Soft Sensor for Net Calorific Value of Coal Based on Improved PSO-SVM ⋆

Hongguang Pan, **Haoqian Song**, **Zheng Wang**

College of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an 710049, China (e-mail: hongguangpan@163.com).

Abstract: The drastic change of coal quality (e.g., net calorific value of coal) is an important factor that reduces the boiler combustion efficiency and stability of thermal power generating units, hence, the accurate and rapid measurement of net calorific value of coal is very critical. Considering the hardware measurement is cumbersome and costly, a soft sensing method based on improved particle swarm optimization - support vector machine (PSO-SVM) is proposed in this paper. Firstly, PSO is improved by dynamically adjusting inertial weights and learning factors, and introducing compression factors, so as to overcome the limitations of traditional PSO. In addition, the improved PSO algorithm is embedded into the process of optimizing parameters of SVM to improve model's accuracy. Secondly, based on the actual production data collected from a power plant in Yulin, Shaanxi, China; five proximate analysis compositions of coal are selected as original variables and preprocessed through gross error analysis, random error analysis. Moreover, combined with mechanism analysis, the invalid data items are eliminated; and based on the results of correlation analysis by using covariance method, the auxiliary variables with larger correlation coefficients are selected. Finally, the soft sensing models based on improved PSO-SVM, SVM, long short term memory (LSTM) and back propagation (BP) neural network are trained and debugged. Compared with SVM, LSTM and BP neural network, the soft sensing model based on improved PSO-SVM has obvious improvement in mean square error and mean square correlation coefficient with higher accuracy.

Keywords: Soft sensor; Auxiliary variable; Net calorific value; Improved PSO-SVM.

1. INTRODUCTION

In the operation of thermal power generating units, the coal calorific value can be used to quickly eliminate combustion disturbance, reduce units load pressure and adjust the air-coal ratio and water-coal ratio. Thus, the accurate measurement of calorific value of coal is very important to improve the stability and economy of thermal power generating units Mohanta et al. (2015). At present, the coal supply is tight and the coal source of power plants is complex, hence, the coal quality is hard to be guaranteed and the net calorific value of coal changes greatly, which have become one of the main factors affecting the stable and economic operation of units Hasan et al. (2017).

The accurate and rapid measurement for net calorific value of coal has become an urgent requirement for actual production. For the measurement of net calorific value of coal, the conventional method is oxygen bomb calorimetry, which needs off-line sampling analysis, and has complex operation, many influencing factors and long analysis period Eze and Oti (2017), Xiong et al. (2017). Although there are many other hardware measuring methods in industrial application, they generally have some disadvantages, such as bulky and expensive equipment, high maintenance cost and so on.

In comparison, the soft sensing method solves the economic problems of hardware measurement, and has lots of advantages, such as simplicity, utility and rapid response Lauri et al. (2017), Pan et al. (2021). The basic idea of soft sensor is as follows: 1) obtain data from the more mature hardware sensors, 2) build model with the related variables to indirectly measure the dominant variables Pan et al. (2020a). In addition, the soft sensing model emphasizes that to obtain the best estimate of the dominant variables Nawaz et al. (2019). Soft sensing method has a wide range of applications and remarkable effectiveness, which have been proved in many literatures Chen et al. (2017), Wang and Liu (2018). For example, Chen et. al adopted the top pressure, blast temperature and air volume of blast furnace in the process of silicon content analysis as input variables, and realized the measurement of silicon content in industrial molten iron Chen et al. (2017); Wang et. al used the speed of air preheater rotor, and the temperature, flow, pressure of the gas at the inlet and outlet of air duct and flue as the input variables, and achieved the measurement of the deformation degree of the air preheater rotor in a thermal power plant Wang and Liu (2018).

There are many modeling methods of soft sensor, which can be divided into mechanism modeling, regression anal-

⋆ This work is supported by the National Science Foundation of China(61422303), Shaanxi Postdoctoral Science Foundation (2018BSHEDZZ124) and China Postdoctoral Science Foundation (2017M623207).
⋆⋆Corresponding author: Hongguang Pan, hongguangpan@163.com.

ysis, state estimation, pattern recognition, artificial neural network (ANN), fuzzy inference, long short term memory (LSTM) network, support vector machine (SVM) and method based on kernel function, etc Khoshnevisan et al. (2015), Pan et al. (2020b). In recent years, there are many achievements on soft sensor for boiler variables in thermal power plants. For example, based on the ANN soft sensing model, Selvi et. al realized the real-time measurement of boiler drum water level during the operation of thermal power generating units Selvi et al. (2017); and based on the LSTM-based soft sensing model, Pan et. al achieved the measurement of oxygen content of flue gas in coal-fired power plant Pan et al. (2021).

In the study of boiler variables, because the measurement of coal calorific value is different from other variables, different research schemes of soft sensor are proposed Lu et al. (2017), Liu et al. (2013). Lu et. al proposed a soft sensing method combining laser-induced breakdown spectroscopy with ANN and genetic algorithm (GA), and realized the on-line measurement of gross calorific value (GCV) of coal in industry Lu et al. (2017). The advantage of ANN modeling method is that it can accurately deal with the complex nonlinear problems, but it is susceptible to over-training and the training speed is slow. Liu et. al proposed a soft sensing model based on three-layer BP neural network, analyzed the actual data of a power plant, and predicted the net calorific value of coal Liu et al. (2013). As a classical neural network, BP network can realize nonlinear mapping between input and output, which is widely used in pattern recognition and adaptive control, but BP network does not have the memory function of data, which causes it to preserve long-term series data features unsuccessfully.

When the boiler is running, due to the strict time series among proximate analysis data and the serious lag in relationship that between the proximate analysis data and coal calorific value data, it is necessary to build the soft sensing model of neural network with memory function and outstanding performance in the series data. On the basis of ordinary multi-layer BP network, the horizontal connection among the units in the hidden layer is added to recurrent neural network (RNN), which makes it have the memory function, but it has the gradient disappearance and explosion problems Czuszynski et al. (2018). LSTM network is improved based on RNN, which solves the long-term dependence problem of RNN and has outstanding performance in time-series data Nugaliyadde et al. (2019).

