Research Article

Hybrid Soft Computing Schemes for the Prediction of Import Demand of Crude Oil in Taiwan

Yuehjen E. Shao,[1] Chi-Jie Lu,[2] and Chia-Ding Hou[1]

[1] *Department of Statistics and Information Science, Fu Jen Catholic University, Xinzhuang, New Taipei 24205, Taiwan*
[2] *Department of Industrial Management, Chien Hsin University of Science and Technology, Zhongli,
Taoyuan 32097, Taiwan*

Correspondence should be addressed to Chia-Ding Hou; stat0002@mail.fju.edu.tw

Crude oil is the most important nonrenewable energy resource and the most key element for the world. In contrast to typical forecasts of oil price, this study aims at forecasting the demand of imported crude oil (ICO). This study proposes different single stage and two-stage hybrid stages of forecasting models for prediction of ICO in Taiwan. The single stage forecasting modeling includes multiple linear regression (MLR), support vector regression (SVR), artificial neural networks (ANN), and extreme learning machine (ELM) approaches. While the first step of the two-stage modeling is to select the fewer but more significant explanatory variables, the second step is to generate predictions by using these significant explanatory variables. The proposed two-stage hybrid models consist of integration of different modeling components. Mean absolute percentage error, root mean square error, and mean absolute difference are utilized as the performance measures. Real data set of crude oil in Taiwan for the period of 1993–2010 and twenty-three associated explanatory variables are sampled and investigated. The forecasting results reveal that the proposed two-stage hybrid modeling is able to accurately predict the demand of crude oil in Taiwan.

1. Introduction

Natural resources are often classified into two groups: renewable and non-renewable resources. Renewable energy is the one which comes from natural resources such as sunlight, wind, and rain, and it is naturally replenished. It is reported that about 16% of global final energy consumption comes from renewable energy. The share of renewable energy in global electricity generation is around 19% [1]. In the United States, renewables provided 12.7% of total domestic electricity in 2011, up from 10.2% in 2010, and 9.3% in 2009. In China, wind power generation increased by more than 48.2% in 2011. In European Union, renewables accounted for more than 71% of total electric capacity additions in 2011 [2]. Non-renewable resources form very slowly or do not naturally form in the environment. A good example of the nonrenewable is fossil fuels. Although many studies have reported that the renewable sources of energy are currently receiving considerable attention, the fossil fuels are still the most needed element of the world energies [3]. In particular, crude oil is the most

important nonrenewable energy resource and the most key element for the world [4, 5].

The relationship between economic growth and oil consumption was addressed [6], and the study determined that the minimum statistical (lower-bound) annual oil consumption for developed countries was 11 barrels per capita. By using autoregressive distributed lag (ARDL) bounds testing approach of cointegration, the study discussed a long-run relationship among quantity of crude oil import, income, and price of the imported crude in India [7]. The results showed that the long-term income elasticity of imported crude in India is 1.97 and there existed a unidirectional long-run causality running from economic growth to crude oil import. The demand for ICO in South Africa as a function of real income and the price of crude oil were studied [8]. The estimated long-run price and income elasticities revealed that import demand for crude oil is price and income inelastic.

The study estimated the short-run and long-run elasticities of demand for crude oil in Turkey for the period of 1980 to 2005 [9]. The comparative study was performed in the

study [10]. A decision support system for forecasting fossil fuel production in Turkey, using a regression ARIMA and SARIMA method, was developed for the period of 1950–2003. Also, the study applied the ARIMA and SARIMA methods to predict primary energy demand in Turkey for the period from 2005 to 2020 [11].

In addition to forecasting the oil demand or consumption, several forecasting methods were applied to predict different types of energies. For example, regression modeling was employed to forecast the coal, oil, and gas electricity requirement [12]. The MR models were also applied to predictions for electricity consumption in Taiwan, Brazil, Delhi, and Hong-Kong, respectively [13–15]. The study [16] used the multiple linear regression techniques to develop simple empirical relations for the estimation of daily and monthly evaporation in Kuwait. While the ARIMA approaches were commonly used for predictions of energy demand [10, 11], the ANN models were also widely used for predictions of energy demand [11, 17–20]. Another promising forecasting technique, SVR, was used to forecast the energy demand [21, 22]. Additionally, the various hybrid modeling approaches were reported in forecasting energy demand in several studies [23–26], and the hybrid modeling schemes appear to be a promising technique. The study provided literature review in detail about the forecasting modeling in energy issue [27].

Extreme learning machine (ELM) proposed by Huang et al. [28, 29] is a novel learning algorithm for single-hidden-layer feedforward neural networks (SLFN). It provides much better generalization performance with much faster learning speed and avoids many issues faced in the traditional algorithms such as stopping criterion, learning rate, number of epochs and local minima, and the over-tuned problems [29]. Moreover, the universal approximation capability of ELM has been analyzed and proven by [30–32] to show the effectiveness of ELM. Thus, ELM has attracted a lot of attentions in recent years and been used for various forecasting issues, such as sales forecasting [33–35], stock price forecasting [36, 37], and electricity price forecasting [38].

In June 1946, Chinese Petroleum Corp. (CPC) was funded, and the headquarters was set up in Taipei under the direction of the Ministry of Economic Affairs. With service facilities covering the whole nation, its operations today include the import, exploration, development, refining, transport, marketing, and sale of petroleum and natural gas. CPC's total capital stands at NT\$130.1 billion, and its total revenues in 2011 amounted to NT\$1.03 trillion.

Since crude oil is extremely important for development of Taiwan's economy, the predictions of the demand of imported crude oil are a must. Accordingly, this study is aimed at proposing single and two-stage forecasting techniques to predict the demand of imported crude oil in Taiwan. The single stage forecasting modeling includes the support vector regression (SVR), artificial neural networks (ANN), extreme learning machine (ELM), and multiple linear regression (MLR) approaches. The two-stage models combine the two modeling components. The first component of the model uses its own feature to capture the significant explanatory variables. Then, the second component generates the predictions based on these explanatory variables. In this study, the combinations of MLR and SVR (i.e., refer to MLR-SVR), MLR and ANN (i.e., refer to MLR-ANN), and MLE and ELM (i.e., refer to MLR-ELM) are used as the two-stage models.

Real data are sampled for the period of 1993–2010 for the ICO in Taiwan. According to the suggestion [39], twenty-three associated variables are collected for serving as the explanatory variables. The predictions for ICO in Taiwan can be made based on the single stage and two-stage forecasting models. This study uses the first 14 years (1993–2006) of data for model building, and the last four years' data are used for the purpose of confirmation. The mean absolute percentage error (MAPE), the root mean square error (RMSE), and the mean absolute difference (MAD) are used as the forecasting accuracy measures.

