



Estimates of residential building energy consumption using a multi-verse optimizer-based support vector machine with k-fold cross-validation

Hamed Tabrizchi¹ · Mohammad Masoud Javidi¹ · Vahid Amirzadeh²

Received: 8 January 2019 / Accepted: 30 March 2019 / Published online: 8 April 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

The ever-increasing human population, building constructions, and technology usages have currently caused electric consumption to grow significantly. Accordingly, some of the efficient tools for more and more energy saving and development are efficient energy management and forecasting energy consumption for buildings. Additionally, efficient energy management and smart restructuring can improve energy performance in different areas. Given that electricity is the main form of energy that is consumed in residential buildings, forecasting the electrical energy consumption in a building will bring significant benefits to the building and business owners. All these means call for precise energy forecast to make the best decisions. In recent years, artificial intelligence, generally, and machine learning methods, in some areas, have been employed to forecast building energy consumption and efficiency. The present study aims to predict energy consumption with higher accuracy and lower run time. We optimize the parameters of a support vector machine (SVM) using a multi-verse optimizer (MVO) without the grid search algorithm, due to the development consequence of residential energy predicting models. This paper presented the MVO-SVM approach for predicting energy consumption in residential buildings. The proposed approach examined a UCI repository dataset. Based on the experimental results MVO can effectively decrease the number of features while preserving a great predicting precision.

Keywords Support vector machine · Energy consumption forecast · Multi-verse optimizer · Cross-validation · Data science

1 Introduction

As we all know energy plays an important role in the modern world so that it ensures human convenience and development of the countries. Currently, the causes of electric consumption rapid growth include hugely increased human population, buildings and technology application (Amasyali and El-Gohary 2018). Statistical reports (World Energy

Trilemma Index 2018) indicate that the pattern of population growth continually rises in many developed countries. The growing pattern of the human population leads to the increased building energy consumption that is more obvious in the industrial field. Faced with population growth as well as the demand for energy needed for human survival, many challenges will arise for expanded or developing countries. Yet, to reduce pollution, carbon emission, and greenhouse effect, environmental themes must be considered in the development process. A significant change in human energy consumption, more environment-friendly products, and identifying are required to reach reduced greenhouse gases. According to Global Energy Statistical Yearbook 2018 (World Power Consumption 2019), due to electrification of energy uses, electricity consumption increases faster compared to other types of energy. Asia experienced most electricity consumption increase in 2017. As in 2016, despite an industrial recovery and widespread energy efficiency improvements, more than half the world electricity consumption rebound is related to the electricity

✉ Mohammad Masoud Javidi
javidi@uk.ac.ir

Hamed Tabrizchi
hamed.tabrizchi@math.uk.ac.ir

Vahid Amirzadeh
v_amirzadeh@uk.ac.ir

¹ Department of Computer Science, Shahid Bahonar University of Kerman, Kerman, Iran

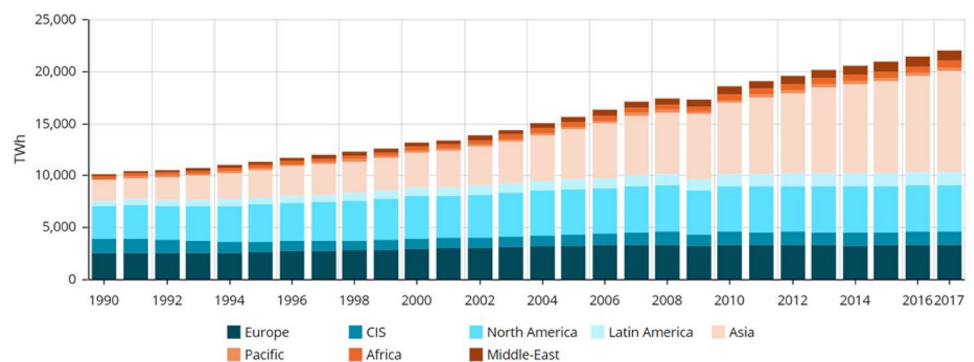
² Department of Statistics, Shahid Bahonar University of Kerman, Kerman, Iran

consumption growth in China. Also, Electricity consumption has considerably augmented in Iran and Egypt. Generally, this is an impression that countries are associated with increasing energy consumption, as shown in Fig. 1.

Nowadays, due to matters like fuel depletion, environmental impact, and carbon dioxide emission, energy consumption forecasting has become a significant issue. In addition, energy consumption forecasting also plays an important role in improving energy performance, saving energy, and reducing the serious environmental impact. It can be briefly mentioned that estimating energy demand of the building is conducive for the optimal control of the building energy infrastructure planning. In addition, forecasting also contributes considerably in decision-making and future planning based on the accurate forecasting (Alobaidi et al. 2018). Some of the many factors effective in energy consumption forecasting in a typical building are outside air temperatures, electrical devices inside the building, Heating, Ventilation and Air Conditioning (HVAC), geographical location of the building and how the building is used (residential, office, classroom, etc.). The HVAC is considerably important for residential buildings because it improves home air quality in terms of temperature and humidity. The common loads in commercial and industrial buildings include lighting, Surveillance Cameras, and air-conditioning system. Furthermore, other devices like computers, printers, televisions, fax machines, etc. also consume energy. However, the most electrical energy consumption is related to the air-conditioning system generally. It should be noted that the geographical location of a building also contributes to the electric energy consumption, and changes the forecasting analysis. Geographical differences lead to different usages of electrical equipment. Several factors are effective in a country's electrical usages that weather condition, the surrounding temperature of the country, and seasons are among them. In a hot weather, the cooling system will consume a lot of energy (Wei et al. 2018). It is clear that in temperate regions, the cooling system uses less electricity. In addition to temperature, some other factors also impress the electricity demand that among them humidity, wind speed, cloudiness, rainfalls

and solar radiation can be mentioned. Furthermore, the different working hours of various buildings can be mentioned in terms of time-scale. As an example, the electricity usage duration is usually 24 h in industrial buildings. However, some industrial buildings operate within working hours (Nikolaou et al. 2015). For office buildings, the operating time is typically set with the specified time interval. In a residential place, the maximum electricity usage is in the evening. However, this electrical energy usage also depends on several conditions, such as holidays, seasons travel, natural disasters, and others. Currently, many forecasting models have been presented using several methods to solve forecasting complexities and to reach maximum forecasting accuracy. There are several popular ways to forecast building energy consumption, divided into four main categories namely engineering calculation, simulation model based, statistical modelling and Artificial intelligence method (Song et al. 2017; Seyedzadeh et al. 2018). The engineering methodologies use physical laws to understand building energy consumption in whole or sub-system levels. These methodologies have complex calculations for different building components and their inputs are internal and external details. In a simulation model methodology, software and computer models are employed for performance simulation with predefined conditions. Computer simulation makes a system which uses a computer to simulate a mathematical model (Nowotarski et al. 2018). In fact, computer simulation has many applications including weather forecasting, price forecasting on financial markets, simulation of electrical circuits, etc. Finally, in a statistical method, mathematical formulas, models and techniques are employed for raw data analysis. Generally, statistical methods extract information from data and create different ways to determine the output accuracy. Also, in a statistical method, building historical data are used and regression is frequently employed to model the energy consumption in buildings. Artificial Intelligence (AI) is a wide science field that uses learning, reasoning, and self-modification to solve problems (Wang et al. 2018). Similarly, AI methods provide the ability to learn from data using a computer algorithm. Artificial intelligence is