However, only when the number of samples is sufficiently large, the model obtained by neural network is stable, and the number of samples in reality is usually limited. SVM can solve this defect of neural network well, and does not depend on the quantity and quality of training samples. It can also guarantee the generalization of small sample problem, and has strict mathematical basis and theoretical derivation, strong approximation and generalization ability. The classical researches based on SVM soft-sensing model include: based on the soft sensing model of SVM, Yang et. al realized the measurement of NO_x concentration produced by different kinds of coal combustion Yang et al. (2017); Chen et. al realized a perfect prediction of monthly evaporation in the Three Gorges Reservoir Area by using the measured meteorological variables Chen et al. (2019).

SVM also has some shortcomings, similar to neural network, there is no unified conclusion on the selection of its super parameters, which generally relies on the experience of process adjustment, leading to difficulty in obtaining the best model effectiveness. Therefore, the SVM-based soft-sensing model is improved by introducing particle swarm optimization (PSO). For example, Li et. al used PSO to optimize the parameters of soft-sensing model based on the least square SVM, and realized the high-precision monitoring of oxygen content in the boiler flue gas Li and Ren (2017). In addition, because traditional PSO is difficult to maintain the balance between global search and local search, it cannot adjust dynamically. Therefore, the traditional PSO in this paper is improved, and the improved PSO is used to optimize the parameters of SVM, and then the soft sensing method based on improved PSO-SVM is used for modeling.

To sum up, a soft sensing method based on improved PSO-SVM is proposed in this paper, and then trained using the proximate analysis data of actual production to quickly and accurately measure the net calorific value of coal. The main works are as follows. Firstly, PSO is improved by dynamically adjusting inertia weight, learning factor and introducing compression factor. Moreover, improved PSO is embedded into the process of optimizing parameters of SVM. Secondly, based on the actual production data, auxiliary variables are initially selected according to mechanism analysis; and the data is preprocessed with the 3δ criteria, average filtering and min-max method. Moreover, the correlation analysis is carried out by using $corrcoef$ Pearson function, and the selection of auxiliary variables is completed by sorting and selecting the best. Finally, the model parameters are initialized according to the requirements of parameter range, the soft sensing models based on improved PSO-SVM, SVM, LSTM and BP neural network are trained and debugged, and the models are tested and compared though test set.

2. ALGORITHM

Learning network is a statistical learning method based on mega data information, its purpose is to establish a model for a specific problem through the statistical learning of a large number of samples, which is also an effective method for soft sensor modeling Yan et al. (2017). Therefore, this section is mainly to introduce the principle of improved PSO-SVM.

2.1 Basic Principles of SVM

Support vector machine (SVM) is a machine learning method based on statistical learning theory, and adopts the principle of structural risk minimization Huang et al. (2018). As shown in Fig. 1, when dealing with nonlinear problems, SVM transforms the nonlinear estimation problem into linear estimation problem of high-dimensional feature space by nonlinearly mapping sample data to high-dimensional space Peng and Chen (2018).

The main idea of SVM is as follows: fitting regression with the function $f(x)$, and achieving the optimal estimation by minimizing the fitting expected risk function (1) Pan and Dias (2017),

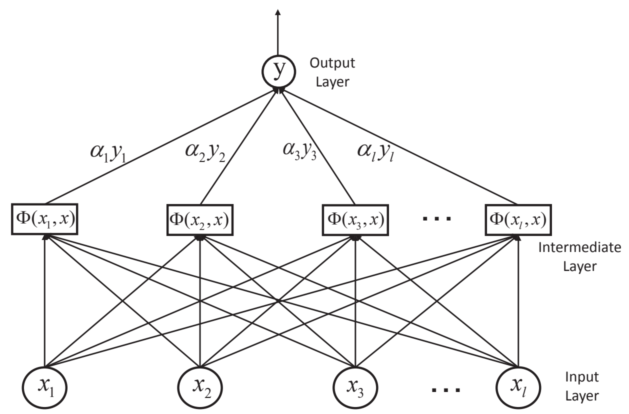


Fig. 1. Architecture of support vector machine

$$R = \int L(y, f(x))dF(x, y) \tag{1}$$

where $\int$ is the integral function, $F(x, y)$ is the joint distribution function, and $L(y, f(x))$ is the error (loss) function. Assume the fitting function is:

$$f = \omega \cdot \Phi(x) + b \tag{2}$$

where $\Phi(x)$ is the high-dimensional mapping function of sample data, ω is the weight vector corresponding to $\Phi(x)$ and it is normal to the hyperplane, and b is the bias factor. Because of the fitting error, the error accuracy ε and relaxation factor $\xi_i \geq 0$, $\xi_i^* \geq 0$ and $\xi_i \xi_i^* = 0$ are introduced, then the fitting problem is transformed into:

$$\min_{\omega,\xi,b} \frac{1}{2} \| \omega \|^2 + C \cdot \sum_{i=1}^{l} (\xi_i + \xi_i^*) \tag{3a}$$

$$\text{s.t.} \quad y_i - \omega \cdot \Phi(x_i) - b \leq \varepsilon + \xi_i \tag{3b}$$

$$\omega \cdot \Phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \tag{3c}$$

where (3b) and (3c) are the constraints of (3a), and C is the penalty coefficient, $i = 1, 2, \cdots, l$. By transforming it into a dual problem, the fitting function can be solved as:

$$f_{\substack{\min \\ \alpha,b}}(x) = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) \langle \Phi(x_i), \Phi(x_j) \rangle + b^* \tag{4a}$$

$$\text{s.t.} \quad \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) = 0, 0 \leq \alpha_i, \alpha_i^* \leq C \tag{4b}$$

where (4b) is the constraint of (4a), and $\langle \Phi(x_i), \Phi(x_j) \rangle$ is the inner product operation between $\Phi(x_i)$ and $\Phi(x_j)$.

2.2 Improved PSO Algorithm

PSO is a swarm intelligence algorithm that searches for the optimal solution. The updated speed of the traditional PSO algorithm is:

$$\begin{aligned} v(t+1) =& \varpi v(t) + c_1(q_{\text{best}}(t) - q(t)) \\ &+ c_2 r(p_{\text{best}}(t) - q(t)) \end{aligned} \tag{5a}$$

$$q(t+1) = q(t) + v(t+1) \tag{5b}$$

among them, $v(t)$ is the particle velocity; ϖ is the inertia weight; $q_{\text{best}}(t)$ is the optimal solution of particle at time t; $q(t)$ is the cross-validation accuracy at time t; $p_{\text{best}}(t)$ is the global optimal solution of all particles at time t; c_1, c_2 are learning factors; r is a random number of $0-1$.