The contents of this study are organized as follows. The following section introduces the proposed forecasting techniques. Section 3 presents the real data of ICO and the forecasting results. The performances for all of the forecasting models are demonstrated and discussed. The final section, Section 4, concludes this study.

2. Research Methodologies

This study considers MLR, SVR, ANN, ELM, and their hybrid modeling schemes as possible forecasting models for import demand of crude oil in Taiwan. These forecasting techniques are introduced in the subsequent sections.

2.1. Multiple Linear Regression. The multiple linear regression analysis is the procedure by which an algebraic equation is formulated to estimate the value of a dependent variable Y_i, given the values of some other independent variables $X_{1i}, X_{2i}, \ldots, X_{pi}, i = 1, 2, \ldots, n$. The relationship between the dependent and independent variables is expressed by a linear algebraic model as follows:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \varepsilon_i, \tag{1}$$
$$i = 1, 2, \ldots, n.$$

Based on the least squares or maximum likelihood criterion, the so-called normal equations and the point estimators of $\beta_0, \beta_1, \ldots, \beta_p$ can be obtained. Under the normal and some mild assumptions, the sampling distributions and hence the statistical inference for the estimators of $\beta_0, \beta_1, \ldots, \beta_p$ can be derived.

To identify significant independent variables, the backward elimination, forward selection, or stepwise regression procedures can be applied. The backward elimination procedure begins with the model which includes all of the available explanatory variables, and successively deletes one variable at a time to the model in such a way that at each step, the variable deleted is the variable contributing the least to the prediction of dependent variable at that step. On the contrary, the forward selection procedure begins with the constant model that includes no explanatory variable, and successively adds one variable at a time to the model in such a way that, at each step,

the variable added is the variable contributing the most to the prediction at that step. The stepwise regression procedure is an admixture of the backward elimination procedure and the forward selection procedure. This selection procedure builds a sequence of models and at each step deletes or adds an explanatory variable according to some selection criterions such as coefficient of partial correlation or error sum of squares reduction.

2.2. Support Vector Regression.

While support vector machine (SVM) is one of the most powerful techniques in machine learning areas [40–44], support vector regression (SVR) is the most common application form of SVMs. Due to its effectiveness, SVR has been used for predictions in many fields [45–49].

Statistical learning theory and structural risk minimization principle have provided a very effective framework for development of support vector regression [48–51]. Based on the computation of a linear regression function in a high dimensional feature space, the inputs for SVR are mapped via a nonlinear function. The modeling of SVR can be described as follows. Suppose that

$$f(x) = (w \cdot \Phi(x)) + b, \qquad (2)$$

where w is the weight vector, x represents the model inputs, b is a bias, and $\Phi(x)$ stands for a kernel function which uses a nonlinear function to transform the nonlinear input to be linear mode in a high dimension feature space.

Typical regression modeling obtains the coefficients through minimizing the square error, which can be considered as empirical risk based on loss function. The ε-insensitivity loss function was introduced [51], and it can be described as follows:

$$L_\varepsilon(f(x), y) = \begin{cases} |f(x) - y| - \varepsilon & \text{if } |f(x) - y| \geq \varepsilon \\ 0 & \text{otherwise,} \end{cases} \qquad (3)$$

where y is the target outputs and ε defines the region of ε-insensitivity; when the predicted value falls into the band area, the loss is zero. However, when the predicted value falls outside the band area, the loss is defined as the difference between the predicted value and the margin.

When empirical risk and structure risk are both considered, the SVR can be setup to minimize the following quadratic programming problem:

$$\begin{aligned} \text{Minimize:} \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ \text{subject to} \quad & (y_i - w \cdot \Phi(x_i)) - b \leq \varepsilon + \xi_i \\ & (w \cdot \Phi(x_i)) + b - y_i \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0, \\ & \text{for } i = 1, \ldots, n, \end{aligned} \qquad (4)$$

where $i = 1, \ldots, n$ is the number of training data, $(\xi_i + \xi_i^*)$ represents the empirical risk, $(1/2)\|w\|^2$ stands for

the structure risk preventing over-learning and lack of applied universality, and C is a modifying coefficient representing the trade-off between empirical risk and structure risk. With an appropriate modifying coefficient C, band area width ε, and kernel function, the optimum value of each parameter can be solved by Lagrange procedure.

The general form of the SVR-based regression function is described as follows [51]:

$$f(x, w) = f(x, r, r^*) = \sum_{i=1}^{N}(r_i - r_i^*)\Phi(x, x_i) + b, \qquad (5)$$

where r_i and r_i^* are Lagrangian multipliers and satisfy the equality $r_i r_i^* = 0$. Additionally, since the radial basis function (RBF) is the most widely used kernel function [49], this study uses it for our experimental study. The RBF can be defined as follows:

$$\phi(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right), \qquad (6)$$

where σ denotes the width of the RBF.

2.3. Artificial Neural Networks.

Artificial neural networks, originally derived from neurobiological models, are massively parallel, computer-intensive, and data-driven algorithmic systems composed of a multitude of highly interconnected nodes, known as neurons as well. Mimicking human neurobiological information-processing activities, each elementary node of a neural network is able to receive an input single from external sources or other nodes and the algorithmic procedure equipped in each node is sequentially activated to locally transforming the corresponding input single into an output single to other nodes or environments.

It was indicated that knowledge is not stored within individual processing units, but is represented by the strength between units [52]. They also stated that neural networks can be classified into two categories: the feedforward networks and the feedback networks. The feedback networks contain neurons that are connected to each other, enabling a neuron to influence other neurons. The feedback networks contain neurons that are connected to themselves, enabling a neuron to influence other neurons and itself. Also, in recent years, there has been a great deal of attention toward the field of ANN [41, 43, 53, 54].

For ANN modeling, the relationship between output (y) and inputs ($x_1, x_2, \ldots, x_m$) can be formed as

$$y = \lambda_0 + \sum_{j=1}^{n}\lambda_j g\left(\delta_{0_j} + \sum_{i=1}^{m}\delta_{ij}x_i\right) + \varepsilon, \qquad (7)$$

where λ_j ($j = 0, 1, 2, \ldots, n$) and δ_{ij} ($i = 0, 1, 2, \ldots, m$; $j = 0, 1, 2, \ldots, n$) are model connection weights, m is the number of input nodes, n is the number of hidden nodes, and ε is the error term. The transfer function in the hidden layer is often represented by a logistic function; that is,

$$g(z) = \frac{1}{1 + \exp(-z)}. \qquad (8)$$

Accordingly, the ANN model in (7) accomplishes a non-linear functional mapping from the inputs $(x_1, x_2, \ldots, x_a)$ to the output y; that is,

$$y = f(x_1, x_2, \ldots, x_m, w) + \varepsilon, \quad (9)$$

where w is a vector of all model parameters and f is a function determined by the ANN structure and connection weights.

