IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 13, 2020, accepted December 29, 2020, date of publication January 1, 2021, date of current version January 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3048714

A Novel Stacked Generalization Ensemble-Based Hybrid PSVM-PMLP-MLR Model for Energy Consumption Prediction of Copper Foil Electrolytic Preparation

ZHIMING LIAO[1], MEI SU[2], (Member, IEEE), GUANGFU NING[2], (Member, IEEE), YONGLU LIU[2], (Member, IEEE), TONG WANG[3], AND JIE ZHOU[3]

[1]School of Computer Science and Engineering, Central South University, Changsha 410083, China
[2]School of Automation, Central South University, Changsha 410083, China
[3]Anhui Tongguan Copper Foil Group Company Ltd., Chizhou 247000, China

Corresponding author: Guangfu Ning (ningguangfu@163.com)

This work was supported in part by the First Batch of Science and Technology Implementation Plan in 2020, and in part by the Copper Nonferrous Holding Plan under Grant 157-2020.

ABSTRACT At present, the energy consuming during the electrolytic copper foil preparation accounts for more than 75% of the total energy consumption. In real-life production, the process parameters are set by the operator empirically and the system may not work at the operating point with minimum energy consumption. Therefore, it is critical to establish an effective model for predicting electrolysis energy consumption to guide the parameters design. In this paper, a novel hybrid model (named PSVM-PMLP-MLR) based on stacked ensemble learning is proposed. The model is divided into two parts: the base-learning model and the meta-learning model. The support vector machine (SVM) model and multilayer perceptron (MLP) model with different input structures are established by the former first. Then the particle swarm algorithm is employed to determine the optimal value of SVM parameters and the optimal weight of MLP by minimizing the mean absolute percentage error (MAPE). The multiple linear regression (MLR) is finally employed as a meta-learning machine to compute the final predictions. Experimental results show that the regression coefficient of this model reached 0.987, and compared with the traditional SVM and MLP models, the accuracy of the model is improved by 10.29% and 8.28%, respectively.

INDEX TERMS Ensemble learning, electrolytic preparation of copper foil, energy consumption, machine learning.

I. INTRODUCTION

With the continuous development of 5G, industrial intelligence, and new energy vehicles, etc., the demand for copper-clad laminate (CCL) and printed circuit board (PCB) is growing rapidly. As the basic electronic material, electrolytic copper foil is not only indispensable but also influences the conductivity of circuits and the interconnection of electronic components in producing CCL and PCB. However, during the whole production process, the electrolytic preparation consumes more than 75% of energy consumption. It is imperative to reduce this energy consumption to save cost.

The associate editor coordinating the review of this manuscript and approving it for publication was Firooz B. Saghezchi.

The energy consumption of electrolysis is proportional to the cell voltage and is inversely proportional to the current efficiency. And electrolytic copper foil is produced at low copper ion concentration and high current density. The former makes the cathode undergo hydrogen evolution, resulting in a decrease in current efficiency. And the later leads to high cell voltage. The combined effects of these two aspects lead to the high energy consumption. In real-life production, the process parameters are set by the operator by virtue of experience. Hence, the energy consumption is not under the specific control. Due to it is difficult to establish a mathematical model to describe the functional relationship between the energy consumption of electrolysis and the process parameters, most researches try to reduce energy consumption by restructuring

the electrolytic cell [1], [2], developing new cathode materials [3], [4] and ion exchange membranes [5], etc.

As suitable materials are hard to find and their costs are high, the consumption is hard to be further reduced by modifying hardware equipment. To address this issue, a soft method to reduce the consumption is proposed in this paper which aims at finding a set of process parameters that can minimize the consumption while ensuring production quality. The key of the method is to get an accurate and reliable energy consumption prediction model.

In the past decade, the machine learning (ML) methods have been adopted to predict complex nonlinear processes [6]–[8]. However, there are two difficult points in this study:

(i) It is difficult to obtain a large amount of data.

(ii) During the whole production process of electrolytic copper foil, the process parameters are characterized by large number, strong coupling and strong nonlinearity.

Support vector machine (SVM), as one of the most classic ML algorithms, is famous for its ability to efficiently solve nonlinear problems with small samples [9], [10]. It has been widely used in regression, classification, nonlinear function approximation and some other analysis [11]–[13]. Related research shows that SVM can be adopted for high-precision energy consumption prediction [14]. Therefore, considering the first difficult point (i), it is expected to use SVM to obtain a higher performance energy consumption prediction model.

Inspired by a mathematical model of biological neural networks, the artificial neural networks (ANN) represents a model that imitates the work way of the brain. It is characterized by strong robustness, memory, nonlinear mapping ability and self-learning ability. As one of the commonly utilized prediction algorithms in ANN, the multilayer perceptron (MLP) algorithm stands out in solving various prediction problems because of its strong fitting ability [15], [16]. Therefore, the MLP is adopted for the second difficult point (ii).

However, problems often occur with these methods, such as the sensitivity of parameter, local optimum and over-fitting phenomenon. As result, it is difficult to determine which one gives the better prediction. Hence, researchers have been working on improving and optimizing these methods, and the ensemble learning provides some new insight. Ensemble learning algorithm overcomes over-fitting and initialization sensitivity by combining homogeneous or heterogeneous ML models, improving the performance of the ML model, which has attracted great attention [17], [18]. In recent years, ensemble learning has been proved to be reliable in improving the performance of the ML model, and it has been applied in many practical projects [19]–[23]. Generally, better prediction results of base-learning machines lead to better counterparts of the ensemble learning model. Therefore, in order to further improve the prediction performance of the base-learning machine, the traditional SVM and MLP needs to be optimized. For SVM, the hyper-parameter selection will affect its performance to a certain extent. The selection of the kernel function, penalty coefficient C and ε-insensitive loss parameter is the key to get a SVM prediction model with superior prediction performance [24], [25]. In the MLP algorithm, hyper-parameter setting is also crucial. Among the hyper-parameters, the number of nodes in the hidden layer and the initial weight are the most important, leading to different prediction results. An optimization algorithm should be adopted to optimize the hyper-parameters of these two types of traditional models to further improve their prediction performance.

The grid algorithm [26], genetic algorithm [27], ant colony algorithm [28] and particle swarm optimization (PSO) are the most popular optimization algorithms. However, each has some limitations. The grid algorithm is time-consuming, computationally expensive and has low learning accuracy. And the genetic algorithm method is complex as different crossover or mutation is needed in different issues [29]. While the search time of ant colony algorithm method is long, and it is easy to lead to local best solution. The PSO has been proved with extensive capability of global optimization for its fast convergence and easy implementation, and it is widely used in various engineering applications [30]–[32]. Its successful application in function minimization [33], [34] and ANN design [35]–[38] proves the promising future of PSO algorithm. It offsets the adverse effects of the MLP algorithm of the feed-forward neural network [39] (such as the slow convergence speed in training and tendency to a local minimum) due to it does not need gradient information and differentiable information.

To the author's knowledge, no one has adopted ML method to predict the energy consumption of electrolytic copper foil preparation before. From the perspective of ML, a novel hybrid model is proposed which based on a stacked generalization ensemble to predict the energy consumption during the preparation process of electrolytic copper foil. The hybrid model takes SVM and MLP optimized by PSO as a base-learning machine and multiple linear regression (MLR) model as a meta-learning machine. This hybrid prediction model is called PSVM-PMLP-MLR.