Fig. 1 Electricity consumption over 1990–2017, breakdown by country (Accessed 2 Jan 2019)



a broader concept of machines being able to carry out tasks in a smart way. Machine learning (ML) is a widespread application of artificial intelligence that uses statistical techniques to give computer systems the ability to learn from data. Among these methods, the most suitable implemented method in forecasting is the ML method, called support vector machine (SVM). Particularly, an SVM includes supervised learning models with associated learning algorithms that analyses data used for classification and regression analysis. SVM, artificial neural networks (ANN), decision trees, and other statistical algorithms are the most commonly-used supervised machine learning algorithms for model training. SVM is a kernel-based machine learning algorithm, which can be used for both regression and classification (Paudel et al. 2017). A predictive modeling method that can be used in several sciences including machine learning, data mining and statistics is decision tree. This method gives the regression and classification models a tree structure in which class labels are indicated by leaves and conjunctions of features are represented by branches. However, using this modeling approach in prediction problems causes important disadvantages. One of these disadvantages is the numerical input attribute that can make trees complex and lead to unstable situations in the learning process. Furthermore, a small change in input data can make trees completely different. One of the computational and mathematical models that has been inspired by biological nervous system is the artificial neural network (ANN). By taking advantage of ANNs, many algorithms can be developed to model and solve a lot of problems including prediction problems. One of the important disadvantages of Neural Networks is their black box nature. In other words, it is hard to realize what happens in the underlying process that comes up with a certain prediction. So, most organizations do not take advantage of the neural network because they cannot explain the reasons for their decision to customers. In this research, the reason for using the SVM method is the robustness that it adds to the model to deal with overfitting. In addition, small changes do not impress boundary significantly and, accordingly, do not make considerable differences. Finally, a comparison between the four main categories mentioned above is given in Table 1.

The rest of paper is organized as follows. Section 2 outlines an overview of some related work. Section 3 presents the methodologies. Section 4 presents the proposed MVO-SVM approach to forecasting energy demand. Section 5 predicts the energy consumption and compares the predictions with predictions from other works. The final section summarizes the main findings and concludes the proposed work.

2 Related work

There are copious methods to forecast energy demand. Currently, the energy demand forecasting studies can be generally divided into two categories, namely, the white box model and the black box model (Guo et al. 2018). The first category is based on the physical method and needs copious detailed features. The second one mainly involves learning methods including neural networks, multiple linear regression, etc. Limanond et al. (2011) employed the log-linear regression model to forecast the transportation energy demand in Thailand. Szoplik (2015) forecasted natural gas consumption in Szczecin using artificial neural networks (ANNs). A weighted support vector regression (SVR) was used by Zhang et al. (2016) to predict building energy usage. It should be noted that linear regression, random forest, and support vector regression (SVR) algorithms are employed to forecast energy demand at the city scale. A building energy prediction model was presented by Jain et al. (2014) in which they use an SVR for multi-family residential buildings. Muralitharan (2018) presented an energy demand forecasting model using a neural network optimized by genetic and particle swarm optimization (PSO) algorithms. An electricity price forecasting model was developed by Yang et al. (2017) in which wavelet and extreme learning machine (ELM) methods are used. The area in which the ELM method has been frequently employed is image and pattern classification. Another application of ELM is the short-term wind speed prediction model in which error correction is also used. Therefore, ELM method is used in the current study to develop some models for studying its potential in the energy demand prediction field. Two electric load forecasting models were developed by Yang et al. (2017) based on ANN and SVR. In fact, they extract features based

Table 1 Comparison between forecasting methods

| Methods | Model complexity | Easy to use | Running speed | Inputs needed | Accuracy |
|-------------------------|------------------|-------------|---------------|-----------------|-------------|
| Engineering calculation | Fairly high | No | Low | Detailed | Fairly high |
| Simulation model based | High | Yes | Low | Detailed | Fair |
| Statistical | Fair | Yes | Fairly high | Historical data | Fair |
| Artificial intelligence | | | | | |
| SVM | Fairly high | No | Low | Historical data | High |
| ANN | High | No | High | Historical data | Fairly high |

on a periodic characteristic to perform the principal component analysis and factor analysis on them. In summary, they develop two electric load forecasting models by ANN and SVR to assess the forecasting efficiency of each forecasting model by k-fold cross-validation and compare the prediction outcome with the real electric load. A clustering method was presented by Moon et al. (2017) that is based on a k-shape algorithm. This method is a relatively novel method to detect shape patterns in time-series data. In fact, in this method, clustering is done for each individual building based on its hourly consumption. In the current study, as an innovation, a new k-shape algorithm is employed to detect building-energy usage patterns at different levels; then, the clustering result is utilized to reach an improved forecasting accuracy. The experimental results indicate that the proposed method can detect building energy usage patterns in different time scales effectively. Furthermore, the results prove that utilizing the results of the proposed clustering method considerably increase the forecasting accuracy of the SVR model.

One of the greatest challenges that all methods face with is the choice of feature variables. It should be noted that one of the important factors for the performance of energy demand prediction models is the selection of feature variables. In some of these models, only the outdoor temperature is used in meteorological parameters. In heating energy demand forecasting, these models do not take into account the effects of other meteorological parameters such as indoor temperature. But successful energy demand prediction models are established by vital meteorological parameters such as operating parameters, time, and indoor temperature. According to the related works, many studies have been conducted on energy demand and consumption forecasting—as shown in Table 2.

The most frequently-used supervised machine learning algorithms for model training are SVM, ANN, decision trees, and other statistical algorithms. As a kernel-based

machine learning algorithm, SVM can be employed for both regression and classification (Pratama et al. 2014). This algorithm is very effective in solving non-linear problems that even have a relatively small amount of training data. SVM transforms the non-linearity between features x_i (e.g., current temperature and current solar flux) and target y (e.g., cooling energy consumption) using linear mapping in two steps to solve a non-linear problem. In the first step, it projects the non-linear problem into a high-dimensional space and specifies a function $f(x)$ which fits best in the high-dimensional space. In the second step, it uses a kernel function to transform the complex nonlinear map into a linear problem. With a closer look at the results of the studies, it can be found that hybrid models of energy consumption forecasting are mostly used to obtain an improved forecasting accuracy from the existing model (Iglesias et al. 2012). The results of the previous researches show that a hybrid model outperforms single and batch models in terms of accuracy. The main goal of both the accuracy of predictability and suitability for using these models is to help users in energy management planning. The performance analyses performed using mean absolute error (MAE) analysis and mean absolute percentage error (MAPE) showed that a hybrid model had less error than a single model. A conventional approach to select SVM parameters is to use an exhaustive grid search algorithm. It should be noted that this method requires a large number of evaluations and, thus, its running time is so very long. Also, feature selection is used to select the features that are not irrelevant, so, it leads to decreased training time and less complex classification methods. Another advantage of feature selection is that it occasionally improves the accuracy of prediction and increase the comprehensibility and generalization of the model. However, selecting the best subset from various possible subsets extends the search space significantly and make feature selection an NP-hard problem. Fortunately, there are