2.4. Extreme Learning Machine.

ELM randomly selected the input weights and analytically determined the output weights of SLFNs. One may randomly choose and fix the hidden node parameters which are the key principle of the ELM. After randomly choosing the hidden nodes parameters, SLFN becomes a linear system where the output weights of the network can be analytically determined using simple generalized inverse operation of the hidden layer output matrices [28, 29].

In general, the concept of ELM is similar to that of the random vector functional-link (RVFL) network where the hidden neurons are randomly selected. However, the main difference between ELM and RVFL is the characteristics of hidden neuron parameters. In ELM, all the hidden node parameters are randomly generated independently of the target functions and the training patterns [55]. In RVFL, the selection of hidden neurons is based on partial randomness and the randomly generated hidden node parameters are not completely independent of the training data [56]. That is, the universal approximation capability of ELM can be linearly extended to RVFL [55, 56].

Consider N arbitrary distinct samples (x_i, t_i) where $x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in R^n$ and $t_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T \in R^m$. SLFNs with $\widetilde{N}$ hidden neurons and activation function $g(x)$ can approximate N samples with zero error. This means that

$$H\beta = T, \quad (10)$$

where

$$H(w_1, \ldots, w_{\widetilde{N}}, b_1, \ldots, b_{\widetilde{N}}, x_1, \ldots, x_N)$$

$$= \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_{\widetilde{N}} \cdot x_1 + b_{\widetilde{N}}) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_{\widetilde{N}} \cdot x_N + b_{\widetilde{N}}) \end{bmatrix}_{N \times \widetilde{N}}, \quad (11)$$

$$\beta_{\widetilde{N} \times m} = \left(\beta_1^T, \ldots, \beta_{\widetilde{N}}^T\right)^t,$$

$$T_{N \times m} = \left(T_1^T, \ldots, T_N^T\right)^t,$$

and $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$, $i = 1, 2, \ldots, \widetilde{N}$, is the weight vector connecting the ith hidden node and the input nodes, $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ is the weight vector connecting the ith hidden node and the output nodes, and b_i is the threshold of the ith hidden node. $w_i \cdot x_j$ denotes the inner product of w_i and x_j. H is called the hidden layer output matrix of the neural network; the ith column of H is the ith hidden node output with respect to inputs $x_1, x_2, \ldots x_N$. Therefore, the determination of the output weights is as simple as finding the least-square solution to the given linear system.

The minimum norm least-square (LS) solution to the linear system is

$$\widehat{\beta} = H^{\Psi} T, \quad (12)$$

where H^{Ψ} is the Moore-Penrose generalized inverse of matrix H. The minimum norm LS solution is unique and has the smallest norm among all the LS solutions.

The first step of ELM algorithm is to randomly assign input weight w_i and bias b_i. Then, the hidden layer output matrix H is calculated. Finally, one can calculate the output weight β, $\widehat{\beta} = H^{\Psi} T$, where $T = (t_1, \ldots, t_N)^t$.

2.5. Hybrid Models.

Recent research indicates that hybrid systems which are integrated with several standard ones can help to achieve a better performance for some applications. For example, the hybrid modeling applications have been reported in forecasting [57–62], credit risk [61], and manufacturing process [41–44]. As a consequence, this study proposes a two-step hybrid data mining mechanisms to predict the demand of ICO in Taiwan. This study proposes three hybrid data mining models, namely, the integration of MLR and ANN (i.e., MLR-ANN), MLR and SVR (i.e., MLR-SVR), and MLR and ELM (i.e., MLR-ELM). The concept of those hybrid modeling is as follows.

In the first stage of hybrid modeling, more significant variables are selected using MLR with forward selection, backward elimination or stepwise regression, say, $X_{i1}^*, X_{i2}^*, \ldots, X_{ir}^*$. In the second stage, the selected significant variables $X_{i1}^*, X_{i2}^*, \ldots, X_{ir}^*$ obtained in the first stage are served as the inputs of ANN, SVR, and ELM in order to establish the two-stage hybrid forecasting models. By providing the ANN, SVR, and ELM with good starting points, it is hoped that more effective models can be developed on the strength of their learning ability. Such two-stage hybrid models then are compared with the single-stage of MLR, ANN, SVR, and ELM models.

3. Results and Analysis

To show the effectiveness of the proposed hybrid modeling, the real data, from the years 1993 to 2010, were sampled for the ICO from Bureau of Energy in Taiwan [62]. Following the suggestion [39] and having discussed with the ICO practitioners, we sample 23 associated influential variables. Table 1 lists these 23 variables. The yearly data were collected for the period of 1993–2010 from the web sites of Bureau of Energy in Taiwan [62]. The first 14 years' data are used for model building, and the last four years' data are used for model confirmation.

3.1. Single-Stage Modeling.

Figure 1 displays the yearly data of ICO in Taiwan for the periods of 1993–2006. For MLR modeling, we first compute the variance inflation factors (VIFs) to examine the existence of collinearity.

Our numerical results reveal that all the values of VIFs of independent variables are greater than 10 except the variables of X_4 and X_5, indicating that there may exists

TABLE 1: Meaning of the influential variables for ICO model building.

| Variable	Meaning
Y	Imported crude oil
X_1	Gross domestic product
X_2	Consumer price index
X_3	Personal disposable income
X_4	Average temperature
X_5	Average sunshine per day (hours)
X_6	Average electricity households
X_7	Average electricity price
X_8	National income
X_9	Population
X_{10}	Foreign trade total
X_{11}	Wholesale prices
X_{12}	Consumer index
X_{13}	GNP deflators
X_{14}	Foreign exchange reserves
X_{15}	Total primary energy supply
X_{16}	Total final consumption
X_{17}	Total domestic consumption
X_{18}	Energy productivity
X_{19}	Energy intensity
X_{20}	Energy consumption of energy intensive industries
X_{21}	Value-added of energy intensive industries
X_{22}	Dependence on imported energy
X_{23}	Electricity average load
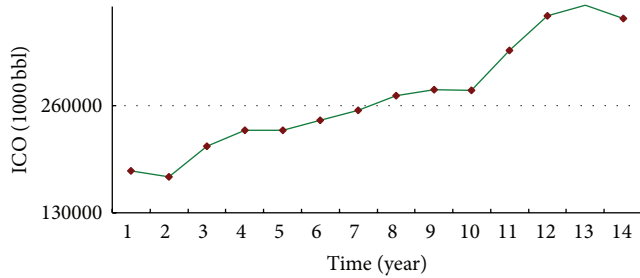
FIGURE 1: Historical yearly data of ICO in Taiwan.

