

# BUILDING AN INTELLIGENT SEASONAL TIME SERIES MODEL FOR FORECASTING BUILDING ELECTRICITY LOAD

## XÂY DỰNG MỘT MÔ HÌNH THEO MÙA THÔNG MINH ĐỂ DỰ BÁO PHỤ TẢI CHO NHÀ DÂN DỤNG

Thuy-Linh Le\*, Ngoc Hoang Tran, Duy Vu Luu, Duc Sy Nguyen, Thi Thu Ha Truong,  
Le Nhat Hoang Tran, Thi Ai Lanh Nguyen

*The University of Danang – University of Technology and Education*

\*Corresponding author: lttlinh@ute.udn.vn

(Received: August 29, 2022; Accepted: November 03, 2022)

**Abstract** - The building sector is a significant energy consumer, and its share of energy consumption is increasing because of urbanization. Forecasting the electricity load for improving building energy efficiency is imperative for reducing energy costs and environmental impacts. This study first builds a seasonal time-series model, then integrates it with IoT in the energy-predict systems. Notably, the built time-series model gives positive results with an R2 training of 0.814 and an R2 test of 0.803, which are much better than the regression model in accuracy and feature cost. Lastly, the proposed system automatically collects data from an IoT platform, predicts energy consumption, and sends results to end users. This system can help the user control their energy consumption or abnormal energy consumption in a home in real time.

**Key words** - Electrical load; energy consumption; forecasting; IoT; machine learning

### 1. Introduction

The faster rate of development of countries around the world leads to an increased demand for energy. Energy is always the big challenge of urbanization and industrialization, thus predicting energy consumption has become crucial for estimating and covering energy usage. As well as providing environmental benefits, buildings with efficient energy systems have much lower operating costs and higher market value. An efficient energy management strategy saves money for owners and stabilizes the project life cycle (including the project life cycle and product life cycle).

The construction sector is a major energy consumer, accounting for approximately 40% of global energy consumption and 30% of CO<sub>2</sub> emissions [1, 2]. Its share is increasing because of urbanization. In the United States, commercial and residential buildings account for 40% of the nation's total energy consumption, and this figure is steadily increasing [3]. In Europe, buildings constitute 40% of the electricity and 36% CO<sub>2</sub> emissions [4]. Accordingly, improving the energy efficiency of buildings is necessary for controlling energy costs, reducing the environmental impact, and increasing the value and competitiveness of buildings.

Buildings have generally long lifespans, thus early designs are extremely important, and adjustments to increase thermal performance, later on, can be very costly and ineffective. Energy costs, on the other hand, account

**Tóm tắt** – Xây dựng là một ngành tiêu thụ năng lượng lớn và chiếm tỉ trọng cao do quá trình đô thị hóa. Dự báo phụ tải tiêu thụ nhằm nâng cao hiệu quả sử dụng năng lượng của các tòa nhà là điều cần thiết để giảm thiểu chi phí năng lượng và tác động tiêu cực đến môi trường. Nghiên cứu này xây dựng mô hình dự báo chuỗi thời gian theo mùa, sau đó tích hợp nó với IoT để tạo ra hệ thống dự đoán năng lượng. Đáng chú ý, mô hình này cho kết quả tích cực với giá trị R2 huấn luyện là 0,814 và R2 thử nghiệm là 0,803, tốt hơn nhiều so với mô hình hồi quy không chỉ ở độ chính xác mà còn cả tính năng chi phí. Hệ thống này sẽ tự động thu thập dữ liệu từ nền tảng IoT, dự đoán và gửi kết quả cho người dùng. Nó có thể giúp người dùng kiểm soát mức tiêu thụ năng lượng và kiểm soát được bất thường về năng lượng tiêu thụ trong nhà theo thời gian thực.

**Từ khóa** - Phụ tải điện; tiêu thụ năng lượng; dự báo; IOT; máy học

for a significant portion of total building running costs, and superior thermal designs can result in significant operating cost savings with quick payback periods. Effective evaluation methods can help reduce building energy use dramatically [4]. In this context, our contribution is to propose a time-series model forecasting system, then integrate it with the IoT for an energy-predicting system. End users can use smart devices such as smart phones or personal computers to control electricity in their homes easily. This will be very useful to reduce the cost of living as well as the operating cost of the building.