Table 2 Studies on energy forecast

| Authors | Methods | Forecast energy type |
|----------------------------|--|--|
| Limanond et al. (2011) | Log-linear regression models and feed-forward neural network models | Natural gas |
| Szoplik. (2015) | Artificial neural networks-MLP (multilayer perceptron model) | Natural gas |
| Zhang et al. (2016) | Partial least square regression (PLSR) | Transportation energy consumption |
| Jain et al. (2014) | Support vector regression (SVR) | Residential buildings Energy consumption |
| Muralitharan et al. (2018) | Neural Network based Genetic Algorithm (NNGA) and d Neural Network based Particle Swarm Optimization (NNPSO) | Electricity, gas and water (Pecan Street Inc.) |
| Yang et al. (2017) | Hybrid approach-the kernel extreme learning machine (KELM) based on self-adapting particle swarm optimization and an auto regressive moving average (ARMA) | Electricity price forecasting |
| Moon et al. (2017) | Artificial neural network, support vector regression | Power consumption |
| Yang et al. (2017) | Clustering method based on k-shape algorithm, SVR | Building energy usage |

other effective solutions that have been presented by different researchers. Metaheuristic algorithms are among the best-known approaches that are used when the complexity of problem increase and search space becomes wider. In the current study, we attempt to present and discuss a robust multi-verse-optimizer-based method to select best feature subset and optimize SVM parameters to predict the electrical energy consumption of residential buildings accurately in the shortest possible time. The main contribution of this study is that it proposes a novel hybrid model based on SVM to forecast the building electrical energy consumption.

3 Methodology

Basically, the majority of learning prediction methods consists of four principal steps. These principal steps are data collection, data pre-processing, model training, and model testing. A lot of learning algorithms and prediction models are employed with these steps to forecast the energy consumption of buildings. Usually, hybrid models use two or more machine learning techniques. Compared to the other methods, these models are more robust because they frequently present the advantages of the incorporated techniques and increase the accuracy of forecasting. This section presents the methods used in a particular area of study.

4 Support vector machine

The support vector machine (SVM) which was proposed by Vladimir Naumovich Vapnik is a kind of statistical learning method (Vapnik 1999). SVM have special benefits for data types with very high dimensions relative to the observations. So far, SVM has been extensively employed in numerous analyses including regression, classification, and nonlinear function approximation. Support vector regression which is used in the regression problem as an SVM application is suitable for a finite sample regression. It also has an outstanding generalization capability in the regression. SVM algorithm fits a boundary to a region of points which are called a hyperplane. Furthermore, the SVM tries to find the optimal hyperplane which separates two classes by maximizing the distance between the margin of the hyperplane and the data points in the dataset. In other words, given labelled training data while the algorithm output is an optimal hyperplane which categorizes new examples. A training set $T = \{(x_i, y_i) | i = 1, 2, \dots, l\}$, where the $x_i \in R^n$ are the input variables and $y_i \in R^n$ is the corresponding output value. As mentioned earlier, solving non-linear problems is one of the significant benefits of SVM algorithm (Wang and Pardalos 2014). In SVR, the support vectors are separated from the other training observations by a discriminating loss function.

This function does not penalize residuals less than a tolerance. ε is set for this purpose and determines a margin of tolerance in which errors are not penalized. In other words, ε defines a tube around the regression function to ignore errors inside. In addition, the value of the ε shows how closely the function fit the data. However, in the case that all training points fit within a tube of width 2ε , the algorithm outputs the function in the middle of the flattest tube which encloses them. The total perceived error is zero in this condition. The basic idea of the SVM attempts to find a hypothesis with not only a small structural risk, but also a reduced complexity. An inductive principle for model selection to learn from finite training data sets is structural risk minimization (SRM) (Vapnik 1992). SRM describes a general model of capacity control and provides a trade-off between hypothesis space complexity and the quality of fitting the training data. We can formulate the regression problem as an optimization problem as follows.

$$\min \left[\frac{1}{2} \|\omega^2\| + C \frac{1}{l} \sum_{i=1}^n (\xi_i + \xi_i^*) \right], \quad (1)$$

where ω is the vector of feature weights, two parameters ξ_i, ξ_i^* capture the magnitude of residuals beyond the prescribed tolerance ε and serve to guarantee a solution for all ε (illustrated in Fig. 2). C is a regularization term that determines the degree of the linear penalty applied to the residual excess ξ_i^* . The basic idea of the support vector regression machine is to select the appropriate nonlinear mapping, map the input variables of the low-dimensional space to the high-dimensional feature space. SVR often employs the *kernel trick* that implicitly maps the input space into a higher dimensional feature space using a kernel function (Bai et al.

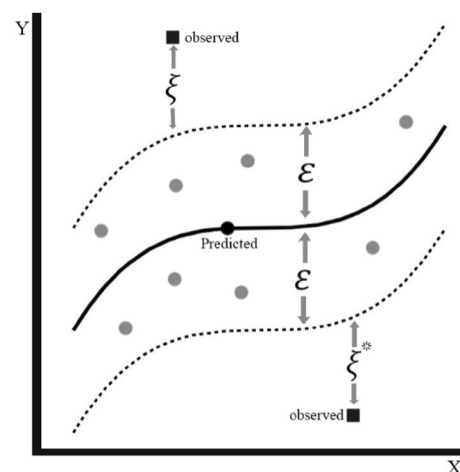


Fig. 2 Parameters of SVR (adapted from de Jesús Rubio et al. 2018 and Gong et al. 2018)

2018; de Jesús Rubio et al. 2018; Gong et al. 2018; de Jesús Rubio 2009, 2017; Cheng et al. 2018; Qiu et al. 2018).

$$K_{RBF}(x_i, x_j) = \exp\left(-\gamma \left\|x_j - x_i^2\right\|\right), \quad \text{where } \gamma > 0. \quad (2)$$

One of the important factors in SVM forecasting performance is the selection of the kernel function and kernel parameters. A kernel function including the linear kernel, the polynomial kernel, the Gaussian kernel, and the RBF kernel is shown as $K(x_i, x_j)$. The most used kernel function is the RBF kernel function. So, the RBF kernel function is used in this study, as given in Eq. (2). Intuitively, the influence radius for each data point is defined by γ . It worth noting that in the RBF implementation, the parameters C , ϵ and γ are user-defined variables with a significant influence on the SVR result (Vapnik 2013). The grid search method and the population intelligent optimization algorithm are among the most frequently parameter optimization methods. The MVO algorithm outperforms other algorithms in terms of optimization results. Accordingly, we take advantage from MVO algorithm in this study is to optimize the parameters of the SVM model.

