R^2 < 0.85$ , and large scattered data, the SVM and KELM models perform much better with  $R^2$  of 0.93 and 0.97, respectively. Of the four models, the KELM performs the best with the highest  $R^2$  value (0.97) and the most convergent data.

From Tables 7 and 8, it can be seen that while the training times of SVM and KELM are just 10% of that of BP and RBF for predicting purity and color value, the training time of KELM is the smallest among the four models. The error indexes, mean and absolute error (MAE), and root mean square error (RMSE) of the KELM are also the smallest among the four models, indicating the high predicting accuracy of the KELM model.

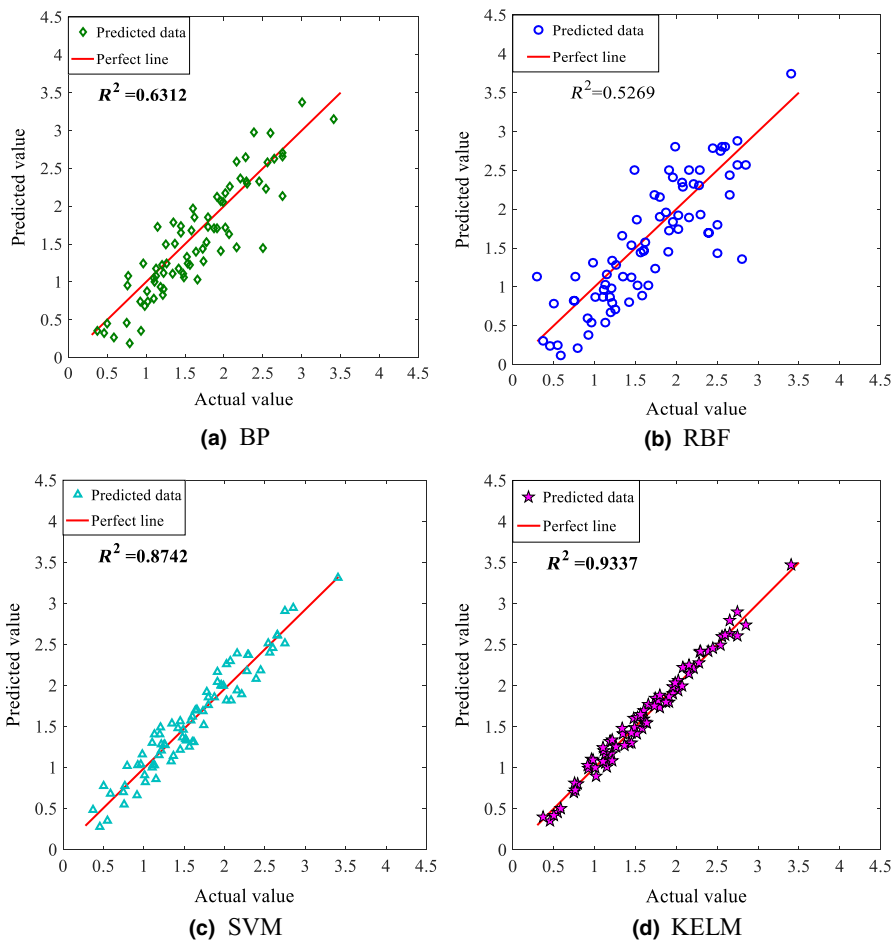
**FIGURE 4** Test set prediction performance of data-driven model with the gravity purity as output: (a) prediction result, (b) prediction error



**FIGURE 5** Test set prediction performance of data-driven model with the color value as output: (a) prediction result, (b) prediction error

Model output	Training time (s)	RMSE	MAE	$R^2$
Gravity purity of juice	0.3325	0.0764	0.0608	0.9337
Color value of clear juice		0.4593	0.2159	0.9722