IoT has evolved from being a vision for the future to being an increasing market reality. Technology companies have dedicated resources and personnel to research IoT and machine-to-machine communication [5].

Engineering problems can be defined using a mathematical model or AI system, which determines the relationship between system outputs and inputs. In recent years, several studies have been carried out in order to identify the most energy-efficient buildings and several research projects have been carried out to identify the best-performing buildings in terms of energy efficiency [6, 7]. Candanedo et al. [6] proposed a data driven prediction model of energy use of appliances in a low-energy house, Soheil Fathi et al. [7] showed a systematic review for machine learning applications in urban building energy performance forecasting.

Building electricity load is provided as time-series data

is sensitive and flexible data, because the operation of electrical appliances is a highly random form of energy use. This work develops an intelligent time series forecast model using a combination of a seasonal variant of ARIMA is SARIMA [8] with Least Square Support Vector Regression (LSSVR) model. The hyperparameters in this model will be optimized by the smart optimize algorithm Particle swarm optimization metaheuristic (PSO) [9]. By doing so, the S-PSO-LSSVR model can efficiently analyze real-time data collected from a smart grid infrastructure. Users can further apply the one-day-ahead forecasts to enhance the efficient energy usage of appliances and electrical equipment in their buildings. The S-PSO-LSSVR model is validated by the data on energy use of appliances in a house, Belgium. The forecasted result will be compared with the previous research. The high performance of the proposed model demonstrates the S-PSO-LSSVR model is a promising model to predict building electricity load.

## 2. Methodology

### 2.1. The S-PSO-LSSVR Proposed Forecast Model

#### 2.1.1. Season AutoRegressive Integrated Moving Average models

In Season AutoRegressive Integrated Moving Average (SARIMA) models, seasonal AR and MA terms predict time-series data  $y_t$  by using data values and errors at periods with lags that are multiples of  $S$  (span of seasonality) [8]. The SARIMA model, denoted as SARIMA  $(p, d, q) \times (P, D, Q)_S$ , incorporates both nonseasonal and seasonal factors into a multiplicative model. This model can be expressed as shown in Eq. (1), as explained in previous studies. Equations (2) to (4) present the formulation of terms in Eq. (1).

$$\theta_p(B)\Theta_p(B^S)(1-B)^d(1-B^S)^D y_t = w_q(B)W_Q(B^S)\alpha_t \quad (1)$$

$$w_q(B) = 1 - w_1 B - w_2 B^2 - w_3 B^3 - \dots - w_q B^q \quad (2)$$

$$\theta_p(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \dots - \theta_p B^p \quad (3)$$

$$W_Q(B^S) = 1 - W_1(B^S) - W_2(B^{2S}) - W_3(B^{3S}) - \dots - W_Q(B^{QS}) \quad (4)$$

where  $p$  represents the nonseasonal AR order,  $d$  represents nonseasonal differencing,  $q$  represents the nonseasonal MA order,  $P$  represents the seasonal AR order,  $D$  represents seasonal differencing,  $Q$  represents the seasonal MA order,  $S$  represents the time span of a repeating seasonal pattern, and  $B$  represents the backward shift operator for a nonstationary time-series data item  $y_t$ . Furthermore,  $w_q(B)$ ,  $\theta_p(B)$ ,  $\Theta_p(B^S)$ , and  $W_Q(B^S)$  are polynomials in  $B$  of degrees  $q$ ,  $p$ ,  $P$ , and  $Q$ , respectively, where  $w_q(B)$  and  $W_Q(B^S)$  indicate that  $y_t$  is a function of the previous forecast error in predicting  $y_t$ , and  $\theta_p(B)$  and  $\Theta_p(B^S)$  indicate that  $y_t$  is a function of its own previous values;  $\alpha_t$  is a current interference. Typically,  $\alpha_t$  is considered the estimated residual at time  $t$ . Moreover,  $p$ ,  $d$ ,  $q$ ,  $P$ ,  $D$ , and  $Q$  are all integers, and  $(1-B)^d y_t$  can be converted to a stationary series by using the difference operator  $1 - B$ ;  $B$  satisfies  $BY_t = y_{t-k}$  and  $B^k y_t = y_{t-k}$ .