II. METHODS AND THEORIES

Six ML models have been introduced in this part, including SVM, MLP, MLR, SVM and MLP based on PSO optimization and PSVM-PMLP-MLR hybrid model. 75% of the production data of electrolytic copper foil were utilized as the training data while the remaining 25% as the test data.

A. SVM ALGORITHM

SVM is one of the most classic ML algorithms, and one of the most robust and accurate methods in many data mining algorithms [40]. In order to solve the linear regression problems in feature space, SVM maps the input data X into high-dimensional feature space Q through nonlinear mapping. In such way, regression approximation solves the problem of estimation function based on a given dataset $X = \{(x_i, y_i)\}_i^n$, where x_i is the input vector, y_i is the desired value, and n is the total amount of dataset. In SVM, the regression

function is expressed as:

$$f(x) = \omega\phi(x) + b \tag{1}$$

where ω and b are the model parameters, and $\phi(x)$ is the feature vector after mapping in the input space x.

The regularization function involving the sum of empirical risk and complexity term $||\omega||^2/2$ is minimized to avoid overfitting. The coefficients ω and b can be estimated by minimizing the regularization risk function.

$$\begin{aligned} &Min\ ||\omega||^2/2 \\ &s.t. \begin{cases} y_i - \phi(\omega, x_i) \le \varepsilon \\ \phi(\omega, x_i) + b - y_i \le \varepsilon \end{cases} \end{aligned} \tag{2}$$

The constraint equation can be expressed as:

$$\begin{aligned} &Min\ \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{n} \xi + \xi^* \\ &s.t. \begin{cases} y_i - \langle \omega, x_i \rangle - b \leqslant \varepsilon + \xi_i \xi_i \geqslant 0 \\ \langle \omega, x_i \rangle + b - y_i \leqslant \varepsilon + \xi_i^* \xi_i^* \geqslant 0 \end{cases} \end{aligned} \tag{3}$$

where the constant C represents the penalty coefficient of the sample, whose error exceeds ε. ξ and ξ^* are two positive slack variables, referring to the distance from the actual value to the corresponding boundary value of the ε-tube.

Optimization methods can be adopted to maximize the function. In this way, the dual problem is derived:

$$\begin{aligned} &Max \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \\ &\quad -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} (\alpha_i - \alpha_i^*)(\alpha_i - \alpha_i^*)K(x_i, y_i) \\ &s.t. \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) = 0\ and\ 0 \leqslant \alpha_i, \alpha_i^* \leqslant C \end{aligned} \tag{4}$$

where α_i and α_i^* are Lagrange multipliers.

Via the above-mentioned maximization function, the SVM for function fitting obtained is thereby expressed as:

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)K(x_i, x) + b \tag{5}$$

In (5), those sampling points with non-zero coefficients are called support vectors. The kernel function $K(x_i, x_j)\phi(x_j) = \phi(x_i)\phi(x_j)$ satisfies Mercer's condition and performs nonlinear mapping.

B. MLP ALGORITHM

Many types of algorithms are included in the neural network architecture which are suitable for various application scenarios with unique properties [41], [42]. Among the various ANN, MLP is a popular feed-forward ANN.

At present, there are several algorithms to train MLP networks: quick back propagation (QPROP), back propagation (BP), quasi-newton back propagation (BFGS) and resilient back propagation (RPROP). The architecture of MLP is shown in Fig. 1.

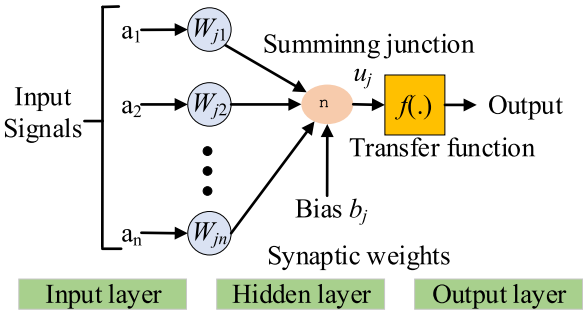

FIGURE 1. General structure of MLP feeds forward neural network.

Each product of the input element (a_i) and the weight (w_{ij}) is input into the summing junction. Meanwhile, the bias of neuron (b_j) is utilized for summation, as shown below:

$$X = (\sum_{i=1}^{n} w_{ij}a_i) + b_j \tag{6}$$

X generates the output through the transfer function F:

$$F(X) = u_j = F[(\sum_{i=1}^{n} w_{ij}a_j) + b_j] \tag{7}$$

Sigmoid, Tansig and Logsig are common activation functions in hidden layers. The sigmoid function is the most used nonlinear activation function, whose output ranges from 0 to 1. The sigmoid activation function can be expressed as:

$$F(X) = \frac{1}{1 + e^{-X}} \tag{8}$$

During the training period, the training algorithm will adjust the weight and biases to minimize the error of neural network model by iteration. The input dataset is trained by learning algorithms. The training process repeats until the error is small enough for acceptance.

This architecture has many advantages, such as more hidden layers and neurons and higher accuracy. However, the number of hyper-parameters is relatively high, and the setting of hyper-parameters affects the prediction performance of the model.

C. PSO ALGORITHM

As a population-based algorithm, PSO imitates group behavior for cooperative learning. Therefore, it leads to better results and faster convergence in global search [43].

The population consists of n particles. The position of the particle i in the D-dimensional space is expressed as a vector $X_i = (x_{i1}, x_{i2}, \ldots, x_{iD})$, and the flying speed is expressed as a vector $V_i = (v_{i1}, v_{i2}, \ldots, v_{iD})$. Each particle can be regarded as a search individual in the D-dimensional search space. The current position of the particle is a candidate solution of the corresponding optimization problem. The flight process of the particle is the search process of the individual. The flying

speed of particles can be dynamically adjusted according to the historical optimal position of the particle and the population. The optimal position of the i-th particle searched so far is called the individual extreme value, denoted as $P_i = (p_{i1}, p_{i2}, \ldots, p_{iD})$.The optimal position searched by the entire particle swarm so far is the global extreme value, denoted as $P_g = (p_{g1}, p_{g2}, \ldots, p_{gD})$. In each iteration, the particle updates itself by tracking two extreme values. This process repeats until the maximum of iterations is set or the best fit condition is reached. The speed and position of particles can be updated using as:

$$v_{id}^{l+1} = wv_{id}^{l} + c_1 rd_1^{l}(p_{id}^{l} - x_{id}^{l}) + c_2 rd_2^{l}(p_{gd}^{l} - x_{id}^{l}) \quad (9)$$

$$x_{id}^{l+1} = v_{id}^{l+1} + x_{id}^{l} \quad (10)$$

where $i = (1,2, \ldots, n)$, $d = (1,2, \ldots, D)$, w is the inertia weight coefficient, a positive number, which adjusts the overall optimization ability. c_1 and c_2 are learning factors; rd_1^l and rd_2^l are positive random numbers in the range of [0,1] under normal distribution; l is the number of the iterations; and x_{id}^l refers to the position of particle i in D-dimensional space. $v_{id} \in [v_{max}, v_{min}]$ represents the velocity of particle i in the D-dimensional space, which determines the direction and distance of the next generation of particles.