4.1 Multi-verse optimizer (MVO)

The multi-verse optimizer (MVO) (Mirjalili et al. 2016) is inspired from three cosmology concepts, namely, white hole, black hole, and wormhole. In fact, the multi-verse theory is a state-of-the-art and well-known theory among the physicists. According to this theory, the number of big bangs is more than one and each causes the birth of a universe. The term multi-verse means there are other universes plus the universe we live in. According to this theory, each of the universes has different physical laws. As for MVO algorithm, three main concepts of the multi-verse theory can be chosen: white holes, black holes, and wormholes. Physicists believe a white hole may be the primary part for the birth of a universe. On the other hand, they believe that black holes, which have been frequently observed, have a completely different behaviour compared to white wholes. Black holes that have an extremely high gravitational force attract everything including light beams. Finally, wormholes connect different parts of a universe together. The wormholes in this theory are like space travel tunnels in which objects are able to travel instantly within a universe. In the modelling of this method, an inflation rate is considered for each world. As seen in other population-based algorithms, these algorithms divide the search process into two phases: exploration versus exploitation. The role of white hole and black hole concepts in MVO is to explore search spaces. According to MVO assumptions, each solution is analogous to a universe; furthermore, each variable in the solution is an object in that universe. Moreover, an inflation rate is assigned to each

solution that is proportional to the solution's corresponding fitness function value.

Accordingly, each variable in the optimization problem is an object in the universe. In the optimization process, these objects comply with the following rules.

1. The higher the inflation rate, the higher the probability of having white holes and the lower the probability of having black holes.
2. When a universe has a high inflation rate it sends objects through white holes, and when a universe has a low inflation rate it receives objects through black holes.
3. No matter the inflation rate is high or low, the objects of all universes may experience random movements towards the best universe by wormholes.

Given the rules listed above, the possibility to move objects from a universe with a high inflation rate to a universe with low inflation rate is always high. This ensures the improved average inflation rates of the whole universes over the iterations. In summary, MVO is a step-by-step approach to solve problems, as described below.

Step 1. Initialize the universe U , the maximum number of iterations Max-iteration , the variable interval $[lb, ub]$, and the universe position.

$$U = \begin{bmatrix} x_1^1 & \cdots & x_1^d \\ \vdots & \ddots & \vdots \\ x_n^1 & \cdots & x_n^d \end{bmatrix} \quad (3)$$

Step 2. Set up a universe through the roulette wheel selection mechanism to choose a white hole according to the universe inflation rate.

Step 3. Update the wormhole existence probability (WEP) and travel distance rate (TDR), and do a boundary check. Two main mentioned coefficients are to define the probability of wormhole existence in universes. To emphasize exploitation as the progress of the optimization process, the linearly should increase over the iterations. TDR is also used to define the distance rate (variation) which an object can be teleported by a wormhole around the best universe found so far. Unlike the WEP, TDR increases over the iterations to have a more accurate exploitation/local search around the best found universe.

$$WEP = \min + I \cdot \left(\frac{\max - \min}{L}\right), \quad (4)$$

$$TDR = 1 - \frac{l^{1/p}}{l^{1/p}}, \quad (5)$$

where Max and Min are the maximum and minimum values of WEP respectively. Moreover, I is the current iteration, L

is the maximum number of iterations, and p is the exploitation accuracy. In the process of working MVO, the values of these two above-mentioned coefficients will cause exploration or exploitation so that a Low WEP and a High TDR implies exploration and avoid local optima, while exploitation is supported by a High WEP and a Low TD. MVO should trade-off these two opposing forces in order to ensure an effective and efficient search. It is clear that in each iteration, the fitness values of different universes, WEPs, and TDRs are different. In the universe with the largest fitness value, the exploration ability should be enhanced, the WEP should decrease, and the TDR should increase. Otherwise, the WEP and TDR should decrease and increase respectively.

Step 4. Calculate the universe’s current inflation rate. In the case that the inflation rate of the universe is better than the current inflation rate of that, the current inflation rate is updated. Otherwise, we should maintain the current universe.

Step 5. Update the universe position based on the following equation

$$x_i^j = \begin{cases} \left\{ \begin{array}{l} X_j + TDR \times ((ub_j - lb_j) \times r4 + lb_j) \\ X_j - TDR \times ((ub_j - lb_j) \times r4 + lb_j) \end{array} \right. & r3 < 0.5 \\ x_i^j & r3 \geq 0.5 \end{cases} \quad \begin{array}{l} r2 < WEP \\ r2 \geq WEP \end{array} \quad (6)$$

where X_j shows the j th parameter of the best universe formed so far. Also, lb_j and ub_j specify the lower and upper bounds of j th variable respectively; x_i^j is the j th parameter of i th universe; and $r2, r3, r4$ are the random numbers drawn from the interval of $[0, 1]$.

Step 6. Termination criterion. The algorithm stops when the termination criterion is realized and the corresponding result is considered as output. Otherwise, after increasing the number of iterations by 1, the algorithm returns to Step 2.

4.2 K-fold cross validation

One of the indispensable tasks to forecast the class values of new instances is classification. Among the methods to evaluate the performance of classification algorithms, k-fold cross validation is a considerably popular one (Fushiki 2011). Most of the time, cross-validation (CV) is usually employed to compare and select a model for a certain predictive modelling problem because of its simplicity and practicality; furthermore, it results in skill estimates which have a lower bias compared to other methods. To study the impact of this randomness mechanism, bias and variance measures are used. In fact, bias represents the expected difference between an accuracy estimate and actual accuracy, and variance represents an accuracy estimate variability (Rohani

et al. 2018; Wong 2015). k-fold cross-validation is used to split the shuffled data into k groups, the estimator is then trained on $k-1$ groups and then tested on the k th group so that k different possibilities can be distinguished to choose which group should be the k th partition. Therefore, you get k results of all k possibilities of your estimator.

In fact, k-fold cross validation is a method used to estimate the skill of the model on new data. The main single parameter called k refers to the number of groups that a given data sample should be split into. After choosing a specific value for k , it may be used in place of k in the reference to the model, such as $k=5$ becoming fivefold cross-validation. It should be noted that the selection of k value for your data sample must be done attentively. A badly chosen value for k may provide incorrect information of the model, the results like a score with a high variance or a high bias. More precisely, four factors can be mentioned that are effective in using k-fold cross-validation, namely, the number of folds (k), the number of instances in a fold, the level of averaging (fold or data set), and the repetition of cross-validation. The average of the k accuracies obtained from

k-fold cross-validation is used to evaluate the classification algorithm performance.

5 Proposed MVO-SVM approach

As mentioned earlier, in this section we use the MVO population-based algorithm to optimize the parameters of the SVM (Faris et al. 2018). In the algorithm, a vector of real numbers is resulted by encoding individuals (universes). The number of elements in each vector is equal to the number of the features in the dataset plus two elements to represent SVM parameters—the Cost (C) and Gamma (γ). The steps of the optimization are as follows (as shown in Fig. 3):

Step 1 (data normalization): to remove the effect of some features which have different range values on the learning process, the values of all features should be mapped into the same scale. Therefore, an equal weight is given to all features. Then, all features are normalized to fall in the interval $[0, 1]$.