#### 2.1.2. Least Square Support Vector Regression

ML techniques have been wide-applied for analyzing time-series data [10, 11]. The SVR approach developed by Vapnik in 1995 is an ML technique that is based on statistical learning theory and the principle of structural risk minimization [12]. Despite its high efficiency, the SVR approach is computationally slow when analyzing large data sets because its speed relies on the number of data samples and quadratic programming solvers [13]. To enhance computational speed, Suykens et al. [14] proposed the LSSVR method. This ML technique has many advanced features that enable high generalization capacity and fast computation [15].

LSSVR solves linear equations instead of a quadratic programming problem. It is preferred for large-scale regression problems that demand fast computation. In a function estimation of the LSSVR, given a training dataset  $\{x_k, y_k\}_{k=1}^N$ , the optimization problem is formulated as Eq.5

$$\min_{(\omega, b, e)} J(\omega, e) = \frac{1}{2} \|\omega\|^2 + \frac{1}{2} C \sum_{i=1}^N e_i^2 \quad (5)$$

subject to

$$y_k = \langle \omega, \varphi(x_k) \rangle + b + e_k, \quad k = 1, \dots, N$$

where  $J(\omega, e)$  is the optimization function;  $\omega$  is the parameter of the linear approximator;  $e_k \in R$  is error variables;  $C \geq 0$  is a regularization constant that represents the trade-off between the empirical error and the flatness of the function;  $x_k$  is input patterns;  $y_k$  is prediction labels; and  $N$  is the sample size.

The resulting LSSVR model for function estimation is expressed as Eq. (6)

$$f(x) = \sum_{k=1}^N \alpha_k K(x, x_k) + b \quad (6)$$

The LSSVR model accuracy depends on its hyperparameters, and especially the regularization constant and kernel parameters. To improve the predictive accuracy of the model, automatic optimization that is integrated with LSSVR should involve the regularization parameter (C) (Eq.(5)) and the sigma of the RBF kernel ( $\sigma$ ) (Eq.(6)). In highly nonlinear spaces, the RBF is often selected as a kernel function of the LSSVR to yields high results for this model.

$$K(x, x_k) = \exp(-\|x - x_k\|^2) / 2\sigma^2 \quad (7)$$

where  $\sigma$  is the kernel parameter which controls the kernel width used to fit the training data.

#### 2.1.3. Particle Swarm Optimization

Particle swarm optimization is a population-based stochastic global optimization technique developed by Kennedy and Eberhart [16]. The PSO consists of a set of particles moving around a search space, and is affected by their own best past location and the best past location of any particle in the swarm or a close neighbor. The velocity of each particle is updated in every iteration.

$$v_i(t+1) = \varphi \times v_i(t) + \left( c_1 \times rand() \times (p_i^{best} - p_i(t)) \right) + \left( c_2 \times rand() \times (p_{gbest} - p_i(t)) \right) \quad (8)$$

where  $v_i(t+1)$  is the new velocity of the  $i^{\text{th}}$  particle;  $\varphi$  is the inertia weight;  $c_1$  and  $c_2$  are the weighting coefficients for the personal best and global best positions respectively;  $p_i(t)$  is the position of the  $i^{\text{th}}$  particle at time  $t$ ;  $p_i^{best}$  is the best known position of the  $i^{\text{th}}$  particle so far, and  $p_{gbest}$  is the best position of any particle in the swarm so far. The  $rand()$  function generates a uniformly random variable  $\in [0,1]$ . Variants on this update equation consider the best positions within the local neighborhood of a particular at time  $t$  and  $p_i(t+1) = p_i(t) + v_i(t)$ . The position of a particular is updated using Eq. (8).

By using tent mapping, the PSO provides a highly diverse initial population. The tent map is a recurrence relation, written as

$$x_{n+1} = \begin{cases} \mu x_n & 0 \leq x \leq \frac{1}{2} \\ \mu - \mu x_n & \frac{1}{2} \leq x \leq 1 \end{cases} \quad (9)$$

for  $0 \leq \mu \leq 2$  and  $0 \leq x \leq 1$

where  $\mu$  is a positive real constant. In the iterating procedure, any point  $x_0$  in the interval is assumed a new subsequent position as described, generating a sequence  $x_n$  in  $[0,1]$ .

The population size is 50, max generation is 25 and PSO learning parameters  $c_1$  and  $c_2$  are 2.05.