Inertia weight w is used to control the influence of previous speed history on current speed. Larger inertia weight value is beneficial to global exploration, while smaller counterparts contribute to local exploration. To reach more balance between the two capabilities, the linear decreasing inertia weight is employed. Generally, $w(k)$ decreases linearly with each iteration from w_{start} to w_{end}. It can be expressed as:

$$w(k) = w_{start} - k \times \frac{w_{start} - w_{end}}{T_{\max}} \quad (11)$$

where w_{start} and w_{end} represent the maximum and minimum values of w, respectively, k represents the current iteration number, and T_{max} represents the maximum iteration number.

D. ENERGY CONSUMPTION PREDICTION MODEL OF ELECTROLYTIC COPPER FOIL PREPARATION BASED ON PSVM

In the SVM regression model, the input data is the five features that affect the energy consumption of electrolytic copper foil, and the corresponding energy consumption is taken as the output. In this way, the correlation coefficient of the regression function can be determined via (5). When the energy consumption of electrolytic copper foil production is predicted, the SVM will present the corresponding energy consumption value according to the determined regression function after the input of the relevant eigenvector. RBF, sigmoid, poly, and linear are the kernel functions that are commonly used in SVM. In order to determine the kernel function with the best prediction performance, ten times cross-validation (CV) method is used to verify the kernel function according to the training dataset. The experimental results are listed in Table 1.

TABLE 1. Result of the trial simulation based on different training algorithms in SVM.

Training algorithm	Estimation index			
	MAE (kW h t^{-1})	RMSE (kW h t^{-1})	MAPE (%)	R^2
RBF	42.1375	59.3210	4.3261	0.9047
Sigmoid	128.7124	146.7375	12.3815	0.5956
Poly	68.92437	80.7642	6.7443	0.8118
Linear	57.4675	66.9346	5.5636	0.8729

Based on the results in Table 1, the Gaussian kernel function RBF is employed to construct SVM regression model. The PSO algorithm not only converges quickly, but also has strong global optimization capabilities. Therefore, in order to determine the penalty term C and ε-insensitive loss parameter that can enable SVM to obtain better prediction results, the PSO method is adopted. Meanwhile, the MAPE (mean absolute percentage error) of the predicted data is utilized as the fitness function. The flow chart is shown in Fig. 2.

PSO algorithm parameters: particle swarm size at 150, 200 iteration times, learning factor $c_1 = c_2 = 2$, penalty term C rang [1-10], and ε-insensitive loss parameter ranging from [0.1-0.0001].

The training process of the PSVM algorithm is shown in Fig. 3. As can be seen, when PSO iterations reach about 155 steps, the fitness value of the best particle tends to be stable while the value of MAPE is 3.736. In this case, penalty term C and ε-insensitive loss parameter are 6.2683 and 0.001327, respectively. It can be seen in Table 1 that the MAPE of traditional SVM with RBF kernel function is 4.3261. Therefore, the MAPE value is reduced by 13.64%.

E. ENERGY CONSUMPTION PREDICTION MODEL OF ELECTROLYTIC COPPER FOIL PREPARATION BASED ON PMLP

In order to select the best basic training algorithm for MLP network, ten times CV method is adopted. The results are shown in Table 2.

TABLE 2. Result of the trial simulation based on different training algorithms in MLP.

Kind of training algorithm	Estimation indexes			
	MAE (kW h t^{-1})	RMSE (kW h t^{-1})	MAPE (%)	R^2
QPROP	71.7324	82.2664	7.0378	0.8134
BP	100.5425	114.8432	9.8522	0.7067
BFGS	39.9624	53.4425	3.7970	0.9118
RPROP	58.1342	68.5054	5.6945	0.8627

Based on the results, BFGS is employed as the kernel function in this paper. In order to determine the optimal number

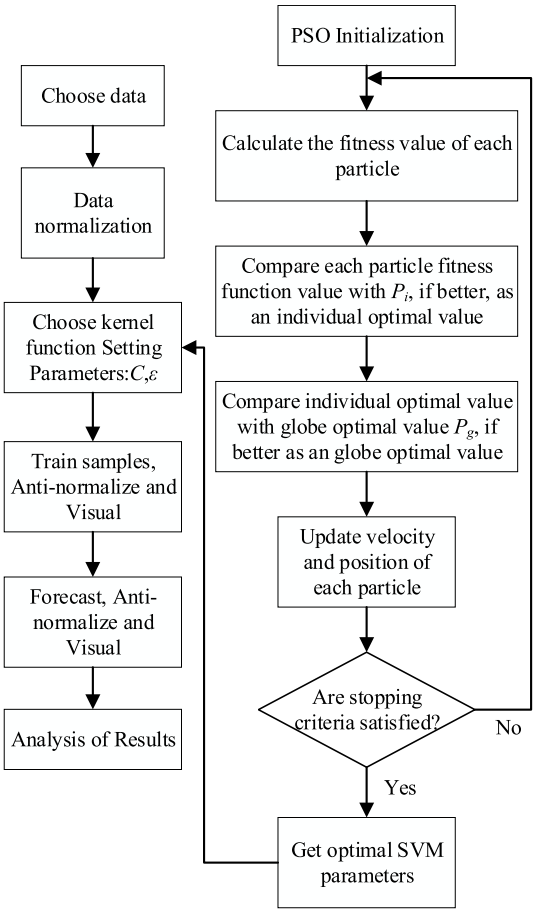

FIGURE 2. The flow diagram of the PSVM optimization.

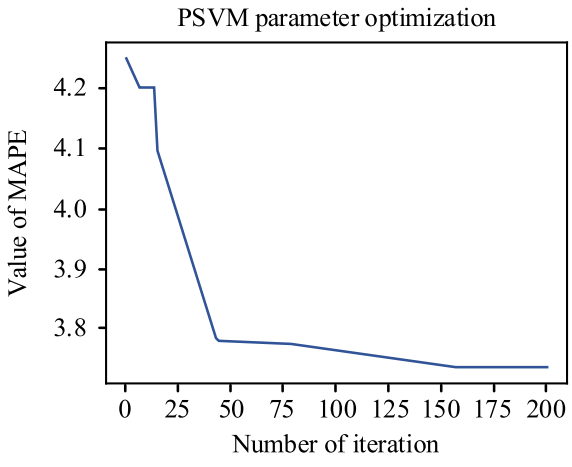

FIGURE 3. The optimization process of the PSVM.

of hidden neurons in the model, a second calculation was performed using samples from the dataset. Finally, according to the minimum MAPE value, thirteen neurons are obtained from models with the best performance.

However, in addition to the structure of the neural network, the weight of each layer is another key parameter that affects the final performance. MLP model tends to have the local minimum error function value in the training process, which leads to the low prediction accuracy of the network. Therefore, the PSO algorithm is utilized to search the best fitness of MLP prediction and get the corresponding particles. The particle itself is used as the initial weight of the MLP.

The main steps of PMLP hybrid calculation are as:

1) Initialize the structure of MLP. Meanwhile, the number of neurons in input and output layers should be set together with network parameters. Via trial calculation, the optimal number of neurons in hidden layer could be calculated.