Step 2 (separate a data set): separating the normalized data into training and testing sets is an important part of the characteristics of the model. Furthermore, by employing k-folds cross-validation, the training part is split again into a number of smaller parts. So, the SVM is trained k

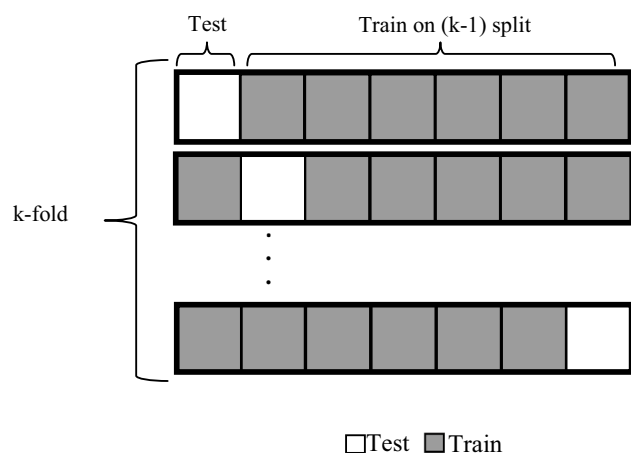


Fig. 3 An illustrated example of k-folds cross validation

times and the average evaluation should be employed again. The training data and validation data are respectively used to estimate the model and to select the final model. As for the testing set, the final model is used to test and compare it with other models.

Step 3 (parameters initialization): initialize the parameters of the maximum number of iterations (Max_iteration), the universe number, and the range of C , γ .

Step 4 (universe position initialization): initialize the universe position. Each universe in the MVO algorithm represents a set of $(C, \gamma, f_1, \dots, f_n)$ initialized according to the parameter range in the previous (Step 3).

Step 5 (function notation and evaluation): calculate the fitness of the universe and sort it. Then, select a white hole via roulette mechanism. The normalized mean square error (NMSE) is used as the evaluation criterion in this study to detect the suitable parameters of the SVM model (Poli and Cirillo 1993; Andonovski et al. 2016). In fact, to estimate the overall deviations between predicted and measured values, NMSE is employed. The NMSE value is always non-negative and if the value of NMSE is closer to zero, the error rate will be lower. In general, NMSE indicates the most notable differences among the models. In fact, a model with a very low NMSE has a good performance. However, a high NMSE value does not presently mean that the model is quite wrong. It worth noting that the determination of each universe's NMSE value is done by the fitness function (de Jesús Rubio et al. 2018; Gong et al. 2018; de Jesús Rubio 2009, 2017; Cheng et al. 2018; Qiu et al. 2018).

$$NMSE = \frac{n-1}{n} \frac{\sum_{i=1}^n [(f_m)_i - (f_p)_i]^2}{\sum_{i=1}^n [(f_m)_i - (\hat{f}_m)_i]^2}, \quad (7)$$

where f represents the model value, and the indices m and p indicates the measured and predicted values, respectively. The average value of the associated variable is shown with a 'tilde' above it, and the total number of training records is shown by n .

Step 6. Update the WEP and TDR based on the Eqs. (4) and (5).

Step 7. In the case that the fitness of the universe outperforms the current fitness of that, update the current fitness of the universe. Otherwise, maintain the current universe.

Step 8. Update the universe position and discover the optimal individual in the optimal universe.

Step 9. Termination criterion. The algorithm stops when the termination criterion is realized and the corresponding result is considered as output. Otherwise, after increasing the number of iterations by 1, the algorithm returns to Step 5.

The process described to optimize SVM parameters using the MVO algorithm considers primary features for residential building energy consumption include the temperature in rooms, number of residents, solar flux, humidity in rooms area, humidity outside the building, and the temperature outside as feature of SVM. At last, the architectures, as shown in Fig. 4, illustrate the steps of the MVO-SVM approach as follows:

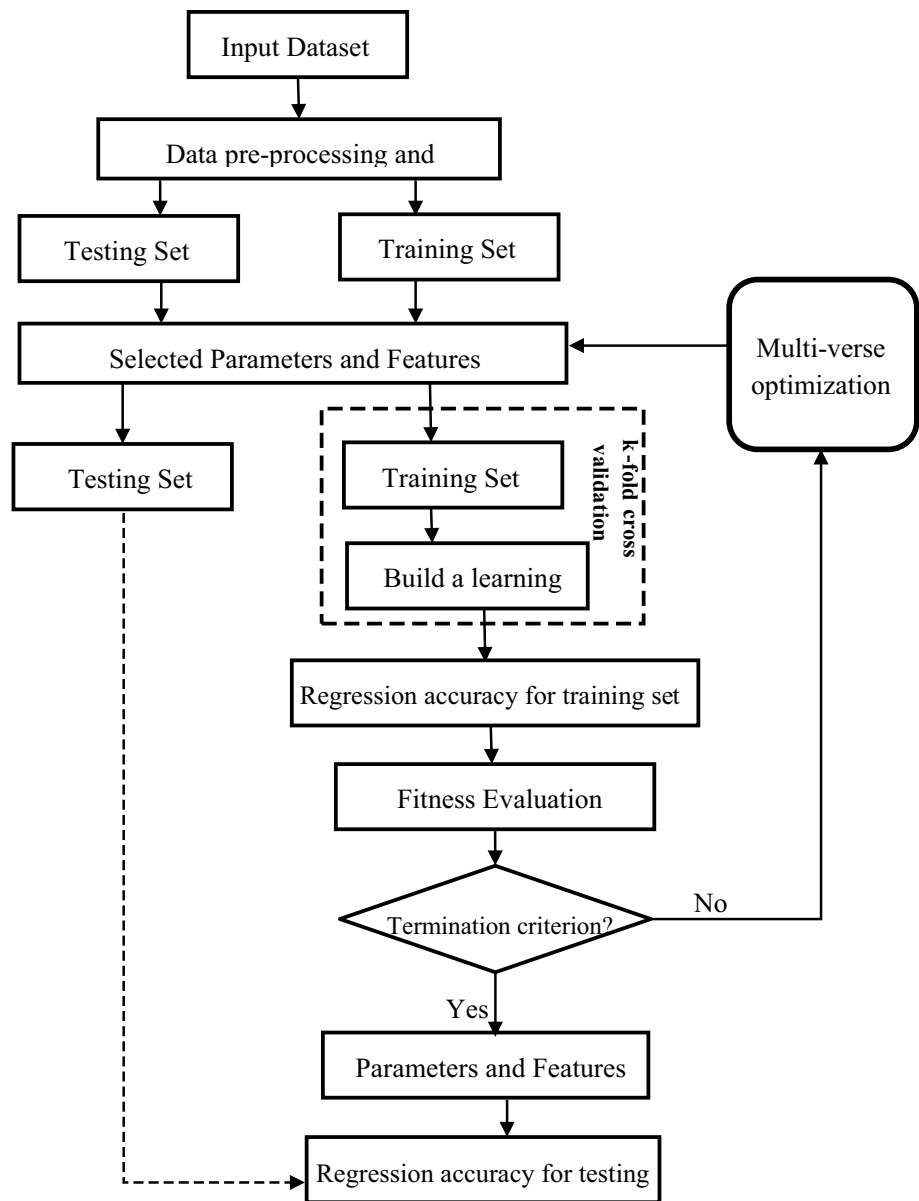
6 Results and discussion

To forecasting the energy consumption patterns in the building, the proposed model was evaluated and examined based on the selected dataset from the UCI repository (UCI Machine Learning Repository 2019). The data for 4.5 months were collected at 10 min. The house temperature, humidity conditions, and wind speed were monitored with a wireless sensor network. Each wireless node transmitted the temperature and humidity conditions which collected from the nearest airport weather station (Chievres Airport, Belgium) and downloaded from a public dataset (Candanedo et al. 2017).