## 2.2. Performance Measures

This study used correlation coefficient (R), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) to evaluate the prediction accuracy of the proposed S-PSO-LSSVR model.

The R is a common measure of how well the curve fits the actual data. A value of 1 indicates a perfect fit between actual and predicted values, meaning that the values have the same propensity. The MAPE is a statistical measure, which is identifying the relative differences between models because it is unaffected by the size or unit of actual and predicted values. The MAE is a quantity used to measure how close forecasts are to the eventual outcomes. The RMSE is computed to find the square error of the prediction compared to actual values and to find the square root of the summation value. Equations (10-13) show the respective formulas used for calculating these measures:

$$R = \frac{n \sum y_i \cdot y_i' - (\sum y_i)(\sum y_i')}{\sqrt{n(\sum y_i^2) - (\sum y_i)^2} \sqrt{n(\sum y_i'^2) - (\sum y_i')^2}} \quad (10)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i')^2} \quad (11)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y_i'| \quad (12)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y_i'}{y_i} \right| \quad (13)$$

In order to be consistent with the previous study, the R value will be squared to compare with [11].

## 2.3. Building data acquisition and notification model on IoT platform

In practice phase, we propose a protocol for acquiring result data and notifying to end user. In our case, Thingspeak is considered as a bridge between end user and Matlab analysis tool. ThingSpeak is a well-known and popular cloud service in the IoT community, allowing users to cloud data by getting data back via HTTP protocol. In our case, we use this IoT analytics platform for visualizing, and analyzing realtime data results on its cloud.

On this platform, the data collection is done using REST API or MQTT that is integrated with Matlab's cloud. Thingspeak acts its database online in which predefined users can access for their observations in graphical form. Through of its channel, they store the data send from various devices (sensors and automated board). The generated API key is provided in order to share visualized data for end users.

The main component of ThingSpeak is its channel which stores data send from various devices. Each channel can save up to eight fields along with device location, url etc. The channel can be made public which can be seen by other users or private which need the API key to view the data. The private channel can be shared for some specific users.

Providing user, a trigger- action protocol for notify analysis result through by SMS or personal contact. Final user monitor in flexible way for the data visualization: predictive indicators which represent predict energy consumption of user's building. From there, their correction is decided for profiting energy-efficient.

## 3. Experimental results

### 3.1. Database

The experimental house is finished in December 2015, Stambuges, Belgium. All the mechanical systems are new. The low-energy house was designed according to the passive house certification - a form of energy-saving housing according to the Belgium policy Passive House Planning project. In this project, the low-energy houses have an annual heating load and cooling load of no more than 15 kWh/m<sup>2</sup>per year. The building air leakage average was 50 Pa per hour, the ventilation unit between 90 and 95% efficiency. The total heated area is 220 m<sup>2</sup> from total floor area is 280 m<sup>2</sup>. There are four occupants in this house, two adults and two teenagers. The mother is a writer and works regularly in the home office.

### 3.2. Experimental results

The energy data is collected from a wireless sensor system called M-BUS energy counters for 137 days (from 11/1/2016 to 27/5/2016). Candanedo [6] was very careful

when taking into room occupancy when combined with relative humidity measurements in order to increase the practicality in forecasting energy consumption. However, when analyzing highly sensitive temporal data sets, the selection of data intervals for analysis is very important to avoid over-fitting and increase forecasting performance (Table 1).

**Table 1.** The correlation coefficient in previous research [6]

Model	R <sup>2</sup> training	R <sup>2</sup> test
LM	0.18	0.16
SVM Radial	0.85	0.52
GBM	0.97	0.57
RF	0.92	0.54

**Table 2.** Result of the best time by hybrid model

Time	R <sup>2</sup> training	R <sup>2</sup> test	CPU Time
29 day	0.83	0.726	1.83
31 day	0.818	0.735	2.05
38 day	0.794	0.805	3.43
<b>45 day</b>	<b>0.814</b>	<b>0.803</b>	<b>5.8</b>
115 day	0.958	0.55	65.11
122 day	0.787	0.794	81.26
129 day	0.801	0.788	117.08
136 day	0.972	0.611	253.73