2) Set the operation parameters of PSO algorithm and initialize the position and speed of particles randomly in a manner. After knowing the size and the individual dimension of particles, map the particles to the corresponding weights.

3) According to the MAPE value, the fitness value of each particle is calculated to find the individual and global extremum.

4) The velocity, position and fitness of each particle should be updated until they meet the conditions. Then, individual and global extremum will be updated according to the new fitness value.

5) The optimal individual position is regarded as the weight of MLP, which is used to train and predict the research object.

PSO algorithm parameters: the particle swarm size is 300, the number of iterations is 200, the learning factor is $c_1 = c_2 = 2$, and the overall weight range is set to $[-2, 2.5]$.

The training process of PMLP algorithm is as shown in Fig. 4, which shows that when PSO iterations reach about 125 steps, the fitness value of the best particle tends to be stable with the MAPE value of 3.24, which has a 14.59% reduction compared with that of traditional MLP whose MAPE is 3.797 (it can be seen in Table 2).

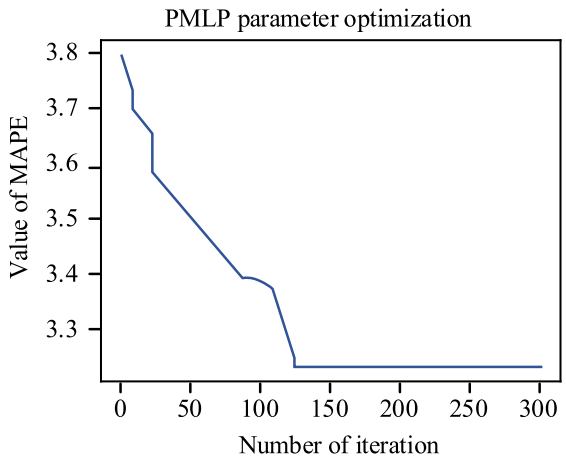

FIGURE 4. The optimization process of the PMLP algorithm.

F. MULTIPLE LINEAR REGRESSION (MLR) MODEL

MLR model is one of the most used estimation models on ML because of its simple structure, calculation, and interpretation. According to the multiple linear regression model, there is a linear relationship between dependent variable y and

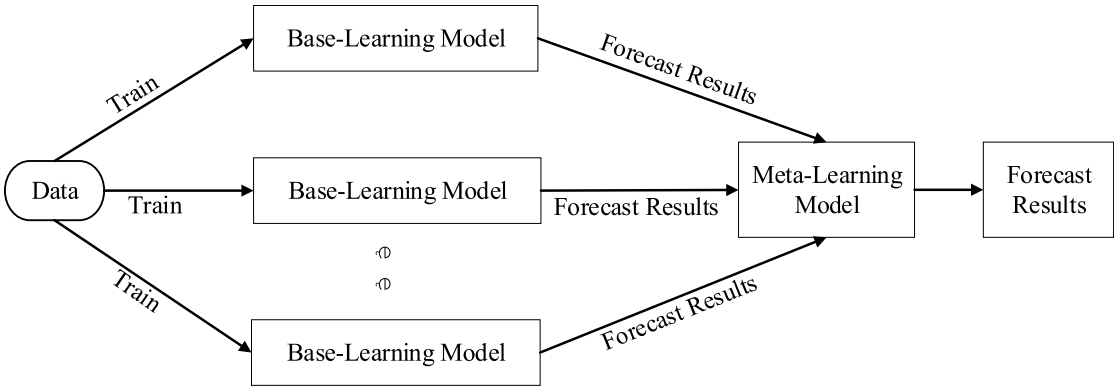

FIGURE 5. A stacked ensemble learning algorithm.

independent variables $x_1, x_2, \ldots, x_n$. MLR is expressed as:

$$Y = a + \sum_{i=1}^{n} b_i x_i + \varepsilon = b_1 x_1 + b_2 x_2 + \cdots + b_i x_i + \varepsilon \quad (12)$$

where Y is the output, x_i ($i = 1, 2, 3, \ldots, n$) is the input variable, a is the intercept, b_i is the regression coefficient, and ε is the error term. In this paper, x_i is the predicted value of PSVM and PMLP model, and Y refers to the final prediction result.

G. STACKED ENSEMBLE LEARNING ALGORITHM

Stacked ensemble learning [44] is a heterogeneous integration strategy. Heterogeneous sets enhance the generalization ability of the strong classifier through integrating several base-learning machines into a strong machine. Stacked ensemble learning algorithm adopts a two-tier framework structure, as shown in Fig. 5. In this paper, the base-learning model is the SVM model and MLP model based on PSO optimization, while the meta-learning model is the MLR model.

The training process is as:

1) The multiple base-learning machines are trained.

2) The prediction results of multiple base-learning machines are employed as the input of meta-learning machines. Afterward, the training should be conducted again.

3) The final ensemble algorithm uses the learning ability of the base-learning machines and meta-learning machine to improve the accuracy of prediction.

The effectiveness of stacked ensemble learning algorithm depends on two factors. One is the prediction result of base-learning machine. Generally, better prediction results of base-learning machines lead to better counterparts of the ensemble learning model. The other is that there are differences between the two base-classifiers as different factors of the machines need to be considered. The stacked ensemble learning algorithm is expressed as:

input: Train Dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$;
Primary learning algorithm $\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_T$
Secondary learning algorithm $\mathcal{L}$
process:
1: *for* $t = 1, 2, \ldots, T$ *do*
2: $h_t = \mathcal{L}_t$ (D);
3: *end for*
4: $D' = \emptyset$
5: *for* $i = 1, 2, \ldots, m$ *do*
6: *for* $t = 1, 2, \ldots, T$ *do*
7: $z_{it} = h_t(x_i)$;
8: *end for*
9: $D' = D' \cup ((z_{i1}, z_{i2}, \ldots, z_{iT}), y_i)$;
10: *end for*
11: $h' = \mathcal{L}(D')$;
Output: $H(x) = h'(h_1(x), h_2(x), \ldots, h_T(x))$

III. CASE STUDY AND SIMULATION RESULTS

A. DATA SOURCES AND INTRODUCTION

1500 pieces of experimental data are employed in this paper, which are provided by Anhui Tongguan Copper Foil Group Co., Ltd of China. These data are exported from the database of the Group and have been verified as valid data.

The experiment plan is roughly divided into the following steps:

(i) According to the data, analyze the importance of related factors affecting energy consumption.

(ii) Normalize the data.

(iii) Among the 1500 experimental data, 75% data are utilized for training, while the remaining 25% are used to predict the results.

The data include six dimensions, respectively being electrolyte temperature, Cu^{2+} concentration, H_2 SO_4 concentration, current density, electrode spacing and power consumption per ton of copper. The unit, maximum and minimum of each index is shown in Table 1. In addition, the information entropy gain of each feature in these data is calculated and normalized to evaluate their importance. As shown in Fig. 6, where A, B, C, D, and E represent Electrolyte temperature, Cu^{2+} concentration, H_2SO_4 concentration, Inter-electrode spacing, and Current density, respectively. More specific score values of the importance are shown in Table 3.