The experiments of the present study executed on a personal machine with an Intel Core i5-2.20 GHz CPU and 16 GB RAM, as well as all approaches implemented by Python using Scikit-Learn libraries.

In this section, the GA-SVM (Shi et al. 2010), and PSO-SVM (Zhang et al. 2012; Han and Bian 2018) considered as two approaches of forecasting energy consumption, and these approaches compared to MVO-SVM with k-fold cross-validation. According to the above mentioned, the SVM objective function used two parameters which define by the user, such as the regularization term C and the radial basis function (RBF) with a value for γ , which identifies the radius of influence in each support vector. To limit overfitting, k-folds cross-validation executed in each combination

Fig. 4 MVO-SVM flow chart



of C and γ . Therefore, to achieve a learning model which have a better performance, requires suitable values for two mentioned parameters, additionally, to achieve these values utilize population-based methods. Table 3 reported the initial parameters of MVO, GA, and PSO algorithms. In all algorithms, the number of universes, individuals and swarm size are similar and equal to 100. In addition, with setting a lower value for maximum iteration, running the algorithm with the small number of iterations leads to decrease the computation time in the metaheuristic algorithms. Accordingly, for the maximum iterations in these algorithms, the value equals to 60.

The initial values for these parameters are selected based on the best available result which achieved by these methods. As the advancement in the optimization process in the

MVO algorithm, the probability of wormhole existence emphasizes exploitation. The value of this probability setting according to three parameters include current iteration, minimum, and maximum of wormhole existence rate. In addition, choosing the roulette wheel method and setting the suitable rates for genetic operators leads to a suitable convergence process and achieve better results, provided that the parameters of the GA algorithm were considered. In the PSO algorithm, the suitable value selected for the inertia weight and acceleration which verified to have the advantage of channel assignment in three aspects such as convergence rate, convergence speed, and the independence in the quality of the initial solution (Gandomi et al. 2013; Angelov 2014; Angelov and Kasabov 2005).

Table 3 Initial parameters of the MVO, GA, PSO

| Algorithm | Parameter | Value |
|-----------|-----------------------------|------------------------|
| MVO | Min wormhole existence rate | 0.2 |
| | Max wormhole existence rate | 1 |
| | Iterations | 60 |
| | Universes | 100 |
| GA | Crossover rate | 0.8 |
| | Mutation rate | 0.1 |
| | elitist ratio | 0.1 |
| | Selection mechanism | Roulette wheel |
| | Population size | 100 |
| | Generations | 60 |
| | PSO | Acceleration constants |
| | Inertia weight | 1 |
| | Generations | 60 |
| | Number of particles | 100 |

In the present study, the cross-validation of proposed comparison approach is equal to 10, therefore, SVM is trained ten times, wherein each time SVM is trained using different ninefolds, and the fitness function returns a fitness value based on the tenth testing fold. The results represented that compared to GA and PSO, the MVO algorithm reached the highest average accuracy rates in introduced Dataset. The MVO-SVM model has the best accuracy for forecasting residential buildings primary energy consumption, for this reason, the mean and standard deviation of R-squared is considered to compare the predictive accuracies of the models (Cameron and Windmeijer 1997; Troy et al. 2007; Angelov et al. 2004).

$$R^2 = 1 - \frac{\sum_1^n (\hat{y}_i - \bar{y}_i)^2}{\sum_1^n (y_i - \bar{y}_i)^2}, \tag{8}$$

where n is the number of data points, \hat{y}_i is the predicted value, y_i is a real value in the testing set and \bar{y}_i is mean of real values. It is clear that R^2 finds a value close to one, which indicates that it is a perfect fit if the real and estimated value were closed together. However, R^2 finds a value close to zero, if there is a large distance between the real and estimated values. Based on Table 4, the obtained results indicating that the AMVO-SVM rolling cross-validation model has more promising results in terms of R-squared, therefore, the training and testing set achieved higher values than the

other optimization methods, respectively. In addition, The MVO-SVM with rolling cross-validation model produced a higher R-squared (93.65%) value in the testing set and has a better fit estimation compared to that of the other models. Results presented that the prediction process in the MVO algorithm is more accurate than the PSO-SVM and GA-SVM algorithms.

The predictive accuracy can significantly improve, when using cross-validation, as well as prediction accuracy was improved with MVO, which is a population-based approach. In this study, three algorithms are evaluated with selected datasets, accordingly, Table 4 and Fig. 5 indicated the average of the accuracy rate and the number of selected features with the standard deviation.

In the present study, other measures of standard evaluation were used to evaluate the prediction model, coefficient of variation (CV), mean squared error (MSE), and R-squared measures commonly used in the energy consumption prediction models. Generally, the CV has been used for two reasons. First, it is a measure of performance which is suggested to evaluate the energy consumption prediction models. Next, it normalizes the prediction error with the average energy consumption and provides a unitless measure which is more convenient for comparison purposes (Edwards et al. 2012). The CV metric is defined by:

$$CV = \frac{\sqrt{\frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2}}{\bar{y}} \times 100, \tag{9}$$

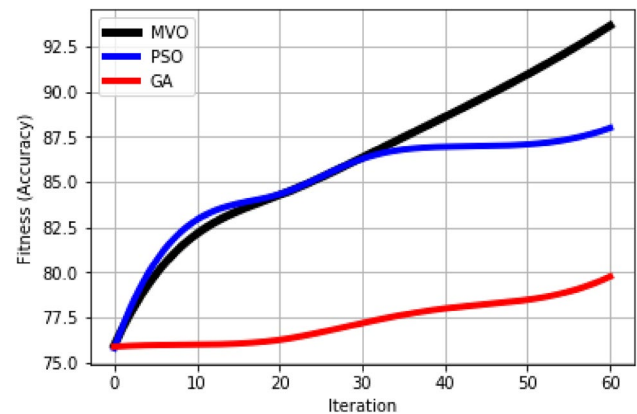


Fig. 5 Convergence curves of MVO, GA, PSO in optimizing SVM and feature selection

Table 4 Comparison of results from various estimation energy consumption models based on the SVM approach

| Algorithms | Whole set R^2 (%) | Training set R^2 | | Testing set R^2 | |
|---------------------------------|---------------------|--------------------|---------|-------------------|---------|
| | | AVG (%) | Std (%) | AVG (%) | Std (%) |
| MVO-SVM k-fold cross-validation | 93.65 | 99.48 | 0.021 | 88.98 | 5.24 |
| PSO-SVM k-fold cross-validation | 87.98 | 99.52 | 0.065 | 72.84 | 24.89 |
| GA-SVM k-fold cross-validation | 79.75 | 99.53 | 0.052 | 59.77 | 30.92 |

Where N is total number of observation, \hat{y}_i is the predicted value, y_i is the observed value and \bar{y} is the mean of observed values.