Due to the parameters of traditional PSO algorithm are set in advance based on experience, such as the inertia weight ϖ and the learning factors c_1, c_2, which reduces the algorithm's optimization ability. By dynamically adjusting the inertia weight, improved PSO can dynamically adjust the search range, which adjusts the balance between global search and local search Tong et al. (2018). The dynamic weight is calculated as:

$$\varpi = \varpi_{\max} - \frac{\varpi_{\max} - \varpi_{\min}}{N_{\max}} \times N \tag{6}$$

where $\varpi_{\max}$ and $\varpi_{\min}$ are the maximum and minimum values of inertia weight; $N_{\max}$ is the maximum number of iterations; N is the number of current iterations.

In addition, asynchronous learning mode is introduced to dynamically adjust c_1 and c_2 Ding et al. (2019), and c_1, c_2 can be calculated as follows:

$$c_1 = c_{1,\max} - \frac{c_{1,\max} - c_{1,\min}}{N_{\max}} \times N \tag{7a}$$

$$c_2 = c_{2,\min} - \frac{c_{1,\min} - c_{1,\max}}{N_{\max}} \times N \tag{7b}$$

where $c_{1,\max}$, $c_{1,\min}$ are the maximum and minimum of learning factor c_1; $c_{2,\max}$, $c_{2,\min}$ are the maximum and minimum of learning factor c_2. In this way, c_1 has the maximum value and c_2 has the minimum value at the beginning of the operation of PSO.

At the same time, the improved PSO introduced a compression factor β to effectively control the particle's flight speed, so as to achieve an effective balance between global detection and local mining Adeli and Broumandnia (2017). The β is calculated as:

$$\beta = \frac{2}{\mid 2 - c_0 - \sqrt{c_0^2 - 4c_0} \mid} \tag{8}$$

where $c_0 = c_1 + c_2$. Therefore, the updated velocity and position of particle in improved PSO algorithm becomes:

$$\begin{aligned} v(t+1) =& \beta(\varpi v(t) + c_1(q_{\text{best}}(t) - q(t))) \\ &+ c_2 r(p_{\text{best}}(t) - q(t)) \end{aligned} \tag{9a}$$

$$q(t+1) = q(t) + v(t+1) \tag{9b}$$

2.3 Improved PSO-SVM Algorithm

The main idea of PSO is to use the accumulated flight experience of particles and learn from exchange information to change the direction and speed of the flight, so as to find the optimal solution Ghodousian and Parvari (2017). Due to the limitation of artificial parameter setting of traditional PSO algorithm, the ability of optimizing and following is reduced Wang et al. (2019). For example, if the inertia weight is too large or too small, the search range is unreasonable, which makes it unable to achieve rapid convergence and avoid falling into a local optimum. Therefore, PSO is improved to dynamically adjust the search range by dynamically adjusting inertia weight ϖ; and asynchronous learning is introduced to dynamically adjust the size of learning factor c_1, c_2 Pan et al. (2020c), which makes the information exchanged between particles more effective, and strengthens the self-learning ability of algorithm Hannan et al. (2017). Furthermore, the compression factor β is introduced to effectively control the flying speed of particles, thereby improving the balance between global mining and local mining.

Because the complexity and stability of the SVM model are greatly affected by the kernel function and hyperparameters, the improved PSO algorithm is embedded into the process of optimizing parameters of SVM model to dynamically find the optimal value of hyperparameters, so as to improve the model optimization speed and accuracy Liu et al. (2018). The modeling process of improved PSO-SVM algorithm is as follows:

Algorithm 1 [Improved PSO-SVM algorithm]

Step 1: initialize the parameters of the improved PSO algorithm: maximum number of iterations $N_{\max}$ and population size d;
Step 2: initialize the particle swarm speed v, and the SVM model is established with the corresponding vector of each particle and training data;
Step 3: using the minimum mean square error of validation set as the fitness function, calculate the fitness value of each particle using (6)(7)(8);
Step 4: update particle swarm speed v with (9);
Step 5: check whether the maximum number of iterations $N_{\max}$ is reached, if it is satisfied, output the optimal value of SVM parameters, and output the improved PSO-SVM soft-sensing model; if not, turn to Step 2.

The modeling based on improved PSO-SVM is different from neural network modeling. With the addition of new training samples, the fitness value corresponding to the support vector of improved PSO-SVM will dynamically adjust, and the number of support vectors will change in real time, which can achieve better modeling effectiveness and accuracy Ding et al. (2019).

3. SELECTION OF AUXILIARY VARIABLES

Due to various factors, the data collected on site will inevitably have errors, and the existence of errors will lead to data failure. Therefore, it is necessary to process the original data. In addition, the selection of auxiliary variables is a key step of soft sensor modeling. However, there are many optional variables and redundant information in data, so it is necessary to filter auxiliary variables. Here, we will introduce the data collection, preprocessing and variables primary, precise selection.

3.1 Data Collection and Variables Primary Selection

The research data is the actual production data of a power plant in Yulin City, Shaanxi Province, China. The collection period of the data is from January $1, 2019$ to July $16, 2019$, the sampling interval is 1 day, and a total of 197 groups of data are collected. Although the amount of data is small, the data sampling time interval is the same, and the time span is large, and because the daily coal of power plants changes very slowly, this data set is not less representative than the large data set. The collected data is mainly the proximate analysis data of coal quality, which includes: total moisture (M_t) (as-received basis moisture), air-dried basis moisture (M_{ad}), air-dried basis ash (A_{ad}), fixed carbon ($F_{c,ad}$), air-dried basis volatile matter (V_{ad}), bomb calorific value ($Q_{b,ad}$), as-received basis net calorific value (MJ/kg) ($Q_{net,ar}$), as-received basis ash (A_{ar}), dry basis sulfur ($S_{t,d}$), dry and ash-free basis volatile matter (V_{daf}), as-received basis sulfur ($S_{t,ar}$), air-dried basis sulfur ($S_{t,ad}$) and as-received basis net calorific value (cal/g) ($Q_{net,ar}$). In addition, these data are necessary for utilizing coal in power plants and are available in real time.