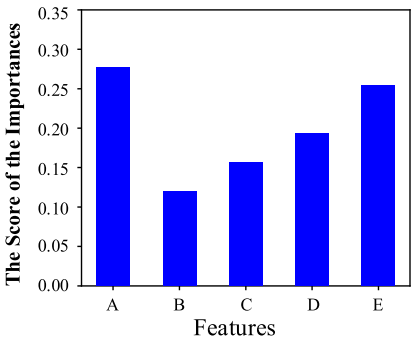

FIGURE 6. Importance score of features.

TABLE 3. Statistical parameters pertaining to each of the attributes of dataset.

Attribute	Unit	Min	Max	Score of features
Electrolyte temperature	℃	35	65	0.27660574
Cu^{2+} concentration	g L^{-1}	7	13	0.12091947
H_2SO_4 concentration	g L^{-1}	110	140	0.15529462
Inter-electrode spacing	mm	28	42	0.19493133
Current density	A m^{-2}	1350	2000	0.25224883
Power consumption	kW h t^{-1}	619.5	1661.5	/

B. DATA PREPROCESSING

1) DATA NORMALIZATIONS

In terms of prediction based on data, it is necessary to normalize training data and test data, which is aimed to avoid the attribute in larger numerical range dominating that in smaller range. In addition, normalization helps avoid numerical difficulties in the calculation process. For a group of electrolytic copper foil production data $\{x_{1k}, x_{2k}, \ldots, x_{ik}\}$ can be normalized to $\{X_{1k}, X_{2k}, \ldots, X_{ik}\}$ according to (16).

$$X_{ik} = \frac{x_{ik} - x_k^{\min}}{x_k^{\max} - x_k^{\min}} \tag{13}$$

where X_{ik} is the scaling value, x_{ik} is the original value, $x_k^{\max}$ is the maximum of feature k in the dataset, and $x_k^{\min}$ is the minimum of feature k in the dataset.

2) CONSTRUCTION OF TRAINING SAMPLES

For a group of energy consumption training data of electrolytic copper foil production, the training sample set is constructed and expressed as:

$$X = \begin{bmatrix} b_1 & b_2 & \ldots & b_m \\ b_2 & b_3 & \ldots & b_{m+1} \\ \vdots & \vdots & & \vdots \\ b_{n-m} & b_{n-m+1} & \ldots & b_{n-1} \end{bmatrix}, \quad Y = \begin{bmatrix} b_{m+1} \\ b_{m+2} \\ \vdots \\ b_n \end{bmatrix} \tag{14}$$

where X is the input vector, Y is the output vector, and m is the dimension of the input vector. In this paper, X refers to the electrolyte temperature, Cu^{2+} concentration, H_2SO_4 concentration, current density, and electrode spacing. Meanwhile, Y is the power consumption per ton of copper. Table 3 summarizes the six attributes in the dataset (five inputs and one output) and the importance of the feature score.

C. PERFORMANCE EVALUATION OF ML MODEL

1) PERFORMANCE EVALUATION INDEX

For quantitative measurement of the prediction performance (relative to test set) of ML model, four accuracy evaluation indexes were selected: MAE (mean absolute error), RMSE (root mean square error), MAPE (mean absolute percentage error), and R^2 (regression coefficient). The formulas for estimating these errors are shown as (15) - (18).

$$MAE: \frac{1}{n}\sum_{i=1}^{n}|yi - \widehat{yi}| \tag{15}$$

$$RMSE: \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - \widehat{yi})^2} \tag{16}$$

$$MAPE: \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{yi - \widehat{yi}}{yi}\right| \tag{17}$$

$$R^2: \frac{\sum_{i=1}^{n}[(yi - \bar{y}_i)(yi - \widehat{y}_i)]^2}{\sqrt{\sum_{i=1}^{n}(yi - \bar{y}_i)^2}.\sqrt{\sum_{i=1}^{n}(yi - \widehat{y}_i)^2}} \tag{18}$$

where $\hat{y}$ is the predicted values, y *is* actual values, and n is the total number of data records in the test dataset.

(i) MAE: As the most basic evaluation method, it is usually used as a reference for the other three methods to compare the pros and cons.

(ii) RMSE: It is mainly adopted to compare the stability of different prediction models due to its characteristics that are easily affected by large deviations.

(iii) MAPE: The overall prediction accuracy of the model can be more intuitively reflected due to its calculation method in percentage form and is not easily affected by large deviations.

(iv) R^2 is often utilized to evaluate how well the model fits the true value.

The higher the values of MAE, RMSE and MAPE, the worse the prediction performance of the model. The value range of R^2 is [0-1], and the closer the value is to 1, the better the fitting optimization effect of the model is.

Theoretically, all of the above four indicators can be employed as a measurement of prediction accuracy in the statistical field. In this study, it is expected to obtain a prediction model with higher accuracy. Generally, the higher the accuracy of a model, the higher its stability. Compared with the other three evaluation indicators, MAPE can not only reflect the accuracy of the model more comprehensively, but also more intuitively reflect the performance of the model due to its percentage expression.

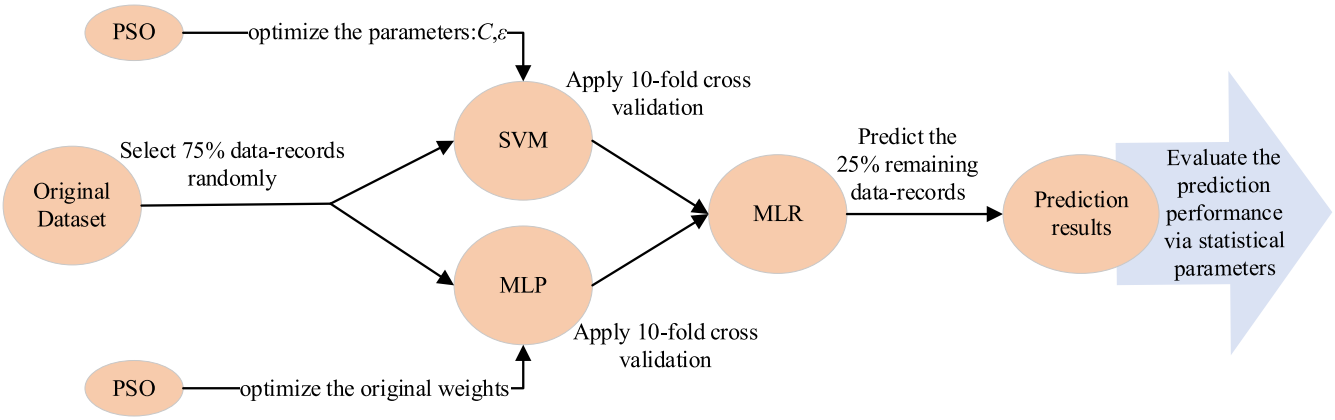

FIGURE 7. Flowchart for training and testing process of PSVM-PMLP-MLR models.