**TABLE 6** Data-driven model performance of sugarcane juice clarification process based on KELM

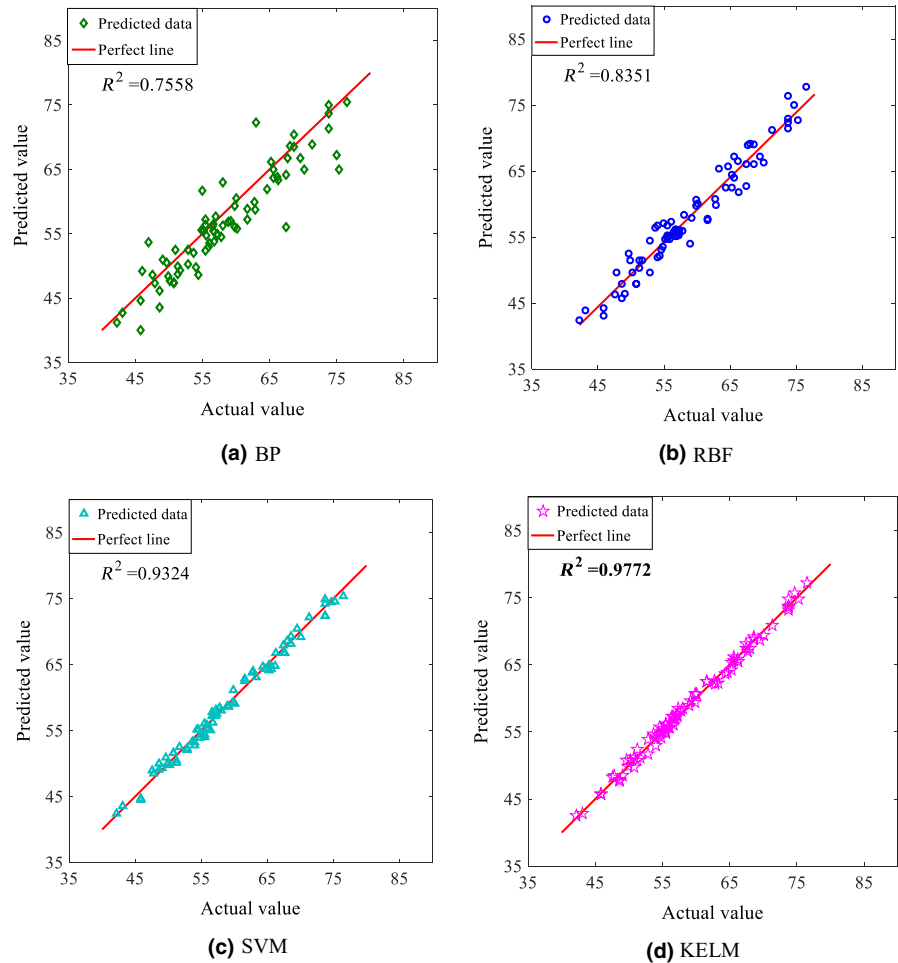


**FIGURE 6** Comparison of coefficient of determination  $R^2$  obtained from different data-driven models with the gravity purity as output

## 4 | CONCLUSION

A data-driven model based on kernel extreme learning machine has been established to predict the gravity purity of juice and the color

value of clear juice which cannot be measured online during the large time-delay nonlinear clarification of sugarcane juice. A comprehensive experiment platform has been built to verify the data-driven model. Using the data obtained from the experiment, the principal



**FIGURE 7** Comparison of coefficient of determination  $R^2$  obtained from different data-driven models with the color value as output

**TABLE 7** Model performance comparison of different models with the gravity purity as output

Algorithms	Training time (s)	RMSE	MAE	$R^2$
BP	3.0956	0.4465	0.2107	0.6312
RBF	4.3215	0.5327	0.2387	0.5269
SVM	0.5742	0.1812	0.1104	0.8742
KELM	0.3325	0.0764	0.0608	0.9337

**TABLE 8** Model performance comparison of different models with the color value as output

Algorithms	Training time (s)	RMSE	MAE	$R^2$
BP	3.0956	1.1872	0.5399	0.7558
RBF	4.3215	1.0371	0.4832	0.8351
SVM	0.5137	0.6742	0.2691	0.9324
KELM	0.3325	0.4593	0.2159	0.9722

component analysis method is used to extract four key variables that have significant influence on the clarification process as input. The two variables, gravity purity and color value, which are difficult to be measured are used as output. Optimizing of model parameters is

achieved by particle swarm optimization. Finally, the model is validated by experimental data. The results show that the data-driven model based on kernel extreme learning machine has good prediction performance and high reliability and efficiency. Comparative analysis of different models indicates the KELM model has superior performance.