high collinearity among the independent variables. To eliminate variables with high collinearity, we apply the Pearson correlation coefficient. If the correlation coefficient between two independent variables is greater than 0.7, we eliminate the variable which has a lower relationship with dependent variable Y. Table 2 lists the analysis results. As shown in Table 2, after excluding the variables with high collinearity, there remained six independent variables, including X_4, X_5, X_7, X_{13}, X_{19}, and X_{22}. In Table 3, it can be found that all the values of VIFs of the remaining variables are smaller than 10. As a consequence, there is no high collinearity among these independent variables. In addition, we apply the backward elimination, forward selection, and stepwise regression procedures to identify significant independent variables, and use

a 0.05 significance level to perform the MLR analysis. All the three procedures have the same selection results. As listed in Table 4, significant independent variables related to imported crude oil in Taiwan include average electricity price (X_7) and dependence on imported energy (X_{22}). Accordingly, the MLR model is derived as follows:

$$\widehat{Y} = -11862349.62 - 75695.53X_7 + 124776.53X_{22}. \quad (13)$$

The regression coefficients in the MLR model indicate that the higher the average electricity price is, the lower the imported crude oil is. On the contrary, the higher the dependence on imported energy is, the lower the imported crude oil is.

For ANN modeling, since the backpropagation neural network (BPNN) structure has been widely used [43, 54], this study employs BPNN as ANN modeling structure. In BPNN structure, we have 23 input nodes and one output node. The hidden nodes range from $i+2$ to $i-2$, where i is the number of input variables. Thus, the hidden nodes were set up as 21, 22, 23, 24, and 25, respectively. The training and testing processes include 14 and 4 data vectors for possible parameter setting. The learning rates are 0.01, 0.005, and 0.001, respectively, according to the suggestions [43].

After applying ANN to ICO data, we have obtained the {23-22-1} topology with a learning rate of 0.01 which provides the best result. The {N_i-N_h-N_o} represents the number of nodes in the input layer, hidden layer, and output layer, respectively. For SVR modeling, same as ANN modeling, we have 23 input variables. The two parameters, C and gamma, were estimated as 2^{-15} and 2^{-15}, respectively.

As mentioned earlier, the most important ELM parameter is the number of hidden nodes and that ELM tends to be unstable in single run forecasting [29, 33]. Therefore, the ELM models with different numbers of hidden nodes varying from 1 to 15 are constructed. For each number of nodes, an ELM model is repeated 30 times and the average RMSE of each node is calculated. The number of hidden nodes that gives the smallest average RMSE value is selected as the best parameter of ELM model. In modeling ELM, the forecasting model with eight hidden nodes has the smallest average RMSE values and is therefore the best ELM model.

3.2. Proposed Hybrid Modeling. A rational strategy for a hybrid modeling is to use the fewer but more informative variables, which were selected by the first stage of modeling approaches, as the inputs for the second stage of classifier approaches. Accordingly, in this study, the significant variables selected, that is, average electricity price (X_7) and dependence on imported energy (X_{22}), are used as the input variables of the ANN and SVR for hybrid modeling.

After completing the first stage of hybrid modeling, the ANN topology settings can be established. This study has found that the {2-3-1} topology with a learning rate of 0.01 provides the best result for the hybrid model. The network topology with the minimum testing RMSE is also considered as the optimal network topology. For the MR/SVR hybrid modeling, the parameters of C and gamma were still estimated as 2^{-15} and 2^{-15}, respectively. In the construction of

TABLE 2: Pearson correlations for pairs of variables.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}	X_{21}	X_{22}	X_{23}	Y
X_1	1.00																							
X_2	**0.97**	1.00																						
X_3	**0.99**	**0.96**	1.00																					
X_4	0.31	0.27	0.32	1.00																				
X_5	0.29	0.26	0.30	0.00	1.00																			
X_6	**0.99**	**0.98**	**0.99**	0.31	0.35	1.00																		
X_7	0.33	0.45	0.31	−0.02	0.13	0.37	1.00																	
X_8	**0.99**	**0.96**	**0.99**	0.32	0.30	**0.99**	0.30	1.00																
X_9	**0.99**	**0.95**	**0.99**	0.34	0.38	**0.99**	0.30	**0.99**	1.00															
X_{10}	**0.92**	**0.90**	**0.91**	0.09	0.16	**0.89**	0.43	**0.91**	**0.87**	1.00														
X_{11}	**0.81**	**0.85**	**0.79**	0.04	0.16	**0.80**	0.55	**0.79**	**0.75**	**0.94**	1.00													
X_{12}	**0.97**	**0.99**	**0.96**	0.27	0.26	**0.98**	0.45	**0.96**	**0.95**	**0.90**	**0.85**	1.00												
X_{13}	0.19	0.18	0.21	0.45	−0.02	0.20	−0.49	0.22	0.23	−0.13	−0.26	0.18	1.00											
X_{14}	**0.90**	**0.86**	**0.89**	0.13	0.40	**0.89**	0.57	**0.89**	**0.88**	**0.92**	**0.86**	**0.86**	−0.22	1.00										
X_{15}	**0.99**	**0.92**	**0.99**	0.31	0.38	**0.97**	0.25	**0.99**	**0.99**	**0.89**	**0.76**	**0.92**	0.17	**0.89**	1.00									
X_{16}	**0.98**	**0.93**	**0.98**	0.30	0.39	**0.98**	0.27	**0.98**	**0.99**	**0.90**	**0.77**	**0.93**	0.15	**0.91**	**1.00**	1.00								
X_{17}	**0.98**	**0.93**	**0.98**	0.30	0.38	**0.98**	0.25	**0.99**	**0.99**	**0.89**	**0.77**	**0.92**	0.16	**0.90**	**1.00**	**1.00**	1.00							
X_{18}	0.11	0.24	0.09	−0.24	−0.31	0.11	0.75	0.08	0.00	0.37	0.54	0.24	−0.53	0.30	−0.02	−0.01	−0.02	1.00						
X_{19}	−0.09	−0.22	−0.07	0.25	0.34	−0.08	−0.72	−0.06	0.02	−0.35	−0.52	−0.22	0.53	−0.27	0.05	0.03	0.05	**−1.00**	1.00					
X_{20}	**0.96**	**0.91**	**0.96**	0.25	0.44	**0.95**	0.38	**0.96**	**0.96**	**0.91**	**0.83**	**0.91**	−0.01	**0.95**	**0.98**	**0.98**	**0.98**	0.07	−0.05	1.00				
X_{21}	**0.98**	**0.92**	**0.98**	0.27	0.33	**0.96**	0.34	**0.98**	**0.96**	**0.94**	**0.82**	**0.92**	0.07	**0.93**	**0.98**	**0.98**	**0.98**	0.12	−0.10	**0.98**	1.00			
X_{22}	**0.98**	**0.93**	**0.98**	0.27	0.38	**0.98**	0.25	**0.98**	**0.99**	**0.88**	**0.75**	**0.93**	0.20	**0.88**	**0.99**	**0.99**	**0.99**	−0.01	0.04	**0.96**	**0.97**	1.00		
X_{23}	**0.99**	**0.94**	**0.99**	0.35	0.35	**0.98**	0.30	**0.99**	**0.99**	**0.91**	**0.80**	**0.94**	0.16	**0.90**	**0.99**	**0.99**	**0.99**	0.06	−0.03	**0.97**	**0.98**	**0.98**	1.00	
Y	0.91	0.84	0.92	0.19	0.41	0.90	0.06	0.92	0.92	0.81	0.68	0.84	0.23	0.82	0.94	0.94	0.94	−0.13	0.15	0.90	0.90	0.95	0.93	1.00