This study presents a one-day-ahead forecasting system, the data is measured directly on a smart energy monitoring system so it has high reliability. However, the amount of data extracted is quite large and contains confounding values, thus pre-processing is needed to increase model

**Table 3.** Performance comparison among predictive models of the previous and proposed model

	Training				Test			
	RMSE	R <sup>2</sup>	MAE	MAPE %	RMSE	R <sup>2</sup>	MAE	MAPE %
LM [6]	93.21	0.18	53.13	61.32	93.18	0.16	51.97	59.93
SVM Radial [6]	39.35	0.85	15.08	15.6	70.74	0.52	31.36	29.76
GBM [6]	17.56	0.97	11.97	16.27	66.65	0.57	35.22	38.29
RF [6]	29.61	0.92	13.75	13.43	68.48	0.54	31.85	31.39
<b>S-PSO-LSSVR</b>	<b>5.001</b>	<b>0.814</b>	<b>25.007</b>	<b>25.87</b>	<b>6.964</b>	<b>0.803</b>	<b>28.496</b>	<b>26.24</b>

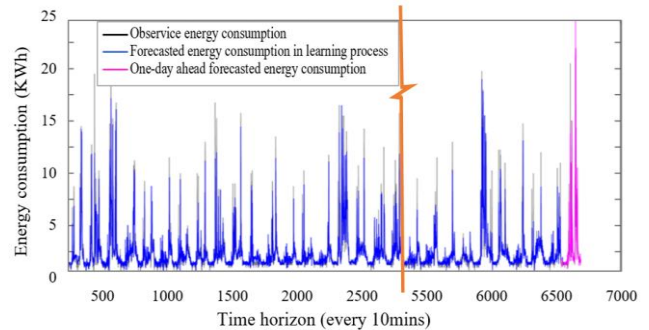
### 3.3. Modeling notification protocol on Thingspeak

We create a channel on Thingspeak for monitoring the predictive data for a day and a google gauge for watching an average of these indicators. Moreover, the visualization and notification of the data note by a smart phone for user is a final step. We propose, therefore, a React App: React app Send a tweet or trigger a ThingHTTP request when the Channel meets a certain condition. In this project the configuration of notification is configured on the late-middle of day. The information results consist of average of energy consumption and highest level in predictive period.

The Figure 2 shows results data on 24 recent times of predictive evolution corresponding to 24 hours/day. Figure 3 shows an automatic text message sent from Thing speak

performance and avoid the possibility of overfitting as [6]. The data set is pre-processed by sensitivity analysis for the correlation coefficient checking. Table 2 shows the highest correlation coefficient is 45 days.

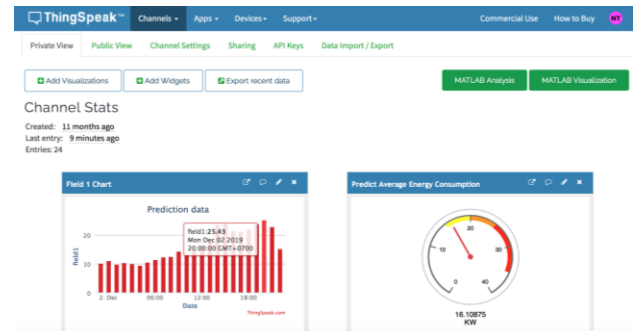
In Table 3 and Figure 1, the forecasting value (colour line) almost coincides with the observance value (black line) demonstrate the high performance of the proposed model. Table 3 shows that the S-PSO-LSSVR gives higher performance and stable training-test pairs than the previous research. These parameters indicate the superiority of the model in predicting the electricity load. S-PSO-LSSVR obtained significantly lower values of RMSE compared to the LM, SVM Radial, GMB, or RF models. MAE training-test pair (25.007-28496) and MAPE training-test pair (25.87-26.24) are highly compatible. The high value of R training and R test are 0.814 and 0.803 respectively showed a better predictive ability and reliability. This confirmed the efficiency of the proposed model in predicting electricity load.



**Figure 1.** The comparison of actual values and predicted values of the electrical load consumption

to user's portable.

This message provides the above necessary information for making decision on consumption energy of building.



**Figure 2.** The predictive results are visualized on Thingspeak

20:39



## REFERENCES

← InfoSMS\_TB

7-8 16:47

Sent from your Twilio trial account - Your predict building energy consumption report: The average consumption is [16.10875 KW](#). The highest indicator of your consumption is 25.43KW at night.