After obtaining the actual production data, because similar data items and directly related data items only different in dimension units or can be calculated from each other, they are should be eliminated and screened according to mechanism analysis Hao et al. (2011). Therefore, the one of two $Q_{net,ar}$, $F_{c,ad}$ and A_{ar} are eliminated. After removal, the data between M_t, M_{ad}, A_{ad}, V_{ad}, $S_{t,ad}$, $S_{t,ar}$, S_d, V_{daf}, and $Q_{b,ad}$ are independent, and can be used as the primary auxiliary variable for model establishment.

3.2 Data Preprocessing

Due to the low quality of data, the data preprocessing must be carried out. The data preprocessing mainly includes the processing of gross error, random error, and data normalization and anti-normalization.

1) The gross error processing adopts the 3δ criterion of statistical discriminant method. For the sampling data $x_1, x_2, \cdots, x_n$, the δ can be calculated as Duarte et al. (2018):

$$\delta = [\frac{{v_i}^2}{n-1}]^{\frac{1}{2}} \tag{10}$$

where n is the number of sampling data, v_i represents the deviation ($i = 1, 2, \cdots, n$) between sampling data value x_i and average value $\bar{x}$.

2) Using the characteristics of random error Duarte et al. (2018), the average filtering method is used for processing random error. Specifically, the sampling data x_i is calculated through:

$$x_i = \sum_{t=i-n}^{i+n} \frac{x_t}{2n} \tag{11}$$

where x_t represents the surrounding time data of sampling time, n is the number of x_t.

3) Due to the difference of collecting data objects, the magnitude of sampling data varies greatly. Therefore, the min-max method is used to normalize the sampling data. The specific steps of min-max method are as follows: find out the maximum value $x_{\max}$ and the minimum value $x_{\min}$, and then calculate the normalized sampling data X_i as follows Oliveira (2016):

$$X_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{12}$$

In addition, the anti-normalized data $[x_i]$ can be calculated as follows:

$$[x_i] = X_i(x_{\max} - x_{\min}) + x_{\min} \tag{13}$$

where $x_{\min}$ and $x_{\max}$ are the same as (12).

After data preprocessing, 193 groups of data were left. Because the soft sensor modeling requires training and testing model, and the sampling data is limited, it is necessary to maintain an appropriate quantity ratio between data sets Hao et al. (2011). Therefore, the first 150 groups of the sampling data are used as the training set, and the last 43 groups of sampling data as the test set, which guarantees the randomness between the data sets.

Table 1. Correlation between Auxiliary and Dominant Variables

No.	Auxiliary variables	Correlation coefficient	Chosen or not
1	air-dried basis ash (A_{ad})	0.9031	Y
2	bomb calorific value ($Q_{b,ad}$)	0.8598	Y
3	air-dried basis volatile matter (V_{ad})	0.6516	Y
4	air-dried basis moisture (M_{ad})	0.2412	Y
5	total moisture (M_t)	0.1859	Y
6	air-dried basis sulfur ($S_{t,ad}$)	0.1853	Y
7	dry and ash-free basis volatile matter (V_{daf})	0.1726	N
8	dry basis sulfur ($S_{t,d}$)	0.1714	N
9	as-received basis sulfur ($S_{t,ar}$)	0.1694	N

3.3 Variables Precise Selection

Based on the data preprocessing, the correlation analysis is performed using the covariance method. According to the size of correlation coefficient between auxiliary variable and target variable, the auxiliary variables with larger correlation coefficients are selected as precise auxiliary variables. The correlation coefficient *corr* can be calculated as follows:

$$\begin{aligned} corr(X,Y) = cov(X,Y) &= \frac{E(X-\mu_X)E(Y-\mu_Y)}{\delta_X\delta_Y} \\ &= \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \end{aligned} \tag{14}$$

the matrix form is as follows:

$$corrcoef(i,j) = \frac{cov(i,j)}{\sqrt{cov(i,i)\cdot cov(j,j)}} \tag{15}$$

The analysis of calculating results (see Table 1) shows that: there is a significant linear correlation between $Q_{b,ad}$, A_{ad}, V_{ad} and $Q_{net,ar}$, a correlation between M_t, M_{ad}, V_{daf} and $Q_{net,ar}$, and there is no obvious correlation between $Q_{net,ar}$ and $S_{t,ar}$, $S_{t,ad}$, $S_{t,d}$.

After sorting the correlation coefficients of auxiliary variables, because too many auxiliary variables will affect the flexibility and timeliness of models, and it is difficult to fully reflect the characteristic information of parameters with too few Wang et al. (2015), so the two larger ones of smaller correlation coefficients are selected. Therefore, the M_t, $Q_{b,ad}$, and M_{ad}, A_{ad}, V_{ad}, $S_{t,ad}$ are selected as precise auxiliary variables, the target variable is the $Q_{net,ad}$.

4. SOFT SENSOR FOR NET CALORIFIC VALUE OF COAL

When modeling, the reasonable initialization of parameters can train the model more effectively, the rationality of parameters directly affects the complexity and accuracy of models Hu and Ding (2019); and the inputs of the model are auxiliary variables, and the output is target variable. Therefore, in this section the initialization of parameters, simulation and debugging will be carried out, and briefly introduce the modeling process.

4.1 Model Training

In the regression modeling of improved PSO-SVM, the choice of kernel function and adjustment of parameters have a great influence on modeling effectiveness. At present, since there is no unified conclusion on the selection of kernel function Zhou et al. (2018), the RBF kernel function is selected through experimental verification and comparison of experimental results. Therefore, the hyper-parameters of improved PSO-SVM model are penalty factor c and parameters of RBF, and they are optimized by using the improved PSO Hu and Ding (2018). As mentioned in 2.3, only the parameters of improved PSO algorithm need to be initialized.

In the process of modeling and simulation, the parameter settings of LSTM neural network should follow the principle that: the number of iterations cannot be too small, because the number of iterations is related to convergence, but there is also an upper limit, and the maximum number of iterations is 5000. The learning rate should range from 0.0001 to 0.001, not be too large, and the number of hidden layers and visible layers cannot differ too much Ergen and Kozat (2018).