Table 3: Collinearity diagnosis for MLR modeling.

Variable X_i	X_4	X_5	X_7	X_{13}	X_{19}	X_{22}
VIF	1.52	2.04	3.91	1.99	4.36	1.49

Table 4: MLR model for ICO in Taiwan.

Variables	Estimated coefficient	Standard error	P value
Constant	−11862349.62	853356.39	<0.01[**]
Average electricity price (X_7)	−75695.53	28564.49	0.02[**]
Dependence on imported energy (X_{22})	124776.53	8782.77	<0.01[**]

[**]Denotes significance at 5% level.

Table 5: Various forecasting models' accuracy measures for ICO.

	MAPE	RMSE	MAD
Single stage models			
ANN	15.681	56190.676	52155.891
SVR	22.304	77973.147	76318.645
MR_{SEL}	11.753	48139.849	39431.691
ELM	10.174	43189.042	33271.068
Proposed hybrid models			
MR_{SEL}-ANN	7.302	33875.635	24572.516
MR_{SEL}-SVR	9.385	35523.521	32465.307
MR_{SEL}-ELM	7.094	28902.330	23271.608

MR/ELM hybrid model, the model with eight hidden nodes has the smallest average RMSE values is the best MR/ELM hybrid model.

3.3. Experimental Results. In this study, we consider the forecasting accuracy measures of MAPE, MSE, and MAD to address the forecasting performance for the five different approaches, ANN, SVR, ELM, MR_{SEL}, MR_{SEL}-ANN, MR_{SEL}-SVR, and MR_{SEL}-ELM, respectively. These prediction measurements are defined as follows:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n} \frac{|e_t|}{Y_t} \times 100,$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(e_t)^2}, \quad (14)$$

$$\text{MAD} = \frac{1}{n}\sum_{t=1}^{n}|e_t|,$$

where e_t stands for the value of the residual at time t. A low MAPE, MSE or MAD is associated with better forecasting accuracy. The results are listed in Table 5. In Table 5, by considering single stage modeling approaches, we note that ELM model has the best performance than the MR, ANN, and SVR models.

Table 6: Improvement of the proposed models in comparison with the single stage models.

Models	MAPE (%)	RMSE (%)	MAD (%)
Proposed hybrid MR_{SEL}-ANN model			
ANN	53.43	39.71	52.89
SVR	67.26	56.55	67.80
MR_{SEL}	37.87	29.63	37.68
ELM	28.23	21.56	26.14
Proposed hybrid MR_{SEL}-SVR model			
ANN	40.15	36.78	37.75
SVR	57.92	54.44	57.46
MR_{SEL}	20.15	26.21	17.67
ELM	7.76	17.75	2.42
Proposed hybrid MR_{SEL}-ELM model			
ANN	54.76	48.56	55.38
SVR	68.19	62.93	69.51
MR_{SEL}	39.64	39.96	40.98
ELM	30.27	33.08	30.05

In comparison to the single stage and our proposed hybrid models in Table 5, one is able to apparently observe that our proposed hybrid models provide more accurate results than the single stage models. In terms of MAPE, MSE, or MAD, the two hybrid models are all lower than the three single stage models. For example, the MAPE percentage improvements of the proposed MR_{SEL}-ANN model over the four-single-stage models, ANN, SVR, MR_{SEL}, and ELM are 53.43%, 67.26%, 37.87%, and 28.23%, respectively. Table 6 lists a comparison with respect to the overall improvement percentage in the single stage models. In addition, we use independent sample t-test to test whether the hybrid models are superior to the single ones. Let μ_s and μ_h be the means of the absolute value of residuals for the single-stage models and hybrid models, respectively. The independent sample t-test is applied to test the following:

$$H_0 : \mu_s - \mu_h \leq 0 \quad \text{versus} \quad H_1 : \mu_s - \mu_h > 0. \quad (15)$$

The P value is 0.0127. Obviously, the hybrid models outperform the single stage models.

Figure 2 shows the actual ICO values for the last four years and the corresponding forecasts by using four-single-stage models. It can be seen that the ELM model provides the best predictions of ICO. Figure 3 displays the actual ICO values and the corresponding forecasts by using three hybrid models. It shows that the hybrid MR-ANN model has the best forecasting capability. Figure 4 plots the actual ICO values and the corresponding forecasts by using the single MR and ANN models and their hybrid technique, MR-ANN. It apparently can be observed that the proposed hybrid model outperforms the single models. The same conclusions can be drawn by observing Figures 5 and 6. That is, the performance of the hybrid models is better than the single models.
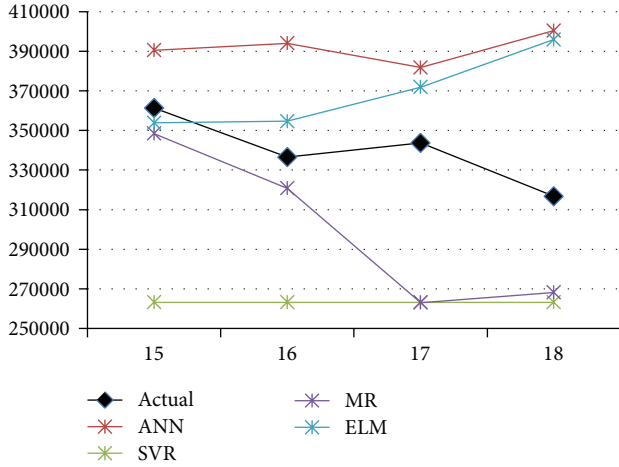
FIGURE 2: Plot of actual ICO values for the last four years and the forecasts by using four single-stage models.
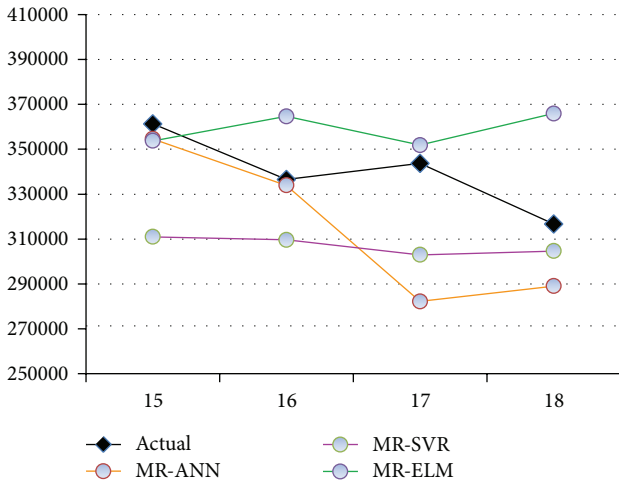