#### ACKNOWLEDGMENT

The project is supported by National Natural Science Foundation of China (No. 61763001).

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### ETHICAL STATEMENTS

This research does not involve any human or animal testing.

#### ORCID

Johnny Qin  <https://orcid.org/0000-0002-2467-948X>



## REFERENCES

- Abderafi, S., & Bounahmidi, T. (1999). Measurement and estimation of vapor-liquid equilibrium for industrial sugar juice using the Peng-Robinson equation of state. *Fluid Phase Equilibria*, 162(1–2), 225–240. [https://doi.org/10.1016/S0378-3812\(99\)00184-3](https://doi.org/10.1016/S0378-3812(99)00184-3)
- Al-Batah, M. S., Mat Isa, N. A., Zamli, K. Z., & Azizli, K. A. (2010). Modified recursive least squares algorithm to train the hybrid multilayered perceptron (HMLP) network. *Applied Soft Computing*, 10(1), 236–244. <https://doi.org/10.1016/j.asoc.2009.06.018>
- Cheng, L., Hou, Z. G., Lin, Y., Tan, M., Zhang, W. C., & Wu, F. X. (2011). Recurrent neural network for non-smooth convex optimization problems with application to the identification of genetic regulatory networks. *IEEE Transactions on Neural Networks*, 22(5), 714–726. <https://doi.org/10.1109/TNN.2011.2109735>
- Farias, F. S., Azevedo, R. A., Rivera, E. C., Herrera, W. E., Filho, R. M., & Lima, L. P. (2014). Product quality monitoring using extreme learning machines and bat algorithms: A case study in second-generation ethanol production. *Computer Aided Chemical Engineering*, 33, 955–960.
- Hamerski, F., da Silva, V. R., Corazza, M. L., Ndiaye, P. M., & de Aquino, A. D. (2012). Assessment of variables effects on sugar cane juice clarification by carbonation process. *International Journal of Food Science & Technology*, 47(2), 422–428. <https://doi.org/10.1111/j.1365-2621.2011.02857.x>
- Huang, G.-B. (2014). An insight into extreme learning machines: Random neurons, random features and kernels. *Cognitive Computation*, 6(3), 376–390. <https://doi.org/10.1007/s12559-014-9255-2>
- Huang, G. B., Zhou, H., Ding, X., & Zhang, R. (2012). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 42(2), 513–529. <https://doi.org/10.1109/TSMCB.2011.2168604>
- Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1), 489–501. <https://doi.org/10.1016/j.neucom.2005.12.126>
- Jian, Y., Huang, D., Yan, J., Lu, K., Huang, Y., Wen, T., ... Xie, Q. (2017). A novel extreme learning machine classification model for e-nose application based on the multiple kernel approach. *Sensors*, 2017(17), 1434. <https://doi.org/10.3390/s17061434>
- Jourani, A., & Bounahmidi, T. (2002). Kinetics of hydroxyapatite heterogeneous nucleation and crystal growth during spontaneous calcium phosphate precipitation in sugar solutions. *Chemical Engineering Journal*, 89(1–3), 185–192.
- Kaya, Y., & Uyar, M. (2013). A hybrid decision support system based on rough set and extreme learning machine for diagnosis of hepatitis disease. *Applied Soft Computing*, 13(8), 3429–3438. <https://doi.org/10.1016/j.asoc.2013.03.008>
- Kennedy, J., & Eberhart, R. C. (1995). Particle swarm optimization. *Proceedings of IEEE International Conference on Neural Networks*, 1942–1948.
- Lin, X., & Yang, J. (2009). HDP for the neutralized pH value control in the clarifying process of sugar cane juice. *AIP Conference Proceedings*, 1127(1), 162–173.
- Lu, H., Du, B., Liu, J., Xia, H., & Yeap, W. K. (2017). A kernel extreme learning machine algorithm based on improved particle swarm optimization. *Memetic Computing*, 9, 121–128. <https://doi.org/10.1007/s12293-016-0182-5>
- Mahor, A., Prasad, V., & Rangnekar, S. (2009). Economic dispatch using particle swarm optimization: A review. *Renewable and Sustainable Energy Reviews*, 13(8), 2134–2141. <https://doi.org/10.1016/j.rser.2009.03.007>