Figure 3. Text message of notification on user's portable

#### 4. Conclusion

This paper proposes an approach "The S-PSO-LSSVR Proposed Forecast Model" by the combination of two classical techniques: Least Square Support Vector Regression (LSSVR) and the Particle swarm optimization metaheuristic (PSO). The high-performance results are compared with the previous model. In addition, data preprocessing has limited the over-fitting and increasing the model reliability.

Unlike prior studies, our contribution focus on developing IoT notification system integrated with AI and optimized artificial intelligence. for monitoring this prediction data and preparing for notification. The final data are sent from the ThingSpeak platform server, using MQTT protocol and visualized in its channel, to the end user's portable for supporting decision making.

Future work will focus to analysis other promising machine learning methods, such as deep learning networks, time-series techniques, and updated metaheuristic optimizer to simulate the prediction models, so as to improve the accuracy and computation time of the system and better enable it to cope with big energy consumption data in the real world.

- [1] A. Costa, M. M. Keane, P. Raftery, and J. O'Donnell, "Key factors methodology—A novel support to the decision making process of the building energy manager in defining optimal operation strategies", *Energy and buildings*, vol. 49, 2012, pp. 158-163.
- [2] A. Allouhi, Y. El Fouih, T. Kousksou, A. Jamil, Y. Zeraouli, and Y. Mourad, "Energy consumption and efficiency in buildings: current status and future trends", *Journal of Cleaner production*, vol. 109, 2015, pp. 118-130.
- [3] Y. M. Lee, L. An, F. Liu, R. Horesh, Y. T. Chae, and R. Zhang, "Applying science and mathematics to big data for smarter buildings", *Annals of the NEW YORK Academy of Sciences*, vol. 1295, no. 1, 2013, pp. 18-25.
- [4] L. Pérez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information", *Energy and buildings*, vol. 40, no. 3, 2008, pp. 394-398.
- [5] T. Malche and P. Maheshwary, "Internet of Things (IoT) for building smart home system", in *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, 2017: IEEE, pp. 65-70.
- [6] L. M. Candanedo, V. Feldheim, and D. Deramaix, "Data driven prediction models of energy use of appliances in a low-energy house", *Energy and buildings*, vol. 140, 2017, pp. 81-97.
- [7] S. Fathi, R. Srinivasan, A. Fenner, and S. Fathi, "Machine learning applications in urban building energy performance forecasting: A systematic review", *Renewable and Sustainable Energy Reviews*, vol. 133, 2020, p. 110287.
- [8] X. Chang, M. Gao, Y. Wang, and X. Hou, "Seasonal autoregressive integrated moving average model for precipitation time series", *Journal of Mathematics & Statistics*, vol. 8, no. 4, 2012, 500-505.
- [9] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization", *Swarm intelligence*, vol. 1, no. 1, 2007, pp. 33-57.
- [10] J.-S. Chou and D.-N. Truong, "Multistep energy consumption forecasting by metaheuristic optimization of time-series analysis and machine learning", vol. 45, no. 3, 2021, pp. 4581-4612.
- [11] J. S. Chou, D. N. Truong, and T. L. Le, "Interval Forecasting of Financial Time Series by Accelerated Particle Swarm-Optimized Multi-Output Machine Learning System", *IEEE Access*, vol. 8, 2020, pp. 14798-14808, doi: 10.1109/ACCESS.2020.2965598.
- [12] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 1999.
- [13] S. Su, W. Zhang, and S. Zhao, "Fault prediction for nonlinear system using sliding ARMA combined with online LS-SVR", *Mathematical Problems in Engineering*, 2014(3-4), 1-9. DOI:10.1155/2014/692848
- [14] J. A. Suykens and J. J. N. p. I. Vandewalle, "Least squares support vector machine classifiers", vol. 9, no. 3, 1999, pp. 293-300.
- [15] J.-S. Chou and N.-T. Ngo, "Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns", *Applied Energy*, vol. 177, 2016, pp. 751-770.
- [16] J. Kennedy and R. Eberhart, "Particle swarm optimization", in *Proceedings of ICNN'95-international conference on neural networks*, vol. 4: IEEE, 1995, pp. 1942-1948.