FIGURE 3: Plot of actual ICO values for the last four years and the forecasts by using three hybrid models.
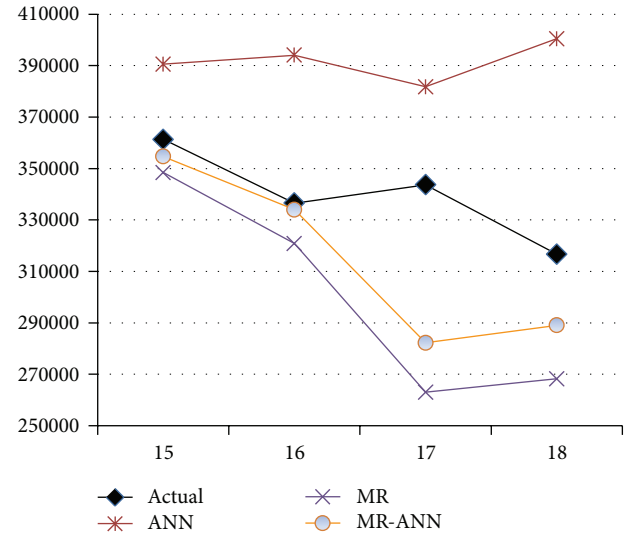

FIGURE 4: Plot of actual ICO values for the last years and the forecasts by using MR, ANN, and the hybrid MR-ANN model.
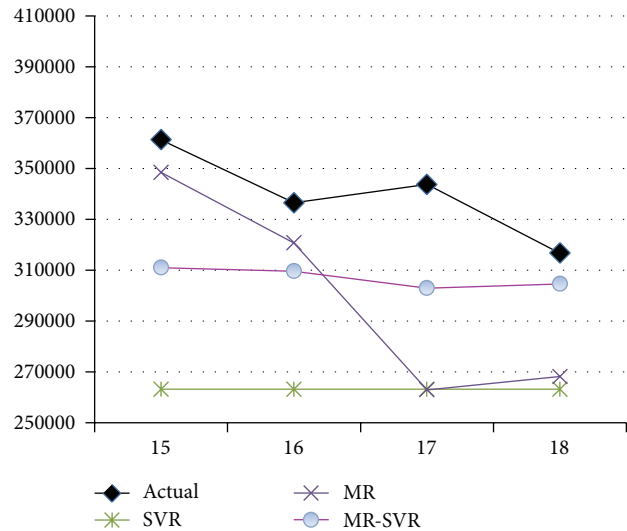

FIGURE 5: Plot of actual ICO values for the last years and the forecasts by using MR, SVR, and the hybrid MR-SVR model.

As shown in Figures 4, 5, and 6 and Tables 5 and 6, the proposed hybrid models outperform all three-single-stage models. As a consequence, the proposed two-stage hybrid approaches are more efficient for forecasting ICO in Taiwan than the typical single stage methods.

4. Conclusions

Oil is not only used to make gas for cars, but for heating homes, producing electricity, making plastics, and other commodities. Oil and its byproducts are ingrained into almost every culture in the world. Therefore, the accurate prediction of the demand of ICO is very important for the economic development of a country.

Because it is difficult to fully capture the characteristics of the real ICO data, the two hybrid prediction models are then proposed to forecast the demand of ICO in Taiwan. Based on our numerical results, it is found that the proposed hybrid

approaches are more accurate than the established single-stage ones. The modeling procedures and results of this work may provide a guidance to develop forecasting models for other energies.

Besides, there are many other two-stage hybrid forecasting models that have been proposed and applied in various fields [63–70]. The proposed hybrid model in this study is not the only hybrid forecasting scheme for the prediction of demand of crude oil, as one can combine other artificial intelligence techniques or traditional multivariate models, like decision tree, multivariate adaptive regression splines, logistic regression, rough set, or independent component analysis, with ANN, SVR, or ELM for further improving the prediction accuracies. Based on our work, further research may be
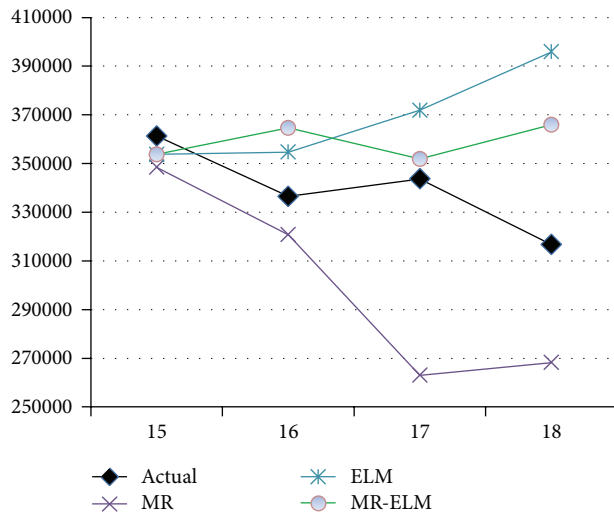
FIGURE 6: Plot of actual ICO values for the last years and the forecasts by using MR, ELM, and the hybrid MR-ELM model.

expanded. For example, extensions of the proposed hybrid prediction methods to other machine learning techniques or statistical prediction methods or to multistage prediction procedures are possible. Such works deserve further research and are our future concern.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] REN21 Renewables 2011: Global Status Report, 2011, http://www.ren21.net/Portals/0/documents/activities/gsr/2011 GSR Press Release-English.pdf.