- Malik, M. B. (2005). Applied linear regression. *Technometrics*, 47(3), 371–372. <https://doi.org/10.1198/tech.2005.s286>
- Mirsaeedghazi, H., Emam-Djomeh, Z., Mousavi, S. M., Enjileha, V., Navidbakhsh, M., & Mirhashemi, S. M. (2010). Mathematical modelling of mass transfer in the concentration polarisation layer of flat-sheet membranes during clarification of pomegranate juice. *International Journal of Food Science & Technology*, 45(10), 2096–2100. <https://doi.org/10.1111/j.1365-2621.2010.02377.x>
- Mohammadi, K., Shamshirband, S., Motamedi, S., Petković, D., Hashim, R., & Gocic, M. (2015). Extreme learning machine based prediction of daily dew point temperature. *Computers and Electronics in Agriculture*, 117, 214–225. <https://doi.org/10.1016/j.compag.2015.08.008>
- Mohammed, A. A., Minhas, R., Jonathan Wu, Q. M., & Sid-Ahmed, M. (2011). Human face recognition based on multidimensional PCA and extreme learning machine. *Pattern Recognition*, 44(10), 2588–2597. <https://doi.org/10.1016/j.patcog.2011.03.013>
- Niu, J. K. (2011). Prediction of demand for iron ores in China based on principal component regression analysis. *Journal of University of Science & Technology Beijing*, 33(10), 1177–1181.
- Sadgrove, E. J., Falzon, G., Miron, D., & Lamb, D. (2017). Fast object detection in pastoral landscapes using a colour feature extreme learning machine. *Computers and Electronics in Agriculture*, 139, 204–212. <https://doi.org/10.1016/j.compag.2017.05.017>
- Sartori, J. A. D., Ribeiro, K., Teixeira, A., Magri, N. T. C., Mandro, J. L., & de Aguiar, C. L. (2017). Sugarcane juice clarification by hydrogen peroxide: Predictions with artificial neural networks. *International Journal of Food Engineering*, 13(2), 1–8. <https://doi.org/10.1515/ijfe-2016-0199>
- Sengupta, S., & Das, A. K. (2017). Particle swarm optimization based incremental classifier design for rice disease prediction. *Computers and Electronics in Agriculture*, 140, 443–451. <https://doi.org/10.1016/j.compag.2017.06.024>
- Song, S., Wu, J., Lin, X., & Liu, H. (2012). Predictive model of production index for sugar clarification process by GDFNN. In J. Wang, G.G. Yen, & M.M. Polycarpou (Eds.), *Advances in neural networks - ISNN 2012* (pp. 585–593). Berlin: Springer-Verlag.
- Suresh, S., & Lal, S. (2017). Multilevel thresholding based on Chaotic Darwinian particle swarm optimization for segmentation of satellite images. *Applied Soft Computing*, 55, 503–522. <https://doi.org/10.1016/j.asoc.2017.02.005>
- Uçar, A., & Özalp, R. (2017). Efficient android electronic nose design for recognition and perception of fruit odors using kernel extreme learning machines. *Chemometrics and Intelligent Laboratory Systems*, 166, 69–80. <https://doi.org/10.1016/j.chemolab.2017.05.013>
- Wang, H., Zheng, B., Yoon, S. W., & Ko, H. S. (2018). A support vector machine-based ensemble algorithm for breast cancer diagnosis. *European Journal of Operational Research*, 267(2), 687–699. <https://doi.org/10.1016/j.ejor.2017.12.001>
- Wong, P. K., Wong, K. I., Vong, C. M., & Cheung, C. S. (2015). Modeling and optimization of biodiesel engine performance using kernel-based extreme learning machine and cuckoo search. *Renewable Energy*, 74, 640–647. <https://doi.org/10.1016/j.renene.2014.08.075>
- Zhang, Y., Gong, D., & Zhang, J. (2013). Robot path planning in uncertain environment using multi-objective particle swarm optimization. *Neurocomputing*, 103, 172–185. <https://doi.org/10.1016/j.neucom.2012.09.019>

**How to cite this article:** Meng Y, Yu S, Wang H, Qin J, Xie Y. Data-driven modeling based on kernel extreme learning machine for sugarcane juice clarification. *Food Sci Nutr*. 2019;7:1606–1614. <https://doi.org/10.1002/fsn3.985>