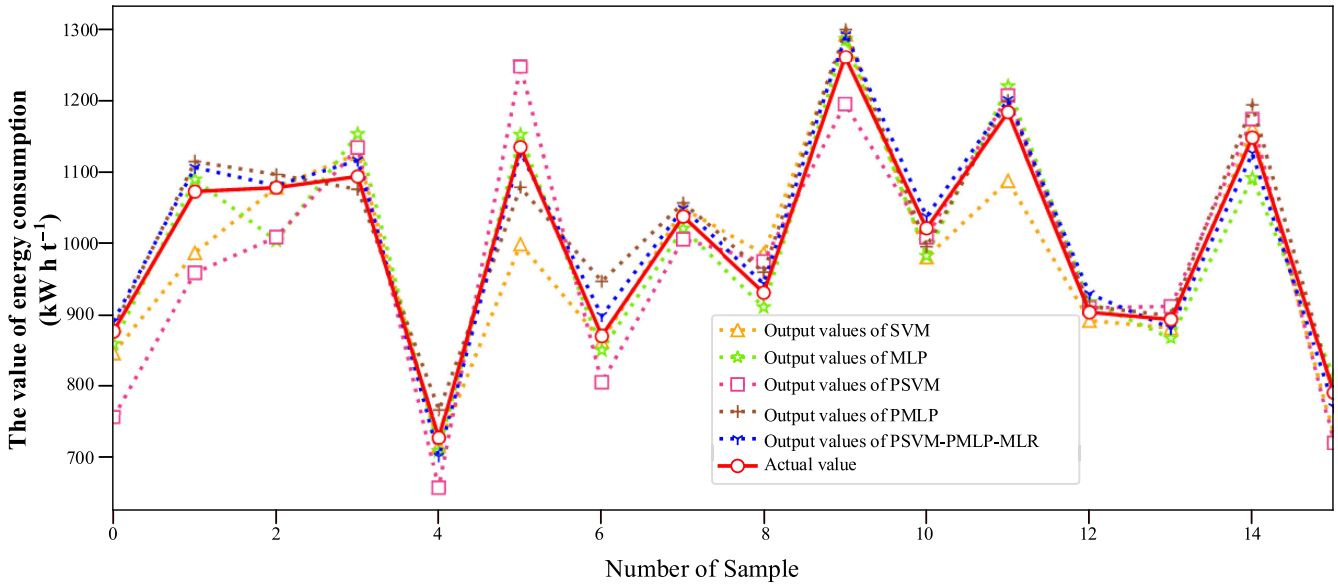

FIGURE 8. Comparison curves between the prediction value and actual value of each model.

Fig. 7 describes the overall training and testing process of the ML model mentioned in this paper. As can be seen, after 75% of the original data is processed, the SVM and MLP based on PSO optimization are trained in a 10-fold cross-validation method. Then, the results obtained by PSVM and PMLP are utilized as input data and passed into the MLR model to obtain the final prediction results.

2) PREDICTION RESULTS AND COMPARATIVE ANALYSIS

Considering that too much data will lead to the problem of insufficient graphics clarity, a small part of the test data is randomly selected so that the effect differences of the five models can be observed more clearly and intuitively, and the results are shown in Fig. 8. To further compare the performance of the five models, a detailed comparison based on the four evaluation indicators as mentioned above is conducted. The comparison results are listed in Table 4. Based on the test data, the residual curve of each model is shown in Fig. 9. The following can be learned from Table 4 and Fig. 8:

1) Compared with the traditional SVM model and MLP model, the MAE and RMSE values of PSVM and PMLP models are lower, while R^2 is closer to 1. From Fig. 9, the predicted values of PSVM and PMLP models are closer to the real values when compared with those of SVM and MLP. Therefore, they have better prediction performance.

2) The residual curve fluctuation of PSVM-PMLP-LR is the slightest among these models. The MAE, RMSE and MAPE of this model are lower than those PSVM and PMLP. Especially its MAPE value is 65.25%, 60.27%, 59.76% and 53.47% less than that of the other four models respectively. Besides, its R^2 reaches 0.9876, indicating that the predicted value of the model is close to the actual one.

In summary, the MAE, RMSE, MAPE, and R^2 of PSVM and PMLP are better than what their corresponding

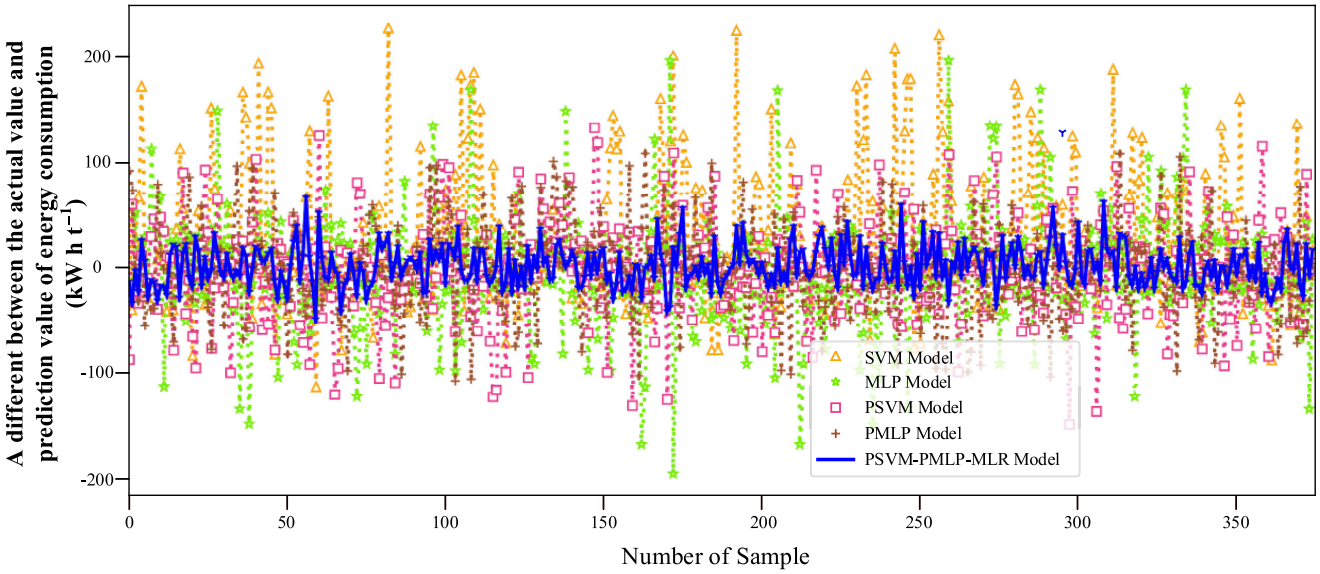

FIGURE 9. Comparison of predicted residual curves of each model.

TABLE 4. Comparison of error indicators of each prediction model.

ML model	Estimation indexes			
	MAE (kW h t^{-1})	RMSE (kW h t^{-1})	MAPE (%)	R^2
SVM	44.5623	63.4502	4.3299	0.8954
MLP	39.1253	52.0346	3.7864	0.9121
PSVM	39.1320	52.4625	3.7391	0.9193
PMLP	32.7452	45.4682	3.2335	0.9377
PSVM-PMLP-MLR	15.4325	19.2046	1.5045	0.9876

base-learning machines, which shows that the performance of the base-learning machine can be improved by adopting PSO. Thereafter, the performance of each base-learning machine will be balanced by the ensemble learning machine based on the strategy of stacked-ensemble generalization, as well as the generalization, stability and accuracy of the prediction model.