[2] REN21, Renewables 2012: Global Status Report, 2012, http://new.ren21.net/Portals/0/documents/activities/gsr/GSR2012_low res_FINAL.pdf.

[3] International Energy Agency (IEA). World Energy Outlook 2004; IEA, Paris, France, 2004, http://www.worldenergyoutlook.org/media/weowebsite/2008-1994/WEO2004.pdf.

[4] International Energy Agency (IEA), World Energy Outlook 2007; IEA, Paris, France, 2007, http://www.worldenergyoutlook.org/media/weowebsite/2008-1994/weo_2007.pdf.

[5] A. Mirchi, S. Hadian, K. Madani, O. M. Rouhani, and A. M. Rouhani, "World energy balance outlook and OPEC production capacity: implications for global oil security," *Energies*, vol. 5, pp. 2626–2651, 2012.

[6] W. P. Nel and C. J. Cooper, "A critical review of IEA's oil demand forecast for China," *Energy Policy*, vol. 36, no. 3, pp. 1096–1106, 2008.

[7] S. Ghosh, "Import demand of crude oil and economic growth: evidence from India," *Energy Policy*, vol. 37, no. 2, pp. 699–702, 2009.

[8] E. Ziramba, "Price and income elasticities of crude oil import demand in South Africa: a cointegration analysis," *Energy Policy*, vol. 38, no. 12, pp. 7844–7849, 2010.

[9] G. Altinay, "Short-run and long-run elasticities of import demand for crude oil in Turkey," *Energy Policy*, vol. 35, no. 11, pp. 5829–5835, 2007.

[10] V. Ş. Ediger, S. Akar, and B. Uğurlu, "Forecasting production of fossil fuel sources in Turkey using a comparative regression and ARIMA model," *Energy Policy*, vol. 34, no. 18, pp. 3836–3846, 2006.

[11] V. Ş. Ediger and S. Akar, "ARIMA forecasting of primary energy demand by fuel in Turkey," *Energy Policy*, vol. 35, no. 3, pp. 1701–1708, 2007.

[12] H. Farahbakhsh, V. I. Ugursal, and A. S. Fung, "A residential end-use energy consumption model for Canada," *International Journal of Energy Research*, vol. 22, no. 13, pp. 1133–1143, 1998.

[13] R. Sadownik and E. P. Barbosa, "Short-term Forecasting of industrial electricity consumption in Brazil," *Journal of Forecasting*, vol. 18, no. 3, pp. 215–224, 1999.

[14] M. Ranjan and V. K. Jain, "Modelling of electrical energy consumption in Delhi," *Energy*, vol. 24, no. 4, pp. 351–361, 1999.

[15] Y. Y. Yan, "Climate and residential electricity consumption in Hong Kong," *Energy*, vol. 23, no. 1, pp. 17–20, 1998.

[16] J. Almedeij, "Modeling pan evaporation for Kuwait by multiple linear regression," *The Scientific World Journal*, vol. 2012, Article ID 574742, 9 pages, 2012.

[17] K. Ermis, A. Midilli, I. Dincer, and M. A. Rosen, "Artificial neural network analysis of world green energy use," *Energy Policy*, vol. 35, no. 3, pp. 1731–1743, 2007.

[18] Z. W. Geem and W. E. Roper, "Energy demand estimation of South Korea using artificial neural network," *Energy Policy*, vol. 37, no. 10, pp. 4049–4054, 2009.

[19] L. Ekonomou, "Greek long-term energy consumption prediction using artificial neural networks," *Energy*, vol. 35, no. 2, pp. 512–517, 2010.

[20] Y. S. Murat and H. Ceylan, "Use of artificial neural networks for transport energy demand modeling," *Energy Policy*, vol. 34, no. 17, pp. 3165–3172, 2006.

[21] S. Fan, L. N. Chen, and W.-J. Lee, "Machine learning based switching model for electricity load forecasting," *Energy Conversion and Management*, vol. 49, no. 6, pp. 1331–1344, 2008.

[22] J. Wang, W. Zhu, W. Zhang, and D. Sun, "A trend fixed on firstly and seasonal adjustment model combined with the ε-SVR for short-term forecasting of electricity demand," *Energy Policy*, vol. 37, no. 11, pp. 4901–4909, 2009.

[23] A. Azadeh, S. F. Ghaderi, S. Tarverdian, and M. Saberi, "Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption," *Applied Mathematics and Computation*, vol. 186, no. 2, pp. 1731–1741, 2007.

[24] M. Amina, V. S. Kodogiannis, I. Petrounias, and D. Tomtsis, "A hybrid intelligent approach for the prediction of electricity consumption," *International Journal of Electrical Power & Energy Systems*, vol. 43, pp. 99–108, 2012.

[25] X. Wang and M. Meng, "A hybrid neural network and ARIMA model for energy consumption forecasting," *Journal of Computers*, vol. 7, no. 5, pp. 1184–1190, 2012.

[26] H. T. Pao, "Forecasting energy consumption in Taiwan using hybrid nonlinear models," *Energy*, vol. 34, no. 10, pp. 1438–1446, 2009.

[27] L. Suganthi and A. A. Samuel, "Energy models for demand forecasting—a review," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 2, pp. 1223–1240, 2012.

[28] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, pp. 985–990, Budapest, Hungary, July 2004.

[29] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.

[30] G.-B. Huang, L. Chen, and C.-K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 879–892, 2006.

[31] G.-B. Huang and L. Chen, "Convex incremental extreme learning machine," *Neurocomputing*, vol. 70, no. 16–18, pp. 3056–3062, 2007.

[32] G.-B. Huang and L. Chen, "Enhanced random search based incremental extreme learning machine," *Neurocomputing*, vol. 71, no. 16–18, pp. 3460–3468, 2008.

[33] Z.-L. Sun, T.-M. Choi, K.-F. Au, and Y. Yu, "Sales forecasting using extreme learning machine with applications in fashion retailing," *Decision Support Systems*, vol. 46, no. 1, pp. 411–419, 2008.

[34] C. J. Lu, T. S. Lee, and C. M. Lian, "Sales forecasting for computer wholesalers: a comparison of multivariate adaptive regression splines and artificial neural networks," *Decision Support Systems*, vol. 54, no. 1, pp. 584–596, 2012.

[35] C. J. Lu and Y. E. Shao, "Forecasting computer products sales by integrating ensemble empirical mode decomposition and extreme learning machine," *Mathematical Problems in Engineering*, vol. 2012, Article ID 831201, 15 pages, 2012.

[36] Z. Fang, J. Zhao, F. Fei, Q. Wang, and X. He, "An approach based on multi-feature wavelet and ELM algorithm for forecasting outlier occurrence in Chinese stock market," *Journal of Theoretical and Applied Information Technology*, vol. 49, no. 1, pp. 369–377, 2013.