When setting the parameters of BP neural network, the parameters that affect the network structure are mainly considered, including the number of nodes in the hidden layer, learning rate and training accuracy. The number of nodes in the hidden layer depends on experience, too few nodes will affect the effectiveness of network, and too many nodes will greatly increase the training time. The learning rate usually set in the range of $0.01-0.09$, which is related to the training times. Generally speaking, the lower learning rate, the more training times. However, if the learning rate is too large, the less training times will affect the stability of network structure. The setting of training accuracy needs to be determined by the output requirements, the lower value of training accuracy, the higher accuracy of output Hasanjania and Sohrabia (2017).

During the debugging of simulation, the population size d of improved PSO-SVM is set to 50, the maximum number of iterations N_{max} is 200, the range of the penalty factor c is $[0.1, 100]$, and the range of kernel parameter g is $[0.01, 1000]$. After continuous debugging, the parameters of LSTM are determined as follows: the hidden layer is 12, learning rate is 0.0006, the number of iterations is 1600. The parameters of BP network are as follows: the hidden layer is 11, learning rate is 0.01, the maximum number of training is 10000, the training accuracy is 0.00001. The parameters of SVM are: the best penalty factor *bestc* is 16, and the best kernel parameter *bestg* is 0.0110. The parameters of improved PSO-SVM are: the best penalty factor *bestc* is 53.8804, and the best kernel parameter *bestg* is 996.1348. Refer to Table 2 for specific parameters.

4.2 Soft Sensor Modeling Algorithm

The soft sensor modeling process based on learning algorithm is as follows:

Algorithm 2 [Soft sensor modeling algorithm]

Step 1: according to the analysis of coal-fired power plant production process, the net calorific value of coal and proximate analysis data are collected;
Step 2: preprocess data according to (10) (11), normalize and anti-normalize data through (12) (13);

Table 2. Network parameters

Name	Hidden layers number	Learning rate	Input number	Output number	Training group number	Test group number	Iteration	Time step	Batch size	Goal error	Bestc	Bestg
LSTM	12	0.0006	6	1	150	43	1600	2	60	–	–	–
BP	11	0.01	6	1	150	43	10000	–	–	0.00001	–	–
SVM	–	–	6	1	150	43	–	–	–	–	16	0.011
Improved PSO-SVM	–	–	6	1	150	43	–	–	–	–	53.8804	996.1348

Step 3: calculate correlation coefficient by (14) (15), select the best as auxiliary variable, and divide data set into training set and test set;
Step 4: according to the requirements of model parameters, preliminarily select relevant parameters and build the model;
Step 5: train the model and optimize the parameters, such as: iteration, learning rate, hidden layers number, goal error;
Step 6: verify the model through the test set.

5. ANALYSIS AND COMPARISON OF SIMULATION RESULTS

The model of this experiment is built in the hardware environment of Intel (R) Core (TM) i7-9750H CPU, DDR4 8G memory and NVIDIA GeForce GTX 1650 graphics card. The LSTM network model is built in PyTorch 0.3.1 of Python framework, the editing environment is PyCharm 5.0.3, and the editing environment of SVM, BP network and improved PSO-SVM model is MATLAB 2019a.

5.1 Evaluation Parameters of Simulation Results

In this experiment, mean square error (MSE) and mean square correlation coefficient (R^2) are used as the evaluation criteria of models. The MSE is calculated as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (16)$$

where y_i is the real value, and $\hat{y}_i$ is the predicted value of the model.

R^2 is the ratio of regression sum of squares (SSR) to total sum of squares (SST). The value of R^2 is within the range of $(0, 1)$, and the closer it is to 1, the higher regression is fitting. The R^2 can be calculated as follows:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (17)$$

where SSE (error sum of squares) is the sum of residual squares, which has the relationship of $SSE + SSR = SST$ with the regression sum of squares and total sum of squares, the SSR, $SST$and SSE are calculated as follows:

$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2 \quad (18)$$

$$SST = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2 \quad (19)$$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (20)$$

where y_i is the sampling data, $\bar{y}_i$ is the mean value of sampling data, and $\hat{y}_i$ is the predicted data.

5.2 Simulation and Analysis of Improved PSO-SVM

Fig.2(a) is the prediction result based on SVM. It can be seen that most of the predicted values basically coincide with the real values, and there is almost no difference between them. Fig.2(b) shows the relative error based on SVM, it can be seen that the overall relative error is pretty small; and the absolute values of relative errors are taken in this paper, so they are all positive. However, the relative error of each point is different, with some changes. From these two figures, it can be found that the overall effectiveness of SVM model is pretty good, but the prediction accuracy of each group of data has deviation. In addition, during the simulation and debugging, the simulation results are relatively stable every time.

Fig.3(a) is the prediction results based on improved PSO-SVM. The predicted values of all points are basically identical with the real values, almost without errors. Moreover, the overall effectiveness of prediction is nearly perfect. Fig.3(b) is the relative error based on improved PSO-SVM. All points are stable at 0.1%, and the whole is a parallel straight line, which proves that the model's dynamic follow-up ability is strong and stable. From these two figures, it can be seen that the simulation results of two figures are better than those based on SVM. Moreover, the two evaluation parameters of model based on improved PSO-SVM are improved some extent, and there is basically no deviation in prediction accuracy.

From Fig.2, 3, it can be seen that the model based on SVM and improved PSO-SVM can obtain a small prediction error for the larger orders of magnitude test data, and the relative error of both is less than 0.3%, which fully meets the needs of industrial applications; and improved PSO-SVM improves the adaptability to data, and has a higher prediction accuracy and stability than the SVM-based model.

5.3 Model Comparison

Fig.4(a) shows the test results of LSTM network model, and the predicted value can be followed with the real value. However, there is a large error between the predicted value and real value. Fig.4(b) is the relative error of LSTM network model. The relative error of all points is bigger, the relative error value of each point is constantly changing, and the overall effectiveness is not as good as the improved PSO-SVM model.