IV. CONCLUSION

In this paper, 1500 energy consumption data of electrolytic copper foil production were provided by Anhui Tongguan Copper Foil Group Co., Ltd of China for training and testing ML model to achieve accurate prediction of energy consumption in the electrolytic copper foil preparation. To further improve the accuracy, a novel PSVM-PMLP-MLR hybrid model is proposed which based on stacked ensemble learning.

Six kinds of ML models are introduced, namely SVM, MLP, PSVM, PMLP, MLR and PSVM-PMLP-MLR. Comparison and analysis have been done on the performance of these ML models in energy consumption prediction in the process of electrolytic copper foil preparation according to four error indexes: R^2, MAE, RMSE and MAPE. The results show that the performance of traditional SVM and MLP models can be improved by PSO optimization. Therefore, SVM and MLP after PSO optimization are combined as the base-learning machine and MLR as a meta-learning machine. They were combined through stacked ensemble strategy to improve the prediction performance of ML model. The R^2 of the proposed PSVM-PMLP-MLR hybrid model can reach 0.9876. Compared with any individual ML model in this experiment, it showed stronger predictive ability.

According to our research, no one has adopted ML method to predict the energy consumption of electrolytic copper foil preparation before. From the perspective of ML, an effective way to predict the energy consumption of the electrolytic copper foil preparation is proposed in this paper, which is of great significance for energy saving and consumption reduction in the production process.

Compared with traditional SVM and MLP, PSVM-PMLP-MLR is inevitably accompanied by some limitations even if it has better generalization ability, stability and prediction accuracy, such as the increased complexity of the model, resulting in higher computing cost and time consumption. Therefore, PSVM-PMLP-MLR is more suitable for solving the time insensitive prediction problem, but not suitable for short-term real-time online prediction. In future work, further improvement of the performance of the model is considered. In addition, extending the model to other industrial electrolysis based on transfer learning has been considered, such as electrolytic aluminum, electrolytic zinc and so on.

REFERENCES

[1] W. T. Xia, X. Y. Xiang, W. Q. Yang, and J. G. Yin, "Effect of flow pattern on energy consumption and properties of copper powder in the electrolytic process," *Solid State Phenomena*, vol. 279, pp. 77–84, Aug. 2018.

[2] Á. Hernández-Gómez, V. Ramirez, and D. Guilbert, "Investigation of PEM electrolyzer modeling: Electrical domain, efficiency, and specific energy consumption," *Int. J. Hydrogen Energy*, vol. 45, no. 29, pp. 14625–14639, May 2020.