[37] J. Zhao, Z. Wang, and D. S. Park, "Online sequential extreme learning machine with forgetting mechanism," *Neurocomputing*, vol. 87, pp. 79–89, 2012.

[38] X. Chen, Z. Y. Dong, K. Meng, Y. Xu, K. P. Wong, and H. W. Ngan, "Electricity price forecasting with extreme learning machine and bootstrapping," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2055–2062, 2012.

[39] Y. E. Shao, C. D. Hou, and C. S. Lin, "Applying artificial neural networks, multiple regression and their hybrid models for predictions of electricity sales in Taiwan," *ICIC Express Letters*. In press.

[40] J. Park, I.-H. Kwon, S.-S. Kim, and J.-G. Baek, "Spline regression based feature extraction for semiconductor process fault detection using support vector machine," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5711–5718, 2011.

[41] Y. E. Shao and B.-S. Hsu, "Determining the contributors for a multivariate SPC chart signal using artificial neural networks and support vector machine," *International Journal of Innovative Computing, Information and Control*, vol. 5, no. 12, pp. 4899–4906, 2009.

[42] C.-J. Lu, Y. E. Shao, and P.-H. Li, "Mixture control chart patterns recognition using independent component analysis and support vector machine," *Neurocomputing*, vol. 74, no. 11, pp. 1908–1914, 2011.

[43] Y. E. Shao, C.-J. Lu, and C.-C. Chiu, "A fault detection system for an autocorrelated process using SPC/EPC/ANN AND SPC/EPC/SVM schemes," *International Journal of Innovative Computing, Information and Control*, vol. 7, no. 9, pp. 5417–5428, 2011.

[44] Y. E. Shao, C. J. Lu, and Y. C. Wang, "A hybrid ICA-SVM approach for determining the fault quality variables in a multivariate process," *Mathematical Problems in Engineering*, vol. 2012, Article ID 284910, 12 pages, 2012.

[45] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6164–6173, 2009.

[46] C.-J. Lu and Y.-W. Wang, "Combining independent component analysis and growing hierarchical self-organizing maps with support vector regression in product demand forecasting," *International Journal of Production Economics*, vol. 128, no. 2, pp. 603–613, 2010.

[47] S. Pang, L. Song, and N. Kasabov, "Correlation-aided support vector regression for forex time series prediction," *Neural Computing and Applications*, vol. 20, no. 8, pp. 1193–1203, 2011.

[48] D. Basak, S. Pal, and D. C. Patranabis, "Support vector regression," *Neural Information Processing-Letters and Reviews*, vol. 11, no. 10, pp. 203–224, 2007.

[49] W. Dai, Y. E. Shao, and C. J. Lu, "Incorporating feature selection method into support vector regression for stock index forecasting," *Neural Computing and Applications*, vol. 23, no. 6, pp. 1551–1561, 2013.

[50] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.

[51] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 2000.

[52] M. Titterington, "Neural networks," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 1, pp. 1–8, 2010.

[53] L. Mi and F. Takeda, "Analysis on the robustness of the pressure-based individual identification system based on neural networks," *International Journal of Innovative Computing, Information and Control*, vol. 3, no. 1, pp. 97–110, 2007.

[54] C.-C. Chiu, Y. E. Shao, T.-S. Lee, and K.-M. Lee, "Identification of process disturbance using SPC/EPC and neural networks," *Journal of Intelligent Manufacturing*, vol. 14, no. 3-4, pp. 379–388, 2003.

[55] G.-B. Huang, M.-B. Li, L. Chen, and C.-K. Siew, "Incremental extreme learning machine with fully complex hidden nodes," *Neurocomputing*, vol. 71, no. 4–6, pp. 576–583, 2008.

[56] G. B. Huang, "An insight into extreme learning machines: random neurons, random features and kernels," *Cognitive Computation*, 2014.

[57] P. G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[58] H. F. Zou, G. P. Xia, F. T. Yang, and H. Y. Wang, "An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting," *Neurocomputing*, vol. 70, no. 16–18, pp. 2913–2923, 2007.

[59] Y. E. Shao, "Body fat percentage prediction using intelligent hybrid approaches," *The Scientific World Journal*, vol. 2014, Article ID 383910, 8 pages, 2014.

[60] H. C. Co and R. Boosarawongse, "Forecasting Thailand's rice export: Statistical techniques vs. Artificial neural networks,"

Computers and Industrial Engineering, vol. 53, no. 4, pp. 610–627, 2007.

[61] S. L. Lin, "A new two-stage hybrid approach of credit risk in banking industry," *Expert Systems with Applications*, vol. 36, no. 4, pp. 8333–8341, 2009.

[62] Bureau of Energy, http://web3.moeaboe.gov.tw/ECW/populace/content/SubMenu.aspx?menu_id=141.

[63] C. J. Lu, Y. E. Shao, and C. C. Li, "Recognition of concurrent control chart patterns by integrating ICA and SVM," *Applied Mathematics & Information Sciences*, vol. 8, no. 2, pp. 681–689, 2014.

[64] Y. E. Shao, C. D. Hou, and C. C. Chiu, "Hybrid intelligent modeling schemes for heart disease classification," *Applied Soft Computing*, vol. 14, pp. 47–52, 2014.

[65] Y. E. Shao and C. D. Hou, "Hybrid artificial neural networks modeling for faults identification of a stochastic multivariate process," *Abstract and Applied Analysis*, vol. 2013, Article ID 386757, 10 pages, 2013.

[66] Y. E. Shao and C.-D. Hou, "Change point determination for a multivariate process using a two-stage hybrid scheme," *Applied Soft Computing Journal*, vol. 13, no. 3, pp. 1520–1527, 2013.

[67] M. Arun Kumar and M. Gopal, "A hybrid SVM based decision tree," *Pattern Recognition*, vol. 43, no. 12, pp. 3977–3987, 2010.

[68] P. Rajendran and M. Madheswaran, "Hybrid medical image classification using association rule mining with decision tree algorithm," *Journal of Computing*, vol. 2, pp. 127–136, 2010.

[69] B. B. Nair, V. P. Mohandas, and N. R. Sakthivel, "A genetic algorithm optimized decision tree-SVM based stock market trend prediction system," *International Journal on Computer Science and Engineering*, vol. 2, pp. 2981–2988, 2010.

[70] B. W. Heumann, "An object-based classification of mangroves using a hybrid decision tree-support vector machine approach," *Remote Sensing*, vol. 3, no. 11, pp. 2440–2460, 2011.