Fig.5(a) is the test results of BP network model, the error between the predicted value and real value of BP network model is also larger, and the overall effectiveness is not very ideal. Fig.5(b) shows the relative error of BP network model, and the whole relative error value is also relatively

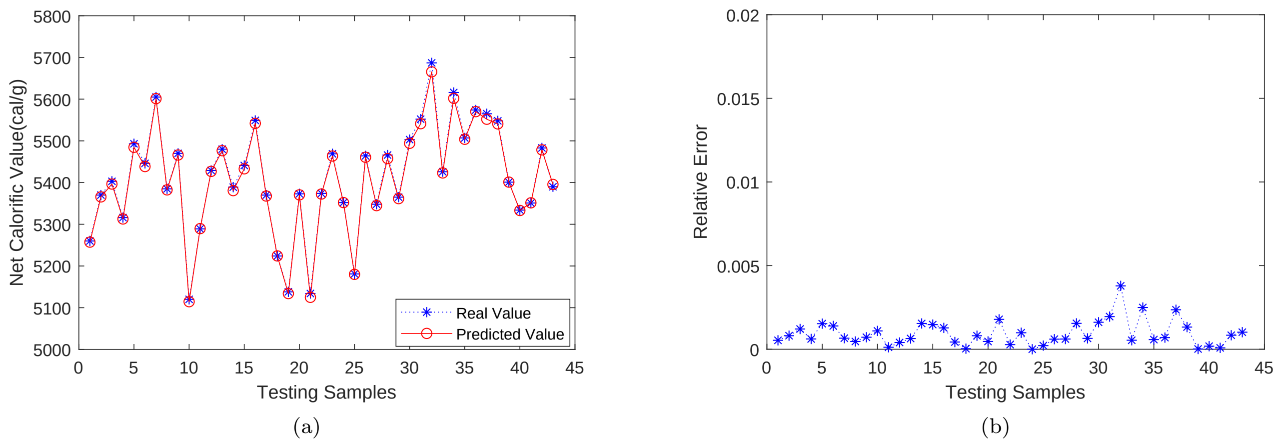


Fig. 2. Prediction results based on SVM: (a) prediction curves; (b) relative error.

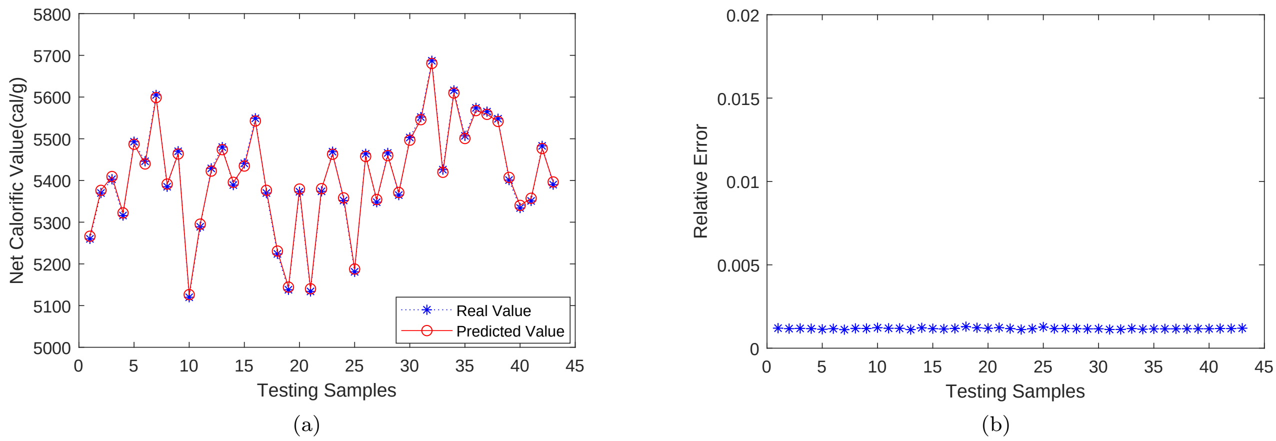


Fig. 3. Prediction results based on improved PSO-SVM: (a) prediction curves; (b) relative error.

large. Moreover, the relative error of each point is also constantly changing and the whole is a polylines.

The prediction results in Fig.4 and Fig.5 are the best results of each network in the debugging process of simulation, and the results are not ideal. In addition, during the simulation and debugging of BP and LSTM network models, whether it is training set or test set, the evaluation parameters MSE and R^2 of the two networks both have relatively large changes.

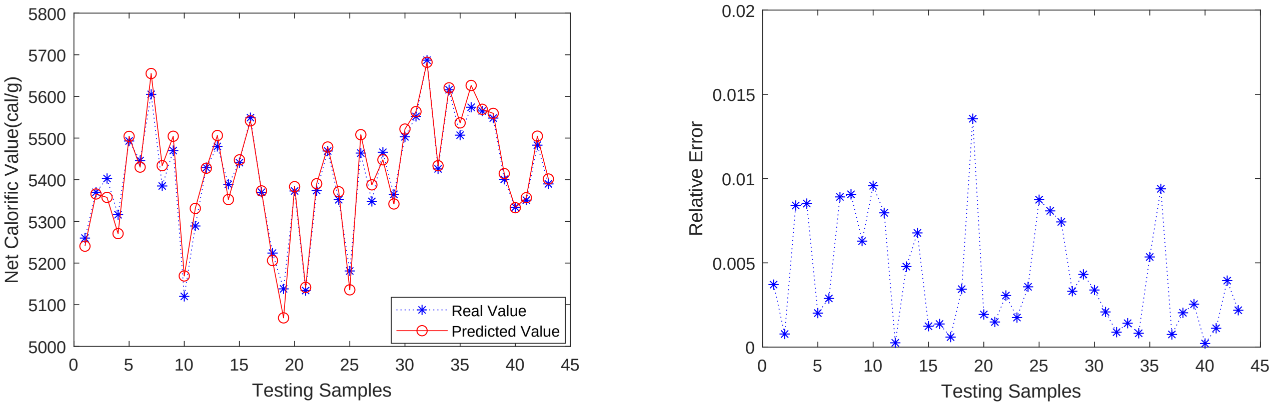
Table 3. Performance comparisons based on LSTM, BP network, SVM and improved PSO-SVM models

Soft sensor	Training		Test	
	MSE($*10^{-4}$)	R^2	MSE($*10^{-4}$)	R^2
BP	81.798	0.97557	93.567	0.97377
LSTM	45.163	0.96395	49.010	0.97549
SVM	9.4495	0.98910	1.0492	0.99902
Im PSO-SVM	0.98990	0.99959	1.0050	0.99897

Table 3 shows the performance comparisons of the four models, in which each parameter value is the corresponding parameter value in Fig.2, Fig.3, Fig.4 and Fig.5; and the effectiveness is improved from top to bottom. It can be clearly seen from the table that the parameters of both LSTM and BP are significantly different from the SVM and improved PSO-SVM models; overall, the performance of the improved PSO-SVM model is the best, and on the basis of SVM, improved PSO-SVM model improves the adaptability to larger sample data while ensuring the same accuracy. In addition, except for SVM, the training parameters of other three models are better than the test parameters; and the special reason for SVM is that its structure is more advantageous for small sample data.