In this study, in order to optimize the SVM parameters, MVO was compared with the grid search. In addition, for a fair comparison, MVO only utilized for parameters optimization due to the grid search does not have the feature selection part, also tenfolds cross-validation used in both techniques. Figure 6 indicates that the result of the comparison which the MVO-SVM model noticeably performed better than the grid search model in daily temporal intervals. Generally, this section conducted that the performance of SVM was improved by the MVO algorithm, and based on the high exploitation of MVO. More accurate results of the were MVO-based SVM was achieved, and a for more accurate SVM, the parameters must be tuned accurately.

The results concluded that the MVO-SVM rolling cross-validation model to predicting energy consumption in residential building has better suitability and predictive ability. In addition, features selected by the multiverse algorithm for residential energy consumption are the temperature in the room, the number of inhabitants, the solar flux, the humidity in the room area, the and the outside temperature as the SVM feature. The global search modified with the MVO algorithm and cross-validation regarded to achieve better learning model. Finally, this modification achieved based on a balance between exploration and exploitation.

7 Conclusions and future work

Forecasting the residential buildings' energy consumption is very crucial for policymakers so that a precise prediction helps to implement energy policies properly. It should be noted that energy consumption is one of the most substantial research areas in today's world. Different models have different goals, cover different areas, involve different datasets, and take advantage of different features for forecasting. This study presents a considerable step in the area of energy forecasting for residential buildings. As a future work, we propose predicting residential energy by additional machine learning approach. Further, in order to select a better subset of features, we suggest applying another feature selection approach. It is important to present other metrics which accurately leads to model performance. State-of-the-art technologies including Big Data and Internet of Things have a special place in building energy usages where massive data obtained from sensors and energy meters require greatly efficient data processing systems. It is obvious that traditional methods of energy prediction cannot meet the needs of modern data mining development. As a result, intelligent models are almost necessary to answer this demand in industry, and it seems that further investigation of using artificial intelligent in building sector emphasizing on industrial data is an inevitable task. This study presented an MVO algorithm to optimize the SVM parameters and used the k-fold cross-validation to enhance its performance. The proposed approach is evaluated using PSO and GA algorithms. The results indicate that the MVO-SVM algorithm

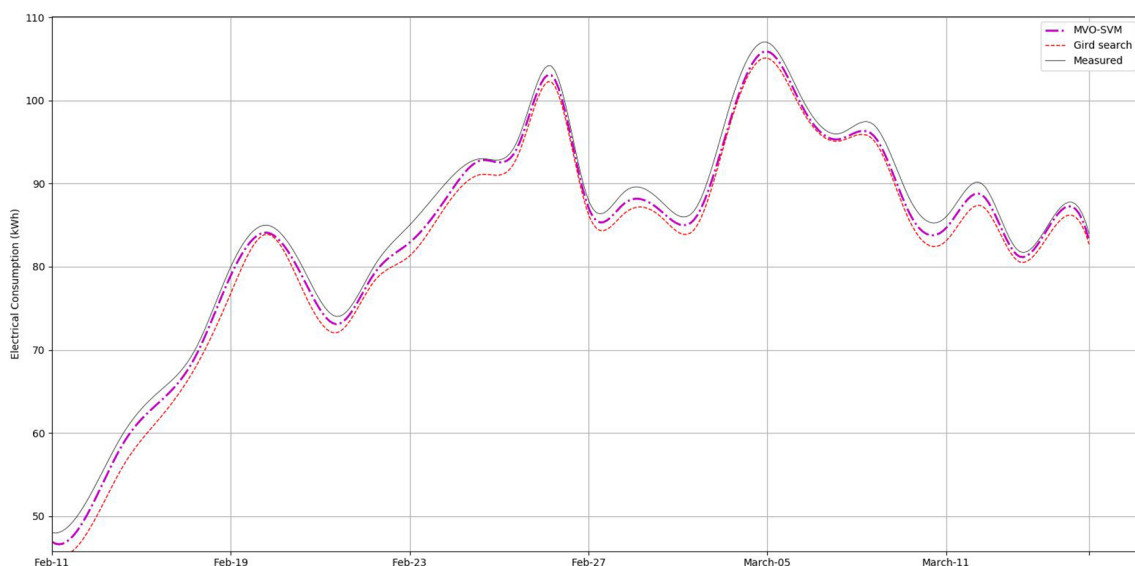


Fig. 6 Daily forecasting results for whole building (predicted by MVO-SVM and grid search)

outperforms others in terms of accuracy. Furthermore, the improved MVO-SVM algorithm has better exploration and exploitation abilities than MVO-SVM algorithm.