[3] Y. Tang, Y. Li, Y. Sun, J. Wang, Y. Chen, X. Yang, and P. Wan, "Energy-saving electrolysis of sodium carbonate with a silver nanoparticles/carbon oxygen reduction cathode," *Electrochemistry Commun.*, vol. 27, pp. 108–111, Feb. 2013.
[4] S. Lu, B. Zhao, M. Chen, L. Wang, X.-Z. Fu, and J.-L. Luo, "Electrolysis of waste water containing aniline to produce polyaniline and hydrogen with low energy consumption," *Int. J. Hydrogen Energy*, vol. 45, no. 43, pp. 22419–22426, Sep. 2020.
[5] J. H. Lee, I. K. Park, D. Duchesne, L. Chen, C. H. Lee, and J. H. Lee, "Saline water electrolysis system with double-layered cation exchange membrane for low-energy consumption and its application for CO_2 mineralization," *J. CO2 Utilization*, vol. 41, Oct. 2020, Art. no. 101269.
[6] S. Ch, N. Anand, B. K. Panigrahi, and S. Mathur, "Streamflow forecasting by SVM with quantum behaved particle swarm optimization," *Neurocomputing*, vol. 101, pp. 18–23, Feb. 2013.
[7] H. Melo and J. Watada, "Gaussian-PSO with fuzzy reasoning based on structural learning for training a neural network," *Neurocomputing*, vol. 172, pp. 405–412, Jan. 2016.
[8] Z. Yang, M. Mourshed, K. Liu, X. Xu, and S. Feng, "A novel competitive swarm optimized RBF neural network model for short-term solar power generation forecasting," *Neurocomputing*, vol. 397, pp. 415–421, Jul. 2020.
[9] S.-W. Fei, M.-J. Wang, Y.-B. Miao, J. Tu, and C.-L. Liu, "Particle swarm optimization-based support vector machine for forecasting dissolved gases content in power transformer oil," *Energy Convers. Manage.*, vol. 50, no. 6, pp. 1604–1609, Jun. 2009.
[10] R. Gang and Z. Zhuping, "Traffic safety forecasting method by particle swarm optimization and support vector machine," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10420–10424, Aug. 2011.
[11] F. Anifowose, J. Labadin, and A. Abdulraheem, "Improving the prediction of petroleum reservoir characterization with a stacked generalization ensemble model of support vector machines," *Appl. Soft Comput.*, vol. 26, pp. 483–496, Jan. 2015.
[12] H. Sharafi, I. Ebtehaj, H. Bonakdari, and A. H. Zaji, "Design of a support vector machine with different kernel functions to predict scour depth around bridge piers," *Natural Hazards*, vol. 84, no. 3, pp. 2145–2162, Dec. 2016.
[13] I. Ebtehaj and H. Bonakdari, "A support vector regression-firefly algorithm-based model for limiting velocity prediction in sewer pipes," *Water Sci. Technol.*, vol. 73, no. 9, pp. 2244–2250, May 2016.
[14] A. S. Ahmad, M. Y. Hassan, M. P. Abdullah, H. A. Rahman, F. Hussin, H. Abdullah, and R. Saidur, "A review on applications of ANN and SVM for building electrical energy consumption forecasting," *Renew. Sustain. Energy Rev.*, vol. 33, pp. 102–109, May 2014.
[15] I. Ebtehaj and H. Bonakdari, "Bed load sediment transport estimation in a clean pipe using multilayer perceptron with different training algorithms," *KSCE J. Civil Eng.*, vol. 20, no. 2, pp. 581–589, 2015.
[16] H. Bonakdari, H. Moeeni, I. Ebtehaj, M. Zeynoddin, A. Mahoammadian, and B. Gharabaghi, "New insights into soil temperature time series modeling: Linear or nonlinear?" *Theor. Appl. Climatol.*, vol. 135, nos. 3–4, pp. 1157–1177, 2019.
[17] Y. Yang and J. Jiang, "Adaptive bi-weighting toward automatic initialization and model selection for HMM-based hybrid meta-clustering ensembles," *IEEE Trans. Cybern.*, vol. 49, no. 5, pp. 1657–1668, May 2019.
[18] M. J. van der Laan, E. C. Polley, and A. E. Hubbard, "Super learner," *Stat. Appl. Genet. Mol. Biol.*, vol. 6, no. 1, p. 1309, Jan. 2007.
[19] H. M. Gomes, J. P. Barddal, F. Enembreck, and A. Bifet, "A survey on ensemble learning for data stream classification," *ACM Comput. Surveys*, vol. 50, no. 2, pp. 1–36, Jun. 2017.
[20] B. Krawczyk, L. L. Minku, J. Gama, J. Stefanowski, and M. Woźniak, "Ensemble learning for data stream analysis: A survey," *Inf. Fusion*, vol. 37, pp. 132–156, Sep. 2017.
[21] F. A. Anifowose, J. Labadin, and A. Abdulraheem, "Ensemble machine learning: An untapped modeling paradigm for petroleum reservoir characterization," *J. Petroleum Sci. Eng.*, vol. 151, pp. 480–487, Mar. 2017.
[22] F. A. Anifowose, J. Labadin, and A. Abdulraheem, "Ensemble model of non-linear feature selection-based extreme learning machine for improved natural gas reservoir characterization," *J. Natural Gas Sci. Eng.*, vol. 26, pp. 1561–1572, Sep. 2015.
[23] F. Anifowose, J. Labadin, and A. Abdulraheem, "Ensemble learning model for petroleum reservoir characterization: A case of feed-forward backpropagation neural networks," in *Trends and Applications in Knowledge Discovery and Data Mining*. Berlin, Germany: Springer, 2013, pp. 71–82.
[24] X. Liang, H. Zhang, J. Xiao, and Y. Chen, "Improving option price forecasts with neural networks and support vector regressions," *Neurocomputing*, vol. 72, nos. 13–15, pp. 3055–3065, Aug. 2009.
[25] C.-C. Liu and K.-W. Chuang, "An outdoor time scenes simulation scheme based on support vector regression with radial basis function on DCT domain," *Image Vis. Comput.*, vol. 27, no. 10, pp. 1626–1636, Sep. 2009.
[26] X. Wang, J. Wen, S. Alam, X. Gao, Z. Jiang, and J. Zeng, "Sales growth rate forecasting using improved PSO and SVM," *Math. Problems Eng.*, vol. 2014, pp. 1–13, Jan. 2014.
[27] E. Avci, "Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using hybrid genetic algorithm–support vector machines: HGASVM," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 1391–1402, Mar. 2009.
[28] Y. Li, J. Xia, S. Zhang, J. Yan, X. Ai, and K. Dai, "An efficient intrusion detection system based on support vector machines and gradually feature removal method," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 424–430, Jan. 2012.
[29] X. Wang, J. Wen, Y. Zhang, and Y. Wang, "Real estate price forecasting based on SVM optimized by PSO," *Optik*, vol. 125, no. 3, pp. 1439–1443, Feb. 2014.
[30] Y. Shi and R. C. Eberhart, "Empirical study of particle swarm optimization," in *Proc. Congr. Evol. Comput.*, Jul. 1999, pp. 1945–1950, doi: 10.1109/CEC.1999.785511.
[31] D. W. Boeringer and D. H. Werner, "Particle swarm optimization versus genetic algorithms for phased array synthesis," *IEEE Trans. Antennas Propag.*, vol. 52, no. 3, pp. 771–779, Mar. 2004.
[32] A. Subasi, "Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders," *Comput. Biol. Med.*, vol. 43, no. 5, pp. 576–586, Jun. 2013.
[33] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw.*, 1995, pp. 1942–1948, doi: 10.1109/ICNN.1995.488968.
[34] A. Ratnaweera, S. K. Halgamuge, and H. C. Watson, "Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients," *IEEE Trans. Evol. Comput.*, vol. 8, no. 3, pp. 240–255, Jun. 2004.
[35] J. Salerno, "Using the particle swarm optimization technique to train a recurrent neural model," in *Proc. IEEE Int. Conf. Tools Artif. Intell.*, Nov. 1997, pp. 45–49, doi: 10.1109/TAI.1997.632235.
[36] W. Z. Lu, H. Y. Fan, and S. M. Lo, "Application of evolutionary neural network method in predicting pollutant levels in downtown area of Hong Kong," *Neurocomputing*, vol. 51, pp. 387–400, Apr. 2003.
[37] Y. Da and G. Xiurun, "An improved PSO-based ANN with simulated annealing technique," *Neurocomputing*, vol. 63, pp. 527–533, Jan. 2005.
[38] D. W. Boeringer and D. H. Werner, "A comparison of particle swarm optimization and genetic algorithms for a phased array synthesis problem," in *Proc. IEEE Antennas Propag. Soc. Int. Symp.*, Jun. 2003, pp. 181–184, doi: 10.1109/APS.2003.1217430.
[39] J. Yu, S. Wang, and L. Xi, "Evolving artificial neural networks using an improved PSO and DPSO," *Neurocomputing*, vol. 71, nos. 4–6, pp. 1054–1060, Jan. 2008.
[40] J. Rutledge, "The top ten algorithms in data mining," *J. Qual. Technol.*, vol. 41, no. 4, p. 441, 2009.
[41] M. Q. Raza and A. Khosravi, "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings," *Renew. Sustain. Energy Rev.*, vol. 50, pp. 1352–1372, Oct. 2015.
[42] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *Int. J. Forecasting*, vol. 32, no. 3, pp. 914–938, Jul. 2016.
[43] F. Tai and H.-T. Lin, "Multilabel classification with principal label space transformation," *Neural Comput.*, vol. 24, no. 9, pp. 2508–2542, Sep. 2012.
[44] C. Sigauke and D. Chikobvu, "Peak electricity demand forecasting using time series regression models: An application to south African data," *J. Statist. Manage. Syst.*, vol. 19, no. 4, pp. 567–586, Jul. 2016.

ZHIMING LIAO was born in Hunan, China, in 1997. He received the B.S. degree from the School of Computers Science and Engineering, Southwest Forestry University, Kunming, China, in 2019. He is currently pursuing the M.S. degree with the School of Computer Science and Engineering, Central South University, Changsha, China. His current research interests include machine learning and data mining.

MEI SU (Member, IEEE) was born in Hunan, China, in 1967. She received the B.S. and M.S. degrees in industrial automation and the Ph.D. degree in control science and engineering from the School of Information Science and Engineering, Central South University, Changsha, China, in 1989, 1992, and 2005, respectively. She has been a Full Professor with the School of Automation, Central South University. Her research interests include matrix converter, adjustable speed drives, and wind energy. She is currently an Associate Editor of the IEEE TRANSACTIONS ON POWER ELECTRONICS.

GUANGFU NING (Member, IEEE) was born in Jiangxi, China, in 1992. He received the B.S. degree in electrical engineering from the University of Shanghai for Science and Technology (USST), Shanghai, in 2014, and the Ph.D. degree in electrical engineering from Southeast University (SEU), Nanjing, in 2020. Since 2020, he has been a Lecturer with the School of Automation, Central South University, Changsha, China. His research interests include the full bridge converters, soft-switching technique, and high-power high-voltage DC/DC converters in MVDC collection systems.

YONGLU LIU (Member, IEEE) was born in Chongqing, China, in 1989. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from Central South University, Changsha, China, in 2012, 2015, and 2017, respectively. He has been an Associate Professor with the School of Automation, Central South University. His research interests include power electronics and renewable energy power conversion systems.

TONG WANG is currently the Deputy General Manager with Anhui Tongguan Copper Foil Group Company Ltd., in charge of production.

JIE ZHOU is currently the Head of the third and fourth workshop of Anhui Tongguan Copper Foil Group Company Ltd., involved in production management.