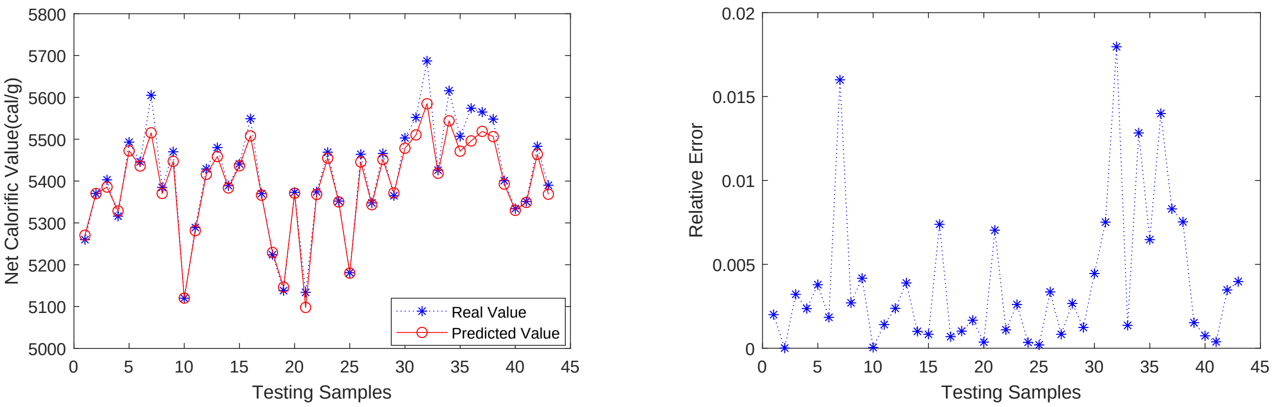
6. CONCLUSIONS

In this paper, the actual production data in thermal power plant are collected and preprocessed, and the auxiliary variables related to the as-received basis net calorific value of coal are found out through the correlation analysis. Without the prior knowledge of the boiler, the soft sensing model is established by using LSTM, BP network and SVM, improved PSO-SVM. Moreover, through the comparison of model evaluation parameters, we found that the dynamic following ability and accuracy of improved PSO-SVM model are better, and determine that the improved PSO-SVM model can be used as a reference model for the actual production. Therefore, this study can provide a theoretical basis for obtaining the reliable reference data



(a) (b)

Fig. 4. Prediction results based on LSTM: (a) prediction curves; (b) relative error.



(a) (b)

Fig. 5. Prediction results based on BP: (a) prediction curves; (b) relative error.

at the lower cost based on the existing equipments, so that the operation of boiler is more reliable and energy-saving.

REFERENCES

Adeli, A. and Broumandnia, A. (2017). Image steganalysis using improved particle swarm optimization based feature selection. *Applied Intelligence*, 48(4), 1–14.

Chen, J., Yang, H., Lv, M., Xiao, Z., and Wu, S. (2019). Estimation of monthly pan evaporation using support vector machine in Three Gorges Reservoir Area, China. *Theoretical and Applied Climatology*, 138, 1095–1107.

Chen, K., Liang, Y., Gao, Z., and Liu, Y. (2017). Just−in−time correntropy soft sensor with noisy data for industrial silicon content prediction. *Sensors*, 17(8), 1830–1842.

Czuszynski, K., Ruminski, J., and Kwasniewska, A. (2018). Gesture recognition with the linear optical sensor and recurrent neural networks. *IEEE Sensors Journal*, 18(13), 5429–5438.

Ding, S., Du, W., Zhao, X., Wang, L., and Jia, W. (2019). A new asynchronous reinforcement learning algorithm based on improved parallel PSO. *Applied Intelligence*, 49, 4211–4222.

Duarte, J., Goosen, R., Lawry, P., and Blackburn, J. (2018). PMA: protein microarray analyser, a user−friendly tool for data processing and normalization. *BMC Research Notes*, 11(1), 156–162.

Ergen, T. and Kozat, S.S. (2018). Efficient online learning algorithms based on LSTM neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8), 3772–3783.

Eze, O. and Oti, W.O. (2017). Characterization of selected nigerian indigenous biomass wastes for their suitability in biofuel production. *International Journal of Pharmaceutical Science Invention*, 6(9), 1–8.

Ghodousian, A. and Parvari, M. (2017). A modified PSO algorithm for linear optimization problem subject to the generalized fuzzy relational inequalities with fuzzy constraints (FRI−FC). *Information Sciences*, 418, 317–345.

Hannan, M.A., Akhtar, M., Begum, R., Basri, H., Hussain, A., and Scavino, E. (2017). Capacitated vehicle−routing problem model for scheduled solid waste collection and route optimization using PSO algorithm. *Waste Management*, 71, 31–41.

Hao, Z., Wang, Q., Zhu, J., and Zhang, X. (2011). Influence of sample data preprocessing on BP neural network−based GPS elevation fitting. *Journal of Geodesy & Geodynamics*, 31(2), 125–128.

Hasan, E., Jahiding, M., Mashuni, Ilmawati, W., Wati, W., and Sudiana, I. (2017). Proximate and the calorific value analysis of brown coal for high−calorie hybrid briquette application. *Journal of Physics Conference Series*, 846(1), 012022–012029.

Hasanjania, H. and Sohrabia, M. (2017). Artificial neural networks (ANN) for the simultaneous spectrophotometric determination of fluoxetine and sertraline in pharmaceutical formulations and biological fluid. *Iranian Journal of Pharmaceutical Research (IJPR)*, 16(2), 478–489.