References

- Alobaidi MH, Chebana F, Meguid MA (2018) Robust ensemble learning framework for day-ahead forecasting of household based energy consumption. *Appl Energy* 212:997–1012
- Amasyali K, El-Gohary NM (2018) A review of data-driven building energy consumption prediction studies. *Renew Sustain Energy Rev* 81:1192–1205
- Andonovski G, Angelov P, Blažič S, Škrjanc I (2016) A practical implementation of Robust Evolving Cloud-based Controller with normalized data space for heat-exchanger plant. *Appl Soft Comput* 48:29–38
- Angelov P (2014) Outside the box: an alternative data analytics framework. *J Autom Mobile Robot Intell Syst* 8:29–35
- Angelov P, Kasabov N (2005) Evolving computational intelligence systems. In: *Proceedings of the 1st international workshop on genetic fuzzy systems*, pp 76–82
- Angelov P, Victor J, Dourado A, Filev D (2004) On-line evolution of Takagi–Sugeno fuzzy models. *IFAC Proc* 37:67–72
- Bai Y, Sun Z, Zeng B, Long J, Li L, de Oliveira JV, Li C (2018) A comparison of dimension reduction techniques for support vector machine modeling of multi-parameter manufacturing quality prediction. *J Intell Manuf*, pp 1–12
- Cameron AC, Windmeijer FAJ (1997) An R-squared measure of goodness of fit for some common nonlinear regression models. *J Econom* 77:329–342
- Candanedo LM, Feldheim V, Deramaix D (2017) Data driven prediction models of energy use of appliances in a low-energy house. *Appl Energy Predict* 140:81–97
- Cheng M-Y, Prayogo D, Wu Y-W (2018) Prediction of permanent deformation in asphalt pavements using a novel symbiotic organisms search–least squares support vector regression. *Neural Comput Appl*, 1–13
- de Jesús Rubio J (2009) SOFMLS: online self-organizing fuzzy modified least-squares network. *IEEE Trans Fuzzy Syst* 17:1296–1309
- de Jesús Rubio J (2017) Interpolation neural network model of a manufactured wind turbine. *Neural Comput Appl* 28:2017–2028
- de Jesús Rubio J, Lughofer E, Meda-Campaña JA, Páramo LA, Novoa JF, Pacheco J (2018) Neural network updating via argument Kalman filter for modeling of Takagi–Sugeno fuzzy models. *J Intell Fuzzy Syst* 35:2585–2596
- Edwards RE, New J, Parker LE (2012) Predicting future hourly residential electrical consumption: a machine learning case study. *Energy Build* 49:591–603
- Faris H, Hassonah MA, Ala’M A-Z, Mirjalili S, Aljarah I (2018) A multi-verse optimizer approach for feature selection and optimizing SVM parameters based on a robust system architecture. *Neural Comput Appl* 30:2355–2369
- Fushiki T (2011) Estimation of prediction error by using K-fold cross-validation. *Stat Comput* 21:137–146
- Gandomi AH, Yang X-S, Talatahari S, Alavi AH (2013) Metaheuristic algorithms in modeling and optimization. In: *Metaheuristic applications in structures and infrastructures*. Elsevier, Amsterdam, pp 1–24
- Gong Y, Yang S, Ma H, Ge M (2018) Fuzzy regression model based on geometric coordinate points distance and application to performance evaluation. *J Intell Fuzzy Syst* 34:395–404
- Guo Y et al (2018) Machine learning-based thermal response time ahead energy demand prediction for building heating systems. *Appl Energy* 221:16–27
- Han B, Bian XJP (2018) A hybrid PSO-SVM-based model for determination of oil recovery factor in the low-permeability reservoir. *Petroleum* 4:43–49
- Iglesias JA, Angelov P, Ledezma A, Sanchis A (2012) Creating evolving user behavior profiles automatically. *Trans Knowl Data Eng* 24:854–867
- Jain RK, Smith KM, Culligan PJ, Taylor JE (2014) Forecasting energy consumption of multi-family residential buildings using support vector regression: investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Appl Energy* 123:168–178
- Limanond T, Jomnonkwo S, Srikaew A (2011) Projection of future transport energy demand of Thailand. *Energy Policy* 39:2754–2763
- Mirjalili S, Mirjalili SM, Hatamlou A (2016) Multi-verse optimizer: a nature-inspired algorithm for global optimization. *Neural Comput Appl* 27:495–513
- Moon J, Park J, Hwang E, Jun S (2017) Forecasting power consumption for higher educational institutions based on machine learning, pp 1–23
- Muralitharan K, Sakthivel R, Vishnuvarthan R (2018) Neural network based optimization approach for energy demand prediction in smart grid. *Neurocomputing* 273:199–208
- Nikolaou T, Kolokotsa D, Stavrakakis G, Apostolou A, Munteanu C (2015) *Managing indoor environments and energy in buildings with integrated intelligent systems*. Springer, Berlin
- Nowotarski J, Weron RJR, Reviews SE (2018) Recent advances in electricity price forecasting: a review of probabilistic forecasting. *Renew Sustain Energy Rev* 81:1548–1568
- Paudel S et al (2017) A relevant data selection method for energy consumption prediction of low energy building based on support vector machine. *Electr Energy Syst* 138:240–256
- Poli AA, Cirillo C (1993) On the use of the normalized mean square error in evaluating dispersion model performance. *Atmos Environ* 27:2427–2434
- Pratama M, Anavatti SG, Angelov PP, Lughofer E (2014) PANFIS: a novel incremental learning machine. *IEEE Trans Neural Netw* 25:55–68
- Qiu B, Zhang Y, Yang Z (2018) New discrete-time ZNN models for least-squares solution of dynamic linear equation system with time-varying rank-deficient coefficient. *IEEE Trans Neural Netw Learn Syst* 29:5767–5776
- Rohani A, Taki M, Abdollahpour MR (2018) A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I). *Renew Energy* 115:411–422
- Seyedzadeh S, Rahimian FP, Glesk I, Roper M (2018) Machine learning for estimation of building energy consumption and performance: a review. *Vis Eng* 6:5
- Shi J, Yang Y, Wang P, Liu Y, Han S (2010) Genetic algorithm-piecewise support vector machine model for short term wind power prediction. In: *2010 8th World congress on intelligent control and automation (WCICA)*. IEEE, pp 2254–2258
- Song H, Qin AK, Salim FD (2017) Multi-resolution selective ensemble extreme learning machine for electricity consumption prediction. In: *International conference on neural information processing*. Springer, Berlin, pp 600–609
- Szoplik J (2015) Forecasting of natural gas consumption with artificial neural networks. *Energy* 85:208–220
- Troy AR, Grove JM, O’Neil-Dunne JP, Pickett ST, Cadenasso ML (2007) Predicting opportunities for greening and patterns of vegetation on private urban lands. *Environ Manag* 40:394–412

- UCI Machine Learning Repository: appliances energy prediction data set. <https://archive.ics.uci.edu/ml/datasets/Appliances+energy+prediction>. Accessed 3 Jan 2019
- Vapnik V (1992) Principles of risk minimization for learning theory. In: Advances in neural information processing systems, pp 831–838
- Vapnik VN (1999) An overview of statistical learning theory. *IEEE Trans Neural Netw* 10:988–999
- Vapnik V (2013) The nature of statistical learning theory. Springer Science and Business Media, New York
- Wang X, Pardalos PM (2014) A survey of support vector machines with uncertainties. *Ann Data Sci* 1:293–309
- Wang Z, Wang Y, Srinivasan RSJE (2018) A novel ensemble learning approach to support building energy use prediction. *Energy Build* 159:109–122
- Wei Y et al (2018) A review of data-driven approaches for prediction and classification of building energy consumption. *Renew Sustain Energy Rev* 82:1027–1047
- Wong T-T (2015) Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognit* 48:2839–2846
- World Energy Trilemma Index 2018. <https://www.worldenergy.org/publications/2018/trilemma-report-2018/>. Accessed 2 Jan 2019
- World Power Consumption | Electricity Consumption. <https://yearbook.enerdata.net/electricity/electricity-domestic-consumption-data.html>. Accessed 2 Jan 2019
- Yang Z, Ce L, Lian L (2017a) Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. *Appl Energy* 190:291–305
- Yang J et al (2017b) k-Shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement. *Energy Build* 146:27–37
- Zhang Y, Zhang X, Tang L (2012) Energy consumption prediction in ironmaking process using hybrid algorithm of SVM and PSO. In: International symposium on neural networks. Springer, pp 594–600
- Zhang F, Deb C, Lee SE, Yang J, Shah KW (2016) Time series forecasting for building energy consumption using weighted support vector regression with differential evolution optimization technique. *Energy Build* 126:94–103

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.