Hu, J. and Ding, B. (2018). Dynamic output feedback predictive control with one free control move for Takagi−Sugeno model with bounded disturbance. *IEEE Transactions on Fuzzy Systems*, 27(3), 462–473.

Hu, J. and Ding, B. (2019). Output feedback robust MPC for linear systems with norm−bounded model uncertainty and disturbance. *Automatica*, 108. https://doi.org/10.1016/j.automatica.2019.07.002.

Huang, L., Ruan, C., Huang, W., Shi, Y., Peng, D., and Ding, W. (2018). Wheat powdery mildew monitoring based on GF−1 remote sensing image and relief−mRMR−GASVM model. *Transactions of the Chinese Society of Agricultural Engineering*, 34(15), 167–175.

Khoshnevisan, B., Rafiee, S., Shamahirband, S., Iqbal, J., Omid, M., Anuar, N., and Wahid, A. (2015). A comparative study between artificial neural networks and adaptive neuro−fuzzy inference systems for modeling energy consumption in greenhouse tomato production − A case study in Isfahan province. *Journal of Agricultural Science & Technology*, 17(1), 49–62.

Lauri, D., Kamyar, R., Hashemian, N., Mehdizadeh, H., and Moshgbar, M. (2017). Moisture soft sensor for batch fluid bed dryers: A practical approach. *Powder Technology*, 326, 69–77.

Li, Y. and Ren, J. (2017). Study on soft−sensors of boiler flue gas oxygen content based on PSO−LSSVM. *Computer Simulation*, 34(11), 58–62.

Liu, H., Tan, H., Xiong, X., Yao, L., Niu, Y., Liu, Y., and Xu, T. (2013). *Prediction of calorific value of coal using real power plant data*. Cleaner Combustion and Sustainable World.

Liu, Z., Wei, H., Liu, K., and Zhong, Q. (2018). Global identification of electrical and mechanical parameters in PMSM drive based on dynamic self−learning P-SO. *IEEE Transactions on Power Electronics*, 33(12), 10858–10871.

Lu, Z., Mo, J., Yao, S., and Lu, J. (2017). Rapid determination of the gross calorific value of coal using laser−induced breakdown spectroscopy coupled with artificial neural networks and genetic algorithm. *Energy & Fuels*, 31(4), 3849–3855.

Mohanta, S., Meikap, B., and Chakraborty, S. (2015). Impact of coal quality on thermal power plant savings: a case study for an Indian thermal power plant. *South African Journal of Chemical Engineering*, 20(2), 56–68.

Nawaz, A., Arora, A., Yun, C., Cho, H., You, S., and Lee, M. (2019). Data authorization and forecasting by proactive soft sensing tool−anammox based process. *Industrial & Engineering Chemistry Research*, 58(22), 9552–9563.

Nugaliyadde, A., Wong, K., Sohel, F., and Xie, H. (2019). Language modeling through long−term memory network. In *2019 International Joint Conference on Neural Networks (IJCNN)*.

Oliveira, J. (2016). Statokinesigram normalization method. *Behavior Research Methods*, 49(1), 1–8.

Pan, H., Mi, W., Lei, X., and Zhong, W. (2020b). A closed−loop BMI system design based on the improved SJIT model and the network of Izhikevich neurons. *Neurocomputing*, 401, 271–280.

Pan, H., Mi, W., Song, H., and Liu, F. (2020a). A universal closed−loop brain−machine interface framework design and its application to a joint prosthesis. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-020-05323-6.

Pan, H., Mi, W., Wen, F., and Zhong, W. (2020c). An adaptive decoder design based on the receding horizon optimization in BMI system. *Cognitive Neurodynamics*, 14(3), 281–290.

Pan, H., Su, T., Huang, X., and Wang, Z. (2021). LSTM-based soft sensor design for oxygen content of flue gas in coal-fired power plant. *Transactions of the Institute of Measurement and Control*, 43(1), 78–87.

Pan, Q. and Dias, D. (2017). An efficient reliability method combining adaptive support vector machine and monte carlo simulation. *Structural Safety*, 67, 85–95.

Peng, X. and Chen, D. (2018). PTSVRs: Regression models via projection twin support vector machine. *Information Sciences*, 435, 1–14.

Selvi, B., Kalpana, D., and Thyagarajan, T. (2017). Modeling and prediction of boiler drum in a thermal power plant. In *2017 Trends in Industrial Measurement and Automation (TIMA)*. IEEE.

Tong, L., Li, X., Hu, J., and Ren, L. (2018). A PSO optimization scale−transformation stochastic−resonance algorithm with stability mutation operator. *IEEE Access*, 6, 1167–1176.

Wang, P., Feng, X., Li, W., Ping, X., and Yu, W. (2019). Robust RHC for wheeled vehicles with bounded disturbances. *International Journal of Robust and Nonlinear Control*, 29(7), 2063–2081.

Wang, X. and Liu, H. (2018). Soft sensor based on stacked auto−encoder deep neural network for air preheater rotor deformation prediction. *Advanced Engineering Informatics*, 36, 112–119.

Wang, Z., He, Q., and Wang, J. (2015). Comparison of variable selection methods for PLS−based soft sensor modeling. *Journal of Process Control*, 26, 56–72.

Xiong, X., Qian, S., Sheng, K., He, Y., Fang, L., and Wu, Z. (2017). Predicting gross calorific value of agricultural feedstock based on proximate/ultimate analysis and visible−near infrared spectroscopy. *Spectroscopy & Spectral Analysis*, 37(5), 1622–1627.

Yan, W., Tang, D., and Lin, Y. (2017). A data−driven soft sensor modeling method based on deep learning and its application. *IEEE Transactions on Industrial Electronics*, 64(5), 4237–4245.

Yang, J., Ye, L., Zhao, M., and Zhao, H. (2017). NO_x prediction model for multi−coal admixing combustion in boiler. *Journal of Chinese Society of Power Engineering*, 37(11), 870–875.

Zhou, Z., Zhang, R., Wang, Y., Zhu, Z., and Zhang, J. (2018). Color difference classification based on optimization support vector machine of improved grey wolf algorithm. *Optik − International Journal for Light and Electron Optics*, 170, 17–29.